JCTC Journal of Chemical Theory and Computation

Insight into the Role of Mg^{2+} in Hammerhead Ribozyme Catalysis from X-ray Crystallography and Molecular Dynamics Simulation

Tai-Sung Lee,*[,†,‡] Carlos Silva-López,[‡] Monika Martick,[§] William G. Scott,[§] and Darrin M. York*[,‡]

Consortium for Bioinformatics and Computational Biology and Department of Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455, and Department of Chemistry and Biochemistry and the Center for the Molecular Biology of RNA, Sinsheimer Laboratories, University of California at Santa Cruz, Santa Cruz, California 95064

Received October 25, 2006

Abstract: Results of a series of 12 ns molecular dynamics (MD) simulations of the reactant state (with and without a Mg^{2+} ion) and early and late transition state mimics are presented based on a recently reported crystal structure of a full-length hammerhead RNA. The simulation results support a catalytically active conformation with a Mg^{2+} ion bridging the A9 and scissile phosphates. In the reactant state, the Mg^{2+} spends significant time closely associated with the 2′OH of G8 but remains fairly distant from the leaving group $O_{5'}$ position. In the early TS mimic simulation, where the nucleophilic $O_{2'}$ and leaving group $O_{5'}$ are equidistant from the phosphorus, the Mg^{2+} ion remains tightly coordinated to the 2′OH of G8 but is positioned closer to the $O_{5'}$ leaving group, stabilizing the accumulating charge. In the late TS mimic simulation, the coordination around the bridging Mg^{2+} ion undergoes a transition whereby the coordination with the 2′OH of G8 is replaced by the leaving group $O_{5'}$ that has developed significant charge. At the same time, the 2′OH of G8 forms a hydrogen bond with the leaving group $O_{5'}$ and is positioned to act as a general acid catalyst. This work represents the first reported simulations of the full-length hammerhead structure and TS mimics and provides direct evidence for the possible role of a bridging Mg^{2+} ion in catalysis that is consistent with both crystallographic and biochemical data.

The hammerhead ribozyme[1] is an archetype system to study RNA catalysis.[2,3] A detailed understanding of the hammerhead mechanism provides insight into the inner workings of more complex cellular catalytic RNA machinery such as the ribosome and ultimately may aid the rational design of new medical therapies[4] and biotechnology.[5,6]

Despite a tremendous amount of experimental and theoretical effort,[1,2,7,8] the details of the hammerhead ribozyme mechanism have been elusive. In particular, one of the main puzzles involves the apparent inconsistency between the interpretation of thio effect experiments[9,10] and mutational data[8] with available crystallographic structural information of the minimal hammerhead sequence.[11−13] Results from the biochemical experiments suggest that a pH-dependent conformational change, inconsistent with crystallographic data,[11−13] must precede or be concomitant with the catalytic chemical step. This includes a possible metal ion bridge between the A9 and scissile phosphates that in previous crystal structures were ∼20 Å apart. Moreover, the function of the 2′OH group of G8 remains unclear.

Very recently, the crystallographic structure of a full length hammerhead sequence has been determined at 2.2 Å resolution.[14] The naturally occurring full-length hammerhead sequence exhibits enhanced catalytic activity and a different metal ion requirement relative to the minimal motif.[15] The crystal structure has the A9 and scissile phosphates in close proximity, consistent with the interpretation of thio effect measurements,[9] and the 2′OH of G8 and N_1 of G12 poised to act as a general acid and base, respectively, consistent with photocrosslinking experiments[16] and mutational data.[8] However, the divalent metal ions required for catalysis were not resolved in this structure. This letter reports the first simulations of the full-length hammerhead ribozyme in the reactant, early, and late transition states along the reaction coordinate. Results support the requirement for a bridging Mg^{2+} ion between the A9 and scissile phosphates in the catalytically active conformation and provide evidence of a role of the metal ion in catalysis that is consistent with both crystallographic and biochemical data.

Simulations were performed with CHARMM[17] (version c32a1) using the all-atom nucleic acid force field[18,19] with extension to reactive intermediates (e.g., transition state mimics)[20] and TIP3P water model.[21] Simulations of the reactant state (with and without a Mg^{2+} ion) and early and late TS mimics were each performed at 298 K and 1 atm in

* Corresponding author e-mail: york@chem.umn.edu.
† Consortium for Bioinformatics and Computational Biology, University of Minnesota.
‡ Department of Chemistry, University of Minnesota.
§ University of California at Santa Cruz.

10.1021/ct6003142 CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/03/2007

Table 1. Key Distances (Å) in the Hammerhead Active Site[a]

	X-ray structure	reactant	reactant w/o Mg^{2+}	early-TS mimic	late-TS mimic
C1.1:O_{P2} ↔ A9:O_{P2}	4.27	3.36(49)	7.16(110)	4.00(06)	4.01(07)
Mg^{2+} ↔ G8:$O_{2'}$	3.08[b]	3.97(102)		2.24(13)	3.21(23)
Mg^{2+} ↔ C1.1:$O_{5'}$	4.04[b]	4.22(21)		3.68(35)	2.09(05)
G8:$H_{O2'}$ ↔ C1.1:$O_{5'}$		4.57(135)	7.61(81)	5.09(74)	2.36(42)
C17:$O_{2'}$ ↔ C1.1:P	3.18	3.61(23)	3.83(19)	1.88(11)	1.75(04)

[a] The simulation results were calculated over the last 10 ns with data collected every 1 ps. Shown are average values and standard deviations in the parentheses. [b] A proposed Mg^{2+} site was assumed directly between the crystallographic positions of C1.1:O_{P2} and A9:O_{P2}.

a rhombododecahedral cell (with PME[22] electrostatics) in the presence of ~10 000 water molecules and 0.14 M NaCl and carried out to 12 ns following 1 ns of solvent equilibration. In three simulations, a single Mg^{2+} ion was positioned so as to bridge the A9 and scissile phosphates that in the crystallographic structure are around 4.3 Å, which is well suited for Mg^{2+}-bridging coordination.[23] The Mg^{2+} ion is critical for stability and adopts different coordination states along the reaction coordinate, verified by preliminary QM/MM calculations (see the Supporting Information) that are supportive of a catalytic role consistent with experiments.

A stable Mg^{2+} ion bridge between the A9 and scissile phosphates is formed in the catalytically active conformation. The simulation results support a catalytic role for a Mg^{2+} ion bridging the A9 and scissile phosphates. In the simulations with a bridging Mg^{2+} ion, the average distance between the A9 and scissile phosphates remain within the crystallographic value of 4.3 Å, whereas in the absence of Mg^{2+} this key contact between stems I and II drifts to over 7 Å (Table 1). In the reactant state, the Mg^{2+} coordination between the C1.1 and A9 phosphate oxygens fluctuates between axial−axial and axial−equatorial modes, resulting in a shorter average oxygen−oxygen distance than that observed in the X-ray structure. This may suggest that in the reactant state the preferred binding mode of Mg^{2+} is different, e.g., between A9 and N_7 of G10.1,[24,25] and that a conformational change brings Mg^{2+} into a bridging position between A9 and the scissile phosphate leading to the transition state.[9] The present simulation results suggest that the close proximity of the A9 and scissile phosphates observed in the new full-length hammerhead structure[14] can be stabilized by a Mg^{2+} ion bridge that brings together stems I and II and facilitates formation of near-attack conformations (see the Supporting Information) in a way different from previous simulations based on the minimal sequence structures.[26,27,28,29]

In the early TS, the Mg^{2+} ion is positioned to shift the pK_a of the 2′OH of G8 to act as a general acid. In the reactant state, the Mg^{2+} spends significant time closely associated with the 2′OH of G8 (Figure 1) but remains fairly distant from the leaving group $O_{5'}$ position. In the early TS mimic simulation, where the nucleophilic $O_{2'}$ and leaving group $O_{5'}$ are equidistant from the phosphorus, the Mg^{2+} ion becomes directly coordinated to the 2′OH of G8, and is positioned closer to the $O_{5'}$ leaving group. The coordination of the Mg^{2+} ion in the early TS mimic simulation is consistent with a role of shifting the pK_a of the 2′OH in G8 so as to act as a general acid (Figure 2, left).

In the late TS, the Mg^{2+} ion can act as a Lewis acid catalyst to stabilize the leaving group and is poised to assist proton transfer from the 2′OH of G8. In the late TS mimic

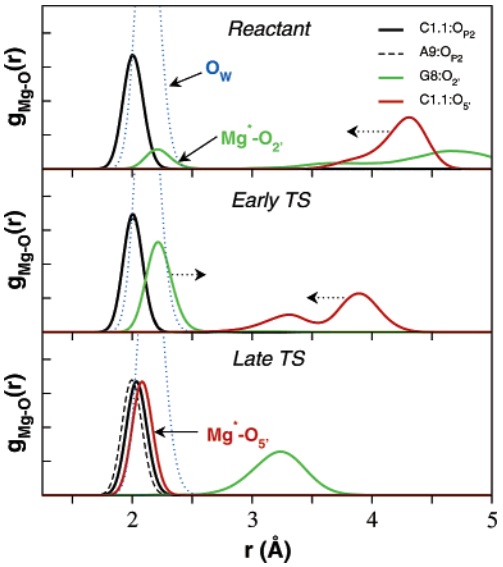


Figure 1. Radial distribution functions of key oxygens around Mg^{2+} in the active site for the reactant, early, and late TS mimic simulations.

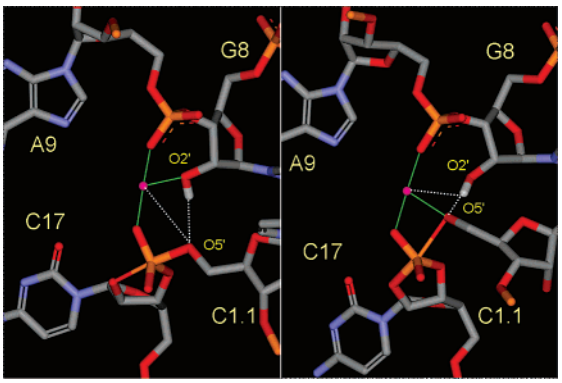


Figure 2. Snapshots of the active site from the early TS mimic (left) and late TS mimic (right) simulations depicting the Mg^{2+} ion direct coordination (green lines) and key hydrogen bonds and indirect Mg^{2+} coordination (dotted lines). For clarity, the water molecules are not shown.

simulation, a transition occurs whereby the Mg^{2+} coordination with the 2′OH of G8 is replaced by direct coordination with the leaving group $O_{5'}$ (Figure 1). In this way, the Mg^{2+} provides electrostatic stabilization of the accumulating charge of the leaving group (i.e., a Lewis acid catalyst).[7] At the same time, the 2′OH of G8 forms a hydrogen bond with the leaving group $O_{5'}$ and is positioned to act as a general acid catalyst (Figure 2, right).

Comparison with Experiment. The present simulation results, together with the crystallographic structure, tie together several key experimental results relating to the role of Mg^{2+} in catalysis. Thio/rescue effect experiments[9] have

suggested that a single metal bound at the P9/G10.1 site (the A9 phosphate in the present work) in the ground state acquires an additional interaction with the scissile phosphate in proceeding to the transition state. Kinetic analysis[30] along with photocross-linking experiments[16] and mutational data[8] have implicated the roles of the 2′OH of G8 and the N_1 of G12 as a general acid and base, respectively, and have been interpreted to be consistent with a transition into an active conformation with appropriate architecture for acid−base catalysis. However, recent kinetic analysis indicates the pK_a of the general acid is downshifted by around 4−7 pK_a units in a metal-dependent manner, correlated with the metal pK_a.[31] The simulation results suggest that the Mg^{2+} interacts strongly with the 2′OH of G8 in the early TS mimic and could contribute to a significant lowering of the pK_a value, and in the late TS mimic the G8 2′OH is hydrogen bonded to the leaving group and poised to act as a general acid catalyst (Figure 2). The Mg^{2+} ion may additionally play a direct role in stabilizing the negative charge accumulated by the leaving group in the late TS, and if a proton from the G8 2′OH is ultimately transferred, the coordination of Mg^{2+} is positioned to revert back to stabilize the resulting G8 2′ alkoxide.

The simulation results presented here are consistent with the direct participation of a single bridging Mg^{2+} ion in hammerhead ribozyme catalysis, although the possibility of involvement of a second ion cannot be definitively precluded.[32,33] The Mg^{2+} preserves the integrity of the active site structure and may serve as an epicenter in the transition state that coordinates the A9 and scissile phosphates, G8 2′OH general acid and $O_{5'}$ leaving group. The present work underscores the need for further investigation of the chemical reaction profile using combined QM/MM models.

ACKNOWLEDGMENT.

The authors are grateful for support from the National Institutes of Health, the IBM-Rochester Life Sciences Group, the Army High Performance Computing Research Center (AHPCRC), and the Minnesota Supercomputing Institute (MSI).

Supporting Information Available: Computational methodology and the rmsd plots of all simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

(1) Scott, W. G. *Q. Rev. Biophys.* **1999**, *32*, 241−294.
(2) Scott, W. G. *Curr. Opin. Struct. Biol.* **1998**, *8*, 720−726.
(3) Doherty, E. A.; Doudna, J. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 457−475.
(4) Rubenstein, M.; Tsui, R.; Guinan, P. *Drugs Future* **2004**, *29*, 893−909.
(5) Vaish, N. K.; Dong, F.; Andrews, L.; Schweppe, R. E.; Ahn, N. G.; Blatt, L.; Seiwert, S. D. *Nat. Biotechnol.* **2002**, *20*, 810−815.
(6) Breaker, R. R. *Curr. Opin. Biotechnol.* **2002**, *13*, 31−39.
(7) Takagi, Y.; Ikeda, Y.; Taira, K. *Top. Curr. Chem.* **2004**, *232*, 213−251.
(8) Blount, K. F.; Uhlenbeck, O. C. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 415−440.
(9) Wang, S.; Karbstein, K.; Peracchi, A.; Beigelman, L.; Herschlag, D. *Biochemistry* **1999**, *38*, 14363−14378.
(10) Suzumura, K.; Takagi, Y.; Orita, M.; Taira, K. *J. Am. Chem. Soc.* **2004**, *126*, 15504−15511.
(11) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065−2069.
(12) Murray, J. B.; Terwey, D. P.; Maloney, L.; Karpeisky, A.; Usman, N.; Beigelman, L.; Scott, W. G. *Cell* **1998**, *92*, 665−673.
(13) Murray, J. B.; Szöke, H.; Szöke, A.; Scott, W. G. *Mol. Cell* **2000**, *5*, 279−287.
(14) Martick, M.; Scott, W. G. *Cell* **2006**, *126*, 309−320.
(15) Canny, M. D.; Jucker, F. M.; Kellogg, E.; Khorova, A.; Jayasena, S. D.; Pardi, A. *J. Am. Chem. Soc.* **2004**, *126*, 10848−10849.
(16) Lambert, D.; Heckman, J. E.; Burke, J. M. *Biochemistry* **2006**, *45*, 7140−7147.
(17) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(18) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(19) MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105−120.
(20) Mayaan, E.; Moser, A.; Mackerell, A. D., Jr.; York, D. M. *J. Comput. Chem.* In press.
(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(22) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Hsing, L.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(23) Mayaan, E.; Range, K.; York, D. M. *J. Biol. Inorg. Chem.* **2004**, *9*, 807−817.
(24) Peracchi, A.; Beigelman, L.; Scott, E. C.; Uhlenbeck, O. C.; Herschlag, D. *J. Biol. Chem.* **1997**, *272*, 26822−26826.
(25) Peracchi, A.; Beigelman, L.; Usman, N.; Herschlag, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11522−11527.
(26) Hermann, T.; Auffinger, P.; Scott, W. G.; Westhof, E. *Nucleic Acids Res.* **1997**, *25*, 3421−3427.
(27) Hermann, T.; Auffinger, P.; Westhof, E. *Eur. Biophys. J.* **1998**, *27*, 153−165.
(28) Torres, R. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11077−11082.
(29) Torres, R. A.; Bruice, T. C. *J. Am. Chem. Soc.* **2000**, *122*, 781−791.
(30) Han, J.; Burke, J. M. *Biochemistry* **2005**, *44*, 7864−7870.
(31) Roychowdhury-Saha, M.; Burke, D. H. *RNA* **2006**, *12*, 1846−1852.
(32) Lott, W. B.; Pontius, B. W.; von Hippel, P. H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 542−547.
(33) Inoue, A.; Takagi, Y.; Taira, K. *Nucleic Acids Res.* **2004**, *32*, 4217−4223.
(34) Leclerc, F.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 3395−3409.

CT6003142

JCTC Journal of Chemical Theory and Computation

Impact of PES on MD Results of the Coalescence of M_2 + M with M = Ir, Pt, Au, Ag

Tiffany Pawluk, Li Xiao, Jennifer Yukna, and Lichang Wang*

Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, Illinois 62901

Received August 8, 2006

Abstract: The accuracy of the Sutton−Chen potential energy surface (PES) for describing atomic interactions in small metal clusters was investigated by comparison with density functional theory (DFT) calculation results. The binding energies calculated using the Sutton−Chen PES for the dimers, trimers, and 8- and 13-atom clusters of four transition metals, Ir, Pt, Au, and Ag, differ from those obtained with DFT calculations. As the DFT results agree well with the available experimental data, the above disagreement indicates that the original Sutton−Chen PES cannot accurately describe the interactions among atoms in the cluster for these metals. The parameters of the Sutton−Chen potential were therefore optimized to the DFT results for each of the metals. Molecular dynamics (MD) simulations were carried out on the coalescence of a dimer with a single atom for these metals. Both the original bulk and the cluster optimized Sutton−Chen PESs were tested with various incident angles and initial kinetic energies. The MD results show that the coalescence is highly dependent on the PES. This demonstrates that use of an accurate PES is critical, particularly at low-energy regime. The kinetic energy, incident angle, and choice of metal were examined for their role in the outcome of the coalescence process.

1. Introduction

Many fascinating properties exhibited by nanomaterials are highly size dependent.[1−3] For example, it is well-known that bulk gold is an inert material. However, temperature-programmed reaction studies revealed that small Au clusters with less than 20 atoms are active for CO combustion, and the catalytic activity varies with cluster size.[4] Small clusters are highly active and may aggregate if they are placed close to each other under suitable conditions.[1,5,6] This formation of large particles due to the aggregation of small clusters, often termed sintering in catalysis, is largely responsible for the complete loss of catalytic activities.[7−24] Despite this consequence, no effective control over this process has been possible due to a lack of understanding of the coalescence or association processes. It is therefore important to understand the coalescence mechanism.

Computationally, Monte Carlo (MC) and molecular dynamics (MD) methods are two common choices used to study the coalescence of large size metal clusters. For example, Tian and Guo performed MC simulations using the Lennard-Jones plus Axilrod-Teller potential to study the growth of 13- and 39-atom clusters for Ag, Au, Ni, and Pd metals.[25] Another example is the use of MD simulations by Hendy et al. to study the coalescence of Pb clusters.[26]

In either MC or MD simulations, a potential energy surface (PES) is used to describe the interactions among atoms in the system. There are four types of PESs, i.e., the Finnis−Sinclair potentials,[27] the embedded atom potentials,[28] the tight binding potentials,[29] and the Murrell−Mottram potentials,[30] which have been widely used for describing interactions among transition metal atoms. These potentials all incorporate parameters fitted to bulk or solution metals. Our goal is to test the reliability of each of these potentials for modeling small metal clusters. We begin our investigation with the Sutton−Chen potential,[31] which is a modified Finnis−Sinclair potential. It seems to be suitable for studying coalescence of transition metal clusters, as it is a many-body

* Corresponding author phone: (618)453-6476; fax: (618)453-6408; e-mail: lwang@chem.siu.edu.

10.1021/ct600257p CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/05/2007

Table 1. Original Sutton−Chen (Bulk) Potential Parameters[a] and Optimized (Cluster) Potential Parameters for Ir, Pt, Au, and Ag[b]

			ϵ (eV)		c		α (Å)	
	m	n	bulk	cluster	bulk	cluster	bulk	cluster
Ir	6	14	0.0024489	0.0025500	334.94	328.22	3.84	3.45
Pt	8	10	0.019833	0.011306	34.408	33.851	3.92	3.90
Au	8	10	0.012793	0.0064582	34.408	33.857	4.08	4.08
Ag	6	12	0.0025415	0.0018828	144.41	144.36	4.09	4.04

[a] See ref 31. [b] The parameters m and n were kept the same, while ϵ, c, and a were optimized.

potential and has been used to search for the global minimum of transition metal clusters.[32]

The parameters used in the Sutton−Chen potential were optimized by fitting to the bulk properties of each metal. Therefore, before carrying out MD simulations of coalescence for transition metal clusters, it is necessary to investigate whether the Sutton−Chen potential can be used to accurately describe the interactions among transition metal atoms in small systems. As the extensive results obtained for transition metal clusters from density functional theory (DFT) calculations[33−36] have just become available, it is possible to investigate the accuracy of the PESs.

The purpose of this current work is twofold. First, we investigated whether the Sutton−Chen potential is accurate for describing the interactions among atoms in a particle. We chose four metals, i.e., Ir, Pt, Au, and Ag, as our systems in this work. We calculated the binding energies for a number of clusters using the Sutton−Chen potential, and compared these results with the DFT results for those metal clusters. As our comparison has shown, the original bulk PES was not accurate; we therefore obtained cluster optimized parameters for the Sutton−Chen potential. Second, we performed MD simulations using the Sutton−Chen potential with both original bulk and cluster parameters in order to investigate the impact of different PESs on the outcome of a coalescence process. The MD simulations were performed to study the reaction dynamics of a single metal atom colliding with a metal dimer, i.e., A + BC with A=B=C = Ir, Pt, Au, or Ag. As those metal atoms are heavy compared to others such as the hydrogen atom, we expect that the quantum effect is negligible in the reaction dynamics. Therefore, we chose to use classical trajectory calculations in our study. Furthermore, the current MD studies focus on comparing individual trajectories at different incident angles and kinetic energies as well as on investigating how the choice of metal affects the outcome of an individual trajectory of the same given conditions. Throughout this work, special attention was given to the impact of PES on the outcome of the MD simulations in an effort to develop a better understanding of the critical role of the PES.

2. Simulation Details

In MD simulations, whether a full quantum or a classical treatment is employed, the PES is the most important part if accurate results are to be obtained. As such, we will discuss the PESs that were used in our simulations first and then provide the details of the MD simulations.

2.1. Potential Energy Surface (PES). In the MD simulations, we used the Sutton−Chen potential[31]

$$V = \epsilon\left[\frac{1}{2}\sum_{ij} v(r_{ij}) - c\sum_{i}\sqrt{\rho_i}\right] \tag{1}$$

where the pair potential is given by

$$v(r_{ij}) = \left(\frac{a}{r_{ij}}\right)^n \tag{2}$$

and the local electron density is given by

$$\rho_i = \sum_j \left(\frac{a}{r_{ij}}\right)^m \tag{3}$$

to describe the interactions among atoms in the system. In eqs 1−3, r_{ij} is the distance between two atoms, and ϵ, a, c, m, and n are parameters, which are constant for each metal but vary for different metals. These potential parameters were obtained by Sutton and Chen[31] for different metals by fitting the bulk properties of each metal. The Sutton−Chen parameters for iridium, platinum, gold, and silver are summarized in Table 1.

Our comparison of binding energies obtained using the Sutton−Chen potential with the parameters originally provided by Sutton and Chen with the DFT results[33−36] for several clusters of Ir, Pt, Au, and Ag showed that it was necessary to improve upon the Sutton−Chen parameters for clusters of those metals. In this work, we therefore obtained cluster optimized parameters for each metal. The details of this are discussed in section 3.1.

2.2. Molecular Dynamics (MD) Simulations. We studied the coalescence of an atom and a dimer by solving the equations of motion for position

$$\frac{dx_i}{dt} = \frac{p_{x_i}}{m_i}, \frac{dy_i}{dt} = \frac{p_{y_i}}{m_i}, \frac{dz_i}{dt} = \frac{p_{z_i}}{m_i} \tag{4}$$

and momentum

$$\frac{dp_{x_i}}{dt} = -\frac{\partial V}{\partial x_i}, \frac{dp_{y_i}}{dt} = -\frac{\partial V}{\partial y_i}, \frac{dp_{z_i}}{dt} = -\frac{\partial V}{\partial z_i} \tag{5}$$

based on the Cartesian coordinates x, y, and z for each atom, i, with $i = 1-3$, where m_i is the mass of the ith atom, and V is the PES described in section 2.1. The above set of coupled equations was solved using the sixth-order Runge−Kutta method.[37] Although other coordinate systems, such as the Jacobi coordinates, can be used here, we adopted Cartesian coordinates. Our intention of building the dynamics code is for studying systems consisting of more than 100 metal atoms where the Cartesian coordinates are the most practical choice.

The coalescence was studied as a single atom approached a dimer, first collinearly, i.e., the incident angle, θ, is zero. A schematic illustration of the system is provided in Figure 1. The starting position of each atom was specified as follows. The bond distance of the dimer equals the equilibrium dimer distance for each metal. The equilibrium distance was obtained by calculating the binding energy of the dimer

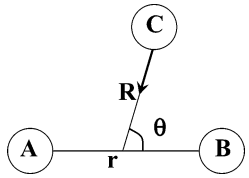


Figure 1. Schematic illustration of the system studied: a single metal atom C colliding with a metal dimer AB with A=B=C = Ir, Pt, Au, or Ag. *R*, *r*, and θ are in Jocobi coordinates. The arrow indicates the direction of the initial momentum of atom C.

as a function of distance using the Sutton−Chen potential. The initial distance between the center of mass of the dimer and the third atom was set to be 10 Å, a sufficient distance to ensure that the single atom and the dimer are in the asymptotic region. Initially, the dimer is in its ground rotational and vibrational states.

In order to study how the coalescence depends on kinetic energy, the initial kinetic energy provided to this collinear system was varied from 0.02 to 2.0 eV. The single atom has the entire kinetic energy, and the direction of its momentum is along the incident angle (see Figure 1). To investigate how the incident angle affects coalescence, the incident angle was varied by 15° from linear, $\theta = 0$, until the third atom was approaching the dimer from a perpendicular direction, $\theta = 90°$. Energies of 0.1 and 1.0 eV were tested for each of the different incident angles.

The positions and momentum of atoms were calculated at each time step of 0.01 fs for 10 ps or until the distance between any two atoms exceeded 10 Å. The products from most trajectories using the above stopping criterion are a dimer and an atom instead of trimer complexes, as will be shown in section 3.2. This indicates that the reaction time is less than 10 ps for the majority of the systems being studied. Furthermore, the focus of the current work is to investigate the impact of different PESs on the outcome of single coalescence processes. The choice of 10 ps is sufficient.

For testing our assumptions to be discussed in section 3.2, we ran some trajectories long enough so that the complex trimers dissociate. During the entirety of the simulations, the total energy and the total linear momentum of the system were conserved.

These MD simulations were carried out using the original Sutton−Chen parameters and again with the improved parameters. The final products were analyzed, and the results are discussed in section 3.2. The vibrational state of the product dimer, where present, was estimated using a harmonic oscillator treatment, in which the vibrational state, v, was calculated by

$$E_{\text{dimer}} = h\omega_e\left(v + \frac{1}{2}\right) \tag{6}$$

where h is the Planck's constant, $\omega_e = 1/2\pi\sqrt{k/\mu}$ with μ being the reduced mass of the product dimer and k being the force constant obtained from the dimer potential curve, and E_{dimer} is the vibrational energy of the product dimer. We note that at high vibrational states, the result will be approximate.

Table 2. Comparison of Binding Energy Obtained Using the Original Bulk and Cluster Optimized Parameters of Sutton−Chen Potential and DFT Calculations[a]

		binding energy (eV)			
		original PES	cluster PES	DFT and references	
Ir	dimer	−2.73	−5.06	−5.06	
	linear trimer	−0.15	−8.09	−9.63	ref 33
	triangle	−8.56	−9.46	−9.09	
Pt	dimer	−7.29	−4.03	−3.52	
	linear trimer	−11.41	−6.32	−6.54	ref 34
	triangle	−12.28	−6.79	−6.99	
Au	dimer	−4.46	−2.20	−2.34	
	linear trimer	−6.91	−3.42	−3.57	ref 35
	triangle	−7.43	−3.67	−3.57	
Ag	dimer	−2.29	−1.68	−1.80	
	linear trimer	−3.73	−2.73	−2.64	ref 36
	triangle	−3.85	−2.74	−2.64	

[a] The reference point of the energy is the asymptotic region of isolated atoms.

3. Results

The results of our study of the potential energy surfaces are presented in section 3.1. The molecular dynamics simulation results are presented and discussed in section 3.2.

3.1. New Parameters for the Sutton−Chen PESs. The quality of the PES is one of the most important factors in determining the accuracy of the results obtained from MD simulations, so we began our investigation here. Using the parameters published by Sutton and Chen,[31] we calculated the PES for each of the metals iridium, platinum, gold, and silver. Throughout this paper, we refer to this PES as the "original bulk PES" using the "original bulk parameters". Using this original PES, we calculated the binding energy of a dimer, linear trimer, triangular trimer, and 8- and 13-atom cluster for each of the metals and compared the results with our previous DFT findings for the binding energy of these clusters. The geometries chosen were the relaxed structures obtained by DFT. The DFT data used has been proven reliable compared to experimental and other theoretical methods through our previous work.[33−36] For example, the DFT binding energy of Ir_2, Pt_2, Au_2, and Ag_2 is −5.06 eV,[33] −3.52 eV,[34] −2.34 eV,[35] and −1.80 eV,[36] respectively, which agrees well with the experimental values, −4.30 eV for Ir_2,[38] −3.14 eV[39] and −3.66 eV[40] for Pt_2, −2.29 eV for Au_2,[41] and −1.65 eV for Ag_2.[42] In contrast, the original bulk Sutton−Chen potential gives a drastically different binding energy of −2.73 eV, −7.29 eV, −4.46 eV, and −2.29 eV for Ir_2, Pt_2, Au_2, and Ag_2, respectively. The huge discrepancy in the dimer binding energy between the experiment/DFT and the Sutton−Chen potential is not surprising, as the original Sutton−Chen potential was only fitted to the bulk properties.

The results of the binding energies from the original Sutton−Chen PES and the DFT calculations for the dimer, linear trimer, and triangular trimer are summarized in Table 2, as they are more relevant to the current work. The conclusions from the comparison on the 8- and 13-atom clusters of these four metals are similar to those on the dimer and trimers. As shown in Table 2, the binding energies calculated from the original PES are considerably different

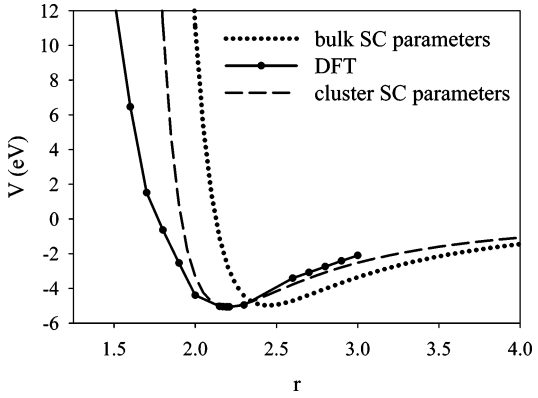


Figure 2. Comparison of the potential energy, *V* (in eV), vs the bond distance, *r* (in Å), for iridium dimers calculated with the original bulk and cluster optimized Sutton−Chen potentials and DFT.

from the DFT results. We use the binding energy as a measure of the stability of the clusters. As it is defined for these calculations, a lower binding energy indicates a more stable structure. Of greater importance is that the relative stability of the trimers does not correspond to the DFT predictions. For iridium, the binding energy predicted by DFT calculations is −9.63 eV for the linear trimer and −9.09 eV for the triangle. However, the binding energies calculated from the original PES are −0.15 eV and −8.56 eV for the linear trimer and triangle, respectively. This indicates the opposite order of stability. Discrepancies are also apparent for the other metals. The binding energies for platinum and gold are about 2 times lower than the DFT values, and the results for silver are similarly problematic.

It was necessary to improve upon the original parameters for these small metal clusters. In our fitting, we fixed *m* and *n* but optimized ϵ, *a*, and *c* so that the binding energies using the optimized parameters are close to the DFT results. Such a choice is somewhat arbitrary, but the optimized results show that the choice is reasonable. For the purpose of this work, we did not choose to use other fitting strategies. In the ongoing research efforts to fit the Sutton−Chen potential including large clusters, we will examine different strategies, such as allowing the parameters *m* and *n* to be optimized as well.

A nonlinear least-squares fitting procedure was used.[33] The parameters were optimized by fitting the DFT results for the dimers at various bond distances and two trimers, linear and triangular, for each metal, as it is directly relevant to the current work. Several sets of parameters were acceptable for each metal, but the optimal parameters are included in Table 1. These cluster optimized parameters were used to construct a new PES for each metal. We will refer to this case throughout this paper as the "cluster optimized PES" and "cluster optimized parameters". The binding energies were calculated for the DFT structure for the dimer at equilibrium distance, and two trimers using this cluster optimized PES. These results for the binding energies are also included in Table 2.

Figure 2 shows the potential curves as a function of dimer bond distance for the iridium dimer calculated by both the bulk and cluster potentials as well as DFT. The original potential gives a reasonable estimate of the global minimum energy for the dimer, at −4.99 eV, but gives very poor agreement with the equilibrium bond distance of 2.45 Å. The DFT results are −5.06 eV and 2.19 Å, respectively. This bulk potential yields a binding energy of −2.73 eV for the DFT structure, which is very different from the DFT binding energy. The cluster optimized potential provides excellent agreement with the DFT results in the minimum region of the curve. We emphasize this region from 2 to 3 Å because the DFT calculations are most reliable here. At bond distances less than 2 Å and longer than 3 Å, i.e. shorter and longer distances, the DFT results are less reliable. The details of the DFT work can be found in refs 33−36.

The new parameters provide very good agreement between the Sutton−Chen potential and DFT results for trimers and similar variations for the 3-atom isomers for each metal. The binding energies from the cluster optimized Sutton−Chen potential for all the metals are much closer to the energies obtained by DFT for all structures.

The binding energies reported in Table 2 were calculated for the DFT geometries. Our comparison of the global minima from the cluster Sutton−Chen potential with these fixed geometry results shows good agreement. For example, the binding energy for the iridium dimer exactly corresponds to the global minimum, and the trimer results agree within 0.3 eV. The equilibrium bond distances of 2.20 Å and 2.31 Å agree within 0.06 Å to the DFT values of 2.16 Å and 2.37 Å for the linear trimer and triangle, respectively. The global minima from the original Sutton−Chen potential occur at an overestimated bond distance of 2.50 Å and 2.53 Å for the linear and trianglar trimer, respectively.

The original and cluster optimized PESs were obtained and compared for each metal for the collinear interaction of three atoms, A, B, and C. As an example, the cluster and bulk PESs for iridium are shown in Figure 3. The potential energy in eV is plotted as a function of the AB distance and BC distance in Å. Overall topology of the PES remains the same. However, as illustrated by the MD simulations to be discussed below, the two PESs produce rather different results.

3.2. Coalescence of AB + C with A=B=C = Ir, Pt, Au, and Ag. We studied the coalescence of a 3-atom system for iridium, platinum, gold, and silver using MD simulations. The initial state of the 3-atom system was a single atom C colliding with a dimer AB at its ground rovibrational state (Figure 1). The MD simulations were performed using the original bulk and cluster optimized PESs described above. A total of 176 trajectories were obtained, and the products are summarized in Tables 3−5. The products from 80% of the total trajectories are a dimer and an atom, and they were formed long before the 10 ps maximum simulation time was reached. In addition to the dimer and atom as the major products, 34 triatomic complexes were still formed at the end of the simulation, i.e. 10 ps. Because the total energy of the system is conserved, these triatomic complexes formed during the coalescence may dissociate into an atom and a dimer if the MD simulation is run for more than 10 ps. To illustrate the possible dissociation of the triatomic complex still being formed at 10 ps, we carried out the MD simulation

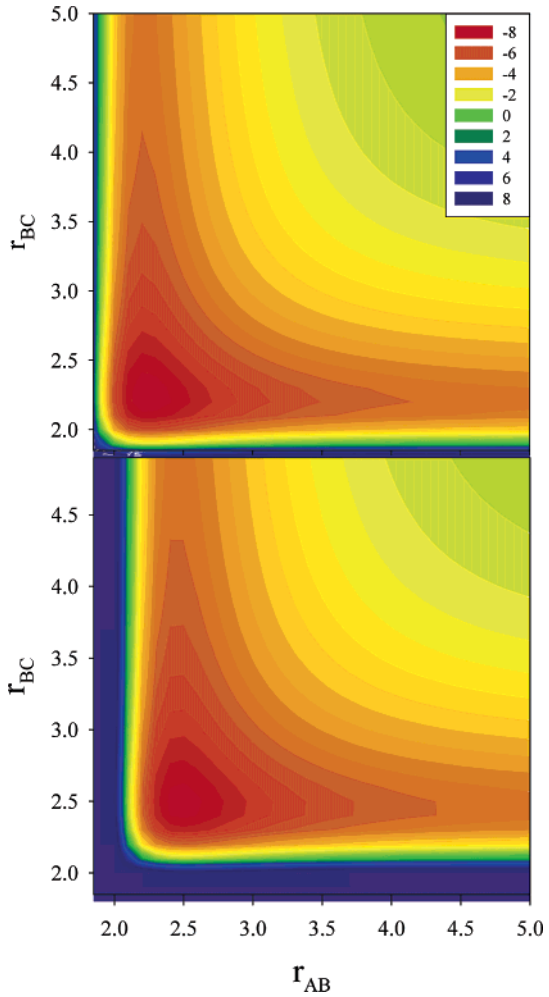


Figure 3. The cluster optimized (top) and original bulk (bottom) Sutton−Chen PES for the collinear AB + C reaction of Ir atoms. The legend shows the potential energy in eV. The interatomic distances, r_{AB} and r_{BC}, are given in angstroms (Å).

for the Ag_2 + Ag system with an initial kinetic energy of 0.1 eV and an incident angle of 45° for longer than 10 ps. The results show that an Ag triatomic complex was formed during the coalescence, and it remained as a complex at 10 ps. At 65.8 ps, the Ag triatomic complex dissociated into Ag_2 + Ag. The formation rate of triatomic complexes, the lifetime of the nascent triatomic complexes, and the energy transfer during a coalescence process are important and interesting issues that need to be addressed in order to fully understand the coalescence processes. Research in this direction is in progress. The work in progress also includes exploring the ways of extracting excess energy from the system so that real sintering processes can be simulated.

In order to investigate how the kinetic energy affects the coalescence of the AB + C system, the initial kinetic energy was varied from 0.02 to 2.0 eV, while the incident angle was kept at 0°. The results are summarized in Table 3. We first discuss the impact of the PES on the results of the collinear reaction dynamics. At lower energies, i.e., the kinetic energy less than 0.2 eV, the cluster PES resulted in more cases of elastic scattering for the lower energies of 0.02, 0.05, and 0.1 eV, as shown by the data in Table 3. In the energy range of 0.2−0.5 eV, the dynamics outcome for

Table 3. Final Product(s) for the Collinear AB + C Reaction as a Function of Initial Kinetic Energy, E_T, for Iridium (Black), Platinum (Red), Gold (Blue), and Silver (Green)[a]

E_T (eV)	Bulk Sutton-Chen parameters			Cluster Sutton-Chen parameters		
	AB+C	A+BC	A-B-C	AB+C	A+BC	A-B-C
0.02		Ir (v=0) Ag (v=0)	Pt (A-B-C) Au (A-B…C)*	Pt (v=0)	Ir (v=2) Ag (v=0)	Au (A-B-C)
0.05	Pt (v=0) Au (v=0)	Ir (v=0) Ag (v=0)		Pt (v=0) Ag (v=0)	Ir (v=1) Au (v=1)	
0.1	Pt (v=0) Au (v=0)	Ir (v=1) Ag (v=1)		Pt (v=2) Ag (v=0)	Ir (v=0)	Au (A-B-C)
0.2	Au (v=0)	Ir (v=0) Pt (v=2) Ag (v=5)			Ir (v=1) Pt (v=1) Ag (v=1)	Au (A-B-C)
0.3		Ir (v=0) Pt (v=0) Ag (v=0)	Au (A-B-C)		Ir (v=0) Pt (v=0) Au (v=1) Ag (v=0)	
0.4		Ir (v=0) Ag (v=0)	Pt (A…B-C)* Au (A-B-C)	Pt (v=3) Au (v=0) Ag (v=0)	Ir (v=0)	
0.5	Pt (v=0)	Ir (v=0) Ag (v=0)	Au (A-B-C)		Ir (v=0) Pt (v=0) Au (v=10) Ag (v=1)	
1.0		Ir (v=0) Pt (v=4) Au (v=0) Ag (v=0)			Ir (v=0) Pt (v=1) Au (v=10) Ag (v=0)	
1.5		Ir (v=0) Pt (v=0) Au (v=2) Ag (v=0)			Ir (v=2) Pt (v=0) Au (v=5) Ag (v=2)	
2.0		Ir (v=3) Pt (v=18) Au (v=10) Ag (v=0)			Ir (v=0) Pt (v=7) Au (v=5) Ag (v=0)	

[a] These results were obtained using the original bulk Sutton−Chen potential parameters as well as the optimized cluster parameters. The final vibrational state is included in the parentheses for cases where a dimer is one of the products. *A...B indicates an interatomic distance greater than 3.0 Å, so that these atoms do not form a real bond but rather a van der Waals complex.

three metals except Ir shows dependence on the PES but to a lesser extent compared to the lower energy range discussed above. At energies higher than 1 eV, the same products were formed, although the final vibrational state is still sensitive to the PES used. This indicates that the difference in PES shows less impact at the higher energies. In conclusion, the results in Table 3 show that the PES is critical in determining the outcome of a coalescence process, particularly at low-energy regime.

We now analyze the impact of kinetic energy on the reaction dynamics of a dimer colliding collinearly with an atom based on the MD simulations. The discussion of results in Table 3 is confined to those obtained using the cluster optimized parameters. For Ir, reactive scattering occurs with the product A + BC in every case. For Pt, Au, and Ag, the results in Table 3 show more variation in the products with different initial energies. The products obtained most frequently for Pt are again A + BC and elastic scattering occurring primarily at the lower energies from 0.02 to 0.1 and 0.4 eV. As for Ag, the products are the same as Au at kinetic energies higher than 0.2 eV. At 0.05−0.2 eV, the Ag system behaves similarly to the Pt system. At 0.02 eV, a reactive scattering occurs with the products of A + BC.

Au forms a trimer more frequently than elastic scattering and more often than any of the other metals. No trimer was formed for Ir, Pt, or Ag. This indicates that energy needs to be dissipated from the reaction complex in order to form a trimer such as using Ar atoms to absorb the energy, which is often used in the formation of transition-metal clusters.[2]

For collinear dynamics at kinetic energies larger than 0.5 eV, the results in Table 3 show that the products of the

Table 4. Final Product(s) for the AB + C Reaction with 1.0 eV of Initial Kinetic Energy as a Function of Incident Angle, θ, for Iridium (Black), Platinum (Red), Gold (Blue), and Silver (Green)[a]

Angle(°)	Bulk Sutton-Chen parameters				Cluster Sutton-Chen parameters			
	AB+C	A+BC	AC+B	ABC[‡], A-B-C, A-C-B, B-A-C	AB+C	A+BC	AC+B	ABC[‡], A-B-C, A-C-B, B-A-C
0		Ir (v=0) Pt (v=4) Au (v=0) Ag (v=0)				Ir (v=0) Pt (v=1) Au (v=10) Ag (v=0)		
15	Pt (v=2)	Ir (v=0) Au (v=3) Ag (v=10)				Ir (v=0) Pt (v=1) Ag (v=7)	Au (v=11)	
30	Ir (v=7) Pt (v=0)	Ag (v=16)		Au (A…B-C)*	Pt (v=2)	Ir (v=13) Au (v=2) Ag (v=0)		
45	Au (v=6) Ag (v=0)	Pt (v=7)	Ir (v=21)		Ir (v=0) Pt (v=3) Au (v=6)		Ag (v=0)	
60	Ir (v=9) Ag (v=0)		Au (v=13)	Pt (B-A-C)	Ir (v=5) Ag (v=10)		Pt (v=9)	Au (A-B-C)
75	Ir (v=4) Au (v=7) Ag (v=0)		Pt (v=1)		Pt (v=3)		Ir (v=3) Au (v=23) Ag (v=6)	
90	Ir (v=7) Ag (v=0)		Pt (v=1) Au (v=2)			Ir (v=4) Pt (v=4) Ag (v=7)		Au (A-C-B)

[a] These results were obtained using the bulk Sutton−Chen potential parameters as well as the optimized cluster parameters. The final vibrational state is included for dimer products, when present. ABC indicates a trimer of triangle shape. *A...B indicates an interatomic distance greater than 3.0 Å, so that these atoms do not form a real bond but rather a van der Waals complex.

Table 5. Final Product(s) for the AB + C Reaction with 0.1 eV of Initial Kinetic Energy as a Function of Incident Angle, θ, for Iridium (Black), Platinum (Red), Gold (Blue), and Silver (Green)[a]

Angle(°)	Bulk Sutton-Chen parameters			Cluster Sutton-Chen parameters			
	AB+C	A+BC	ABC[‡], A-B-C, A-C-B, B-A-C	AB+C	A+BC	AC+B	ABC[‡], A-B-C, A-C-B, B-A-C
0	Ir (v=0) Pt (v=0) Au (v=0)	Ag (v=1)		Ir (v=0) Pt (v=2) Ag (v=0)			Au (A-B-C)
15	Ir (v=2) Au (v=0) Ag (v=0)		Pt (A-B-C)	Ir (v=0) Au (v=1)	Pt (v=1) Ag (v=0)		
30	Au (v=1) Ag (v=0)		Ir (A-B-C) Pt (B-A-C)		Pt (v=3) Ag (v=5)		Ir (B-A-C) Au (A-B-C)
45		Ag (v=0)	Ir (A-C…B)* Pt (A-B-C) Au (A-B-C)	Pt (v=1)	Au (v=1)		Ir (A-B-C) Ag (A…B-C)*
60	Ir (v=1)		Pt (B-A-C) Au (A-B…C)* Ag (ABC)	Pt (v=1)	Ir (v=1) Ag (v=1)		Au (A-C-B)
75		Ag (v=0)	Ir (B-A-C) Pt (A-B-C) Au (A-B…C)*	Ag (v=0)	Pt (v=2)	Ir (v=2)	Au (ABC)
90	Ir (v=31) Pt (v=0) Ag (v=11)		Au (A…B-C)*	Pt (v=1) Ag (v=0)	Ir (v=6)		Au (A-B-C)

[a] These results were obtained using the bulk Sutton−Chen potential parameters as well as the optimized cluster parameters. The final vibrational state is included for dimer products, when present. ABC indicates a trimer of triangle shape. *A...B indicates an interatomic distance greater than 3.0 Å, so that these atoms do not form a real bond but rather a van der Waals complex.

dynamics are not sensitive to the type of metals. This indicates that the details of PESs are less critical at this energy range. This can be well understood by the fact that the high-energy possessed by the system allows the dynamics to not be constrained within the valley of the PES as the low-energy dynamics would. Therefore, the subtle difference between PESs is no longer critical at high energies.

To gain a better understanding of the coalescence mechanism, the AB + C reaction was further studied for each metal at the energies of 0.1 and 1.0 eV by varying the incident angle from 0° to 90° as discussed in section 2.2. These calculations were performed using the bulk and cluster PESs. The results are summarized in Table 4 for 1.0 eV and Table 5 for 0.1 eV.

Before discussing the coalescence mechanism, we summarize the sensitivity of reaction dynamics on the PES for different metals. The products for gold are quite varied between the two PESs. Pt and Ag are also more sensitive to the PES. At the lower energy of 0.1 eV, the MD results are

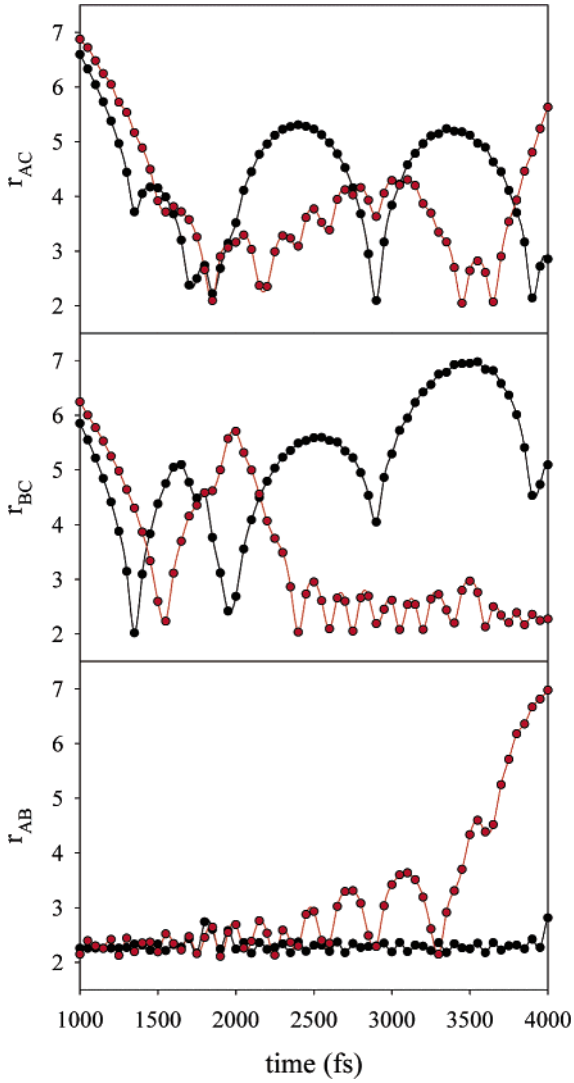


Figure 4. Interatomic distances (in Å) as a function of time (fs) for the products of the AB + C reaction of Pt atoms with initial kinetic energy of 0.1 eV and an incident angle of 75°. Results are shown for the bulk (black) and cluster (red) PESs for atoms (a) AC, (b) BC, and (c) AB.

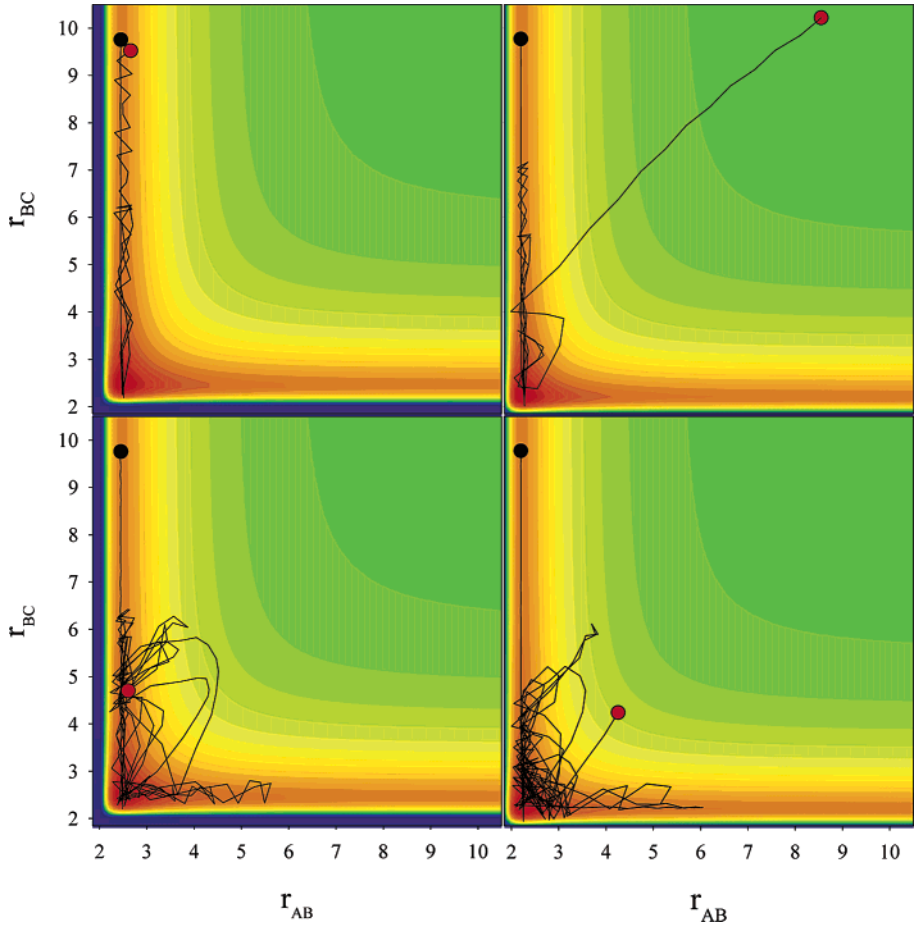


Figure 5. Trajectories (black curves) for the AB + C reaction of Ir atoms with initial kinetic energy of 1.0 eV (top figures) and 0.1 eV (bottom figures) obtained with the original bulk (left figures) and cluster optimized (right figures) PESs. The interatomic distances, r_{AB} and r_{BC}, are given in Å. The incident angle is 45° in all cases. The black and red circles indicate the starting and ending distances (in Å), respectively. The contour plots, which are the same as shown in Figure 3, serve as guidance only as they are the collinear case of the corresponding PESs on which these trajectories were obtained.

very sensitive to the PES, as shown in Table 5. Different products were obtained with the bulk and cluster PESs for all incident angles. In fact, the products are completely different for all metals at the incident angles of 30°, 45°, and 60°.

In order to further investigate the differences between the bulk and cluster PESs, the interatomic distances were plotted vs time for the case of Pt_2 + Pt with initial kinetic energy of 0.1 eV and an incident angle of 75°. The r_{AC}, r_{BC}, and r_{AB} distances are shown in Figure 4 (parts a−c, respectively) over the time of 1000−4000 fs. At the beginning of the calculation, very little difference is seen using the bulk and cluster PESs. However, as the time progresses, the differences become more pronounced. By 4000 fs, different products are formed with the different PESs.

Figure 5 shows the trajectories for the Ir_2 + Ir reaction with an incident angle of 75° and initial kinetic energy of 1.0 eV (top) and 0.1 eV (bottom) obtained with the original (left) and optimized (right) PESs. We note that the contour plots of PES illustrated in Figure 5 correspond to the collinear case. These contours are for guidance only as they are not the actual PESs that the trajectories were formed from. For 1.0 eV, the Sutton−Chen PES with original parameters predicts an AB dimer remains, while with the cluster parameters an AC dimer is formed. The difference between the left and right black curves in the top figures shows clearly that the trajectory that these atoms follow is impacted greatly by the PES. At the lower energy of 0.1 eV, a nearly triangular structure of Ir_3 is formed with the original PES, while a dimer is the product with the cluster parameters. In summary, the results discussed above demonstrate that the accuracy of PES is essential in order to obtain meaningful results for coalescence.

As in our discussion of the impact of kinetic energy to coalescence, here we confine our discussion of the incident angle effect to the MD results based on the cluster optimized PES, shown in Tables 4 and 5. At the higher energy of 1.0 eV (Table 4), the collision results in reactive scattering in all cases at the incident angles of 0° and 15°. As the incident angle is increased toward perpendicular, the collision becomes elastic in many cases. These results show that the incident angle plays a key role in the coalescence process.

At the lower energy of 0.1 eV (Table 5), the collision for Ir systems becomes increasingly more reactive and complex with the increase of incident angle. A trimer is still formed at the end of our simulation at incident angles of 30° and 45°. Many Au trimers were also still formed at the end of MD simulations. These results further illustrated that coalescence mechanism is very different at low kinetic energies for different metals, while it is less so at high kinetic energies.

Finally, we mention the effect of metal on the coalescence. The results in Tables 3−5 have shown that more similarities exist between Ir and Ag and between Pt and Au. This may be due to the similarities of the Sutton−Chen parameters between Ir and Ag and between Pt and Au.

4. Conclusions

We compared the binding energy obtained from the original Sutton−Chen potential to our DFT results for the dimers, trimers, and 8- and 13-atom clusters of iridium, platinum, gold, and silver. It was found that the original Sutton−Chen potential was not accurate for describing small metal clusters. This can be understood by the fact that the Sutton−Chen potential parameters were optimized to the bulk properties for each of the metals studied. In this work, we optimized these parameters to our DFT results for the small clusters and constructed a new PES for each metal.

Molecular dynamics simulations were performed using the bulk and cluster PESs to study the coalescence of a 3-atom system, AB + C. The reaction AB + C was investigated with various incident angles and initial kinetic energy. The MD results demonstrated that the accuracy of a PES is critical to the outcome of the dynamics simulations, especially at lower energies, for all four metals. The MD results also show that the products of coalescence were greatly impacted by both the incident angle and the energy. The choice of metal will also influence the outcome of coalescence, though more similarities were observed between Ir and Ag and between Pt and Au.

Acknowledgment. We gratefully acknowledge the support from the American Chemical Society Petroleum Research Fund under Grant No. 41572-G5 and the Materials Technology Center of Southern Illinois University at Carbondale.

References

(1) Campbell, C. T.; Parker, S. C.; Starr, D. E. *Science* **2002**, *298*, 811.

(2) Baletto, F.; Ferrando, R. *Rev. Mod. Phys.* **2005**, *77*, 371.

(3) Lisiecki, I. *J. Phys. Chem. B* **2005**, *109*, 12231.

(4) Sanchez, A.; Abbet, S.; Heiz, U.; Schneider, W.-D.; Hakkinen, H.; Barnett, R. U.; Landman, U. *J. Phys. Chem. A* **1999**, *103*, 9583.

(5) Mason, T. G. Lin, M. Y. *J. Chem. Phys.* **2003**, *119*, 565.

(6) Mitchell, C. E. J.; Howard, A.; Carney, M.; Egdell, R. G. *Surf. Sci.* **2001**, *490*, 196.

(7) Xu, L. J.; Henkelman, G.; Campbell, C. T.; Jonsson, H. *Phys. Rev. Lett.* **2005**, *95*, 146103.

(8) Zhu, H. H.; Lu, L.; Fuh, J. Y. H. *J. Mater. Process. Technol.* **2003**, *140*, 314.

(9) Zhang, Y. M.; Chen, Y. W.; Li, P. J.; Male, A. T. *J. Mater. Process. Technol.* **2003**, *135*, 347.

(10) Teixeira, A.; Giudici, R. In *Catalyst Deactivation 2001, Proceedings*; 2001; Vol. 139, p 495.

(11) Smejkal, Q.; Linke, D.; Bentrup, U.; Pohl, M. M.; Berndt, H.; Baerns, M.; Bruckner, A. *Appl. Catal,. A* **2004**, *268*, 67.

(12) Tang, Y.; Loh, H. T.; Wong, Y. S.; Fuh, J. Y. H.; Lu, L.; Wang, X. *J. Mater. Process. Technol.* **2003**, *140*, 368.

(13) Santra, A. K.; Goodman, D. W. *J. Phys.: Condens. Matter* **2003**, *15*, R31.

(14) Ojeda, M.; Rojas, S.; Garcia-Garcia, F. J.; Granados, M. L.; Terreros, P.; Fierro, J. L. G. *Catal. Commun.* **2004**, *5*, 703.

(15) Minay, E. J.; Boccaccini, A. R.; Veronesi, P.; Cannillo, V.; Leonelli, C. *Adv. Appl. Ceram.* **2005**, *104*, 49.

(16) Kwon, Y. S.; Dudina, D. V.; Korchagin, M. A.; Lomovsky, O. I. *J. Mater. Sci.* **2004**, *39*, 5325.

(17) Kim, B. K.; Ha, G. H.; Lee, D. W. *J. Mater. Process. Technol.* **1997**, *63*, 317.

(18) Hsieh, Y. Z.; Chen, J. F.; Lin, S. T. *J. Mater. Sci.* **2000**, *35*, 5383.

(19) Greco, A.; Licciulli, A.; Maffezzoli, A. *J. Mater. Sci.* **2001**, *36*, 99.

(20) Heinrichs, B.; Noville, F.; Pirard, J. P. *J. Catal.* **1997**, *170*, 366.

(21) Duck, J.; Niebling, F.; Neesse, T.; Otto, A. *Powder Technol.* **2004**, *145*, 62.

(22) Barbero, J.; Pena, M. A.; Campos-Martin, J. M.; Fierro, J. L. G.; Arias, P. L. *Catal. Lett.* **2003**, *87*, 211.

(23) Banhart, J. *Prog. Mater. Sci.* **2001**, *46*, 559.

(24) Kwon, Y. S.; Savitskii, A. *J. Mater. Synth. Process.* **2001**, *9*, 299.

(25) Tian, D.-X.; Guo, X.-Y. *Comput. Mater. Sci.* **2005**, *34*, 14.

(26) Hendy, S.; Brown, S. A.; Hyslop, M. *Phys. Rev. B* **2003**, *68*, 241403.

(27) Finnis, M. W.; Sinclair, J. E. *Philos. Mag. A* **1984**, *50*, 45.

(28) Daw, M. S.; Baskes, M. I. *Phys. Rev. B* **1984**, *29*, 1285.

(29) Mrovec, M.; Nguyen-Manh, D.; Pettifor, D. G.; Vitek, V. *Phys. Rev. B* **2004**, *69*, 094115.

(30) Murrell, J. N.; Mottram, R. E. *Mol. Phys.* **1990**, *62*, 571.

(31) Sutton, A. P.; Chen, J. *Philos. Mag. Lett.* **1990**, *64*, 139.

(32) Doye, P. K.; Wales, D. J. *New J. Chem.* **1998**, 733.

(33) Pawluk, T.; Hirata, Y.; Wang, L. *J. Phys. Chem. B* **2005**, *109*, 20817.

(34) Xiao, L.; Wang, L. *J. Phys. Chem. A* **2004**, *108*, 8605.

(35) Xiao, L.; Wang, L. *Chem. Phys. Lett.* **2004**, *392*, 452. Xiao, L.; Tollberg, B.; Hu, X.; Wang, L. *J. Chem. Phys.* **2006**, *124*, 114309.

(36) Yukna, J.; Tang, H.; Wang, L. *J. Chem. Phys.* Submitted for publication.

(37) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes [Fortran Version]*; Cambridge University Press: 1990.

(38) Miedema, A. R.; Gingerich, K. A. *J. Phys. B: At., Mol. Opt. Phys.* **1979**, *12*, 2081.

(39) Taylor, S.; Lemire, G. W.; Hamrick, Y. M.; Fu, Z.; Morse, M. D. *J. Chem. Phys.* **1988**, *89*, 5517.

(40) Gupta, S. K.; Nappi, B. M.; Gingerich, K. A. *Inorg. Chem.* **1981**, *20*, 966.

(41) Bishea, G. A.; Morse, M. D. *J. Chem. Phys.* **1991**, *95*, 5646.

(42) Morse, M. D. *Chem. Rev.* **1986**, *86*, 1049.

CT600257P

JCTC Journal of Chemical Theory and Computation

Atom-Centered Density Matrix Propagation Calculations on the Methyl Transfer from CH_3Cl to NH_3: Gas-Phase and Continuum-Solvated Trajectories

Ashley N. Jay,† Kelly A. Daniel, and Eric V. Patterson*

Truman State University, Division of Science, 100 E. Normal St., Kirksville, Missouri 63501

Received September 8, 2006

Abstract: Atom-centered density matrix propagation (ADMP) calculations have been carried out to determine gas-phase and continuum-solvated (aqueous) trajectories for the Menshutkin reaction of methyl chloride with ammonia. The gas-phase trajectories reveal an exit channel that has not been previously reported. The aqueous trajectories give the expected results, indicating that solvated ADMP trajectories may be successfully computed using implicit solvation models. The solvated trajectories demonstrate the same stability and convergence qualities as the gas-phase trajectories.

Introduction

The Born−Oppenheimer approximation,[1] wherein the nuclear degrees of freedom are separated from the electronic degrees of freedom in the solution of the Schrödinger equation, has long been a centerpiece of quantum chemical calculations. While modern quantum chemical methods, particularly those employing density functional theory, efficiently provide accurate geometries and electronic energies for stationary points along a potential energy surface, the Born−Oppenheimer approximation eliminates all information about nuclear motion and its effect upon the potential energy surface. Thermodynamic parameters determined from harmonic vibrational analysis may be used to correct the electronic energies to enthalpies or free energies at a given temperature, yet the potential energy surface remains static. While the Born−Oppenheimer approximation has proved to be quite satisfactory, there are instances when it is necessary to account for nuclear motion and the effect it has on the potential energy surface.

In many circumstances, nuclear motion may be adequately described through classical mechanics. Fully classical methods, wherein the potential energy of the system is determined by a classical force field, have been used with great success.[2−7] For processes that require a quantum description of the potential energy, such as bond formation or dissociation, methods have been developed that combine a quantum chemical potential energy with classical propagation of the nuclear degrees of freedom, often collectively referred to as ab initio molecular dynamics (AIMD),[8] including the very popular Car−Parrinello methods,[9−11] Born−Oppenheimer MD,[12] and atom-centered density matrix propagation (ADMP).[13−15] While classical MD and AIMD methods are applicable to a wide variety of chemical problems, they are not appropriate for every case. If quantum effects such as tunneling or electronic excitations are known to be important, then the nuclei can no longer be treated as classical particles.[16−18] Of course, with the progression from fully classical to fully quantal approaches, the system size for efficient simulations dwindles from tens of thousands of atoms to two or three.

Given that most chemical reactions do not occur in the gas phase, it is important to account for solvation effects in order to accurately model a chemical process. Two limiting methods for applying solvation corrections exist: explicit solvation and implicit solvation. In the former, sufficient solvent molecules are added to the system to reproduce the effect of bulk solvation. Because of the size of the resulting supermolecular cluster, bulk explicit solvation is largely limited to classical MD. In principle, it is possible to use the bare minimum number of solvent molecules to reproduce

* Corresponding author e-mail: epatters@truman.edu.

† Former Truman State University undergraduate student. Current address: Department of Chemistry, University of Minnesota, Minneapolis.

10.1021/ct6002803 CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/24/2007

important stabilizing interactions, and this type of "microsolvation" has seen considerable use in MD simulations. Of course, care must be taken in the placement of the solvent molecules so that a balanced description of the entire potential energy surface is obtained.[7]

Implicit solvation models mimic bulk solvation effects through the use of a dielectric continuum or some similar means.[19] Continuum methods have seen wide use in the field of computational chemistry.[7,20] The advantage of continuum solvation is that it adds modestly to the cost of the computation. The primary disadvantage is that there is no way to recover structural information about the solvent molecules, so any structural role that the solvent plays in the course of a chemical reaction will remain undiscovered. A more subtle difficulty lies in the creation of the solute cavity within the continuum. Depending on how the cavity is constructed, numerical instabilities may arise for processes such as bond dissociation that significantly alter the shape of the cavity.[7]

A number of methods have been developed that seek to combine implicit and explicit solvation models to take advantage of the strengths of each. For example, it is possible to enclose a multilayered supermolecular cluster within a continuum model.[19] While this may be a reasonable thing to do, using the two theoretical approaches simultaneously combines not only their strengths but also their weaknesses. Multilayer methods, such as the popular ONIOM formalism,[21] reduce the computational time required to perform an explicitly solvated calculation by modeling the solvent at a lower level of theory than the solute. The combination of ONIOM with ADMP has been reported in the literature using a development version of Gaussian 03.[22]

Among the more promising recent developments in AIMD is the ADMP method.[13−15] The primary advantage of ADMP is that it provides results similar in accuracy with Born−Oppenheimer MD[12] but at a substantially reduced computational cost. Recent applications of ADMP demonstrate the accuracy and efficiency of the method.[22−31] Given the promise of ADMP, it is worthwhile to explore whether accurate solvated trajectories can be obtained simply by applying a standard continuum solvation model. If so, the combined efficiencies of ADMP and continuum solvation could make it possible to determine solvated dynamic trajectories for reasonably large systems where structural data regarding the solvent are not essential. In fact, a recent study demonstrated the reliability of ADMP coupled with the conductor-like polarizable continuum model (CPCM) in a mixed explicit/implicit solvent system.[32] Even greater efficiency and flexibility could be obtained if no explicit solvent molecules were used. Therefore, we have chosen to explore ADMP in conjunction with the integral equation formalism−polarizable continuum model (IEF−PCM) solvation model[33] in the absence of any explicit solvent. Furthermore, both have been used as implemented in the standard binary release of Gaussian 03,[34] meaning no code modifications are necessary.

To test the general applicability of the ADMP/IEF−PCM combination, trajectories for the S_N2 Menschutkin reaction[35,36] of ammonia with methyl chloride have been determined in both the gas phase and in aqueous solution. This reaction was chosen since it is well-known to have exceptionally different gas-phase and solution-phase potential energy surfaces. Accordingly, the ADMP/IEF−PCM combination should produce dramatically different trajectories than those determined in the gas phase. Menshutkin reactions in general have been extensively studied experimentally,[36] and the specific reaction explored herein has been the subject of a number of theoretical studies.[37−48] Early simulations of the reaction pathway employed statistical trajectory[37] and Monte Carlo calculations[38] in an explicit water solvent; however a symmetry constraint was employed to keep the N, C, and Cl atoms collinear. A study employing second-order reaction path following in conjunction with the implicit generalized conductor-like screening model water model examined the trajectory with no symmetry constraint.[44] Monte Carlo simulations in implicit solvent have also been performed,[48] as have continuum-solvated[46] and microsolvated[45] variational transition state calculations, which can account for quantum-mechanical tunneling. The consensus of this previous work is that the gas-phase reaction has a barrier of approximately 30 kcal/mol and an overall energy change of approximately 117 kcal/mol, in excellent agreement with the experimental value of 111 ± 5 kcal/mol.[49] In aqueous solution, the consensus barrier height is approximately 25 kcal/mol, while the overall reaction is now exothermic by approximately 20 kcal/mol. There is some variability in these latter numbers, due in no small part to the various approximations employed in the different studies. The reaction with methyl bromide in place of methyl chloride has also been examined computationally,[50,51] with similar results.

In the current work, we report ADMP trajectory calculations for the Menschutkin reaction of ammonia with methyl chloride in both the gas phase and an implicit aqueous solution. Previously, ADMP has been shown to work well with CPCM for studying the solvation energy of chloride anions in aqueous solution.[32] We now show that continuum-solvated ADMP trajectories are equally suited for the study of nucleophilic reactions in aqueous solution. While the primary purpose of the current work is to explore the stability of continuum-solvated ADMP trajectories for computing nucleophilic reaction pathways, the data have revealed a novel product channel in the gas phase. The present results demonstrate that ADMP may be reliably paired solely with a continuum solvation model for the exploration of solvated reaction dynamics. The combination of these two efficient methods holds great promise for the study of larger and more complex systems than the simple test case presented herein.

Computational Methods

Stationary-point geometries were determined using the MPW1K hybrid density functional[52] and the MIDIY+ basis set,[53] both due to the Truhlar group. This combination was chosen as it has been shown to provide excellent estimates for forward and reverse barrier heights while still giving good results for overall energy changes during the course of a chemical reaction.[53] Furthermore, MPW1K has performed well for capturing nonbonded interactions.[54] Aqueous-phase

geometries were optimized using the IEF−PCM of Tomasi.[33] The all-atom UFF topological model was used in place of the default united-atom UA0 topological model in order to avoid difficulties with proton dissociation. All other default parameters for an aqueous solvent were employed. Harmonic frequencies were determined for all stationary points to ascertain the correct number of imaginary eigenvalues (i.e., 0 for minima and 1 for transition states). Intrinsic reaction coordinate calculations[55] were employed to confirm that each transition state connects to the minima as described below.

Given that noncovalent complexes figure heavily in the potential energy surface for this reaction (see below), the quality of the MPW1K results was checked against results obtained from second-order Møller−Plesset theory.[56] All stationary points were reoptimized at the MP2/MIDIY+ level of theory, both in the gas phase and in aqueous solution as described above. In every case, the results from the MPW1K calculations compare favorably with the MP2 results, yielding relative energies of the stationary points that differ by less than 1 kcal/mol in most cases (see the Supporting Information). Thus, we have confidence that the MPW1K density functional captures noncovalent interactions with accuracy similar to that of MP2, and trajectories computed at the MPW1K/MIDIY+ level should reliably sample the potential energy surfaces of the gas-phase and solvated reactions.

ADMP trajectories[13−15] were initiated from both the gas-phase and solution-phase transition states for backside nucleophilic attack. The MPW1K/MIDIY+ level of theory was employed for the ADMP calculations as described above for the stationary-point calculations. A time step of 0.25 fs was used for the majority of the trajectories, while the default time step of 0.1 fs was required in a few cases. The default fictitious electron mass of 0.1 amu was used throughout. A thermostat was employed to maintain a constant temperature of 298 K.[57,58] A total of 100 random trajectories were determined in each phase. Each trajectory was integrated for a minimum of 200 fs, with more data collected as necessary (see below). A smaller number of trajectories was initiated from other reactant complexes.

All calculations were performed with the standard binary release of G03M.[34]

Results and Discussion

Given that no zero-point or thermal corrections are applied during the determination of an ADMP trajectory, all discussion will be limited to relative self-consistent field energies with no further correction. Standard corrections are provided in the Supporting Information.

Stationary Points. Selected geometries and relative energies are presented in Figure 1. The complete set of stationary-point data is available in the Supporting Information.

As is typical of S_N2 reactions in the gas phase, separated reactants initially collapse to an ion-dipole complex that is 2.3 kcal/mol lower in energy than the reactants. This proceeds to a transition state with a relative energy of 32.0 kcal/mol, which continues to an ion-dipole product complex that lies 30.6 kcal/mol above isolated starting materials. The isolated ionic products are found to be 117.0 kcal/mol higher in energy than the reactants, indicating an impossibly unfavorable reaction in the gas phase. These results are all in substantial agreement with previous experimental[49] and theoretical[37−48] studies on this system. A much lower-energy hydrogen-bonded product complex, which has not been previously reported, is found to be 9.7 kcal/mol below starting materials. In this complex, the chlorine and amine share the acidic proton in a nearly equitable fashion (Figure 1). Of course, to reach this minimum, the chloride has to migrate a significant distance from its location in the ion-dipole product complex, implying that the minimum energy reaction pathway may not involve complete dissociation to the ionic products. Complete abstraction of this proton by chloride would lead to a charge-neutral product pair (methylamine plus hydrogen chloride), which is 5.6 kcal/mol higher in energy than the separated reactants.

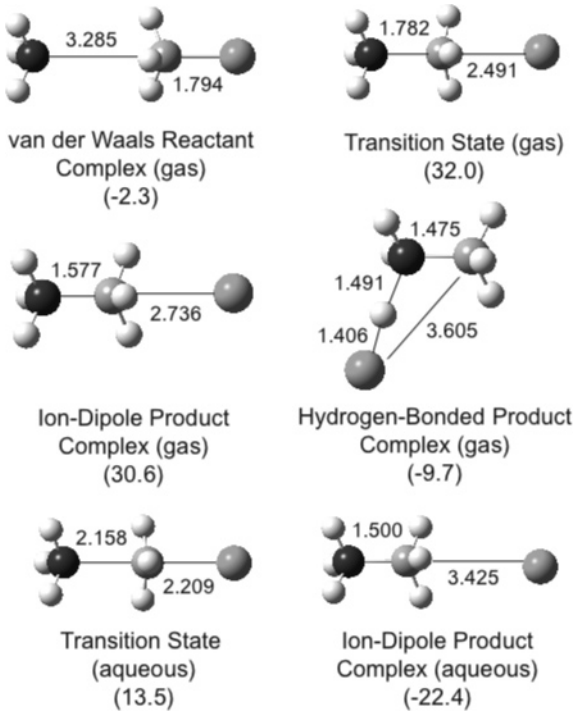


Figure 1. Stationary point geometries and important bond lengths (Å) for selected species considered in this study. Electronic energies relative to separated reactants are given in parenthesis (kcal/mol).

The aqueous-phase relative energies are also in good agreement with previously published work.[37,38,40−42,44−48] The transition state for attack is found 13.5 kcal/mol above isolated reactants and leads to an ion-dipole product complex 22.4 kcal/mol below the reactants (no van der Waals reactant complex could be located in the aqueous phase). Separated ionic products are nearly isoenergetic with the ion-dipole product complex, while the charge-neutral product pair is 6.1 kcal/mol higher in energy than the reactants. No aqueous-phase stationary point corresponding to the hydrogen-bonded complex could be located. Thus, the aqueous reaction closely resembles a textbook example of an exothermic S_N2 mechanism, and the ionic products are predicted to be quite favorable.

It is instructive to examine the differences between corresponding gas-phase and solution-phase structures. It has been suggested that, in general, the differences in geometry

Table 1. Important Bond Distances and Total Energies of Selected Gas-Phase and Aqueous-Phase Geometries

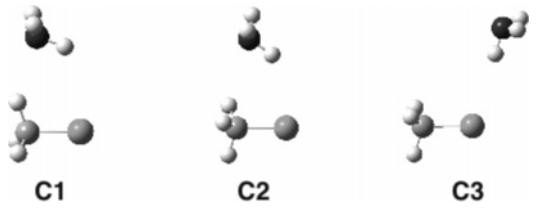
	bond distance[a]		PCM E(SCF)[b] (ΔG_{solv})[c]	
species	gas phase	aqueous phase	gas-phase geometry	aqueous-phase geometry
		Methyl Chloride		
C−Cl	1.788	1.800	−497.77868 (0.4)	−497.77874 (0.4)
		Ion-Dipole Reactant Complex		
C−Cl	1.794	N/A	−554.02853 (2.1)	N/A
C−N	3.285	N/A		
		TS		
C−Cl	2.491	2.209	−554.03120 (−35.0)	−554.00889 (−11.1)
C−N	1.782	2.158		
		Ion-Dipole Product Complex		
C−Cl	2.736	3.425	−554.05526 (−48.7)	−554.06601 (−62.7)
C−N	1.577	1.500		
		Hydrogen-Bonded Product Complex		
C−N	1.475	N/A	−554.05037 (−5.3)	N/A
N−H	1.491	N/A		
H−Cl	1.406	N/A		
		Methylammonium		
C−N	1.512	1.497	−95.78453 (−65.8)	−95.78470 (−66.1)
		Methylamine		
C−N	1.466	1.473	−95.32144 (−0.2)	−95.32163 (−0.4)

[a] Units of angstroms. [b] Units of hartrees. [c] Units of kcal/mol.

between the gas phase and solution phase are insignificant enough that applying solvation corrections to gas-phase geometries is sufficient and that the extra computational expense associated with determining solution-phase geometries would be time wasted. Geometric data is provided in Figure 1 and summarized in Table 1, which also provides a comparison between the self-consistent field (SCF) and free energies of solvation from single-point IEF−PCM calculations on the gas-phase structures to the corresponding energies of the solution-phase structures.

An examination of Table 1 reveals that the bond distances for the minima vary by less than 1% between the gas-phase and solution-phase structures for minimum-energy species. Similarly, the IEF−PCM SCF energies agree within 0.1 kcal/mol for these same species. However, there are dramatic differences in the transition state structures, where the bond distances differ by 11% for the C−Cl bond and 17% for the C−N bond and the energies differ by 14 kcal/mol. The gas-phase transition state structure more closely resembles products than does the solution-phase transition state. This trend has been observed previously for this reaction[37,38,41,42,44−46,48] and is as predicted by the Hammond Postulate.[59] Clearly, then, if the single-point solvation corrections on the gas-phase transition state structure had been used to predict the barrier for this particular reaction, the barrier would have been underestimated by a significant margin. It would, in fact, be estimated to have a barrier of −0.7 kcal/mol (the IEF−PCM SCF energy for the gas-phase ammonia structure is −56.25144 hartrees). The same general conclusion is reached if the free energies of solvation, which are given in parentheses under the SCF energies in Table 1, are examined. A similar trend is seen for the initial ion-dipole product complex, except that the aqueous geometry is now more highly solvated than the gas-phase geometry. While the reaction under study is a severe test due to the ionic nature of the products, it is clear that a good degree of caution must be exercised when estimating solvated surfaces using gas-phase transition state structures.

Chart 1



Alternate modes of nucleophilic attack were considered. A series of relaxed potential energy surface scans was conducted in both the gas-phase and solution, fixing either the N−C distance or the N−Cl distance while systematically altering the N−C−Cl angle. In this way, concentric potential energy plots were constructed at N−C and N−Cl distances ranging from 2.0 to 4.5 Å at 0.5 Å steps. These scans led to the discovery of additional noncovlant reaction complex stationary points, shown in Chart 1 (details are provided in the Supporting Information). In the gas phase, **C1** was found to be a minimum while **C2** is a transition state for methyl rotation leading to **C1**. In the aqueous phase, **C1** is found to be a third-order saddle point while **C3** is a second-order saddle point. These imaginary modes correspond to methyl rotations, rotations of the ammonia moiety relative to the substrate, or slipping of the nucleophile along the van der Waals contact radius. Complex **C1** is not a stationary point in aqueous solvent, while **C3** is not a stationary point in the gas phase. In the gas phase, complexes **C1** and **C2** are essentially isoenergetic with the hydrogen-bonded reactant complex discussed above. In the aqueous phase, complexes **C1** and **C3** are essentially isoenergetic with separated reactants. Thus, there is predicted to be a very flat potential energy profile for the nucleophile orbiting the substrate in both phases, and attack could potentially initiate from any direction.

Accordingly, several attempts were made to locate additional attack transition states for approach of the nucleophile. A relaxed potential energy surface scan starting with **C1** and forcing the N−C bond distance to become shorter did eventually lead to the formation of the hydrogen-bonded product complex shown in Figure 1. However, the activation barrier along this pathway was estimated to be greater than 50 kcal/mol. Furthermore, every attempt to optimize a transition state along this pathway failed, with the optimizations reverting to complex **C1**. Similar calculations beginning from **C2** and **C3** yield the same result. Likewise, no attack transition states could be located from any of the three complexes when shrinking the N−Cl distance. Therefore, the only confirmed attack transition states are the traditional S_N2 backside attack transition states discussed above (Figure 1). It appears clear, then, that stationary-point calculations predict only a single entrance channel for this reaction.

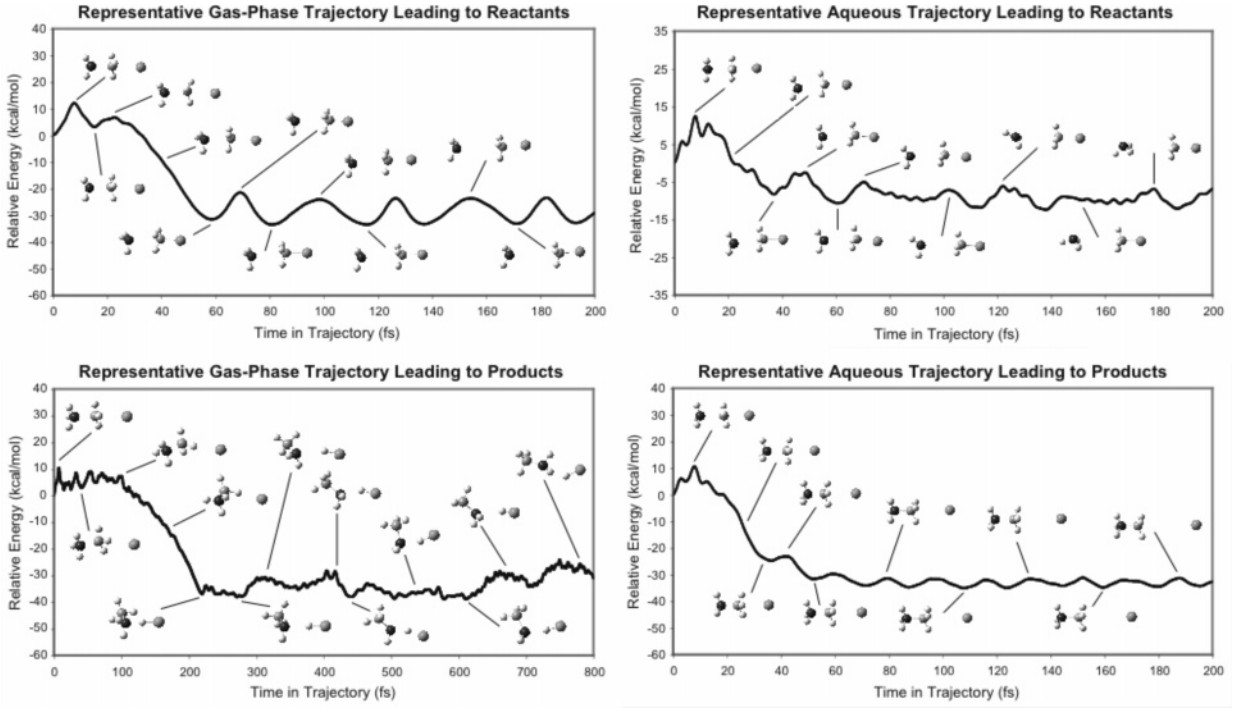


Figure 2. Representative trajectories in the gas and aqueous phases and selected geometries along each trajectory.

Trajectory Calculations. A total of 100 random trajectories were initiated each in the gas and aqueous phases, starting from the transition state stationary structures. Of the gas-phase trajectories, six failed to complete, while seven failed in the aqueous phase. The remaining trajectories were characterized as progressing toward products or reactants by visualizing the results in GaussView[60] and analyzing the changes in the C−N and C−Cl bond distances. Representative trajectories are presented in Figure 2, while a typical bond distance analysis is provided in Figure 3.

Gas-Phase Trajectories. Of the gas-phase trajectories, 84% returned to separated starting materials while 16% proceeded to the hydrogen-bonded product complex. None were found to lead to complete dissociation to ionic products. Those that returned to reactants did so quickly, with the C−N bond distance exceeding that found in the van der Waals reactant complex no later than 120 fs into the trajectory. By the end of 200 fs, the C−N bond distance exceeds 4.5 Å while the C−Cl distance has settled into an oscillation about the equilibrium bond distance as determined by the stationary-point calculations (Figure 3). The potential energy has settled on a value around 30 kcal/mol below the transition state (Figure 2), as is expected from the stationary point data presented in Figure 1.

Much more interesting to consider is the fate of those gas-phase trajectories that formed the hydrogen-bonded complex. In the typical example shown in Figure 2, the chloride appears to be dissociating rapidly by 50 fs. However, between 100 and 200 fs, the chloride completes a partial orbit of the methylammonium moiety and is positioned to abstract a proton from the nitrogen. Given the unusual nature of these product trajectories, they were continued for an additional 600 fs to give a total of 800 fs in the trajectory. During this

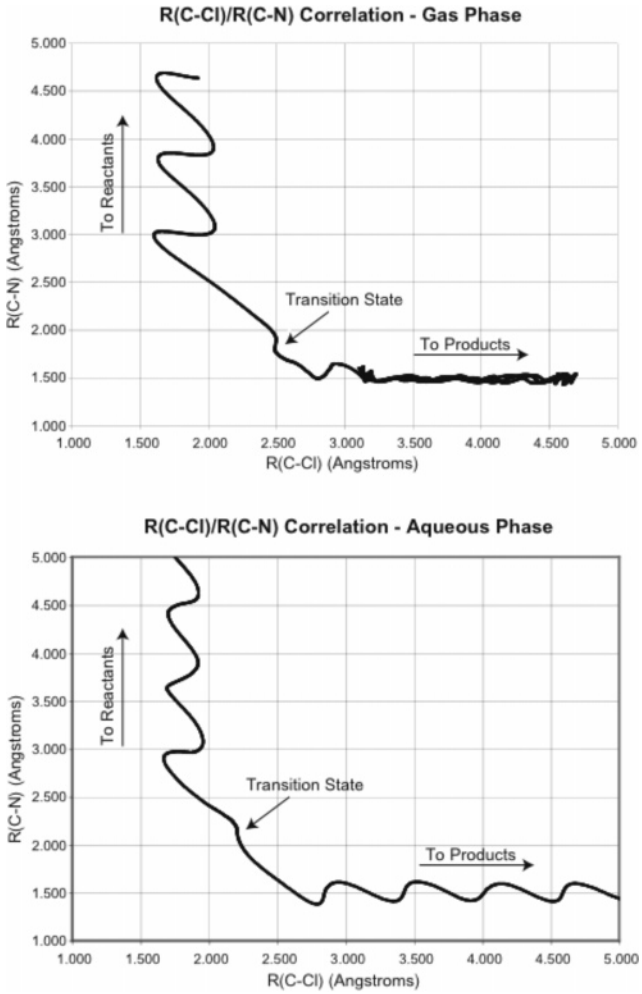


Figure 3. Correlation of C−Cl and C−N bond distances for those trajectories presented in Figure 2.

additional simulation time, the hydrogen-bonded complex never dissociates but is quite dynamic. When the product quadrant of the gas-phase plot in Figure 3 is examined, it can be seen that the C−Cl bond distance fluctuates between 3.5 and 4.7 Å while the C−N bond distance remains nearly constant at 1.5 Å. The potential energy has settled on a value nearly 40 kcal/mol below the transition state, which agrees well with the relative energy of the separated products (see above).

A trajectory of this sort has never been reported for this reaction and indicates that there may be a viable product channel available at 298 K. It is noteworthy that the potential energy appears to be rising over the last 150 fs of the 800 fs simulation, which may indicate that additional dissociation is occurring. Accordingly, one trajectory was integrated for a total of 2 ps. Indeed, between 900 and 1400 fs, the complex becomes much more dynamic, leading to dissociation by 1600 fs. By the end of the 2 ps trajectory, the methylamine and hydrogen chloride molecules have separated by more than 9 Å (see the Supporting Information).

Aqueous Trajectories. In the aqueous simulations, 52% of the trajectories returned to separated reactants while 48% continued to separated ionic products. None remained associated in any way or followed an unexpected trajectory as discussed above for the gas-phase simulations. Those trajectories that returned to reactants did so at a slightly faster pace than seen in the gas-phase simulation, with the C−N bond distance exceeding 3.5 Å by 100 fs. By 150 fs, there is little interaction between the two fragments and the lone pair of the ammonia is no longer oriented coincident with the C−Cl axis. By the end of the 200 fs simulation, the C−N bond distance has exceeded 5 Å while the C−Cl distance has settled into an oscillation about 1.7 Å (Figure 3), and the energy has settled on the expected value (Figure 2). There are few significant differences between those gas-phase and aqueous-phase trajectories that return to reactants.

The aqueous-phase product trajectories are dramatically different from the gas-phase product trajectories. First, similar numbers of aqueous trajectories follow the product and reactant channels, which is reasonable given that both the products and reactants are steeply downhill from the transition state. This contrasts with the gas-phase results, where the vast majority of the trajectories return to reactants. Second, the aqueous product trajectories rapidly expel the chloride ion and form the expected ionic products with no intervention of ion-dipole or hydrogen-bonded complexes. The chloride is effectively expelled within 100 fs, and the C−Cl distance exceeds 5 Å by the end of the 200 fs simulation, while the C−N bond distance and potential energy oscillate around their expected values (Figures 2 and 3).

Alternative Trajectories. In addition to the trajectories discussed above, which all originated from the traditional S_N2 backside attack transition states, a limited number of trajectories were initiated from complexes **C1**, **C2**, and **C3** (see above). Regardless of solvation, the general finding from these trajectories is that the nucleophile remains within van der Waals contact of the substrate or drifts away to separated reactants. In no case was an attack trajectory observed. We therefore conclude that the backside attack trajectories discussed above are the only energetically feasible reaction manifold for this system under the specific conditions considered.

Stability of Simulations. Several parameters may be considered when determining whether a simulation has remained stable. First, smooth fluctuations in energy and geometric parameters as indicated in Figures 2 and 3 show that there are no discontinuities. This is particularly important to note for the solvated trajectories, where it is reasonable to be concerned about fluctuations in the shape and size of the solute cavity. The continuous results indicate that those concerns are unfounded, at least on the time and distance scale considered in this work. For 10 of the 100 aqueous trajectories, a step size of 0.1 fs was required to obtain smoothly continuous results. Even with reduced step sizes, a handful of trajectories failed. A total of seven aqueous- and six gas-phase trajectories failed to propagate smoothly. The fact that the failure rate is nearly identical for the aqueous- and gas-phase trajectories implies that the reason for failure has little to do with the solvation model and is more likely due to the initial random velocities leading to unstable results.

In addition to the qualitative measures just described, the quality of a trajectory can be evaluated on theoretical grounds by examining the change of the fictitious Hamiltonian with respect to time, the idempotency, and ensuring proper initial conditions for the kinetic energy. With respect to the first two, there is no discernible difference in the Hamiltonian traces or idempotency of the gas-phase and aqueous trajectories, and all are well within acceptable limits.[14] Such stability of ADMP simulations in a continuum solvent has been previously demonstrated.[32] In order to ensure that the simulation remains on the ground-state electronic surface, initial conditions must be chosen such that the supplied kinetic energy is much less than the highest occupied molecular orbital−lowest unoccupied molecular orbital (HOMO−LUMO) gap. For these simulations, the initial kinetic energy is 0.011 hartrees, while the HOMO−LUMO gap in the transition state structures is approximately 0.3 hartrees.

Conclusions

In one of the first applications of its type, ADMP trajectory simulations have been coupled with the IEF−PCM implicit solvation model to successfully simulate the Menschutkin reaction of ammonia with methyl chloride. It is gratifying that there are no surprises in the aqueous-phase trajectories and that the solvated simulations are just as stable as the gas-phase simulations, as has been previously shown for the simulation of chloride anions in aqueous solution.[32] It is clear that the combination of ADMP and IEF−PCM has efficiently and accurately modeled the potential energy surface for the reaction of ammonia with methyl chloride. The aqueous product trajectories are dramatically different from the gaseous product trajectories, as is expected. It is readily apparent that ADMP/IEF−PCM simulations may be useful when solvation is essential, but detailed structural information about explicit solvent molecules is not, and when computational efficiency is of concern.

Acknowledgment. The authors thank Truman State University for supporting this research.

Supporting Information Available: Stationary point geometries and their associated energies and data for the 2 ps gas-phase trajectory leading to methylamine and hydrogen chloride. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Born, M.; Oppenheimer, R. *Ann. Phys.* **1927**, *84*, 457−484.

(2) Brooks, C. L., III; Case, D. A. *Chem. Rev.* **1993**, *93*, 2487−2502.

(3) Cheatham, T. E., III; Brooks, B. R. *Theor. Chem. Acc.* **1998**, *99*, 279−288.

(4) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182−196.

(5) Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589−1615.

(6) Jensen, F. *Introduction to Computational Chemistry*; John Wiley and Sons: Chichester, U. K., 1999; pp 6−51.

(7) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley and Sons Ltd: West Sussex, England, 2004; pp 69−102, 385−455.

(8) Marx, D.; Hutter, J. In *Procedings of Modern Methods and Algorithms of Quantum Chemistry, Julich*; Grotendorst, E., Ed.; John von Neumann Institute for Computing: Jülich, Germany, 2000.

(9) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.

(10) Hartke, B.; Carter, E. A. *J. Chem. Phys.* **1992**, *97*, 6569−6578.

(11) Lippert, G.; Hutter, J.; Parrinello, M. *Theor. Chem. Acc.* **1999**, *103*, 124−140.

(12) For a discussion of several recent developments in this field, see: *Modern Methods for Multidimensional Dynamics Computation in Chemistry*; Thompson, D. L., Ed.; World Scientific: Singapore, 1998.

(13) Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2001**, *114*, 9758−9763.

(14) Iyengar, S. S.; Schlegel, H. B.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2001**, *115*, 10291−10302.

(15) Schlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2002**, *117*, 8694−8704.

(16) Jungwirth, P.; Gerber, R. B. *Chem. Rev.* **1999**, *99*, 1583−1606.

(17) Jasper, A. W.; Nangia, S.; Zhu, C.; Truhlar, D. G. *Acc. Chem. Res.* **2006**, *39*, 101−108.

(18) Fernández-Ramos, A.; Miller, J. A.; Killenstein, S. J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 4518−4584.

(19) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3093.

(20) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2160−2200.

(21) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419−1432.

(22) Rega, N.; Iyengar, S. S.; Voth, G. A. *J. Phys. Chem. B* **2004**, *108*, 4210−4220.

(23) Sailer, W.; Pelc, A.; Limao-Vieira, P.; Mason, N. J.; Limtrakul, J.; Scheier, P.; Probst, M.; Mark, T. D. *Chem. Phys. Lett.* **2003**, *381*, 216−222.

(24) Tian, W. Q.; Wang, Y. A. *J. Chem. Theory Comput.* **2005**, *1*, 353−362.

(25) Iyengar, S. S. *J. Chem. Phys.* **2005**, *123*, 084310/9.

(26) Liu, L. V.; Tian, W. Q.; Wang, Y. A. *J. Phys. Chem. B* **2006**, *110*, 13037−13044.

(27) Joubert, L.; Adamo, C. *J. Chem. Phys.* **2005**, *123*, 211103/1−211103/4.

(28) Iyengar, S. S.; Petersen, M. K.; Day, T. J. F.; Burnham, C. J.; Teige, V. E.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 084309/1−084309/9.

(29) Zhang, L.; Li, H.; Hu, X.; Shijuin, H. *J. Phys. Chem. A* **2006**, *110*, 7690−7695.

(30) Ciofini, I.; Adamo, C. *J. Mol. Struct.* **2006**, *762*, 133−137.

(31) Joubert, L.; Pavone, M.; Barone, V.; Adamo, C. *J. Chem. Theory Comput.* **2006**, *2*, 1220−1227.

(32) Rega, N.; Brancato, G.; Barone, V. *Chem. Phys. Lett.* **2006**, *422*, 367−371.

(33) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032−3041.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03M*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(35) Menshutkin, N. *Z. Phys. Chem.* **1890**, *5*, 589.

(36) Abboud, J.-L. M.; Notario, R.; Bertran, J.; Sola, M. *Prog. Phys. Org. Chem.* **1993**, *19*, 1−182.

(37) Gao, J. *J. Am. Chem. Soc.* **1991**, *113*, 7796−7797.

(38) Gao, J.; Xia, X. *J. Am. Chem. Soc.* **1993**, *115*, 9667−9675.

(39) Shaik, S.; Ioffe, A.; Reddy, A. C.; Pross, A. *J. Am. Chem. Soc.* **1994**, *116*, 262−273.

(40) Maran, U.; Pakkanen, T. A.; Karelson, M. *J. Chem. Soc., Perkin Trans. 2* **1994**, 2445−2452.

(41) Dillet, V.; Rinaldi, D.; Bertrán, J.; Rivail, J.-L. *J. Chem. Phys.* **1996**, *104*, 9437−9444.

(42) Fradera, X.; Amat, L.; Torrent, M.; Mestres, J.; Constans, P.; Besalú, E.; Matrí, J.; Simon, S.; Lobato, M.; Oliva, J. M.; Luis, J. M.; Andrés, J. L.; Solà, M.; Carbó, R.; Duran, M. *J. Mol. Struct.* **1996**, *371*, 171−183.

(43) Maran, U.; Karelson, M.; Pakkanen, T. A. *J. Mol. Struct.* **1997**, *397*, 262−272.

(44) Truong, T. N.; Truong, T.-T. T.; Stefanovich, E. V. *J. Chem. Phys.* **1997**, *107*, 1881−1889.

(45) González-Lafont, A.; Jordi, V.; Lluch, J. M.; Bertrán, J.; Steckler, R.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 3420−3428.

(46) Chuang, Y.-Y.; Cramer, C. J.; Truhlar, D. G. *Int. J. Quantum Chem.* **1998**, *70*, 887−896.

(47) Amovilli, C.; Mennucci, B.; Floris, F. M. *J. Phys. Chem. B* **1998**, *102*, 3023−3028.

(48) Castejon, H.; Wiberg, K. B. *J. Am. Chem. Soc.* **1999**, *121*, 2139−2146.

(49) Pedley, J. B. *Thermochemical Data and Structures of Organic Compounds*; Thermodynamics Research Center: College Station, TX, 1994; Vol. 1.

(50) Sola, M.; Lledos, A.; Duran, M.; Bertran, J.; Abboud, J.-L. M. *J. Am. Chem. Soc.* **1991**, *113*, 2873−2879.

(51) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293−307.

(52) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908−6918.

(53) Lynch, B. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2004**, *111*, 335−344.

(54) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364−382.

(55) Gonzales, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154.

(56) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622.

(57) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511−519.

(58) Hoover, W. G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695−1697.

(59) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334−338.

(60) Dennington, R., II; Keith, T.; Millam, J. M.; Eppinnett, K.; Hovell, W. L.; Filliland, R. *GaussView*; Semichem, Inc.: Shawnee Mission, KS, 2003.

CT6002803

JCTC Journal of Chemical Theory and Computation

Frozen Gaussian Wavepacket Study of the Ground State of the He Atom

Ling Wang and Eli Pollak*

Chemical Physics Department, Weizmann Institute of Science, 76100 Rehovoth, Israel

Received November 9, 2006

Abstract: The Rayleigh−Ritz functional is used in conjunction with an *approximate* time evolution to improve ab initio estimates of ground-state energies. The improvement is due in part to the introduction of a novel variational "normalization function" for the approximate propagator. An additional variational parameter was introduced in the form of a constant shift energy of the Hamiltonian. The approximate propagator used was the frozen Gaussian propagator; however, the trajectories evolved on the coherent-state averaged Hamiltonian (Q representation). For Coulombic forces, this removes the singularity, easing the computation. An additional variational parameter was the width parameter used for the coherent states appearing in the frozen Gaussian propagator. Using an initial combination of nine Gaussian functions for He, with an initial energy of −2.5115 au, the variational method, with a very short time interval of integration, led to an improved energy of −2.81 ± 0.04 au.

I. Introduction

The quantum mechanics of the He atom in its lowest electronic state are well-understood. One can employ any number of standard ab initio chemistry packages to create a Gaussian basis set which is large enough to accurately diagonalize the Hamiltonian of the He atom and provide accurate estimates of the energies of the lowest eigenstates. Alternatively, one may employ quantum Monte Carlo (MC) simulation methods.[1] This does not mean that the solution is trivial. The He atom has two electrons whose motions are strongly coupled to each other and the He nucleus. For this reason, the He atom is a "good" testing ground for different approaches to the solution of the quantum dynamics of complex systems.

The real-time quantum dynamics of complex systems can at least in principle shed light also on structure. The simplest route is to Fourier-transform the time-dependent overlap of a wavefunction with itself. In principle, the transform will give a series of peaks whose energies are the eigenenergies of the system under study. The Fourier transform however provides accurate estimates only if the time interval used is very long. It must be much larger than $2\pi\hbar/\Delta E$, where ΔE is the level spacing between adjacent levels. It is notoriously difficult to compute numerically the long time quantum dynamics of complex systems. This led to the invention of the filter diagonalization method (FDM)[2,3] where the characteristic time needed is $2\pi\hbar/\bar{E}$, where $\bar{E}$ is the local average level energy. This time scale is much shorter than the time scale arising from the level spacing. The central object in the FDM is the overlap function of an initial state $|\Psi\rangle$ with its time evolved form:

$$c(t) = \langle\Psi|\hat{K}(t)|\Psi\rangle \tag{1.1}$$

where $\hat{K}(t) = \exp(-iHt/\hbar)$. Harmonic inversion is then used to represent the correlation function in terms of the eigenvalues of the Hamiltonian. The FDM method is not foolproof; as noted by Mandelshtam,[3] "the degree of convergence will always be a delicate issue".

Using a related approach, we have recently shown[4] how one can employ the Rayleigh−Ritz functional

$$E[\Psi] = \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} \geq E_0 \tag{1.2}$$

(where E_0 denotes the exact ground-state energy) together with time propagation to obtain improved estimates of the ground-state energy. By using as a trial function the linear combination of $|\Psi\rangle + \hat{K}(t)|\Psi\rangle$, the functional becomes time-dependent, and one can use the time as a variational parameter to improve upon the initial estimate for the energy. Here too, the characteristic time needed is on the order of $2\pi\hbar/\bar{E}$, much shorter than the Fourier time.

The time-dependent methods are, though, predicated on the solution of the real-time quantum dynamics. Even for

* Corresponding author e-mail: eli.pollak@weizmann.ac.il.

10.1021/ct600332v CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/16/2007

short times, this is not easy. Although significant progress has been made during the past decade,[5−16] the challenge of creating a general methodology remains formidable. The difficulty has created a flurry of more approximate quantum propagation methods of which perhaps the semiclassical initial value representation (SCIVR) and its variants is one of the more promising approaches.[17−19,21−26] The SCIVR methodology has a long history; here, we note two complementary approximations. Heller[18] invented the frozen Gaussian propagation, which was then improved upon by Herman and co-workers,[19,20] who introduced a prefactor to the frozen Gaussian. Although the prefactor significantly improves the quality of the approximation, there is a heavy price to pay; instead of having to solve $2N + 1$ equations of motion (where N is the number of degrees of freedom and the extra equation is for the action), one has to solve an additional $4N^2$ equations of motion to obtain the time-dependent monodromy matrix elements. The difficulty is that, without the prefactor, the frozen Gaussian approximation rapidly loses normalization.[27]

Interestingly, if one wants to employ the time as a variational parameter within the Rayleigh−Ritz functional, one really does not need the exact real-time propagation. Any approximate propagator $\hat{K}_0(t)$ will still give a time-varying functional, and if the energy is lower than that obtained from the initial wavefunction, then one will have improved upon the original choice. In this paper, we report results for the time-dependent variational functional using a variationally optimized frozen Gaussian SCIVR for the propagator. To overcome the loss of normalization which would render the method useless, we introduce an additional variational time-dependent function into the linear combination. We find that a very short time propagation reduces the energy from -2.5115 to -2.81 ± 0.04 au (the ground-state energy is -2.9038 au). However, due to the fact that the frozen Gaussian propagation is effected through a Monte Carlo procedure, the statistical error is rather large, and it is expensive to reduce it to acceptable chemical accuracy.

In section II, we describe the variational method used to propagate the initial wavefunction. Numerical results are presented in section III; we end in section IV with a discussion of the results and their implication for future studies using approximate propagation methods.

II. Theory

A. Time-Dependent Variational Theory. We assume that some initial normalized wavefunction $|\Psi\rangle$ has been chosen. The estimate for the ground-state energy associated with the wavefunction is

$$E_\Psi = \langle\Psi|\hat{H}|\Psi\rangle \tag{2.1}$$

The wavefunction is time-evolved under the action of some approximate propagator, such that

$$|\Psi_0(t)\rangle = \hat{K}_0(t)|\Psi\rangle \tag{2.2}$$

In our previous work,[4] we chose the linear combination $|\phi(t)\rangle = |\Psi\rangle + K(t)|\Psi\rangle$ as the time-dependent wavefunction to be inserted into the Rayleigh−Ritz upper-bound expression for the ground-state energy. In this paper, we will not use the exact propagator but the frozen Gaussian propagator (see also below), which rapidly loses normalization. This would suggest that one should compute the normalization function

$$N_\Psi(t) = \langle\Psi_0(t)|\Psi_0(t)\rangle \tag{2.3}$$

and renormalize the propagator to $\hat{K}_0(t)/\sqrt{N_\Psi(t)}$.

Apart from normalization, there exist other identities. For example, $E_\Psi = \langle\Psi|\hat{K}^\dagger(t)\,\hat{H}\,\hat{K}(t)|\Psi\rangle$ [where $\hat{K}(t) = \mathrm{e}^{-i\hat{H}t}$ is the exact quantum propagator] since the exact propagator commutes with the Hamiltonian. One could thus equally well renormalize the approximate propagator to $\hat{K}_0(t)\sqrt{E_\Psi(0)/E_\Psi(t)}$ where

$$E_\Psi(t) = \frac{\langle\Psi_0(t)|\hat{H}|\Psi_0(t)\rangle}{\langle\Psi_0(t)|\Psi_0(t)\rangle} \equiv \frac{H_\Psi(t)}{N_\Psi(t)} \tag{2.4}$$

This suggests that the best result will be obtained by renormalizing the approximate propagator by an unknown real function $f(t)$ to be determined variationally.

Additional freedom and better results may be obtained by demanding that the function be complex. However, the same result may be obtained in a more physically transparent and numerically stable way by shifting the Hamiltonian by an arbitrary constant energy ϵ. This shifts the ground-state energy by the same constant value. The approximate propagator then becomes $\mathrm{e}^{-i\epsilon t/\hbar}\,\hat{K}_0(t)$ so that the trial time-dependent wavefunction will be

$$|\phi(t)\rangle = |\Psi\rangle + f(t)\,\mathrm{e}^{-i\epsilon t}\,\hat{K}_0(t)|\Psi\rangle \tag{2.5}$$

The time-dependent Rayleigh−Ritz functional for the ground-state energy, obtained by subtracting out the constant shift energy, then becomes eq 2.6 (see Chart 1). This functional

Chart 1

$$E(t) = \frac{\langle\Psi|\hat{H}|\Psi\rangle + fe^{-i\epsilon t}\langle\Psi|\hat{H}\hat{K}_0(t)|\Psi\rangle + fe^{i\epsilon t}\langle\Psi|\hat{K}_0^\dagger(t)\hat{H}|\Psi\rangle + f^2\langle\Psi|\hat{K}_0^\dagger(t)\hat{H}\hat{K}_0(t)|\Psi\rangle}{1 + fe^{-i\epsilon t}\langle\Psi|\hat{K}_0(t)|\Psi\rangle + fe^{i\epsilon t}\langle\Psi|\hat{K}_0^\dagger(t)|\Psi\rangle + f^2\langle\Psi|\hat{K}_0^\dagger(t)\hat{K}_0(t)|\Psi\rangle} \geq E_0 \tag{2.6}$$

Chart 2

$$f(\varepsilon,t) = \frac{-(H_\Psi(t) - N_\Psi(t)E_\Psi)}{H_\Psi(t)S_\Psi(t,\varepsilon) - N_\Psi(t)h_\Psi(t,\varepsilon)} \pm \frac{\sqrt{(H_\Psi(t) - N_\Psi(t)E_\Psi)^2 - (h_\Psi(t,\varepsilon) - E_\Psi S_\Psi(t,\varepsilon))(H_\Psi(t)S_\Psi(t,\varepsilon) - N_\Psi(t)h_\Psi(t,\varepsilon))}}{H_\Psi(t)S_\Psi(t,\varepsilon) - N_\Psi(t)h_\Psi(t,\varepsilon)} \tag{2.7}$$

now depends on four unknowns, the time, the width parameter of the coherent states (see below), the shift energy, and the yet to be determined function $f(t)$.

Variation of the energy with respect to the function $f(t)$ leads to a quadratic equation whose two solutions are given in eq 2.7 (see Chart 2), where we used the additional notation

$$h_\Psi(t,\epsilon) = e^{i\epsilon t/\hbar}\langle\Psi|\hat{H}\,\hat{K}_0|\Psi\rangle + \text{cc} \tag{2.8}$$

$$S_\Psi(t,\epsilon) = e^{-i\epsilon t/\hbar}\langle\Psi|\hat{K}_0|\Psi\rangle + \text{cc} \tag{2.9}$$

The dependence of the variational energy $E(t)$ on the shift energy ϵ is known analytically. For each fixed time, one varies the shift energy to determine the time-dependent minimal energy. One then finds the minimal energy with respect to the remaining time variable. This process may then be repeated for different width parameters to obtain the best minimum.

B. Frozen Gaussian Propagation. The coordinate representation of a coherent state in one dimension is

$$\langle x|g(p,q)\rangle = \left(\frac{\Gamma}{\pi}\right)^{1/4}\exp\left[-\frac{\Gamma}{2}(x-q)^2 + \frac{i}{\hbar}p(x-q)\right] \tag{2.10}$$

where p and q are respectively the coherent-state momentum and coordinate and Γ is the width parameter. The He atom has two electrons; in Cartesian space, these correspond to six degrees of freedom, three for each electron. We will thus employ a multidimensional coherent state defined as a product of the six one-dimensional functions. Because of symmetry, the width parameters for all functions are taken to be identical.

The Hamiltonian of the He atom (in atomic units so that $\hbar = 1$, $m_e = 1$, etc.) is

$$\hat{H} = \sum_{i=1}^{6}\frac{\hat{p}_i^{\,2}}{2} - \frac{2}{\hat{r}_1} - \frac{2}{\hat{r}_2} + \frac{1}{\hat{r}_{12}} \tag{2.11}$$

with $\hat{r}_i^2 = \hat{x}_i^2 + \hat{y}_i^2 + \hat{z}_i^2$, $i = 1$ and 2, and $\hat{r}_{12}^2 = (\hat{r}_1 - \hat{r}_2)^2$. The coherent-state averaged Hamiltonian (also known as the Q representation of the Hamiltonian[28]) is then readily found to be[29]

$$H(\mathbf{p},\mathbf{q}) \equiv \langle g(\mathbf{p},\mathbf{q})|\hat{H}|g(\mathbf{p},\mathbf{q})\rangle = \sum_{i=1}^{6}\frac{p_i^{\,2}}{2} + \frac{3}{2}\Gamma + V_\Gamma(r_1,r_2,r_{12}) \tag{2.12}$$

where the coherent-state averaged potential is

$$V_\Gamma(r_1,r_2,r_{12}) \equiv -2\frac{\text{erf}(\sqrt{\Gamma}r_1)}{r_1} - 2\frac{\text{erf}(\sqrt{\Gamma}r_2)}{r_2} + \frac{\text{erf}\left(\sqrt{\frac{\Gamma}{2}}r_{12}\right)}{r_{12}} \tag{2.13}$$

All classical trajectories $\mathbf{p}_t$ and $\mathbf{q}_t$ will be propagated using Hamilton's equations as derived from the Q-representation Hamiltonian. Note that in this representation the singularity associated with the Coulomb potential at the origin is removed.

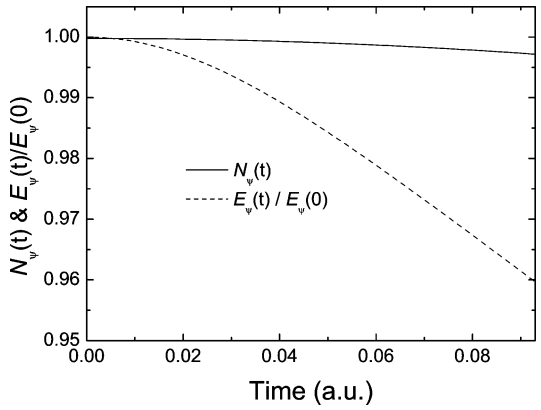


Figure 1. Normalization $N_\Psi(t)$ and the normalized energy function $E_\Psi(t)/E_\Psi(0)$ plotted as a function of time. The reduced energy function varies more rapidly than the normalization. Note that the decrease of the reduced energy function implies an *increase* in the energy $E_\Psi(t)$ with time, since $E_\Psi(0) < 0$.

The frozen Gaussian propagator is defined as

$$\hat{K}_0(t) = \int\frac{d\mathbf{p}\,d\mathbf{q}}{(2\pi)^6}\exp[i\,W(\mathbf{p},\mathbf{q};t)]|g(\mathbf{p}_t,\mathbf{q}_t)\rangle\langle g(\mathbf{p},\mathbf{q})| \tag{2.14}$$

and the classical action is

$$W(\mathbf{p},\mathbf{q};t) = \int_0^t dt'\left[\sum_{i=1}^{6}\frac{p_i^{\,2}(t')}{2} - \frac{3}{2}\Gamma - V_\Gamma(r_1,r_2,r_{12};t')\right] \tag{2.15}$$

The energy functional is then obtained by evaluating the various integrals, using a Monte Carlo methodology which takes advantage of the Gaussian structure of the coherent states as described in the Appendix.

III. Numerical Results

To simulate a typical result obtained from a Gaussian program, we first optimized with the Gaussian program the energy for a single-electron wavefunction consisting of three Gaussians. Our initial wavefunction was thus chosen to be the normalized product of the two single one-electron wavefunctions:

$$\langle x|\Psi\rangle = \sum_{i=1}^{3}\sum_{j=1}^{3}a_i a_j\exp(-b_i r_1^{\,2})\exp(-b_j r_2^{\,2}) \tag{3.1}$$

The parameters are $a_1 = 0.074\,877$, $a_2 = 0.259\,881$, $a_3 = 0.171\,081$, $b_1 = 6.362\,42$, $b_2 = 1.158\,23$, and $b_3 = 0.313\,650$. The initial energy $\langle\Psi|H|\Psi\rangle = -2.5115$ au is much higher than the exact ground-state energy of the He atom -2.9038 au. This choice of wavefunction mimics to some extent the complexity of ab initio Gaussian wavefunctions. It contains nine terms and thus implies that the various overlap functions include 81 integrals. Even if one can evaluate accurately with Monte Carlo methods a single integral, it is not clear a priori that one could obtain reasonable accuracy for the sum on the order of 10^2 terms whose sign may change.

The Monte Carlo methodology used to evaluate the various integrals is described in Appendix A. In the numerical results presented in Figures 1−3, we used 10^8 Monte Carlo sample points and a time step of $\Delta t = 0.003$ au for integrating the

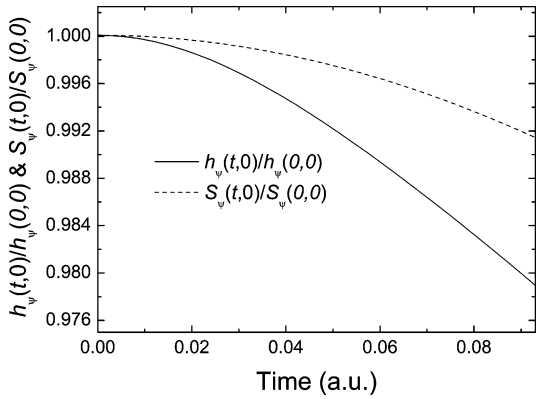


Figure 2. Time-dependent normalized energy overlap function and normalized overlap function vs time for the He atom. These functions are needed to estimate the Rayleigh−Ritz expression. Note the smoothness of the functions. For further details, see the text.

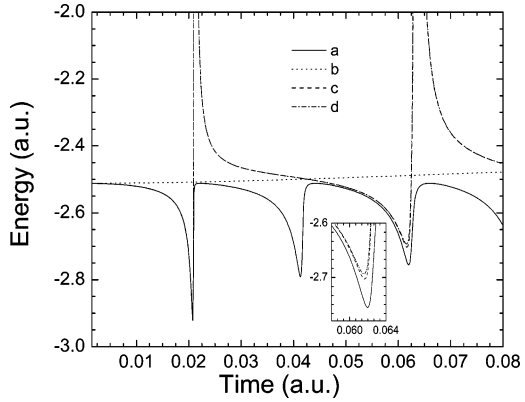


Figure 3. Time-dependent energy as a function of time for fixed values of the shift energy ϵ. The solid line (a) shows the time-dependent energy obtained by the quadratic eq 2.7 optimization of $f(\epsilon,t)$ at the fixed scaling energy $\epsilon = -148$ au. The dotted line (b) is obtained by setting $\epsilon = 0$ and $f(\epsilon,t) = 1$. The dashed line (c) shows the energy obtained after renormalizing the frozen Gaussian propagator with the time-dependent normalization $[\hat{K}_0(t)/\sqrt{N_\Psi(t)}]$, while the dashed dotted line (d) shows the result obtained after renormalization of the propagator with the time-dependent energy $[\hat{K}_0(t)\sqrt{E_\Psi(0)/E_\Psi(t)}]$. Both lines c and d are also obtained at the fixed scaling energy $\epsilon = -148$ au. The inset shows the improvement obtained by using the variational solution for the function $f(\epsilon,t)$.

classical equations of motion. The width parameter Γ appearing in the coherent state was varied so as to minimize the energy; the optimal value we found was $\Gamma = 0.5$ (au).

In Figure 1, we plot the normalization function $N_\Psi(t)$ (eq 2.3) as a function of time. For the exact quantum propagator, this function is a constant equal to unity. We also plot (dashed line) the normalized energy function $E_\Psi(t)/E_\Psi(0)$ (see eq 2.4) as a function of the time. Here too, for the exact quantum propagator, this function equals unity at all times. Interestingly, the variation of the normalization is slower than the variation in time of the energy. Therefore, the normalized energy function $E_\Psi(t)/N_\Psi(t)$ increases monotonically in time. The normalized time-dependent energy does not lead to any improvement in the estimate of the ground-state energy.

The computation of the normalization function $N_\Psi(t)$ and the energy function $E_\Psi(t)$ is substantially more expensive than the computation of the overlap functions. Here, one has a product of two propagators, and so it is necessary to perform a double integration over the system phase space. The integrand is much more oscillatory, and so the accuracy of the results is lower. However, the Monte Carlo sample used was sufficiently large to ensure that the plots shown in the figure are as accurate as the width of the line.

The functions $h_\Psi(t,0)/h_\Psi(0,0)$ and $S_\Psi(t,0)/S_\Psi(0,0)$ are plotted in Figure 2. It is significant to note that also these functions vary smoothly with the energy. It is therefore not difficult to solve the quadratic equation for the variational function $f(\epsilon,t)$. As noted earlier, employing the function within a variational context is the optimal strategy with respect to different renormalizations of the frozen Gaussian propagator. In Figure 3, we plot the time-dependent energy for four different scenarios. The solid line shows the time-dependent energy obtained by the quadratic eq 2.7 optimization of $f(\epsilon,t)$ at the fixed scaling energy $\epsilon = -148$ au. The dotted line is obtained by setting $\epsilon = 0$ and $f(\epsilon,t) = 1$. As already noted above, this gives a monotonically increasing function of the energy. Using the frozen Gaussian propagation without any further optimization does not lead to any improvement relative to the initial result for the ground-state energy. The dashed line shows the energy obtained after renormalizing the frozen Gaussian propagator with the time-dependent normalization $[\hat{K}_0(t)/\sqrt{N_\Psi(t)}]$, while the dashed dotted line shows the result obtained after renormalization of the propagator with the time-dependent energy $[\hat{K}_0(t)\sqrt{E_\Psi(0)/E_\Psi(t)}]$. Both lines are obtained at the fixed scaling energy $\epsilon = -148$ au. Both of these results are similar to each other and to the optimal result; however, the optimal result does provide a significant improvement in the minima obtained as a function of time.

Although it would seem that the first minimum shown in the figure comes very close to the exact ground-state energy, this is misleading. The minimal energy found at $t = 0.021$ au is $E_{\min} = -2.92 \pm 0.73$. The reason for the huge error bar comes from the fact that, at the minimum, the denominator of the energy functional eq 2.6 almost vanishes. Even though each term in the denominator separately is evaluated with an accuracy of at least 4×10^{-4}, the close cancellation of all terms leaves us with an unacceptable statistical error. Reducing the error to acceptable limits turned out to be too costly.

The results shown in Figure 3 are, though, instructive. They show that the lowest minimum is obtained at very early times. Only a small amount of time propagation seems to go a long way in obtaining an improved result. With this in mind, as well as the observed smoothness of the various functions, we repeated the computation, using a much larger time step in the integration of the classical equations of motion and only the three times $t = 0.021$, 0.0315, and 0.042 au. The Monte Carlo sample used for $t = 0.021$ was 1.305×10^{10} MC steps; for $t = 0.0315$ it was 6.48×10^9 MC steps and for $t = 0.042$, 3.5×10^9 MC steps. For the three times $t = 0.021$, 0.0315, and 0.042 au, we used 1, 2, and 2 integration time steps, respectively. This computation led to

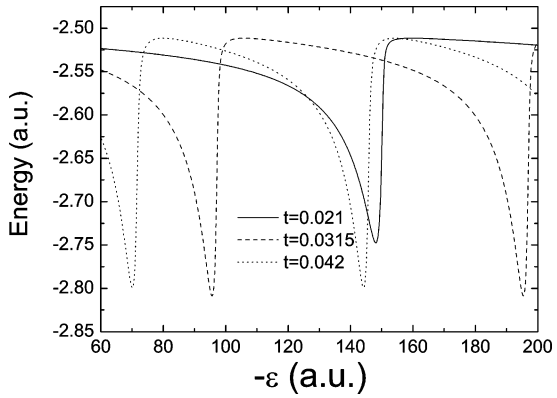


Figure 4. Energy as a function of the energy shift ϵ at fixed times. The solid line shows the energy functional as a function of the energy shift at the fixed time $t = 0.021$ au. The minimum energy $E_{min} = -2.748 \pm 0.035$ au is found at $\epsilon = -148$. The dashed lined shows the same but at the fixed time $t = 0.0315$ au. Here, the minimum energy $E_{min} = -2.809 \pm 0.036$ au is found at $\epsilon = -95.2$ au. The dotted line gives the results for $t = 0.042$ au where the minimum energy $E_{min} = -2.799 \pm 0.026$ au is found for $\epsilon = -70$ au.

a sufficiently small error, as shown below. The solid line in Figure 4 shows the energy functional as a function of the energy shift ϵ at the fixed time $t = 0.021$ au. The minimum energy $E_{min} = -2.748 \pm 0.035$ au is found at the shift energy $\epsilon = -148$. The dashed line shows the same but at the fixed time $t = 0.0315$ au. Here, the minimum energy $E_{min} = -2.809 \pm 0.036$ au is found at $\epsilon = -95.2$ au. Finally, the dotted line is for $t = 0.042$ au, where the minimum energy $E_{min} = -2.799 \pm 0.026$ au is found for $\epsilon = -70$ au. We thus conclude that only a very short time propagation is needed to significantly improve the original estimate of the energy from -2.51 to -2.81 au. These results should be compared with the exact ground-state energy of He which is -2.9038 au.

IV. Discussion

This paper presents some new elements in the time-dependent variational determination of ground-state energies. The Rayleigh−Ritz functional was used in conjunction with an *approximate* time evolution. We showed that one can significantly improve the ground-state energy, even when the time propagation is not precise. A second new element was the introduction of the variational "normalization function" $f(\epsilon,t)$ for the approximate propagator. The use of this variational function compensated for the lack of normalization in the approximate propagator. The third element was the introduction of the shift energy, or equivalently allowing the "normalization function" to be complex. Numerically, we found that more stable results were obtained by introducing the shift energy and varying it.

The model studied was the He atom. The approximate propagator was the frozen Gaussian propagator, where we also introduced a novel element; namely, the trajectories were propagated on the Q-representation Hamiltonian rather than the bare Hamiltonian as was usually done thus far. As in the coupled coherent-states method,[29] this removes the singularity in the potential, making it easier to propagate the classical trajectories.

The methodology led to a significant improvement relative to the initial energy. Only very short time propagation was needed to improve the accuracy of the ground state by roughly 75%. The computation itself was not trivial. Not only was it necessary to perform 12 fold integrals, the final results came from a summation of close to 100 terms. This points out the complexity of carrying out even the simplest SCIVR type of computation for strongly coupled systems such as the two electrons of the He atom. It is this complexity which necessitated averaging over $\sim 10^9$ samples, making the computation very costly. Even with this extensive computation, the error bars in the final energy were large on a chemical scale; an error of 0.02 au in energy is an error of ~ 13 kcal/mol. Reducing the error by an additional factor of 10 would imply increasing the sample size to 10^{11}.

The conclusion from all of this is that perhaps the time-dependent variational method will ultimately turn out to be useful; however, the SCIVR route, even for the short times needed, is not. Present variational and diffusion Monte Carlo methods converge with chemical accuracy to the ground state of He.[30,31] The purpose of this paper was not to present a method which can already compete with such well-known and well-tested methods. Rather, we demonstrated the kind of insight and quality of answer that one may obtain from a very different approach, namely, real-time propagation coupled with the variational theorem. It remains to be seen in future work whether perhaps basis set approaches such as the coupled coherent-states method or the multiconfiguration time-dependent Hartree method can lead to better and more accurate results.

Acknowledgment. This work has been supported by grants of the U.S.−Israel Binational Science Foundation, the Israel Science Foundation, the Donors of the American Chemical Society Petroleum Research Fund, and the German−Israel Foundation for Basic Research.

Appendix A: Monte Carlo Sampling

A Monte Carlo sampling is needed to carry out the integration over the 12-dimensional phase space of the two electrons. In any of the integrals, one always has matrix elements of the projection of the initial wavefunction onto the coherent states $\langle g(\mathbf{p},\mathbf{q})|\Psi\rangle$. The wavefunction is composed of nine terms, three for each electron, so that this overlap has the following form:

$$\langle g(\mathbf{p},\mathbf{q})|\Psi\rangle = \left(\frac{\Gamma}{\pi}\right)^{3/2} \sum_{i=1}^{3}\sum_{j=1}^{3} a_i a_j \left(\frac{\pi}{\Gamma/2+b_i}\right)^{3/2}\left(\frac{\pi}{\Gamma/2+b_j}\right)^{3/2} \prod_{k=1}^{3} \exp\left[-\frac{\Gamma b_i (q_1^{\,k})^2}{2(\Gamma/2+b_i)} + \frac{i b_i p_1^{\,k} q_1^{\,k}}{(\Gamma/2+b_i)\hbar} - \frac{(p_1^{\,k})^2}{4\hbar^2(\Gamma/2+b_i)}\right] \prod_{k=1}^{3} \exp\left[-\frac{\Gamma b_j (q_2^{\,k})^2}{2(\Gamma/2+b_j)} + \frac{i b_j p_2^{\,k} q_2^{\,k}}{(\Gamma/2+b_j)\hbar} - \frac{(p_2^{\,k})^2}{4\hbar^2(\Gamma/2+b_j)}\right] \quad \text{(A.1)}$$

where the superscript k denotes the three Cartesian directions x, y, and z and the subscripts 1 and 2 denote the phase-space variables for electrons 1 and 2, respectively. Since the wavefunction $|\Psi\rangle$ is composed of Gaussian functions, one naturally obtains the Gaussian weightings in phase space which come from the overlap. These Gaussian functions

$$\prod_{k=1}^{3} \exp\left[-\frac{\Gamma b_i (q_1^{\,k})^2}{2(\Gamma/2 + b_i)} - \frac{(p_1^{\,k})^2}{4\hbar^2(\Gamma/2 + b_i)}\right]$$

and

$$\prod_{k=1}^{3} \exp\left[-\frac{\Gamma b_j (q_2^{\,k})^2}{2(\Gamma/2 + b_j)} - \frac{(p_2^{\,k})^2}{4\hbar^2(\Gamma/2 + b_j)}\right]$$

are then used to implement the Box−Muller method for Gaussian probability distributions, replacing the respective coordinates and momenta by random variables according to

$$q_1^{\,k} = \sqrt{\frac{\Gamma/2 + b_i}{\Gamma b_i/2}[-\ln(1-\zeta)]}\cos(2\pi\eta)$$

$$p_1^{\,k} = \sqrt{4\hbar^2(\Gamma/2 + b_i)[-\ln(1-\zeta)]}\sin(2\pi\eta)$$

$$q_2^{\,k} = \sqrt{\frac{\Gamma/2 + b_j}{\Gamma b_j/2}[-\ln(1-\zeta)]}\cos(2\pi\eta)$$

$$p_2^{\,k} = \sqrt{4\hbar^2(\Gamma/2 + b_j)[-\ln(1-\zeta)]}\sin(2\pi\eta) \quad \text{(A.2)}$$

where, $k = 1-3$, and ζ and η are random numbers varying in the interval (0,1).

References

(1) Schmidt, K. E.; Ceperley, D. M. In *The Monte Carlo Method in Condensed Matter Physics*; Springer: Berlin, 1992; p 203.

(2) Wall, M. R.; Neuhauser, D. *J. Chem. Phys.* **1995**, *102*, 8011.

(3) Mandelshtam, V. A. *Proc. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 159.

(4) Saltzer, M.; Pollak, E. *J. Chem. Theory Comput.* **2005**, *1*, 439.

(5) Topaler, M.; Makri, N. *J. Chem. Phys.* **1994**, *101*, 7500.

(6) Topaler, M.; Makri, N. *J. Phys. Chem.* **1996**, *100*, 4430.

(7) Stockburger, J.; Mak, C. H. *J. Chem. Phys.* **1999**, *110*, 11.

(8) Stockburger, J.; Grabert, H. *Phys. Rev. Lett.* **2002**, *88*, 170407.

(9) Shalashilin, D. V.; Child, M. S. *J. Chem. Phys.* **2000**, *113*, 10028.

(10) Shalashilin, D. V.; Child, M. S.; Clary, D. C. *J. Chem. Phys.* **2004**, *120*, 5608.

(11) Burant, J. C.; Batista, V. S. *J. Chem. Phys.* **2002**, *116*, 2748.

(12) Beck, M. H.; Jaeckle, A.; Worth, G. A.; Meyer, H.-D.; Kuehn, O. *Phys. Rep.* **2000**, *324*, 1.

(13) Wang, H.; Thoss, M. *J. Chem. Phys.* **2003**, *119*, 1289.

(14) Worth, G. A.; Robb, M. A.; Burghardt, I. *Faraday Discuss.* **2004**, *127*, 307.

(15) Burghardt I. In *Quantum Dynamics of Complex Systems*; Micha, D. A., Burghardt, I., Eds.; Springer: New York, 2006; Springer Series in Chemical Physics, Vol. 83, p 135.

(16) Kuleff, A. I.; Breidbach, J.; Cederbaum, L. S. *J. Chem. Phys.* **2005**, *123*, 044111.

(17) Miller, W. H. *J. Chem. Phys.* **1970**, *53*, 3578.

(18) Heller, E. J. *J. Chem. Phys.* **1981**, *75*, 2923.

(19) Herman, M. F.; Kluk, E. *Chem. Phys.* **1984**, *91*, 27.

(20) Kluk, E.; Herman, M. F.; Davis, H. L. *J. Chem. Phys.* **1986**, *84*, 326.

(21) Grossmann, F. *Comments At. Mol. Phys.* **1999**, *34*, 141.

(22) Tannor, D. J.; Garaschuk, S. *Annu. Rev. Phys. Chem.* **2000**, *51*, 553.

(23) Baranger, M.; de Aguiar, M. A. M.; Keck, F.; Korsch, H. J.; Schellhaas, B. *J. Phys. A: Math. Gen.* **2001**, *34*, 7227.

(24) Miller, W. H. *J. Phys. Chem. A* **2001**, *105*, 2942.

(25) Kay, K. G. *Annu. Rev. Phys. Chem.* **2005**, *56*, 255.

(26) Pollak, E. In *Quantum Dynamics of Complex Systems*; Micha, D. A., Burghardt, I., Eds.; Springer: New York, 2006; Springer Series in Chemical Physics, Vol. 83, p 259.

(27) Harabati, C.; Rost, J. M.; Miller, W. H. *J. Chem. Phys.* **2004**, *120*, 26.

(28) Martin-Fierro, E.; Gomez Llorente, J. M. *Chem. Phys.* **2006**, *322*, 13.

(29) Shalashilin, D. V.; Child, M. S. *J. Chem. Phys.* **2005**, *122*, 224108.

(30) Kole, J. S.; De Raedt, H. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2001**, *64*, 016704.

(31) Ma, A.; Drummond, N. D.; Towler, M. D.; Needs, R. J. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2005**, *71*, 066704.

CT600332V

JCTC Journal of Chemical Theory and Computation

Monte Carlo Simulations of an Isolated *n*-Octadecane Chain Solvated in Water−Acetonitrile Mixtures

Li Sun,[†] J. Ilja Siepmann,*[,†] and Mark R. Schure[‡]

Departments of Chemistry and of Chemical Engineering and Materials Science, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455, and Theoretical Separation Science Laboratory, Rohm and Haas Company, 727 Norristown Road, P.O. Box 0904, Spring House, Pennsylvania 19477

Received July 21, 2006

Abstract: To investigate conformational properties of an isolated *n*-octadecane chain solvated in water−acetonitrile mixtures, configurational-bias Monte Carlo simulations in the isobaric−isothermal ensemble were performed at $T = 323$ K and $p = 10$ atm. The united-atom version of the transferable potentials for phase equilibria force field was used to represent *n*-octadecane and acetonitrile, and the TIP-4P model was used for water. In all four environments (neat water, 33 and 67 mole percent acetonitrile, and neat acetonitrile), similar conformational distributions are observed as in a previous study for water−methanol solvent mixtures; that is, the *n*-octadecane chain is found to predominantly adopt extended but not all-trans conformations, and only a small fraction of more collapsed conformations is observed for aqueous hydration, water-rich solvent environments. Analysis of the local solvation structures in the water−acetonitrile mixtures shows an enrichment of the acetonitrile molecules near the methylene and methyl segments of the *n*-octadecane chain. However, upon increasing the concentration of acetonitrile, the enhancement of acetonitrile and the depletion of water is more pronounced than for water−methanol mixtures because of the weaker interactions between acetonitrile and water.

1. Introduction

Water−acetonitrile and water−methanol mixtures are the most commonly used mobile phases in reversed-phase liquid chromatography (RPLC). When the fraction of organic modifier in the mobile phase falls below a certain threshold, a loss of retention is usually observed. It was suggested that the retention loss is due to the collapse of alkyl chains in the stationary phase.[1,2] It was also observed that some solutes have different thermodynamic behavior in the two different mobile phases,[3−5] and the formation of acetonitrile pockets in the organic-rich mobile phase was suggested to be a very important factor.[3,4,6,7] Isotherm measurements show that acetonitrile forms a much thicker adsorbed layer (with dimensions of four to five molecules) near the hydrophobic surface of the RPLC bonded phase than that observed for methanol (only a monolayer), and retention can occur in the interfacial environment.[8]

Particle-based molecular simulations have also been applied to study water−acetonitrile mixtures.[9−12] It was found that the water structure is enhanced upon adding acetonitrile, whereas the acetonitrile structure remains relatively intact in the aqueous solution.[9] Other simulations point to a broad size distribution for water clustering that fills in the spaces naturally occurring in the acetonitrile liquid. However, the clustering is not thought to be an important feature influencing retention; the properties of long chain molecules are considered to be more important.[10,11] Some molecular dynamics studies of simplified RPLC model systems with C_8 or C_{18} chains as a stationary phase yield a picture with the chains densely packed near the substrate and little solvent penetration.[13,14] In contrast, recent simulations for RPLC

* Corresponding author fax: (612) 626-7541; e-mail: siepmann@chem.umn.edu.
[†] University of Minnesota.
[‡] Rohm and Haas Co.

10.1021/ct600239z CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/05/2007

Table 1. Simulation Details and Ensemble Averages: Numbers of *n*-Octadecane and Solvent Molecules, Number of Monte Carlo Production Cycles for Each Independent Simulation, Average Linear Dimension of the Simulation Box (in Units of Å), Average End-to-End Length (in Units of Å), and Folding Equilibrium Constant for Comparison[a]

system	N (C_{18})	N (H_2O)	N (ACN/MeOH)	N_{MC}	$\langle L \rangle$	$\langle r_E \rangle$	$\langle K_{UN} \rangle$
WAT	1	900		2.3×10^5	30.40 ± 0.06	14.8 ± 0.4	0.11 ± 0.03
33A	1	600	300	2.7×10^5	35.83 ± 0.02	14.7 ± 0.8	0.10 ± 0.03
67A	1	300	600	2.6×10^5	40.01 ± 0.01	15.3 ± 0.2	0.027 ± 0.008
ACN	1		900	2.3×10^5	43.47 ± 0.01	15.6 ± 0.2	0.032 ± 0.012
33M	1	600	300	3.4×10^5	33.94 ± 0.06	14.1 ± 0.3	0.13 ± 0.03
67M	1	300	600	3.2×10^5	37.05 ± 0.07	15.4 ± 0.2	0.035 ± 0.03
MET	1		900	3.5×10^5	39.94 ± 0.07	16.1 ± 0.2	0.013 ± 0.005

[a] The results from previous simulations[17] for water−methanol mixtures are also listed.

systems with an explicit silica substrate show penetration for water−methanol solvent mixtures and chains in relatively extended conformations.[15,16] Furthermore, in our previous work probing the chain conformation in aqueous methanol solutions,[15,17] it was observed that an isolated *n*-octadecane chain prefers rather extended but not all-trans conformations over the entire concentration regime for water−methanol solvent mixtures. In contrast, simulation studies for more hydrophobic chains (with 22 or 25 segments, stronger dispersive segment−segment interactions, and/or less torsional rigidity) show a preference for folded structures.[18,19]

The purpose of this paper is to extend our computational study on the solvation of *n*-octadecane to water−acetonitrile mixtures. In particular, we focus on the chain conformation and the local solvent environment to elucidate the difference between alkane/water−methanol and alkane/water−acetonitrile interactions.

II. Molecular Models and Simulation Details

The simulation setup is very similar to our previous study on water−methanol solvation.[17] It consists of a single *n*-octadecane chain solvated either in water−acetonitrile solutions with an acetonitrile mole fraction of 33% or 67% (systems 33A and 67A, respectively) or in neat acetonitrile (system ACN). All three systems contained 900 solvent molecules (see Table 1), and the simulations were carried out at a temperature of 323.15 K and at a pressure of 1015 kPa (10 atm).

The united-atom version of the transferable potentials for phase equilibria (TraPPE-UA) force field[20−22] was used for *n*-octadecane and acetonitrile, and water was represented by the TIP4P model.[23] The Lennard-Jones parameters for unlike interactions were determined from the Lorentz−Berthelot combining rules.[24] A site−site-based, spherical cutoff at 14.0 Å and analytical tail corrections[25] were used for the Lennard-Jones interactions, and the Ewald summation technique[25] was employed to compute the Coulombic interactions arising from the partial charges on the acetonitrile and water molecules.

Monte Carlo simulations for the three systems were carried out in the isobaric−isothermal ensemble.[26] In addition to the usual translational[27] and rotational displacements[28] and volume moves,[26] the conformational degrees of freedom of the *n*-octadecane chain were sampled using a combination of coupled−decoupled configurational-bias Monte Carlo (CBMC) moves for regrowing multiple segments including at least one terminal group[29−31] and self-adapting fixed-endpoint CBMC moves for regrowing multiple interior segments.[32]

For each of the systems, four independent simulations were carried out, and the standard deviations were estimated from the results of these independent simulations. The production periods consisted of up to 270 000 Monte Carlo cycles where one cycle involves N randomly selected moves, where N (= 901) is the total number of molecules. The statistical uncertainties listed in the tables and shown in the figures are the standard errors of the mean computed from the four independent simulations.

III. Results and Discussion

A. Conformation of Chains in Different Environments. Figure 1 shows the evolution of the end-to-end distance, r_E, for the solvated *n*-octadecane chain. For all solvent environments, one can observe that the alkyl chain mostly undergoes minor fluctuations around $r_E \approx 17$ Å. Occasionally, rapid transitions from this relatively extended state to a more collapsed state with $r_E \approx 5$ Å are observed, but the chain conformation does not remain in the collapsed state for extended periods.

In order to compare the results for the whole concentration range of water−acetonitrile mixtures, data from our simulations using neat water as the solvent (system WAT)[17] are also included in the following discussion. The probability distribution of the end-to-end distances of the solvated *n*-octadecane chains averaged over the four independent simulations are compared in Figure 2. A bin width of 2 Å was used for this analysis. Similar as for solvation in water−methanol mixtures,[17] the main peak appears at 17 Å and is highly asymmetric with an extended tail toward a shorter end-to-end distance. The minor peaks at 5 Å correspond to a folded state. For system WAT, the main peak is somewhat sharper and the minor peak is more pronounced than for the three water−acetonitrile systems (33A, 67A, and ACN). The average end-to-end distances, $\langle r_E \rangle$, are listed in Table 1. For the water-rich solvent environments (WAT, 33A, and also 33M[17]), $\langle r_E \rangle$ falls close to 14.5 Å, whereas the organic-rich phases (67A, ACN, and also 67M and MET[17]) yield a slightly larger value of about 15.6 Å.

The folding equilibrium constants, K_{UN},[17] between extended (unfolded) states (with $r_E > 8$ Å) and collapsed (native) states (with $r_E \leq 8$ Å) are listed in Table 1. The values for K_{UN} range from 0.03 (for systems 67A and ACN)

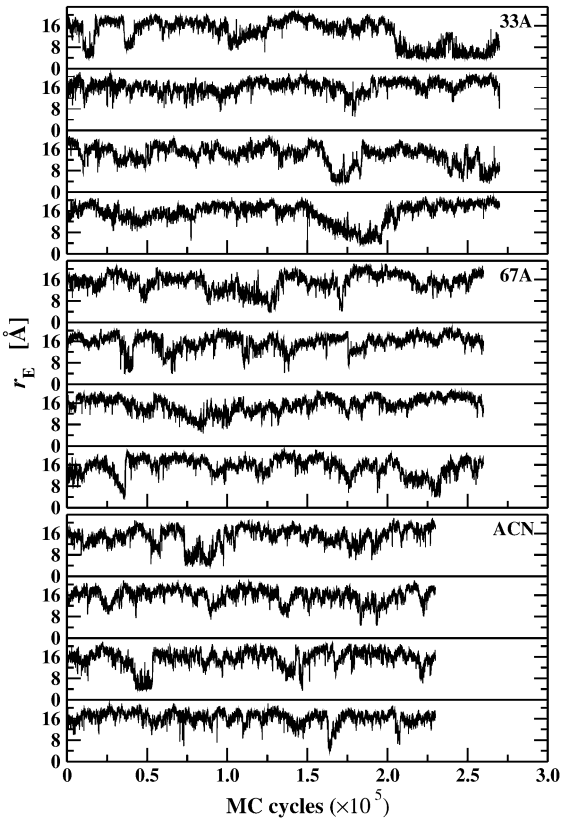


Figure 1. Evolution of the end-to-end distance of the solvated *n*-octadecane chains. Data are shown separately for the four independent simulations for each system: 33A (top), 67A (middle), and ACN (bottom).

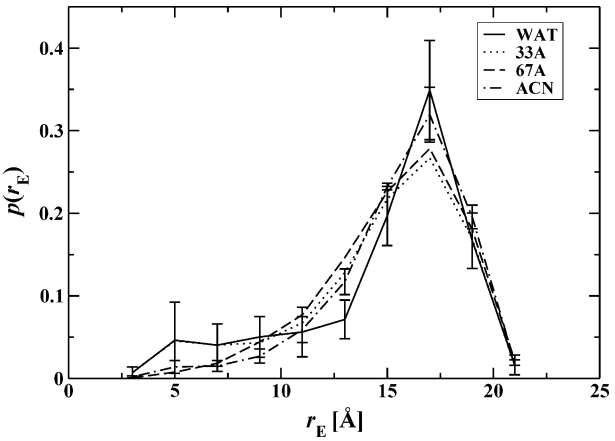


Figure 2. Probability density of the end-to-end distance for *n*-octadecane in four different environments. The solid, dotted, dashed, and dashed-dotted lines show the data for solvation in WAT, 33A, 67A, and ACN, respectively. For clarity, the standard errors of the mean are only shown for systems WAT and ACN.

to about 0.1 (for systems 33A and WAT), that is, in all solvent environments investigated here (and also for water−methanol mixtures);[17] extended states are strongly preferred, but folded states are sampled with significantly higher frequency in the water-rich environments. A comparison between the values for water−methanol solutions and those for water−acetonitrile solutions shows that the K_{UN} values are quite similar in the water-rich solvent environment

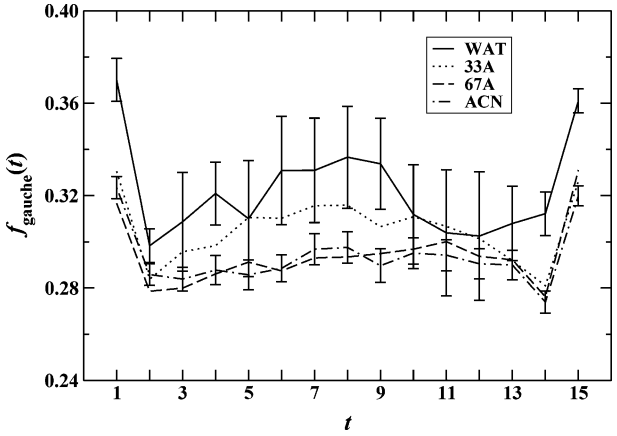


Figure 3. Fraction of gauche defects along the carbon backbone. Line styles are as in Figure 2.

(WAT, 33A and 33M) and very close in the organic-rich environment (67A, ACN, and 67M). Except for MET, the K_{UN} is smaller; that is, more extended conformations exist in the MET system.

Figure 3 depicts the fraction of gauche defects as a function of the position along the chain backbone. For all four solvents, the overall gauche fractions fall well-below 36%, the value that would be expected on the basis of a Boltzmann population analysis of the torsional potential for an ideal chain at the same temperature.[17,33] This shows that steric hindrance is more important than chain segment−segment attraction and solvent-induced forces that might favor folded states with more gauche defects than for an ideal chain. Only the terminal dihedral angles in system WAT have a gauche fraction of about 36%. Comparing the different solvation environments, one can see that, as the concentration of acetonitrile increases, the fraction of gauche defects decreases, though the difference is not dramatic. The *n*-octadecane chain in system WAT contains about 33% gauche defects, followed by that in system 33A with 31%, and then system 67A and 33M with about 29%. It is clear that the gauche defect distribution is related to the end-to-end distance distribution: chains that have higher probabilities in folded states also have a relatively higher fraction of gauche defects near the chain center. The reason for this is that gauche defects near the chain center are required to allow a chain to fold back on itself and bring the two chain ends into close contact.

B. Preferential Solvation. In order to investigate the local preferential solvation of the *n*-octadecane chain, radial distribution functions (RDFs), radial number densities (RNDs), and their corresponding number integrals (NIs) of solvent functional groups around the methyl or methylene groups of *n*-octadecane are compared in Figures 4 and 5. Compared with the local solvation environment found in water−methanol mixtures,[17] similar trends are observed for the RDFs and NIs between $CH_2/CH_3(C_{18})$ and ACN functional groups. The height of the first peak increases as the acetonitrile concentration decreases, while the NIs show that the number of ACN functional groups (N or CH3) increases with increasing concentration; that is, there is an enhancement in the composition but not in the absolute number of

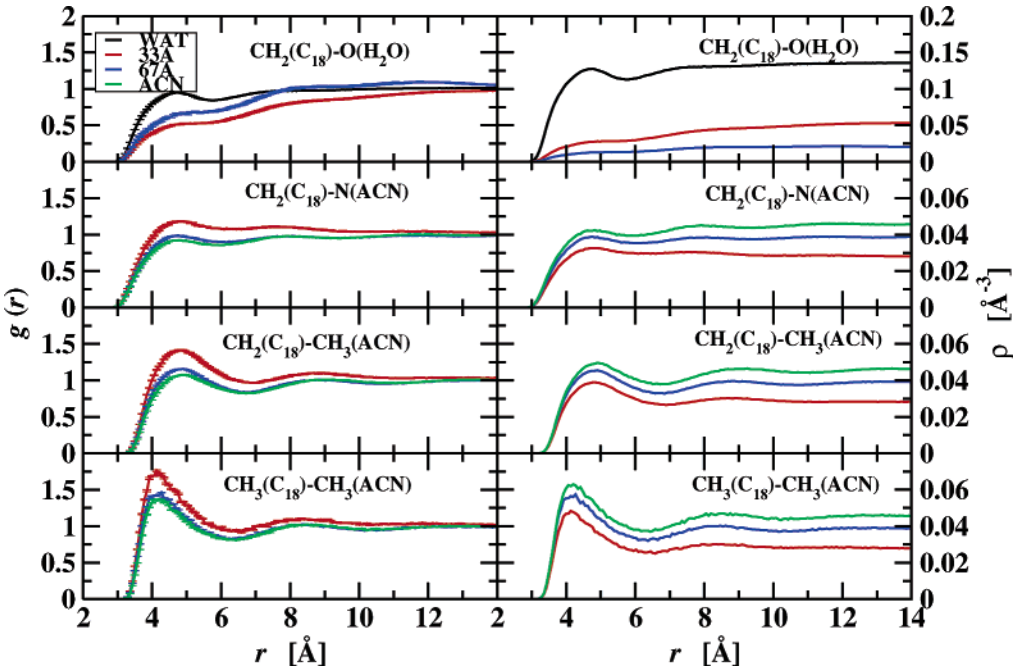


Figure 4. Radial distribution functions (left column) and radial number densities (right column) for solute segment−solvent segment pairs: top row, $CH_2(C_{18})$−$O(H_2O)$; second row, $CH_2(C_{18})$−N(ACN); third row, $CH_2(C_{18})$−CH_3(ACN); bottom row, $CH_3(C_{18})$−CH_3(ACN). The black, red, blue, and green lines represent WAT, 33A, 67A, and ACN systems, respectively.

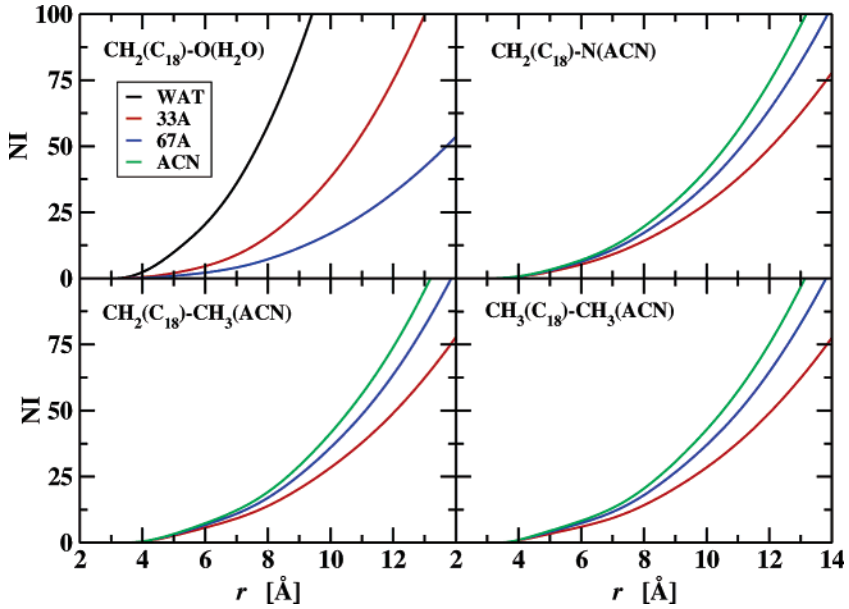


Figure 5. Number integrals for solute segment−solvent segment pairs: top left, $CH_2(C_{18})$−$O(H_2O)$; top right, $CH_2(C_{18})$−N(ACN); bottom left, $CH_2(C_{18})$−CH_3(ACN); bottom right, $CH_3(C_{18})$−CH_3(ACN). Line styles are as in Figure 4.

ACN molecules in the solvation shell. The first peak height in the RDFs for the $CH_2(C_{18})$−CH_3(ACN) pairs exceeds those for the $CH_2(C_{18})$−N(ACN) pair, but the difference is small compared to the corresponding RDFs involving the methyl group and oxygen atom of methanol.[17] Thus, the orientational distribution of the ACN molecules in the first solvation shell is more uniform than that for methanol, which has a clear preference for pointing with its methyl tail toward the chain.[17] Because of the larger accessible surface area and, hence, larger hydrophobicity of a methyl group compared to a methylene group, the local ACN structuring is further enhanced around the terminal methyl group of the C_{18} chain. Overall, the local solvation environment for the *n*-octadecane chain in systems ACN and 67A are quite similar, whereas that for system 33A differs substantially. This jump agrees with the folding equilibrium constants discussed above. Overall, it is apparent that, at typical chromatographic conditions, the organic modifier concentrations are sufficiently high that significant nonlinearities are observed for the RDFs, RNDs, and NIs of acetonitrile around the solute (i.e., RNDs and NIs do not show equal spacing). A similar observation was made for the mutual solubilities of ACN/water mixtures in *n*-hexadecane.[34] On the side, one may note that many experimental and simulation studies of bio-

Table 2. Numbers of Molecules, Mole Fractions, and Acceptor and Donor Hydrogen Bonds Per Molecule and Water's Tetrahedral Order Parameter Found in the Hydration Shell and Bulk Solvent

system	WAT	33A	67A	ACN
$N_{shell}(H_2O)$	90.0 ± 0.5	23.9 ± 0.6	10.9 ± 0.3	
N_{shell}(ACN)		26.2 ± 0.2	31.7 ± 0.1	35.7 ± 0.1
x_{shell}(ACN)		0.52 ± 0.01	0.74 ± 0.02	
$n^{acc}_{shell}(H_2O)$	1.80 ± 0.01	1.11 ± 0.01	0.69 ± 0.02	
n^{don}_{shell} (H_2O)	1.80 ± 0.01	1.78 ± 0.01	1.78 ± 0.01	
n^{acc}_{shell} (ACN)		0.84 ± 0.03	0.46 ± 0.01	
q_{shell} (H_2O)	0.13 ± 0.01	−0.46 ± 0.03	−0.47 ± 0.01	
x_{bulk} (ACN)		0.326 ± 0.001	0.668 ± 0.002	
n^{acc}_{bulk} (H_2O)	1.87 ± 0.01	1.33 ± 0.01	0.82 ± 0.01	
n^{don}_{bulk} (H_2O)	1.87 ± 0.01	1.8 ± 0.01	1.81 ± 0.01	
n^{acc}_{bulk} (ACN)		1.05 ± 0.01	0.50 ± 0.01	
q_{bulk} (H_2O)	0.477 ± 0.001	0.201 ± 0.001	0.038 ± 0.001	

molecule conformation and aggregation have shown linear behavior between bulk and vicinal concentrations at a low additive concentration.[35−37]

The $CH_2(C_{18})-O(H_2O)$ RDFs for both water−acetonitrile mixtures show a depletion of water near the C_{18} chain. However, as measured by the depressed height of the shoulder for the first solvation shell, the chain appears to be more dewetted for system 33A than for 67A. The opposite was observed for the water−methanol mixture; that is, the height of the peak/shoulder in the $CH_2(C_{18})-O(H_2O)$ RDFs decreased in the order WAT, 33M, to 67M.[17] On the other hand, the corresponding NIs (see Figure 5) show that, as the water concentration increases, the number of oxygen atoms (i.e., water molecules) around the chain segments increases from 67A to 33A. The unexpected order of the shoulder heights for the water−acetonitrile mixtures can be explained by the number of solvent molecules that, on average, populate the solvation shell. A simple distance-based criterion is used here to select solvent molecules as part of the first solvation shell; that is, a solvent molecule belongs to the solvation shell of *n*-octadecane if the separation between any solute segment and any heavy atom of the solvent molecule (including the methyl group and the nitrogen and α carbon atoms for acetonitrile and the oxygen atom for water) is less than 6 Å. The numerical values are listed in Table 2. The number of solvent molecules ranges from 90 for system WAT to 36 for system ACN, and the total numbers of solvent molecules for system 33A and 67A are 51 and 43, respectively. For both systems 33A and 67A, the solvation shells show enhanced acetonitrile mole fractions, which are 52% and 74%, respectively. It is clear that the enhancement by a factor of 1.6 for system 33A is much greater than that for system 67A, which is only 1.1. Therefore, the water concentration is much more depleted in system 33A, and even at this low concentration, the number of acetonitrile molecules in the first solvation shell has already reached about $^3/_4$ of the number found for neat acetonitrile. In contrast, for the water−methanol systems, the mole fractions of methanol in the first solvation shell are 46% and 76% for systems 33M and 67M, respectively.[17]

The local mole fraction enhancements[38] of solvent molecules around solute methyl or methylene groups for systems 33A and 67A are depicted in Figure 6. For both solvent mixtures, the local enhancement of acetonitrile peaks is at a separation of about 5.3 Å. The difference in molecular size is the reason for the lower enhancement at even smaller separation; that is, the smaller water molecule can approach somewhat closer than the larger acetonitrile molecule, but the number of solvent molecules at distances less than 4.5 Å is very small. We also should note a difference in the behavior of the local mole fraction enhancements between the water−acetonitrile and the water−methanol solvent mixtures. For the latter, the mole fraction enhancement is more pronounced around the methylene group, while the opposite is true for the water−acetonitrile mixture. The likely reason is that the large, linear acetonitrile molecule is not able to pack very well around the high-curvature contact surface of the methyl group.

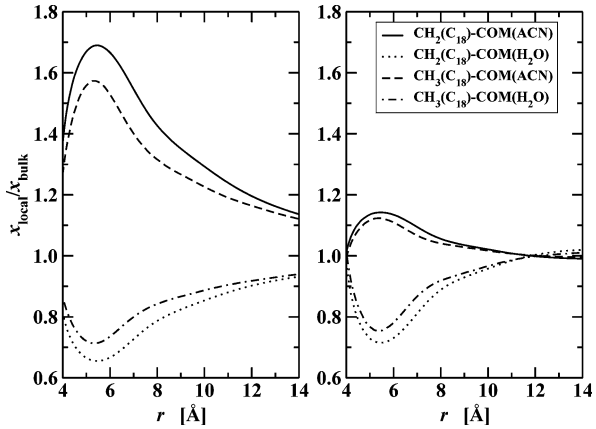


Figure 6. Local mole fraction enhancements in the vicinity of the *n*-octadecane chain for systems 33A (left) and 67A (right). The solid, dotted, dashed, and dotted−dashed lines represent $CH_2(C_{18})-COM(ACN)$, $CH_2(C_{18})-COM(H_2O)$, $CH_3(C_{18})-COM(ACN)$, and $CH_3(C_{18})-COM(H_2O)$, respectively.

C. Hydrogen-Bonding around the Solute. Hydrogen bonding around a nonpolar solute is also a topic of considerable interest.[18,19,39−41] Chandler and co-workers have pointed out that the disruption of the hydrogen-bonding network by a hydrophobic solute is scale-dependent. When the hydrophobic species is small, then the hydrogen-bonding network can persist, whereas the network is significantly disrupted

near large solutes (on the scale of nanometers).[40] More recently, Rajamani et al. have shown that the crossover length from small to large solute solvation is reduced by the addition of additives that lower the surface tension.[41] In Table 2, the numbers of hydrogen-bond donors and acceptors are compared for the solvation shell and the bulk solvent region. Here, any solvent molecule that is at least 12 Å away from any solute segment is considered to be in the bulk region (i.e., there is a 6-Å-thick buffer region between the shell and solvent). After checking that the geometric and energetic distributions are rather similar for hydrogen bonds involving oxygen or nitrogen as the acceptor, we continue to use the same hydrogen-bonding criteria as those for water/alcohol mixtures:[42,43] $r_{OX} \leq 3.3$ Å (where X can be the oxygen atom belonging to a second water molecule or the polar nitrogen atom), $r_{O\cdots H} \leq 2.5$ Å, $\cos \theta_{O-H\cdots X} \leq -0.1$, and $U_{pair} \leq -13$ kJ/mol, where $\theta_{O-H\cdots X}$ and U_{pair} are the hydrogen-bond angle and the interaction energy for a pair of solvent molecules, respectively. Because the methyl group in acetonitrile is modeled as a united atom, there cannot be a "weak" hydrogen bond involving a methyl hydrogen as the donor. Such interactions between a methyl hydrogen and a water oxygen were observed in a previous simulation study.[9]

For system WAT, it can be seen that the number of hydrogen bonds is about 4% higher for molecules belonging to the bulk region than that for molecules belonging to the solvation shell; that is, there is a minor disruption of the hydrogen-bond network. The fact that acetonitrile can only act as a hydrogen-bond acceptor leads to a donor−acceptor imbalance for the solvent mixtures. As a result, the number of hydrogen bonds accepted by a water molecule decreases in the order WAT, 33A, to 67A, with the decrease being more pronounced for the solvation shell than for the bulk region because of the acetonitrile mole fraction enhancement for the latter.

As discussed in a previous section, an acetonitrile molecule can only serve as a hydrogen-bond acceptor, a methanol molecule can serve as one donor and two acceptors, and a water molecule can contribute as two donors and two acceptors. Therefore, from Table 2, one can see that water molecules are involved in a higher number of hydrogen bonds as donors than as acceptors. Compared to systems 33M and 67M,[16] water molecules in systems 33A and 67A are involved in lower numbers of hydrogen bonds as acceptors. The number of acetonitrile molecules as acceptors is also lower than that for methanol molecules. It is also shown in Table 2 that, in the solvation shell, the number of molecules involved in hydrogen bonds is lower than that in the bulk. This difference is larger than in water−methanol systems. For acceptors (either water or acetonitrile), the average decrease is about 15%, while for donors, the decrease is only 2%. This could be due to the greater enrichment of acetonitrile in the solvation shell. Although the total number of hydrogen bonds donated by water molecules in the solvation shell decreases by only 1% from system WAT to 67A, there is a larger decrease for those water molecules in the bulk region; that is, the presence of acetonitrile leads to a disruption of the hydrogen-bond network in the bulk region of these systems.

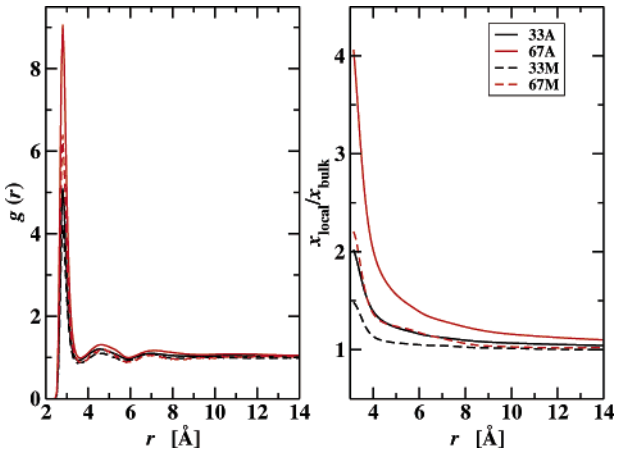


Figure 7. $O(H_2O)-O(H_2O)$ radial distribution functions (left) and local mole fraction enhancement for water (right). The black solid, red solid, black dashed, and red dashed lines depict data for systems 33A, 67A, 33M, and 67M, respectively.

D. Water Structure. As mentioned in the Introduction, it is well-known that water−acetonitrile mixtures show composition heterogeneities on small length scales, such as the preferential aggregation of water molecules.[9,10] The height of the first peak in the $O(H_2O)-O(H_2O)$ RDFs (see Figure 7) is significantly enhanced, with the enhancement being stronger for lower water concentration (systems 67A and 67M) and being stronger for water−acetonitrile mixtures than for water−methanol mixtures. A similar observation can also be made on the basis of the local mole fraction enhancement (see Figure 7). Although the composition enhancement is unambiguous, one should use care when interpreting the strong first peak in the $O(H_2O)-O(H_2O)$ RDF (e.g., with a value of about 9 for system 67A) as more structured water. Once the first peak in the RDF is divided by the local mole fraction enhancement at $r = 2.8$ Å, then the composition adjusted peak heights are 2.25, 2.55, 2.89, and 2.85 for systems 67A, 33A, 67M, and 33M. These rescaled peak heights for the water−methanol mixtures yield a value very close to that found for neat water, whereas those for the water−acetonitrile mixtures clearly fall below that for neat water.

A similar picture is also obtained by computing the tetrahedral order parameter for water.[44] Here, the nitrogen atom of acetonitrile is included in the search for the coordination shell. As can be seen from the data listed in Table 2, the tetrahedral order is significantly disrupted by the presence of acetonitrile. Thus, the composition enhancement for water goes hand-in-hand with a structural disordering.

IV. Conclusions

This computational study of a single *n*-octadecane chain solvated in water, acetonitrile, and their mixtures is an extension of our previous study in water−methanol systems. Simulation results show statistically significant but not dramatic differences in chain conformation. In all cases, including the water−methanol mixtures, the chain prefers unfolded conformations and the distribution of end-to-end distances peaks at 17 Å, a value that falls about 20% below

the end-to-end distance for the all-trans conformation. For both organic modifiers, the average end-to-end distance increases and the folding equilibrium constant decreases as the concentration of the organic modifier is increased.

The enhancement of acetonitrile with a concomitant depletion of water around the *n*-octadecane chain shows a different behavior as that in water−methanol mixtures due to the different hydrogen-bonding nature of the organic modifier. At a low organic concentration, water is significantly more depleted in system 33A than in 33M, and the concentration enhancement for acetonitrile extends to larger distances than that for methanol. However, acetonitrile does not pack well near methyl groups, and the enhancement of the acetonitrile mole fraction is more pronounced near methylene groups, whereas the opposite is true for methanol. Although water−water contacts are enhanced in water−acetonitrile mixtures compared to water−methanol mixtures, the presence of acetonitrile is more disruptive for water's tetrahedral hydrogen-bonding network. Overall, it appears that, for a given mole fraction, acetonitrile causes more significant changes than methanol in the solubility characteristics of aqueous solutions, which agrees with chromatographic observations.[3−8]

Acknowledgment. Financial support from the National Science Foundation (CHE-0213387) and the Advanced Biosciences Division of the Rohm and Haas Company is gratefully acknowledged. Part of the computer resources were provided by the Minnesota Supercomputing Institute.

References

(1) Przybyciel, M.; Majors, R. E. Phase Collapse in Reversed-Phase Liquid Chromatography. *LCGC North Am.* **2002**, *20*, 516.

(2) Walter, T. H.; Iraneta, P.; Capparella, M. Mechanism of retention loss when C8 and C18 HPLC columns are used with highly aqueous mobile phases. *J. Chromatogr., A* **2005**, *1075*, 177.

(3) Alvarez-Zeeda, A.; Barman, B. N.; Martire, D. E. Thermodynamic Study of the Marked Differences between Acetonitrile/Water and Methanol/Water Mobile-Phase Systems in Reversed-Phase Liquid Chromatography. *Anal. Chem.* **1992**, *64*, 1978.

(4) Guillaume, Y. C.; Gunichard, C. Retention Mechanism of Weak Polar Solutes in Reversed-Phase Liquid Chromatography. *Anal. Chem.* **1996**, *68*, 2869.

(5) Stalcup, A. M.; Martire, D. E.; Wise, S. A. Thermodynamic Comparison of Monomeric and Polymeric C_{18} Phases Using Aqueous Methanol and Acetonitrile Phases. *J. Chromatogr., A* **1988**, *442*, 1.

(6) Balakrishnan, S.; Easteal, A. J. Intermolecular Interactions in Water + Acetonitrile Mixtures : Evidence from the Composition Variation of Solvent Polarity Parameters. *Aust. J. Chem.* **1981**, *34*, 943.

(7) Carr, P. W.; Doherty, R. M.; Kamlet, M. J.; Taft, R. W.; Melander, W.; Horvath, C. Study of Temperature and Mobile-Phase Effects in Reversed-Phase High-Performance Liquid Chromatography by the Use of the Solvatochromic Comparison Method. *Anal. Chem.* **1986**, *58*, 2674.

(8) Kazakevich, Y. V.; LoBrutto, R.; Chan, F; Patel, T. Interpretation of the Excess Adsorption Isotherms of Organic Eluent Components on the Surface of Reversed-Phase Adsorbents Effect on the Analyte Retention. *J. Chromatogr., A* **913**, *2001*, 75.

(9) Kovacs, H.; Laaksonen, A. Molecular Dynamics Simulation and NMR Study of Water−Acetonitrile Mixtures. *J. Am. Chem. Soc.* **1991**, *113*, 5596.

(10) Mountain, R. D. Molecular Dynamics Study of Water−Acetonitrile Mixtures. *J. Phys. Chem. A* **1999**, *103*, 10744.

(11) Mountain, R. D. Molecular Dynamics Study of Thin Water−Acetonitrile Films. *J. Phys. Chem. B* **2001**, *105*, 6556.

(12) Satoh, Y; Nakanishi, K. Theoretical Studies of Acetonitrile−Water Mixtures/Monte Carlo Simulation. *Fluid Phase Equilib.* **1995**, *104*, 41.

(13) Klatte, S. J.; Beck, T. L. Microscopic Simulation of Solute Transfer in Reversed Phase Liquid Chromatography. *J. Phys. Chem.* **1996**, *100*, 5931.

(14) Slusher, J. T.; Mountain, R. D. A Molecular Dynamics Study of a Reversed-Phase Liquid Chromatography Model. *J. Phys. Chem. B* **1999**, *103*, 1354.

(15) Zhang, L.; Sun, L.; Siepmann, J. I.; Schure, M. R. A Moleculer Simulation Study of the Bonded-Phase Structure in Reversed-Phase Liquid Chromatography with Neat Aqueous Solvent. *J. Chromatogr., A* **2005**, *1079*, 127.

(16) Zhang, L.; Rafferty, J. L.; Siepmann, J. I.; Schure, M. R. A Moleculer Simulation Study of the Bonded-Phase Structure in Reversed-Phase Liquid Chromatography with Neat Aqueous Solvent. *J. Chromatogr., A* **2006**, *1126*, 219.

(17) Sun, L.; Siepmann, J. I.; Schure, M. R. Conformation and Solvation Structure for an Isolated *n*-Octadecane Chain in Water, Methanol, and Their Mixtures. *J. Phys. Chem. B* **2006**, *110*, 10519.

(18) Mountain, R. D.; Thirumulai, D. Molecular Dynamics Simulations of End-to-End Contact Formation in Hydrocarbon Chains in Water and Aqueous Urea Solution. *J. Am. Chem. Soc.* **2003**, *125*, 1950.

(19) Ghosh, T.; Kalra, A.; Garde, S. On the Salt-Induced Stabilization of Pair and Many-Body Hydrophobic Interactions. *J. Phys. Chem. B* **2005**, *109*, 642.

(20) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of *n*-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569.

(21) Wick, C. D.; Stubbs, J. M.; Rai, N.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 7. United-Atom Description for Nitrogen, Amines, Amides, Nitriles, Pyridine and Pyrimidine. *J. Phys. Chem. B* **2005**, *109*, 18974.

(22) Siepmann, J. I. TraPPE: Transferable Potentials for Phase Equilibria Force Field. http://www.chem.umn.edu/groups/siepmann/trappe/intro.php (accessed July 20, 2006).

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(24) Maitland, G. C.; Rigby, M.; Smith, E. B.; Wakeham, W. A. *Intermolecular Forces: Their Origin and Determination*; Oxford University Press: Oxford, U. K., 1987; p. 155.

(25) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U. K., 1987; pp 65, 156ff.

(26) McDonald, I. R. *NpT*-Ensemble Monte Carlo Calculations for Binary Liquid Mixtures. *Mol. Phys.* **1972**, *23*, 41.

(27) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087.

(28) Barker, J. A.; Watts, R. O. Structure of Water: A Monte Carlo Calculation. *Chem. Phys. Lett.* **1969**, *3*, 144.

(29) Martin, M. G.; Siepmann, J. I. Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description for Branched Alkanes. *J. Phys. Chem. B* **1999**, *103*, 4508.

(30) Siepmann, J. I.; Frenkel, D. Configurational-Bias Monte Carlo - A New Sampling Scheme for Flexible Chains. *Mol. Phys.* **1992**, *75*, 59.

(31) Vlugt, T. J. H.; Martin, M. G.; Smit, B.; Siepmann, J. I. ; Krishna, R. Improving the Efficiency of the Configurational-Bias Monte Carlo Algorithm. *Mol. Phys.* **1998**, *94*, 727.

(32) Wick, C. D.; Siepmann, J. I. Self-Adapting Fixed-Endpoint Configurational-Bias Monte Carlo Method for the Regrowth of Interior Segments of Chain Molecules with Strong Intramolecular Interactions. *Macromolecules* **2000**, *33*, 7207.

(33) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106*, 6638.

(34) Sun, L.; Siepmann, J. I.; Schure, M. R. Monte Carlo Studies of Partitioning Between *n*-Hexadecane and Water−Acetonitrile Mixtures. *J. Phys. Chem. B*, Submitted for publication.

(35) Timasheff, S. N. Control of Protein Stability and Reaction by Weakly Interacting Cosolvents: The Simplicity of the Complicated. *Adv. Protein Chem.* **1998**, *51*, 355.

(36) Parsegian, V. A.; Rand, R. P.; Rau, D. C. Osmotic Stress, Crowding, Preferential Hydration, and Binding: A Comparison of Perspectives. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3987.

(37) Kalra, A.; Tugcu, N.; Cramer, S. M.; Garde, S. Salting-In and Salting-Out of Hydrophobic Solutes in Aqueous Salt Solutions. *J. Phys. Chem. B* **2001**, *105*, 6380.

(38) Chen, B.; Potoff, J. J.; Siepmann, J. I. Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Tertiary Alcohols. *J. Phys. Chem. B* **2001**, *105*, 3093.

(39) Huang, D. M.; Chandler, D. The Hydrophobic Effect and the Influence of Solute-Solvent Attractions. *J. Phys. Chem. B* **2002**, *106*, 2047.

(40) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999**, *103*, 4570.

(41) Rajamani, S.; Truskett, T. M.; Garde, S. Hydrophobic Hydration from Small to Large Lengthscales: Understanding and Manipulating the Crossover. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9475.

(42) Stubbs, J. M.; Siepmann, J. I. Aggregation in Dilute Solutions of 1-Hexanol in *n*-Hexane: A Monte Carlo Simulation Study. *J. Phys. Chem. B* **2002**, *106*, 3968.

(43) Chen, B.; Siepmann, J. I. Microscopic Structure and Solvation in Dry and Wet Octanol. *J. Phys. Chem. B* **2006**, *110*, 3555.

(44) Errington J. R.; Debenedetti, P. G. Relationship between Structural Order and the Anomalies of Liquid Water. *Nature* **2001**, *409*, 318.

CT600239Z

JCTC Journal of Chemical Theory and Computation

Conceptual Density-Functional Theory for General Chemical Reactions, Including Those That Are Neither Charge- nor Frontier-Orbital-Controlled. 1. Theory and Derivation of a General-Purpose Reactivity Indicator

James S. M. Anderson,[†] Junia Melin,[‡] and Paul W. Ayers*[,†]

Department of Chemistry, McMaster University, Hamilton, Ontario, Canada L8S 4M1, and Department of Chemistry, Kansas State University, Manhattan, Kansas 66506-3701

Received May 10, 2006

Abstract: A new general-purpose reactivity indicator is derived. Unlike existing indicators, this indicator can describe the reactivity of molecules that lie between the electrostatic (or charge) control and electron-transfer (or frontier-orbital) control paradigms. Depending on the parameters in the indicator, it describes electrostatic control (where the electrostatic potential is the appropriate indicator), electron-transfer control (where the Fukui function's potential is the appropriate indicator), and intermediate cases (where linear combinations of the electrostatic potential and the Fukui function's potential are appropriate indicators). Our analysis gives some insight into the origins of the local hard/soft−acid/base principle. The "minimum Fukui function" rule for hard reagents also emerges naturally from our analysis: if (1) a reaction is strongly electrostatically controlled and (2) there are two sites that are equally favorable from an electrostatic standpoint, then the most reactive of the electrostatically equivalent sites is the site with the smallest Fukui function. An analogous electrostatic potential rule for soft reagents is also introduced: if (1) a reaction is strongly electron-transfer-controlled and (2) there are two sites where the Fukui function's potential are equivalent, then the most reactive of the Fukui-equivalent sites will be the one with greatest electrostatic potential (for electrophilic attack on a nucleophile) or smallest electrostatic potential (for nucleophilic attack on an electrophile).

I. Introduction

Many qualitative and semiquantitative methods have been developed for predicting how and whether a reaction will take place. Perhaps the most popular method of prediction is frontier molecular orbital theory (FMO).[1] This method uses the shapes and symmetries of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) to indicate whether a reaction will occur. If the HOMO of the electron donor and the LUMO of the electron acceptor have the same shape and phase, then electron transfer from the HOMO of the first molecule to the LUMO of the second can occur, often forming a bond between the reagents. Kenichi Fukui and Roald Hoffmann shared the 1981 Nobel Prize in chemistry for precisely this: showing that the shape[2−5] (Fukui et al.) and phase[6−10] (Hoffmann et al.) of the HOMO and LUMO orbitals are indicative of chemical reactivity.

A primary limitation of the frontier molecular orbital theory approach is that it presupposes the validity of the orbital model and thus fails to incorporate the effects of electron correlation or orbital relaxation. This motivated the

* Corresponding author fax: (905) 522-2509; e-mail: ayers@mcmaster.ca.
[†] McMaster University.
[‡] Kansas State University.

10.1021/ct600164j CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/16/2007

definition of a "Fukui function" in the context of density-functional theory (DFT), a function that encapsulates the essence of FMO[11−14] but, in principle, includes both electron correlation[11,15] and orbital relaxation.[12,16] The Fukui function from below, $f^-(r)$, is defined as the change in density that one observes when one goes from N to $N-1$ electrons (with the nuclear positions fixed); in simple molecular orbital theory, this would give precisely the density of the HOMO orbital. A similar function, $f^+(r)$, can be defined as the difference between the electron densities of the $N+1$ and N electron systems; this is analogous to the LUMO orbital density.

The Fukui function is labeled according to whether the system is acting as an electron acceptor or an electron donor. $f^+(r)$ says where an electron (received from a perfect electron donor) will add to the molecule. $f^-(r)$ says where an electron given to an electron acceptor (a perfect one, if you like) will come from. Electron donors tend to attack the molecule where $f^+(r)$ is large because this is where the molecule "wants electrons." Electron acceptors tend to attack the molecule where $f^-(r)$ is large because this is where the molecule has electrons that it is "willing to give up."[11,13] This reasoning, of course, is only valid when the transition state lies early enough on the reaction path for the reacting fragments to still resemble the isolated reagents.[13] The fact that conceptual DFT tends to work even when the transition state is not especially early may be attributed to the empirically observed utility of "hill-climbing" methods for locating transition states.[17,18]

In 1989, Dewar[19] listed several nucleophiles where FMO failed to describe electrophilic aromatic substitution, including isoquinoline, 10-hydroxy-10,9-borazarophenanthrene, and 10-methyl-10,9-borazarophenanthrene. We attempted to explain the reactivity of these molecules using the tools of conceptual DFT, but it does not seem possible to describe the reactivity of these molecules without considering *both* electron-transfer effects *and* electrostatic interactions. This spurred us to develop a new reactivity indicator that combines the Fukui function and the electrostatic potential. The indicator we developed provides a general-purpose model for chemical reactivity. The purpose of this paper is to derive and discuss this reactivity indicator. The second paper in this series will discuss the application of the indicator to Dewar's problematic molecules.[20]

Before deriving the general-purpose reactivity indicator, we present a brief overview of FMO and conceptual DFT in section II. The new reactivity indicator is then derived, step-by-step, in section III. Our model provides a unified picture of chemical reactivity and elucidates, among other things, the "minimum Fukui function rule."[21,22] Section IV summarizes our findings.

II. Theoretical Background

A. Overview of Frontier Molecular Orbital Theory. FMO arises as a simplification of the treatment proposed by Coulson and Longuet-Higgins,[23−26] who used second-order perturbation theory to describe the interactions between the filled molecular orbitals of one reactant and the empty molecular orbitals of the other. The interaction energy between the fragments is then

$$E_{AB}^{FMO} = 2\left(\sum_{i\in O_A}\sum_{b\in U_B}\frac{|\langle\phi_i^{(A)}|\hat{h}_{AB}|\phi_b^{(B)}\rangle|^2}{\epsilon_i^{(A)}-\epsilon_b^{(B)}} + \sum_{j\in O_B}\sum_{a\in U_A}\frac{|\langle\phi_j^{(B)}|\hat{h}_{AB}|\phi_a^{(A)}\rangle|^2}{\epsilon_j^{(B)}-\epsilon_a^{(A)}}\right) \quad (1)$$

Here, $\{\phi_i^{(A)}\}$ and $\{\phi_j^{(B)}\}$ are the molecular orbitals of fragments A and B; $\{\epsilon_i^{(A)}\}$ and $\{\epsilon_j^{(B)}\}$ are their respective orbital energies; O_A and O_B are the sets of occupied molecular orbitals, and U_A and U_B are the sets of unoccupied (virtual) orbitals in fragments A and B, respectively. $\hat{h}_{AB}$ denotes the one-electron Hamiltonian for the "supermolecule" (A + B; e.g., the Fock operator in Hartree−Fock or the Kohn−Sham Hamiltonian in DFT). The numerators in eq 1 are analogous to resonance integrals in Hückel theory.

In general, the most important terms in eq 1 are those with the smallest denominator. This suggests ignoring terms that do not depend on the frontier molecular orbitals, so that

$$E_{AB}^{FMO} \approx 2\left(\frac{|\langle\phi_{HOMO}^{(A)}|\hat{h}_{AB}|\phi_{LUMO}^{(B)}\rangle|^2}{\epsilon_{HOMO}^{(A)}-\epsilon_{LUMO}^{(B)}} + \frac{|\langle\phi_{HOMO}^{(B)}|\hat{h}_{AB}|\phi_{LUMO}^{(A)}\rangle|^2}{\epsilon_{HOMO}^{(B)}-\epsilon_{LUMO}^{(A)}}\right) \quad (2)$$

If A (the Lewis acid) is the electron acceptor and B (the Lewis base) is the electron donor, then we expect that $\epsilon_{LUMO}^{(A)}$ is small and $\epsilon_{HOMO}^{(B)}$ is large. By this argument, the second term in eq 2 should be larger than the first term. Neglecting the first term gives

$$E_{AB}^{FMO} \approx 2\left(\frac{|\langle\phi_{HOMO}^{(B)}|\hat{h}_{AB}|\phi_{LUMO}^{(A)}\rangle|^2}{\epsilon_{HOMO}^{(B)}-\epsilon_{LUMO}^{(A)}}\right) \quad (3)$$

This equation is the essential basis for the frontier molecular orbital theory. It is only valid when the neglected terms in eq 1 (and especially the neglected first term in eq 2) are negligible.[19] The numerator of eq 3 is a sort of "generalized resonance integral", and so, one can infer, by the usual arguments (e.g., from the justification of the Wolfsberg−Helmolz approximation[27]), that a large overlap between electron-donating and electron-accepting orbitals is favorable. This result can also be inferred from a Hölder inequality:

$$|\langle\phi_{HOMO}^{(B)}|\hat{h}_{AB}|\phi_{LUMO}^{(A)}\rangle|^2 \le \|\hat{h}_{AB}\|^2\|\phi_{HOMO}^{(B)}\phi_{LUMO}^{(A)}\|^2 = \|\hat{h}_{AB}\|^2\{\int|[\phi_{HOMO}^{(B)}(r)]^*\phi_{LUMO}^{(A)}(r)|\,dr\}^2 \quad (4)$$

B. The Fukui Function. As mentioned before, because the Fukui function contains similar information to the frontier molecular orbitals, it can be used to provide a DFT-based alternative to the standard rationalization of FMO theory. The Fukui function, $f(r)$, is defined as[11−13]

$$f(r) = \left[\frac{\delta\mu}{\delta v(r)}\right]_N = \left[\frac{\partial\rho(r)}{\partial N}\right]_v \quad (5)$$

Here, μ is the electronic chemical potential (equal to minus

the electronegativity), $v(r)$ is the "external" potential due to the atomic nuclei, $\rho(r)$ is the electron density, and N is the number of electrons. The equality between the functional derivative of the chemical potential and the ordinary derivative of the electron density arises as a "Maxwell relation."[28]

The slope of $\rho(r)$ as a function of N has discontinuities,[29−31] and so, the derivative must be evaluated from above and below (and averaged if necessary). This results in a Fukui function appropriate for describing nucleophilic attack[11]

$$f^+(r) = \left[\frac{\partial \rho(r)}{\partial N}\right]^+_{v(r)} \tag{6}$$

and a Fukui function appropriate for describing electrophilic attack

$$f^-(r) = \left[\frac{\partial \rho(r)}{\partial N}\right]^-_{v(r)} \tag{7}$$

If one writes the electron density as a function of the Kohn−Sham orbitals, $\phi_i(r)$, and orbital occupation numbers, n_i,

$$\rho(r) = \sum_{i=1}^{\infty} n_i |\phi_i(r)|^2 \tag{8}$$

$$n_i = \begin{cases} 1 & i \le \text{HOMO} \\ 0 & i \ge \text{LUMO} \end{cases} \tag{9}$$

then, from eqs 6 and 7, one has[12,32]

$$f^+(r) = |\phi_{\text{LUMO}}(r)|^2 + \sum_{i=1}^{\text{HOMO}} \left(\frac{\partial |\phi_i(r)|^2}{\partial N}\right)^+_{v(r)} \tag{10}$$

$$f^-(r) = |\phi_{\text{HOMO}}(r)|^2 + \sum_{i=1}^{\text{HOMO}} \left(\frac{\partial |\phi_i(r)|^2}{\partial N}\right)^-_{v(r)} \tag{11}$$

The link to frontier molecular orbital theory is obtained by neglecting the orbital relaxation terms, so that

$$f^+(r) \approx |\phi_{\text{LUMO}}(r)|^2 \tag{12}$$

$$f^-(r) \approx |\phi_{\text{HOMO}}(r)|^2 \tag{13}$$

C. Theoretical Description of Electrostatic and Electron-Transfer Effects. We have not yet provided a mathematical reason, comparable to eq 3, for interpreting the Fukui function as a reactivity indicator. To do this, consider how the energy of a molecule changes in response to an attacking electrophile. The attacking electrophile will take electrons from the molecule ($\Delta N_{\text{molecule}} < 0$). Additionally, the electrons that remain in the molecule will be perturbed by the presence of the reagent: that is, the electrons in the nucleophile will feel an additional external potential due to the electrons and nuclei of the electrophile. When these two effects are combined, the change in the total energy of the molecule undergoing attack is

$$\begin{aligned} \Delta U = & \int \frac{\delta V_{nn}[v]}{\delta v(r)} \Delta v(r)\, dr + \left(\frac{\partial E}{\partial N}\right)^-_{v(r)} \Delta N \\ & + \int \left\{\frac{\delta E_v[v;N]}{\delta v(r)}\right\}_N \Delta v(r)\, dr \\ & + \left\{\int \left[\frac{\partial}{\partial N}\left(\frac{\delta E_v[v;N]}{\delta v(r)}\right)\right]^-_{v(r)} \Delta v(r)\, dr\right\} \Delta N + \ldots \end{aligned} \tag{14}$$

Here, we are using U to denote the potential energy surface for the atomic nuclei. Within the Born−Oppenheimer approximation, $U = E + V_{nn}$; that is, U is sum of the electronic energy and the nuclear−nuclear repulsion energy.

The terms in eq 14 represent (a) the change in nuclear−nuclear repulsion energy, (b) the change in electronic energy due to electron transfer, (c) the change in electronic energy due to the change in external potential, and (d) the cross term linking electron transfer to changes in external potential. We will work within an exact formulation for the isolated system (e.g., the zero-temperature grand canonical ensemble[29]), so second- and higher-order derivatives with respect to the number of electrons vanish.[29,30] We are neglecting terms including higher-order responses to the external potential even though the first term (which models the polarization of the system by the approaching electrophile) might be important. Polarization effects are commonly neglected in the DFT-based approach to chemical reactivity. This is mostly because it is difficult to compute the polarizability kernel;[33] the success of DFT-based reactivity methods even when the polarization term is neglected suggests that the approximation in eq 14 is often sufficient for qualitative considerations. This success may be rationalized by noting that the dominant contribution to polarization arises from induced dipoles on the atomic centers, and the resulting interaction is short-ranged and weak[34] compared to the interaction from atomic charges. As such, the polarization contribution will ordinarily be an order of magnitude smaller than the energetic contributions from the molecular electrostatic potential, so it is often negligible for qualitative and semiquantitative purposes. Further evidence for this assertion comes from the broad (but not universal!) success of classical molecular dynamics that do not include polarization.

Subject to the approximations inherent in eq 14, the change in the energy of the nucleophile due to the approaching electrophile is

$$\begin{aligned} \Delta U_{\text{nucleophile}} = & (-I_{\text{nucleophile}})\Delta N \\ & - \int \left(\sum_{\alpha \in \text{nucleophile}} [Z_\alpha \delta(r - R_\alpha)] - \rho_{\text{nucleophile}}(r) \right. \\ & \left. - \Delta N f^-_{\text{nucleophile}}(r) \right) \Delta v(r)\, dr \end{aligned} \tag{15}$$

Here, $I_{\text{nucleophile}}$ is the vertical ionization potential of the molecule, $\{Z_\alpha\}$ and $\{R_\alpha\}$ denote the nuclear charges and their positions, respectively, and the summation is over all of the atomic nuclei in the nucleophile. Because the first term in

eq 15 depends only on ΔN and not on the position of the electrophile, the second term controls the regioselectivity of the reaction.

There is an exact model for $\Delta v(r)$, but it is not useful for computational purposes.[35,36] For reactants that are far apart, exchange and correlation between electrons on different subsystems should be negligible, so we can approximate $\Delta v(r)$ with minus the electrostatic potential of the attacking reagent.[35,36] The resulting model for the change in external potential at the point r_p is

$$\Delta v(r_p) \approx -\int \frac{\sum_{\beta\in\text{electrophile}} Z_\beta \delta(r-R_\beta) - [\rho_{\text{electrophile}}(r) + (-\Delta N) f^{+}_{\text{electrophile}}(r)]}{|r-r_p|}\,dr$$
$$= -\Phi_{\text{electrophile}}(r_p) - \Delta N \int \frac{f^{+}_{\text{electrophile}}(r)}{|r-r_p|}\,dr \quad (16)$$

The first term in the third line is the electrostatic potential of the isolated electrophile, and the second term is the correction to the electrostatic potential from electron transfer to the electrophile. In our model, $\Delta N < 0$ is the change in the number of electrons on the nucleophilic substrate, so the change in electron number for the electrophilic reagent is $-\Delta N$, which is greater than zero. This sign convention anticipates the second paper of this series,[20] where we will use these results to describe where nucleophiles are most susceptible to electrophilic attack.

Just as the energy of the nucleophile is changed by the approaching electrophile, the energy of the electrophile is changed by the presence of the nucleophile. In analogy to eq 15, one has

$$\Delta U_{\text{electrophile}} = (-A_{\text{electrophile}})(-\Delta N) - \int \left[\sum_{\beta\in\text{electrophile}} Z_\beta \delta(r-R_\beta) - \rho_{\text{electrophile}}(r) - (-\Delta N) f^{+}_{\text{electrophile}}(r)\right] \Delta v(r)\,dr \quad (17)$$

Here, the change in external potential is due to the electrons and nuclei in the nucleophile. Again, we approximate the change in external potential with the negative electrostatic potential of the nucleophile as corrected for electron transfer

$$\Delta v(r_p) \approx -\int \frac{\sum_{\alpha\in\text{nucleophile}} Z_\alpha \delta(r-R_\alpha) - [\rho_{\text{nucleophile}}(r) + \Delta N f^{-}_{\text{nucleophile}}(r)]}{|r-r_p|}\,dr$$
$$= -\Phi_{\text{nucleophile}}(r_p) + \Delta N \int \frac{f^{-}_{\text{nucleophile}}(r)}{|r-r_p|}\,dr \quad (18)$$

The replacement of ΔN in eqs 15 and 16 with $-\Delta N$ in eqs 17 and 18 is dictated by charge conservation: electrons are transferred from the nucleophile to the electrophile. Equations 16 and 18 lead to the identification of the electrostatic potential

$$\Phi(r_p) = \int \frac{\sum_{\beta\in\text{electrophile}} Z_\beta \delta(r-R_\beta) - \rho(r)}{|r-r_p|}\,dr \quad (19)$$

and the Fukui potential

$$v^f(r_p) = \int \frac{f(r)}{|r-r_p|}\,dr \quad (20)$$

as key reactivity indicators. Clearly, the Fukui potential is only relevant in cases where electron transfer is important. The Fukui potential also plays a key role in the reactivity model proposed by Berkowitz.[37]

Combining eqs 15 and 17 and correcting for the double counting of interactions gives a model for the interaction energy between the nucleophilic substrate and the electrophilic reagent:

$$U_{\text{int}} = (A_{\text{electrophile}} - I_{\text{nucleophile}})\Delta N + \int \left(\sum_{\alpha\in\text{nucleophile}} Z_\alpha \delta(r-R_\alpha) - \rho_{\text{nucleophile}}(r)\right) \Phi_{\text{electrophile}}(r)\,dr + \Delta N \int \left[f^{+}_{\text{electrophile}}(r)\,\Phi_{\text{nucleophile}}(r) - f^{-}_{\text{nucleophile}}(r)\,\Phi_{\text{electrophile}}(r)\right] dr - (\Delta N)^2 \int\int \frac{f^{-}_{\text{nucleophile}}(r)\, f^{+}_{\text{electrophile}}(r')}{|r-r'|}\,dr\,dr' \quad (21)$$

The first term in this equation is a constant and does not affect site selectivity. The next term reflects the electrostatic interactions between electrophiles and nucleophiles: because the active sites of electrophiles are usually positively charged while the reactive sites of nucleophiles are usually negatively charged, this term is usually negative. That is, electrostatic effects are usually attractive. This is in marked contrast to the terms in the third line of eq 21. Because $\Delta N < 0$, if $f^{+}_{\text{electrophile}}(r) > 0$ at the reactive site of the electrophile and the reactive site of the nucleophile is negatively charged [so that $\Phi_{\text{nucleophile}}(r) < 0$], then the first term in the third line of eq 21 is usually positive. Similarly, the second term in the third line of eq 21 is usually positive because $f^{-}_{\text{nucleophile}}(r) > 0$ at the active site of the nucleophile and the active site of the electrophile is usually positively charged [so that $\Phi_{\text{electrophile}}(r) > 0$]. The third line of eq 21, then, reflects the fact that charge transfer from the nucleophile to the electrophile helps to equalize the charges of the reagents, which reduces the electrostatic attraction between them. The fourth line of eq 21 will generally be negative. The expression in the fourth line plays a key role in the theory of electron transfer proposed by Berkowitz.[37] Of all the terms in eq 21, this is the term that most closely resembles the results from frontier molecular orbital theory. In fact, an expression very similar to eq 4 can be derived:

$$\begin{aligned}\int\int \frac{f^-_{\text{nucleophile}}(r)\, f^+_{\text{electrophile}}(r')}{|r-r'|}\,\mathrm{d}r\,\mathrm{d}r' &\le \left|\left|\frac{1}{|r-r'|}\right|\right| \left|\int f^-_{\text{nucleophile}}(r)\, f^+_{\text{electrophile}}(r)\,\mathrm{d}r\right| \\ &\approx \left|\left|\frac{1}{|r-r'|}\right|\right| \left|\int |\phi^{(\mathrm{B})}_{\mathrm{HOMO}}(r)|^2 |\phi^{(\mathrm{A})}_{\mathrm{LUMO}}(r)|^2\,\mathrm{d}r\right| \\ &\le \left|\left|\frac{1}{|r-r'|}\right|\right| \left|\int [\phi^{(\mathrm{B})}_{\mathrm{HOMO}}(r)]^*\, \phi^{(\mathrm{A})}_{\mathrm{LUMO}}(r)\,\mathrm{d}r\right|^2 \\ &\le \left|\left|\frac{1}{|r-r'|}\right|\right| \{\int |[\phi^{(\mathrm{B})}_{\mathrm{HOMO}}(r)]^*\, \phi^{(\mathrm{A})}_{\mathrm{LUMO}}(r)|\,\mathrm{d}r\}^2 \end{aligned} \quad (22)$$

Reactants will approach each other in a way that minimizes the interaction energy. That is, the more negative the interaction energy, the stronger the attraction between reagents and the greater their susceptibility to reaction. Coulson and Longuet-Higgins used perturbation theory and molecular orbital theory to derive a formula for the interaction energy between two separated reagents; this led them to eq 1. Equation 21 is a just a density-functional-theory-inspired reformulation of perturbation theory about the separated reagent limit. Reactivity indicators based on these formulas might be unreliable when the transition state occurs late in the reaction, because in those cases the molecular geometry in the transition state may not resemble the isolated reagents.

Some readers may find it surprising that the electrostatic potential enters into eq 21 in such a natural way. Traditionally, the electrostatic potential has not been considered a reactivity index associated with conceptual DFT.[38] However, it is evident from the preceding analysis that, whenever the change in the potential energy surface due to the external potential is addressed[35]

$$\int \left(\frac{\delta U}{\delta v(r)}\right)_N \delta v(r)\,\mathrm{d}r \quad (23)$$

the electrostatic potential enters into conceptual DFT in a very natural way. Evaluating expressions like eq 23 requires evaluating the functional derivative of the nuclear−nuclear repulsion energy with respect to changes in the external potential. Evaluating that functional derivative is a nontrivial mathematical exercise; details can be found in Appendix A. This analysis helps provide a theoretical foundation for recent work relating "conventional" DFT-based reactivity indicators to the electrostatic potential.[39,40]

III. A General-Purpose Model for Chemical Reactivity

A. The Reactive Site Interaction Model. In general, the regioselective preferences of a nucleophile undergoing electrophilic attack are preserved across a broad range of electrophilic partners. Because the specific identity of the electrophile is not critical, we can replace the electrophile with a "model perturbation".[14,35,41] Presuming that the reactivity is dominated by the properties at the active site, we will represent the electrostatic potential of the electrophile as the electrostatic potential due to the charge on the reactive site

$$\Phi_{\text{electrophile}}(r) \approx \frac{q^{(0)}_{\text{electrophile}}}{|r - R_{\text{electrophile}}|} \quad (24)$$

and, similarly, replace the Fukui function with its "condensed" value[42,43]

$$\begin{aligned} f^+_{\text{electrophile}}(\boldsymbol{r}) &\approx f^{(+)}_{\text{electrophile}}\delta(r - R_{\text{electrophile}}) \\ &= (q^{(0)}_{\text{electrophile}} - q^{(+)}_{\text{electrophile}})\,\delta(\boldsymbol{r} - \boldsymbol{R}_{\text{electrophile}}) \end{aligned} \quad (25)$$

Here, $q^{(0)}_{\text{electrophile}}$ and $q^{(+)}_{\text{electrophile}}$ denote the effective charges on the reactive site of the electrophile and the electrophile with an additional electron, respectively. If the Fukui function is approximated by eq 25, then the equation for the Fukui potential becomes

$$v^{f^+}_{\text{electrophile}}(r) \approx \frac{f^{(+)}_{\text{electrophile}}}{|r - R_{\text{electrophile}}|} = \frac{q^{(0)}_{\text{electrophile}} - q^{(+)}_{\text{electrophile}}}{|r - R_{\text{electrophile}}|} \quad (26)$$

Models based on eqs 24−26 should be accurate for atomic cations (where $q^{(0)}_{\text{electrophile}} = f^{(+)}_{\text{electrophile}} = 1$) and reasonable for molecular electrophiles, though in that case, it will be important that the "effective charge" and "effective condensed Fukui function" on the reactive site might not equal the atomic quantities but might instead represent a partial sum over several atoms in the vicinity of the reactive site. With that caveat, this model should be reasonably realistic whenever the reagents are very far apart, so that the overlaps between their electron distributions and Fukui functions are negligible. When these approximations are used, the interaction energy between a nucleophile and a model electrophile whose reactive site is (a) located at the point r_p, (b) has charge $q^{(0)}_{\text{electrophile}}$, and (c) has condensed Fukui function $f^{(+)}_{\text{electrophile}}$ becomes

$$\begin{aligned} U_{\text{int}}(r_p) \approx\ & (A_{\text{electrophile}} - I_{\text{nucleophile}})\Delta N \\ & + (q^{(0)}_{\text{electrophile}} + \Delta N f^{(+)}_{\text{electrophile}})\,\Phi_{\text{nucleophile}}(r_p) \\ & - \Delta N(q^{(0)}_{\text{electrophile}} + \Delta N f^{(+)}_{\text{electrophile}})\, v^{f^-}_{\text{nucleophile}}(r_p) \end{aligned} \quad (27)$$

This "single reactive site interaction model" is easily generalized to cases where multiple reactive sites need to be considered: simply sum over the effective charges and effective condensed Fukui functions[42] on all of the relevant sites

$$\begin{aligned} U_{\text{int}} \approx\ & (A_{\text{electrophile}} - I_{\text{nucleophile}})\Delta N \\ & + \sum_{\beta\in\text{electrophile}} \sum_{\alpha\in\text{nucleophile}} \frac{(q^{(0)}_\beta + \Delta N f^{(+)}_\beta) q^{(0)}_\alpha}{|R_\alpha - R_\beta|} \\ & - \Delta N \sum_{\beta\in\text{electrophile}} \sum_{\alpha\in\text{nucleophile}} \frac{(q^{(0)}_\beta + \Delta N f^{(+)}_\beta) f^{(-)}_\alpha}{|R_\alpha - R_\beta|} \end{aligned} \quad (28)$$

Again, a smaller (more negative) U_{int} represents greater attraction between reagents and indicates greater reactivity.

Insofar as we are modeling the attacking reagent with a point charge, these expressions for $U_{\text{int}}(r_p)$ are similar to the indicator of Brønsted−Lowry acidity introduced in ref 35, though the Taylor series expansion in that work includes

higher-order terms in the external potential and does not include the electron-transfer contribution. The present model is not restricted to charges of unit magnitude and, in that sense, is more like the single-interaction-site point-charge model that recently was used to elucidate the hard/soft−acid/base (HSAB) principle.[44] More generally, models resembling this one are commonly encountered when the so-called perturbative perspective on conceptual density-functional theory is utilized.[14,33,35,41,43,45] This sort of analysis, with its fundamental link to the Taylor series expansion to the energy and the use of "model perturbations" to define reactivity indicators, grew out of the work of De Proft, Liu, Nalewajski, Parr, and Senet, among others.[28,46−50]

B. Deriving a General-Purpose Reactivity Indicator for Nucleophiles. Equation 27 provides the basis for a general-purpose reactivity indicator for nucleophiles. Note, first of all, that the key parameter

$$\tilde{\kappa} \equiv q^{(0)}_{\text{electrophile}} + \Delta N f^{(+)}_{\text{electrophile}} \tag{29}$$

modulates the electrostatic and Fukui function contributions to the interaction energy. Insert the definition of $\tilde{\kappa}$ into eq 27, and note that the first term in eq 27 does not depend on the position, r_p. One obtains

$$\tilde{\Xi}^{\tilde{\kappa}}_{\Delta N\le 0}(r_p) \equiv \tilde{\kappa}[\Phi_{\text{nucleophile}}(r_p) - \Delta N v^{f^-}_{\text{nucleophile}}(r_p)] \tag{30}$$

which is a regioselectivity indicator for electrophilic attack. A nucleophile will be most reactive in places where $\tilde{\Xi}^{\tilde{\kappa}}_{\Delta N\le 0}(r_p)$ is small (most negative) because electrophilic attack at those sites is energetically favorable.

Equation 30 uses the reaction site interaction model to provide simple approximations to the complicated integral expressions in eq 21. But this is *too* simple. In particular, the reactive site interaction model is more accurate for some integrals than it is for others. The model typically underestimates the integrals because interactions between the asymptotic tails of the electron densities and the Fukui functions are neglected. Moreover, this error is largest for the integrals containing Fukui functions (which are concentrated on the periphery of the reagents). This implies that the reactive site interaction model is better at describing the interaction of the electrophile with the electrostatic potential (which contains the potential due to the electron density) than it is at describing the interaction of the electrophile with the Fukui potential. If we correct eq 30 for the errors associated with the reaction site interaction model, then we obtain an expression with the form

$$\tilde{\Xi}^{\tilde{\kappa}}_{\Delta N\le 0}(r_p) \equiv (\tilde{\kappa} + \epsilon_\rho)\,\Phi_{\text{nucleophile}}(r_p) - (\tilde{\kappa} + \epsilon_f)\Delta N v^{f^-}_{\text{nucleophile}}(r_p) \tag{31}$$

Here, ϵ_ρ corrects for the errors incurred by the reactive-site approximation for the electron density and ϵ_f corrects for the error incurred by the reactive-site approximation for the Fukui function. According to the preceding arguments

$$\epsilon_f < \epsilon_\rho \tag{32}$$

Usually, both ϵ_f and ϵ_ρ are negative. Appendix B provides a full accounting of the error terms.

We cannot evaluate these error terms within the context of the reactive site interaction model. It is certainly true that the errors are small when the molecules are far apart and that the error "correction" terms dominate when the molecules are close together. When the molecules are in van der Waals contact, one suspects that the error terms make significant corrections to $\tilde{\kappa}$.

Our goal is to derive a *qualitative* indicator of chemical reactivity. Note that (i) $\tilde{\kappa} \equiv q^{(0)}_{\text{electrophile}} + \Delta N f^{(+)}_{\text{electrophile}}$ has units of electric charge because the charge and the condensed Fukui function of the electrophile's reactive site both have units of electric charge. (ii) $\tilde{\kappa}$ is of order unity if one measures it in terms of the magnitude of charge on the electron, e. This is because e is the natural unit for expressing the charge and the condensed Fukui function of the electrophile's reactive site. (iii) We can eliminate the unknown error terms by introducing a new scale of electric charge, with the new unit of charge defined by

$$1 = \frac{\epsilon_\rho - \epsilon_f}{2} \tag{33}$$

We then define

$$\kappa = \tilde{\kappa} + \epsilon_\rho - 1 \tag{34}$$

This definition "sets the zero" of a scale for κ. Specifically, $\kappa = 0$ occurs when electrostatic and electron-transfer effects are perfectly balanced.

Using these relations, we can eliminate the unknown error terms from eq 31. This results in a general-purpose regioselectivity indicator for nucleophiles:

$$\Xi^{\kappa}_{\Delta N\le 0}(r_p) \equiv (\kappa + 1)\,\Phi_{\text{nucleophile}}(r_p) - (\kappa - 1)\Delta N v^{f^-}_{\text{nucleophile}}(r_p) \tag{35}$$

There is another way to derive eq 35: starting with eq 31, choose the unit of energy and the zero of energy so that the unknown error terms are eliminated. (Notice, however, that the energy scale one defines depends on what position, r_p, in the molecule is being considered.)

Equation 35 is our general-purpose reactivity indicator. Because $\Xi^{\kappa}_{\Delta N\le 0}(r_p)$ is a qualitative measure of the interaction energy of the nucleophile with an electrophile at the point r_p, the nucleophile will be most reactive where $\Xi^{\kappa}_{\Delta N\le 0}(r_p)$ is most negative.

Equation 35 is the most conceptually transparent form for our indicator. However, for computational applications, we find it convenient to compute the indicator from the electrostatic potential of the nucleophile, $\Phi^{(0)}_{\text{nucleophile}}(r)$, and the electrostatic potential of the nucleophile with one electron removed, $\Phi^{(-)}_{\text{nucleophile}}(r)$:

$$\Xi^{\kappa}_{\Delta N\le 0}(r_p) = [1 + \kappa + \Delta N(\kappa - 1)]\,\Phi^{(0)}_{\text{nucleophile}}(r_p) - \Delta N(\kappa - 1)\,\Phi^{(-)}_{\text{nucleophile}}(r_p) \tag{36}$$

When this expression is used, our reactivity indicator can be evaluated using any popular quantum chemistry package.

Remember that $\Xi^{\kappa}_{\Delta N\le 0}(r_p)$ models the interaction energy, so the nucleophile is most reactive where $\Xi^{\kappa}_{\Delta N\le 0}(r_p)$ has the smallest (most negative) values.

The key to deriving eq 36 is to note that the Fukui potential (and the Fukui function) can be computed from the electrostatic potentials of the nucleophile, $\Phi^{(0)}_{\text{nucleophile}}(r)$ and the electrostatic potential of the nucleophile with one electron removed, $\Phi^{(-)}_{\text{nucleophile}}(r)$. Specifically,

$$\upsilon^{f^-}_{\text{nucleophile}}(r) \equiv \int \frac{f^-_{\text{nucleophile}}(r)}{|r - r_p|}\,dr = \int \frac{\rho^{(0)}_{\text{nucleophile}}(r) - \rho^{(-)}_{\text{nucleophile}}(r)}{|r - r_p|}\,dr = \Phi^{(-)}_{\text{nucleophile}}(r_p) - \Phi^{(0)}_{\text{nucleophile}}(r_p)$$

$$f^-_{\text{nucleophile}}(\boldsymbol{r}) = \frac{1}{4\pi}\nabla^2[\Phi^{(0)}_{\text{nucleophile}}(r) - \Phi^{(-)}_{\text{nucleophile}}(r)] \quad (37)$$

Equation 35 or, equivalently, eq 36 is our general-purpose reactivity model for nucleophiles. Because of the way we have accommodated the error terms in eq 31, this is only a qualitative model for reactivity. One could derive other models with qualitatively similar behavior but somewhat different functional forms. We selected this form because it has an appealing symmetry and because it is easy to interpret (using eq 35) and apply (using eq 36).

$\Xi^{\kappa}_{\Delta N\le 0}(\boldsymbol{r}_p)$ depends on two parameters, ΔN (measuring the extent of electron donation) and κ (measuring the relative importance of electrostatic effects and electron-transfer effects). $\kappa = 1$ corresponds to pure electrostatic control; $\kappa = -1$ corresponds to pure electron-transfer (or Fukui-function) control. $\kappa = 0$ corresponds to a perfect balance between electrostatic and electron-transfer control.

The change of units that accompanies the elimination of the error terms means that $\kappa \neq \tilde{\kappa}$. However, if the error terms are not too big, κ will be approximately proportional to $\tilde{\kappa}$. We denote this $\kappa \approx \tilde{\kappa}$. This insight gives us a working approximation for κ (cf. eq 29), namely,

$$\kappa \approx q^{(0)}_{\text{electrophile}} + \Delta N f^{(+)}_{\text{electrophile}} \quad (38)$$

When the charge on the electrophile is very large, we still expect to see electrostatically controlled reactivity. Similarly, when the charge on the electrophile's reactive site is small compared with the extent of electron transfer to the reactive site ($q^{(0)}_{\text{electrophile}} \le |\Delta N| f^{(+)}_{\text{electrophile}}$), the reaction is electron-transfer controlled.

Like $\tilde{\kappa}$, κ has units of electric charge and has order of magnitude unity. Equation 38 is an approximate proportionality, and it is useful for elucidating how a nucleophile's reactivity depends on the charge and condensed Fukui function of an electrophile's reactive site. However, eq 38 is not a quantitative formula for κ. Determining the value of κ that is most appropriate for a given reaction is sensitive to (a) the inherent errors in the reactive site interaction model and (b) molecular polarization and other effects that are neglected in this analysis. However, eq 38 should be sufficient to indicate whether κ is "large and positive", "small and positive", "almost zero", "small and negative, or "large and negative." Section III.F. contains a detailed discussion of each of these cases.

C. A General-Purpose Reactivity Indicator for Electrophiles. There is clearly an analogous indicator for electrophiles. Specifically, one has

$$\Xi^{\kappa}_{\Delta N\ge 0}(r_p) \equiv -(\kappa + 1)\,\Phi_{\text{electrophile}}(r_p) + \Delta N(\kappa - 1)\,\upsilon^{f^+}_{\text{electrophile}}(r_p) \quad (39)$$

where the Fukui potential for the electrophile is given by

$$\upsilon^{f^+}_{\text{electrophile}}(r) \equiv \int \frac{f^+_{\text{electrophile}}(r)}{|r - r_p|}\,dr = \int \frac{\rho^{(+)}_{\text{electroophile}}(r) - \rho^{(0)}_{\text{electroophile}}(r)}{|r - r_p|}\,dr = \Phi^{(0)}_{\text{electrophile}}(r_p) - \Phi^{(+)}_{\text{electrophile}}(r_p)$$

$$f^+_{\text{electrophile}}(r) = \frac{1}{4\pi}\nabla^2[\Phi^{(+)}_{\text{electrophile}}(r) - \Phi^{(-)}_{\text{electrophile}}(r)] \quad (40)$$

The appropriate values of κ are the same as before: $\kappa \ge 1$ (electrostatic control), $\kappa \le -1$ (electron-transfer control), and $-1 < \kappa < 1$ (intermediate). Corresponding to eq 38, we can say that κ is approximately proportional to $-q^{(0)}_{\text{nucleophile}} - \Delta N f^{(-)}_{\text{nucleophile}}$. Here, $q^{(0)}_{\text{nucleophile}}$ is the charge on the reactive site of the nucleophile and is typically negative. $f^{(-)}_{\text{nucleophile}}$ is the condensed Fukui function from below at the reactive site of the nucleophile. It would be shocking to observe a negative value for $f^{(-)}_{\text{nucleophile}}$.[43,51−55]

D. The Condensed General-Purpose Reactivity Indices. Because local reactivity indicators vary on a point-by-point basis, it is often convenient to "condense" their values to atomic sites.[42,43] A condensed indicator related to $\Xi^{\kappa}_{\Delta N}(r_p)$ follows directly from the fitting of the electrostatic potential to atomic charges. Specifically, the electrostatic potential can be expanded in an asymptotic series of atomic multipoles. Truncating the multipole expansion after the monopoles yields an expression for the electrostatic potential in terms of atomic charges

$$\Phi^{(0)}_{\text{nucleophile}}(r) \sim \sum_{\alpha\in\text{nucleophile}} \frac{q^{(0)}_{\text{nucleophile},\alpha}}{|r - R_\alpha|}$$

$$\Phi^{(-)}_{\text{nucleophile}}(r) \sim \sum_{\alpha\in\text{nucleophile}} \frac{q^{(-)}_{\text{nucleophile},\alpha}}{|r - R_\alpha|} \quad (41)$$

These expressions are not very accurate close to the molecule. For example, these expressions do not reproduce the correct singularity at the atomic nuclei. However, the relevant values of a local reactivity indicator occur far from the molecule, on a "reactivity surface" that represents how closely two reagents can approach each other and still retain their separate identities. Not only do expressions like eq 41 suffice for this purpose, condensed expressions are actually preferable because they "average over" the irrelevant fine structure of the spatially varying indicators.[43]

"Condensed" expressions for the Fukui potential are easily constructed from eqs 37, 40, and 41

$$v^{f^-_{\text{nucleophile}}}(r) \sim \sum_{\alpha\in\text{nucleophile}} \frac{f^-_{\text{nucleophile},\alpha}}{|r-R_\alpha|}$$

$$f^-_{\text{nucleophile},\alpha} = q^{(-)}_{\text{nucleophile},\alpha} - q^{(0)}_{\text{nucleophile},\alpha} \quad (42)$$

$$v^{f^+_{\text{electrophile}}}(r) \sim \sum_{\alpha\in\text{electrophile}} \frac{f^+_{\text{electrophile},\alpha}}{|r-R_\alpha|}$$

$$f^+_{\text{electrophile},\alpha} = q^{(0)}_{\text{electrophile},\alpha} - q^{(+)}_{\text{electrophile},\alpha} \quad (43)$$

$f^\pm_\alpha$ are called the condensed Fukui functions.[42] We can now write asymptotic expressions for the reactivity indicators proposed in this paper. For nucleophiles,

$$\Xi^\kappa_{\Delta N\le 0}(r_p) \sim \sum_{\alpha\in\text{nucleophile}} \frac{\Xi^\kappa_{\Delta N\le 0,\alpha}}{|r-R_\alpha|}$$

$$\begin{aligned}\Xi^\kappa_{\Delta N\le 0,\alpha} &= (\kappa+1)q^{(0)}_{\text{nucleophile},\alpha} - \Delta N(\kappa-1)f^-_{\text{nucleophile},\alpha}\\ &= [1+\kappa+\Delta N(\kappa-1)]q^{(0)}_{\text{nucleophile},\alpha}\\ &\quad - \Delta N(\kappa-1)q^{(-)}_{\text{nucleophile},\alpha}\end{aligned} \quad (44)$$

For electrophiles,

$$\Xi^\kappa_{\Delta N\ge 0}(r_p) \sim \sum_{\alpha\in\text{electrophile}} \frac{\Xi^\kappa_{\Delta N\ge 0,\alpha}}{|r-R_\alpha|}$$

$$\begin{aligned}\Xi^\kappa_{\Delta N\ge 0,\alpha} &= -(\kappa+1)q^{(0)}_{\text{electrophile},\alpha} + \Delta N(\kappa-1)f^+_{\text{electrophile},\alpha}\\ &= [\Delta N(\kappa-1)-\kappa-1]q^{(0)}_{\text{electrophile},\alpha}\\ &\quad - \Delta N(\kappa-1)q^{(+)}_{\text{electrophile},\alpha}\end{aligned} \quad (45)$$

Nucleophiles will be susceptible to electrophilic attack at the atomic sites where $\Xi^\kappa_{\Delta N\le 0,\alpha}$ is small, and ideally negative. Electrophiles will be susceptible to nucleophilic attack at the atomic sites where $\Xi^\kappa_{\Delta N\ge 0,\alpha}$ is small.

E. Single-Parameter Variants of the General-Purpose Reactivity Indicator. Because our general-purpose reactivity indicator, $\Xi^\kappa_{\Delta N}(r)$, depends on two parameters (κ and ΔN), it is sometimes difficult to visualize the wealth of information it contains. In some contexts, then, it would be convenient to use a variant of $\Xi^\kappa_{\Delta N}(r)$ that depended only on the amount of electron transfer, because that is easily computed using electronegativity equalization schemes. Because the appropriate value of κ varies depending on the amount of electron transfer, one could introduce a one-parameter model by setting

$$\kappa \approx \begin{cases} 1+2(\Delta N) & \Delta N \le 0 \\ 1-2(\Delta N) & \Delta N \ge 0 \end{cases} \quad (46)$$

This model correctly predicts electrostatic control when $\Delta N \approx 0$ and electron-transfer control when $\Delta N \approx \pm 1$. Using eq 46, one obtains one-parameter models for the reactivity of nucleophiles

$$\Xi^{1+2\Delta N}_{\Delta N\le 0}(r_p) = 2\left[(1+\Delta N)\,\Phi_{\text{nucleophile}}(r_p) - (\Delta N)^2\, v^{f^-_{\text{nucleophile}}}(r_p)\right] \quad (47)$$

and electrophiles

$$\Xi^{1-2\Delta N}_{\Delta N\ge 0}(r_p) \equiv 2\left[(\Delta N-1)\,\Phi_{\text{electrophile}}(r_p) - (\Delta N)^2\, v^{f^+_{\text{electrophile}}}(r_p)\right] \quad (48)$$

While these single-parameter models are simpler than their two-parameter counterparts, our preliminary investigations indicate that, in molecules with multiple reactive sites, the single-parameter models sometimes fail to identify one or more reactive sites. For this reason, we will focus on the more general two-parameter models.

F. General-Purpose Reactivity Indicators: Discussion. We now discuss the interpretation of the general-purpose reactivity indicators for nucleophiles

$$\Xi^\kappa_{\Delta N\le 0}(r) \equiv (\kappa+1)\,\Phi_{\text{nucleophile}}(r) - \Delta N(\kappa-1)\, v^{f^-_{\text{nucleophile}}}(r) \quad (49)$$

and electrophiles

$$\Xi^\kappa_{\Delta N\ge 0}(r) \equiv -(\kappa+1)\,\Phi_{\text{electrophile}}(r) + \Delta N(\kappa-1)\, v^{f^+_{\text{electrophile}}}(r) \quad (50)$$

These indicators are designed to reproduce the qualitative features of the molecular interaction energy expression that we derived in section II.C, namely,

$$\begin{aligned}U_{\text{int}} &= (A_{\text{electrophile}} - I_{\text{nucleophile}})\Delta N\\ &\quad + \int\Big[\sum_{\alpha\in\text{nucleophile}} Z_\alpha\delta(r-R_\alpha) - \rho_{\text{nucleophile}}(r)\Big]\,\Phi_{\text{electrophile}}(r)\,dr\\ &\quad + \Delta N\int[f^+_{\text{electrophile}}(r)\,\Phi_{\text{nucleophile}}(r) - f^-_{\text{nucleophile}}(r)\,\Phi_{\text{electrophile}}(r)]\,dr\\ &\quad - (\Delta N)^2\int\int\frac{f^-_{\text{nucleophile}}(r)\,f^+_{\text{electrophile}}(r')}{|r-r'|}\,dr\,dr'\end{aligned} \quad (51)$$

Negative values of the interaction energy, U_{int}, are associated with favorable interactions between the electrophile and the nucleophile. Similarly, molecules will be most highly reactive where $\Xi^\kappa_{\Delta N}(r)$ is the smallest.

In accord with a suggestion by Langenaeker and coworkers,[56] each of our reactivity indicators is a linear combination of an appropriate reactivity index for hard−hard interactions (the electrostatic potential) and an appropriate reactivity index for soft−soft interactions (the Fukui function). The relative importance of these two contributions is controlled by the extent of charge transfer and the charge on the reactive site of the attacking reagent, as it should be. These are the key ingredients that make $\Xi^\kappa_{\Delta N}(r_p)$ a true general-purpose reactivity indicator: it can model the extreme cases of electrostatic and electron-transfer control, but it can also model the "in-between" cases. We will now explore each of these cases in more detail. In the interest of specificity and brevity, the following discussion is focused on the reactivity of nucleophiles. The results for electrophiles are broadly similar.

Case 1: Electrostatically Controlled Reactions; $\kappa \geq 1$. Electrostatically controlled (also called charge-controlled) reactions are typified by the following: (a) The reagents have large charges [so $q^{(0)}_{\text{electrophile}} >> 0$ and $\Phi_{\text{nucleophile}}(r_p) \ll 0$]. (b) Reactivity does not necessarily occur in places where the Fukui functions are large. ($f^{(+)}_{\text{electrophile}}$ may not be very big.) (c) The charge transfer between reagents is minimal ($\Delta N \approx 0$). Under these conditions, κ is greater than, or approximately equal to, 1. We call the case $\kappa = 1$ "pure electrostatic control" and the case $\kappa > 1$ "strong electrostatic control."

Case 1A: Pure Electrostatic Control; $\kappa = 1$. When $\kappa = 1$

$$\Xi^{\kappa=1}_{\Delta N\leq 0}(r_p) = 2\Phi_{\text{nucleophile}}(r_p) \tag{52}$$

Thus, for purely electrostatically controlled reactions, the appropriate reactivity indicator is the electrostatic potential. Furthermore, using the condensed version of the indicator

$$\Xi^{\kappa=1}_{\Delta N\leq 0,\alpha} = 2q^{(0)}_{\text{nucleophile},\alpha} \tag{53}$$

we infer that, in the limit of pure electrostatic control, nucleophiles react at the most negatively charged atomic site. (For electrophiles, the most reactive sites are those where the electrostatic potential/atomic charge is most positive.)

Case 1B: Strong Electrostatic Control; $\kappa > 1$. When $\kappa > 1$, the second term in eq 49 is usually very small because $\kappa - 1 < \kappa + 1$ and because the extent of electron transfer is very small ($\Delta N \approx 0$). For this reason, there can be no doubt that the most appropriate reactivity indicator for electrostatically controlled reactions is the electrostatic potential.[22,57]

What happens if there are two sites that are equivalent from an electrostatic perspective? Because $-\Delta N(\kappa - 1) > 0$ whenever $\kappa > 1$, $\Xi^{\kappa>1}_{\Delta N\leq 0}$ will be smallest in those places where the Fukui potential is the smallest. Insofar as the Fukui potential is smallest in places where the Fukui function is also small, this suggests that, *for reactions that are strongly electrostatically controlled, if there are two reactive sites with the same electrostatic favorability, then the site with minimum Fukui function is the most favorable.* This rule can also be derived from the condensed version of our indicator. In that case, *for electrophilic attack occurring under conditions associated with strong electrostatic control of chemical reactivity, a nucleophile possessing two sites with equal negative charge will be most reactive at the site with the smallest condensed Fukui function.* These rules might be called "minimum Fukui function tie-breaking rules" because they indicate that, if a reaction is strongly electrostatically controlled and there are two sites with equivalent electrostatic properties (and also other properties, notably including polarizability), then minimizing the Fukui function serves as a "tie breaker" between the two equivalent sites.

In 1995, Li and Evans proposed that hard−hard interactions occur where the Fukui function is small.[21] This counterintuitive result started a debate in the literature.[22,58−60] Our analysis establishes that it *is* true that sometimes having a small value for the Fukui function is favorable for reactivity, but the conditions under which this rule is valid are rather limited. There might be cases where the minimum Fukui function tie breaker is decisive in determining the reactivity. However, the second term in eq 49 is usually *much smaller* than the first term, so the minimum Fukui function tie breaker will *only* be operative when two hard reagents interact (so that $\Delta N \approx 0$), the electrophile is highly charged (so that $q^{(0)}_{\text{electrophile}}$, and thus κ, is large), and the nucleophile has two reactive sites with *very similar* electrostatic profiles. Unless the possible reaction sites are very similar electrostatically, the weak dependence of $\Xi^{\kappa>1}_{\Delta N\leq 0}$ on the Fukui function will not be decisive.

Even when it appears that the tie-breaking rule should apply, it may not. Because the dependence on the Fukui function is so weak, some of the interactions that were neglected in our model might be more important for determining the nucleophile's reactivity. The most important of these neglected interactions is probably the polarizability. Consider the Berkowitz−Parr formula for the polarizability kernel[33,61]

$$\left(\frac{\delta E}{\delta v(r)\,\delta v(r')}\right)_N = \left(\frac{\delta E}{\delta v(r)\,\delta v(r')}\right)_\mu + \left(\frac{\partial N}{\partial \mu}\right)_{v(r)} f(r)\,f(r') \tag{54}$$

or the approximate formula for the polarizability kernel based on the Kohn−Sham noninteracting reference system[33,62]

$$\begin{aligned}\left(\frac{\delta E}{\delta v(r)\,\delta v(r')}\right)_N &\approx \sum_i \sum_{j\neq i} \frac{n_j - n_i}{\epsilon_j - \epsilon_i}\phi_i^*(r)\,\phi_j(r)\,\phi_j^*(r')\,\phi_i(r') \\ &\approx \frac{\left[\begin{array}{l}-\phi_{\text{HOMO}}^*(r)\,\phi_{\text{LUMO}}(r)\,\phi_{\text{LUMO}}^*(r')\,\phi_{\text{HOMO}}(r') \\ -\phi_{\text{LUMO}}^*(r)\,\phi_{\text{HOMO}}(r)\,\phi_{\text{HOMO}}^*(r')\,\phi_{\text{LUMO}}(r')\end{array}\right]}{\epsilon_{\text{HOMO}} - \epsilon_{\text{LUMO}}}\end{aligned} \tag{55}$$

From these formulas, it is clear that a molecule tends to be most polarizable where the Fukui function (eq 54) and the frontier orbitals (eq 55) have a large amplitude. Given that strong electrostatic control will only occur for highly charged electrophiles, the second-order response to the change in external potential might not be especially small. Consequently, polarization effects may well cancel out, or even reverse, the predictions that would be obtained by naïvely applying the minimum Fukui function tie breaker.

The idea that the minimum Fukui function might correspond to the most reactive site goes against conventional wisdom, but it can be simply explained using the fundamental equation for the interaction energy, eq 51. Note that (i) the first line of this equation is a constant and does not influence regioselectivity. (ii) The second line of eq 51 models the electrostatic attraction between the electrophile and the nucleophile. If we are in a "tie-breaking" situation, this term is the same for electrophilic attack at both reactive sites. Thus, although the electrostatic interaction in the second line of this equation makes the biggest contribution to the interaction energy, it does not determine the regioselectivity. (iii) Recall from our discussion of eq 51 that both terms on the third line are positive. (iv) The first term on the third line of eq 51 does not distinguish between electrostatically equivalent sites on the nucleophile. (v) The term on the fourth

line of eq 51 is negligible. If we are in a strongly electrostatically controlled regime, the extent of electron transfer is small, and so, the $(\Delta N)^2$ term is entirely negligible. (The term on the fourth line might be important if a "second tie breaker" was needed, though.)

The remaining term in eq 51 must determine the regioselectivity. This term—the second term on the third line of eq 51—can be rewritten as

$$-\Delta N \int f^{-}_{\text{nucleophile}}(r)\, \Phi_{\text{electrophile}}(r)\, \mathrm{d}r \tag{56}$$

Equation 56 models how electron transfer quenches the electrostatic attraction between the nucleophile and the electrophile. The quenching arises because, after the nucleophile donates electrons to the electrophile, the electron density on the nucleophile decreases by $\Delta N f^{-}_{\text{nucleophile}}(r)$. Because the reactive sites of the nucleophile are less negatively charged after the electron transfer, the electrostatic attraction between the nucleophile and the electrophile is weaker after electron transfer than it was before. If electron transfer quenches the electrostatic attraction at one interaction site more than the other $[f^{-}_{\text{nucleophile}}(r_1) > f^{-}_{\text{nucleophile}}(r_2)]$, then the second interaction site—the one with "minimum Fukui function"—will be the most favorable reactive site.

It should be stressed, again, that *the minimum Fukui function rule is a "tie-breaking" rule; it is not a general-purpose reactivity rule.* The minimum Fukui function rule applies only when (1) the reaction is a strongly electrostatically controlled reaction (the term in the fourth line of eq 51 must be negligible!); (2) multiple reactive sites are equivalent electrostatically (the first term in eq 51 and the first term on the third line in eq 51 must fail to distinguish between the reactive sites.); (3) the Fukui-function term (the second term on the third line in eq 51) is more important than the effects, like polarization, that are neglected by our model.

A minimum Fukui function tie breaker is also operative in strongly electrostatically controlled nucleophilic attacks on electrophiles. In that case, it is the first term on the third line in eq 51 that determines the regioselectivity.

Case 2: Electron-Transfer-Controlled Reactions; $\kappa \leq -1$. Electron-transfer-controlled (also called Fukui-function-controlled and frontier-orbital-controlled) reactions are typified by the following: (a) The reactive sites have small charges (so $q^{(0)}_{\text{electrophile}} \approx 0$ and $\Phi_{\text{nucleophile}}(r_p) \approx 0$). (b) Reactivity occurs where the Fukui functions are large ($f^{(+)}_{\text{electrophile}} >> 0$). (c) Electron transfer between reagents is significant ($\Delta N \approx -1$). Under these conditions, κ is less than, or approximately equal to, -1. We call the case $\kappa = -1$ "pure electron-transfer control" and the case $\kappa < -1$ "strong electron-transfer control."

Case 2A: Pure Electron-Transfer Control; $\kappa = -1$. When $\kappa = -1$

$$\Xi^{\kappa=-1}_{\Delta N \leq 0}(r_p) = 2\Delta N \upsilon\, f^{-}_{\text{nucleophile}}(r_p) \tag{57}$$

Thus, for pure electron-transfer-controlled reactions, the appropriate reactivity indicator is the Fukui function's potential. Because $\Delta N < 0$, the preferred reactive site is the location where the Fukui potential is the largest. Because the Fukui potential is usually large in the same places that the Fukui function is large, electron-transfer-controlled reactions tend to occur where the Fukui function is large.

For the condensed version of the general-purpose reactivity indicator

$$\Xi^{\kappa=-1}_{\Delta N \leq 0,\alpha} = 2\Delta N f^{-}_{\text{nucleophile},\alpha} \tag{58}$$

On the basis of this, we infer that, for pure electron-transfer control, the atomic site with the largest condensed Fukui function is the most reactive. The situation for electrophiles is essentially the same, but in that case, it is the Fukui function from above, $f^{+}_{\text{electrophile},\alpha}$, that is relevant.

Case 2B: Strong Electron-Transfer Control; $\kappa < -1$. When $\kappa < -1$, the magnitude of the first term in eq 49 is usually much smaller than the second term because $|\kappa - 1| > |\kappa + 1|$ and the nucleophile is not highly charged in electron-transfer-controlled reactions ($q^{(0)}_{\text{nucleophile},\alpha} \approx 0$). For this reason, there can be no doubt that the most appropriate reactivity indicator for electron-transfer-controlled reactions is the Fukui potential or, alternatively, the Fukui function.[22,57]

What happens if there are two sites that have equivalent values of the Fukui potential? Because $\kappa + 1 < 0$ whenever $\kappa < -1$, $\Xi^{\kappa<-1}_{\Delta N \leq 0}$ will be smallest in those places where the electrostatic potential, $\Phi_{\text{nucleophile}}(r_p)$, is the largest. That is, reactivity will be favored at the more positive (or less negative) reactive sites. This counterintuitive electrostatic potential tie breaker is the analogue of the minimum Fukui function tie breaker for strongly electron-transfer-controlled reactions.

For nucleophiles, the "electrostatic potential tie-breaking rule" states that, *if an electrophilic attack reaction on a nucleophile is strongly electron-transfer-controlled and if there are two reactive sites with the same Fukui potential, then the site with the greatest electrostatic potential is the most favorable.* This principle can be restated in terms of condensed reactivity indicators: *for electrophilic attack occurring under conditions associated with strong electron-transfer control of the chemical reactivity, a nucleophile possessing two sites with equal condensed Fukui functions will be the most reactive at the site with the greatest charge.*

The analogous tie-breaking rule for electrophiles is, *if a nucleophilic attack reaction on a electrophile is strongly electron-transfer-controlled and if there are two reactive sites with the same Fukui potential, then the site with the smallest electrostatic potential is the most favorable.* The condensed version of this rule is, *for nucleophilic attack occurring under conditions associated with strong electron-transfer control of the chemical reactivity, an electrophile possessing two sites with equal condensed Fukui functions will be the most reactive at the site with the least charge.*

Just as we did for the minimum Fukui function tie breaker, we can elucidate the origins of the electrostatic potential tie breaker by studying the fundamental equation for the interaction energy, eq 51. Note that (i) the first line of this equation is a constant and does not influence regioselectivity. (ii) The fourth line of this equation models electron transfer from the nucleophile to the electrophile. Although this term usually dominates the regioselectivity, when we are in a "tie-

breaking" situation, the Fukui potential of the nucleophile is the same at both reactive sites. Consequently, this term does not determine the regioselectivity. (iii) Because the nucleophile's Fukui potential is similar at the two reactive sites, the second term on the third line does not contribute to the regioselectivity preference. (iv) Because we are in the strong electron-transfer-control limit, neither the electrophile nor the nucleophile are highly charged. The second line of eq 51 models electrostatic effects, but if magnitudes of the atomic charges on the electrophile and the atomic charges on the nucleophile are both small, the mutual attraction between these charges

$$E_{\text{int}}^{\text{electrostatic}} \approx \sum_{\alpha \in \text{nucleophile}} \sum_{\beta \in \text{electrophile}} \frac{q_{\text{nucleophile},\alpha} q_{\text{electrophile},\beta}}{|R_\alpha - R_\beta|} \tag{59}$$

should be very small.

The remaining term in eq 51

$$\Delta N \int f^{+}_{\text{electrophile}}(r)\, \Phi_{\text{nucleophile}}(r)\, \mathrm{d}r \tag{60}$$

must determine the regioselectivity. As electrons move from the nucleophile to the electrophile, the electron density on the electrophile increases by $-\Delta N f^{+}_{\text{electrophile}}(r)$. This causes the atomic charges on the electrophile to become less positive (or more negative) and decreases the electrostatic attraction between the nucleophile and the electrophile. (Because the electrophile was not highly charged to begin with, it is even conceivable that, after electron transfer, some key sites on the electrophile might be negatively charged. This corresponds to the extreme case where the electron transfer is so dramatic that the polarity of the nucleophile−electrophile bond is reversed.) Given a choice between two reactive sites with $\Phi_{\text{nucleophile}}(r_1) > \Phi_{\text{nucleophile}}(r_2)$, the most favorable site will be the first site, because this site is associated with a more favorable (or at least a less unfavorable) electrostatic interaction with the electrophile.

Case 3: Joint Electrostatic and Electron-Transfer Control, $-1 < \kappa < 1$. When κ is between -1 and $+1$, a nucleophile will be most reactive at places where the electrostatic potential is negative or the Fukui potential is big.

When κ is close to $+1$ or the extent of electron transfer is small ($\Delta N \approx 0$), reactivity preferences are predominately determined by the electrostatic potential. However, when two sites have similar electrostatic potentials, the most reactive site will be the site with the *largest* value of the Fukui potential. This should be contrasted with the extreme case of strong electrostatic control ($\kappa > 1$), where the site with the smallest value of the Fukui potential was the most favorable.

When κ is close to -1 and the extent of electron transfer is large ($\Delta N \approx -1$), reactivity preferences are predominately determined by the Fukui function. However, when two sites have similar Fukui functions, the site with the minimum electrostatic potential will be favored. This should be contrasted with the extreme case of strong electron-transfer control ($\kappa < -1$), where the site with maximum electrostatic potential would be favored.

In the intermediate regime, where $\kappa \approx 0$, the importance of electrostatic effects and electron-transfer effects are nearly balanced. In such cases, a reactive site that is negatively charged and has a reasonably positive value for the Fukui function might be favored over an uncharged site with a larger Fukui function. Similarly, such a site might be favored over an even more negatively charged site if that site was associated with a negligible value for the Fukui function.

Joint electrostatic and electron-transfer control seems to be one of the most common situations in chemical reactivity. For example, one might expect that the protonation of aminoethanol, $(H_2N)H_2CCH_2OH$, would be strongly electrostatically controlled: the proton is the prototypical hard acid,[63] and aminoethanol is not an especially soft base. It is observed that the electrostatic potential around the oxygen atom and that around the nitrogen atom are about the same.[41] Assuming strong electrostatic control, one would then infer that protonation occurs on the oxygen atom, because this is the site with minimum Fukui function. This is not the case; protonation occurs on the nitrogen atom, where the Fukui function is the largest.[41] Thus, even in cases where strong electrostatic control might be expected, one frequently observes joint electrostatic and electron-transfer control.

It is interesting to notice what happens if one neglects the error-correction terms in eq 31 and returns to the oversimplified reactivity indicator in eq 30. In that model, $\tilde{\kappa} > 0$ corresponds to strong electrostatic control and $\tilde{\kappa} < 0$ corresponds to strong electron-transfer control of the reactivity. This shows that, in a single reactive site interaction model, the possibility of joint electrostatic and electron-transfer control of the reactivity arises *because* the point-charge representation of the electrophile provides a more accurate approximation to the electrostatic effects (involving the electrostatic potential of the nucleophile) than it does to the electron-transfer effects (involving the Fukui potential of the nucleophile). If this was not the case, then eq 32 would not be valid. If this equation was not valid, then every reaction would be either strongly electrostatically controlled or strongly electron-transfer-controlled and a simple indicator like $\tilde{\Xi}^{\tilde{\kappa}}_{\Delta N \le 0}(r_p)$ (cf. eq 31) would be qualitatively correct. However, joint electrostatic and electron-transfer control is commonly observed, while situations that require the minimum-Fukui function and electrostatic potential tie-breaker rules are uncommon. This observation provides experimental evidence for the validity of eq 32 and strongly supports the error analysis in Appendix B.

G. Implications for the Local HSAB Principle. Our analysis provides a measure of support for the local hard/soft−acid/base principle.[44b,64−67] The local hard/soft−acid/base principle indicates that ambidentate ligands react with soft reagents in locations where the local softness is large and hard reagents in places where the local softness is small. [The local softness is just the global softness, S, times the Fukui function, $s(r) = Sf(r)$.] Our analysis indicates that soft reagents ($\kappa \approx -1$) should react with the molecule in the places where the Fukui function (and thus the local softness) is the largest. For hard reagents, the Fukui function is not an important indicator, and reactions could occur where the Fukui function is small. Hard reagents might also react with

the molecule in places where the Fukui function is large, however, because large values of the Fukui function are neither favorable nor unfavorable in electrostatically controlled reactions. The (probably extremely rare) exception is strongly electrostatic-controlled ($\kappa > 1$) reactions of molecules that have multiple electrostatically favorable sites. In that case, a small value of the Fukui function (and thus the local softness) would be preferred over a large value.

The overall picture is entirely consistent with the work of Klopman:[57] electrostatic effects [$\Phi(r_p)$ is the appropriate reactivity indicator] are typically dominant in reactions between hard reagents (where charge transfer is minimal); electron-transfer effects [$f^-(r_p)$ is the appropriate reactive indicator] are typically dominant in reactions between soft reagents (where substantial charge transfer occurs). For the intermediate cases, both effects are important.

IV. Recapitulation

It seems desirable to review what we have accomplished. Starting from the Taylor expansion for the potential energy surface of interacting electrophiles and nucleophiles (eq 21), we developed a "reactive site interaction model" based on the assumption that the interaction between the electrophile and the nucleophile is dominated by the interaction between their active sites. This led to a simplified indicator for the regioselectivity of the nucleophile (eq 27). This model was then subjected to a detailed error analysis and parametrized to obtain our final indicators: eq 35 (for the regioselectivity of nucleophiles) and eq 39 (for the regioselectivity of electrophiles). These expressions could then be condensed into indicators for the reactivity of different atoms in the molecule, giving eq 44 (for nucleophiles) and eq 45 (for electrophiles).

All of our reactivity indicators represent models for the interaction energies between the electrophile and the nucleophile; because of this, highly reactive sites are associated with negative values of the reactivity indicator, which we denote $\Xi^{\kappa}_{\Delta N}$. This model interaction energy clearly depends on two parameters. The first parameter, ΔN, is the amount of electron transfer. ΔN could be computed from the chemical potential and the hardnesses of the reagents[68] or, alternatively, on the basis of a quantum mechanical calculation of the product state (when the product of the chemical reaction is known). The second parameter, κ, quantifies whether the reaction is electrostatically controlled ($\kappa \geq 1$), electron-transfer-controlled ($\kappa \leq -1$), or somewhere in between ($-1 < \kappa < 1$). The relative values of κ for different reagents can be compared using the approximate proportionalities

$$\kappa_{\Delta N \leq 0} \propto q^{(0)}_{\text{electrophile}} + \Delta N f^{(+)}_{\text{electrophile}} \tag{61}$$

$$\kappa_{\Delta N \geq 0} \propto -(q^{(0)}_{\text{nucleophile}} + \Delta N f^{(-)}_{\text{nucleophile}}) \tag{62}$$

For highly charged electrophiles and small amounts of electron transfer, $\kappa \approx 1$ (electrostatic control). For weakly charged nucleophiles and significant electron transfer, $\kappa \approx -1$.

In most cases, the $\kappa \approx 1;|\Delta N| \approx 1$ case (electrostatic control and large amounts of electron transfer) is chemically irrelevant unless electrostatic effects are very, very strong. (This case could be important, for example, when the electrophile being reduced is a metal cation in a high oxidation state.) Similarly, the $\kappa \approx -1;\Delta N \approx 0$ case (electron-transfer control and negligible amounts of electron transfer) is chemically irrelevant unless electrostatic effects are weak. The $\kappa \approx 1;\Delta N \approx 0$ and $\kappa \approx -1;|\Delta N| \approx 1$ cases are very important limiting cases; they are associated with "classic" electrostatic and electron-transfer control, respectively. Most chemical reactions fall between those extremes; in these "jointly electrostatically and electron-transfer-controlled" reactions, the present indicator is preferable to existing approaches.

Using this reactivity indicator, we were able to gain some insight into appropriate indicators for different types of reactions. In the strong electrostatic control limit, $\kappa > 1$, it is observed that, given two sites with similar electrostatic potential, the reactive site with the smallest Fukui function is favored. Similarly, in the strong electron-transfer-control limit, $\kappa < 1$, it is observed that, given two sites with similar Fukui potentials, the reactive site with the greatest electrostatic potential (if the molecule is a nucleophile) or the least electrostatic potential (if the molecule is an electrophile) is most reactive. Both of these results are counterintuitive, and it is reassuring that, in the cases of greatest chemical relevance, the electrostatic potential is the dominant indicator for electrostatically controlled reactions and the condensed Fukui function is the dominant indicator for electron-transfer-controlled reactions. When the reactivity is between these two extremes, the most reactive site will be determined by a balance between the most favorable electrostatic potential and the most favorable Fukui potential, as one would expect.

When applying this model, the key assumptions that were made during its derivation are important: (i) the reactive site interaction model, that is, we assumed that the attacking reagent can be modeled as a point charge with a specified condensed Fukui function, and (ii) neglect of the polarization and other terms from higher-order derivatives with respect to the external potential.

The first assumption is required for any reactivity indicator. Because we are seeking a qualitative reactivity indicator, it is imperative that our model depend only on the coarsest details of the attacking reagent. This is in keeping with experimental evidence: most molecules react at only one or two places, regardless of the choice of reagent. The "details" of the reagents cannot be very important for determining the reactivity of the molecule that is attacked.[41]

The second assumption is merely pragmatic. We hope to incorporate polarization effects in our future work, but it is difficult to concoct a simple atom-condensed reactivity indicator that depends on a two-point quantity like the polarizability kernel. In addition, we believe that the most useful reactivity indicators are those that are easily evaluated using the output of standard quantum chemistry programs. The present reactivity indicator is easily computed from the atomic charges, which is a standard feature in quantum chemistry codes. In contrast, we do not know any simple

way to extract a condensed polarizability kernel from the output file of a quantum chemistry program.

Acknowledgment. Helpful discussions with Dr. David C. Thompson are acknowledged. NSERC, the Canada Research Chairs, and PREA provided funding for the Canadian authors. This research was performed when the second author visited McMaster University in the winter of 2005, and she wishes to thank the chemistry department at McMaster University for their hospitality.

Appendix A. Derivation of the Electrostatic Potential Contribution

The key formula is the expression for the nuclear−nuclear repulsion energy in ref 35

$$V_{nn}[v] = \frac{1}{32\pi^2}\int\int_{\mathbf{r}\neq\mathbf{r}'} \frac{[\nabla_r{}^2 v(r)][\nabla_{r'}{}^2 v(r')]}{|r-r'|}\,\mathrm{d}r\,\mathrm{d}r' \quad (63)$$

We can derive $\delta V_{nn}[v]/\delta v(r)$ by finding the coefficient of $\delta v(r)$ in the expression $V_{nn}[v+\delta v] - V_{nn}[v]$.

$$\begin{aligned} &V_{nn}[v+\delta v] - V_{nn}[v] \\ &= \frac{1}{32\pi^2}\left[\int\int_{r\neq r'} \frac{\nabla_r{}^2[v(r)+\delta v(r)]\nabla_{r'}{}^2[v(r')+\delta v(r')]}{|r-r'|}\,\mathrm{d}r\,\mathrm{d}r' \right. \\ &\left. - \int\int_{r\neq r'} \frac{\nabla_r{}^2 v(r)\,\nabla_{r'}{}^2 v(r')}{|r-r'|}\,\mathrm{d}r\,\mathrm{d}r'\right] \\ &= \left\{\frac{1}{32\pi}\left[\int\int_{r\neq r'} \frac{\nabla_r{}^2\,\delta v(r)\,\nabla_{r'}{}^2\,v(r')}{|r-r'|}\,\mathrm{d}r\,\mathrm{d}r' \right.\right. \\ &\left.\left. + \int\int_{r\neq r'} \frac{\nabla_{\mathbf{r}}{}^2 v(\mathbf{r})\nabla_{\mathbf{r}'}{}^2\delta v(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\right]\right\} \\ &+ \left[\frac{1}{32\pi^2}\int\int_{r\neq r'} \frac{\nabla_r{}^2\,\delta v(r)\,\nabla_{r'}{}^2\,\delta v(r')}{|r-r'|}\,\mathrm{d}r\,\mathrm{d}r'\right] \quad (64) \end{aligned}$$

The first term (in braces) represents the electrostatic interaction between the change in the external potential and the pre-existing charge density. The second term (in brackets) represents the self-repulsion energy of the perturbation. In an application such as ours, this term is neglected because it is already included in the energy expression for the attacking reagent. (If we included this term, we would make a "double-counting" error.) In any event, the term in brackets is second-order in the perturbing potential and will not contribute to the first functional derivative. Neglecting this term and noting that the two integrals in the braces have identical values, we can simplify eq 64 to

$$\begin{aligned} &V_{nn}[v+\delta v] - V_{nn}[v] \\ &= \frac{1}{16\pi^2}\left(\int\int_{r\neq r'} \frac{\nabla_r{}^2\,\delta v(r)\,\nabla_{r'}{}^2\,v(r')}{|r-r'|}\,\mathrm{d}r\,\mathrm{d}r'\right) \quad (65) \end{aligned}$$

This equation is simplified by writing the functional variation as a nested integration

$$\delta_v[V_{nn},\delta v] = \frac{1}{16\pi^2}\int \nabla_{r'}{}^2 v(r')\left(\int \frac{\nabla_r{}^2\delta v(r)}{|r-r'|}\,\mathrm{d}r\right)\mathrm{d}r' \quad (66)$$

and then simplifying the inner integral using Green's theorem

$$\begin{aligned} \int\int\int \frac{1}{|r-r'|}[\nabla_r{}^2\delta v(r)]\,\mathrm{d}r &= \int\int\int\left(\nabla_r{}^2\frac{1}{|r-r'|}\right)\delta v(r)\,\mathrm{d}r \\ &+ \int\int \frac{1}{|r-r'|}[\nabla_r\delta v(r)]\cdot n\mathrm{d}a - \int\int \delta v(r)\left(\nabla_r\frac{1}{|r-r'|}\right)\cdot n\mathrm{d}a \end{aligned} \quad (67)$$

Our system is defined over all space, so in eq 67 we choose the surface to be a sphere infinitely far from the origin. We assume that the change in charge density associated with the change in external potential, $\delta q(r) = -{}^1/_{4\pi}\nabla_r{}^2\delta v(r)$, is relatively localized, so that we can use the asymptotic form $\delta v(r) \sim (\langle\delta q\rangle/r)$. With this assumption, the surface integrals give

$$\begin{aligned} &\lim_{r\to\infty}\int\int \frac{1}{|r-r'|}[\nabla\delta v(r)]\cdot\hat{n}r^2 d\Omega = \lim_{r\to\infty}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{\langle\delta q\rangle}{r}\right)\right](4\pi r^2) = 0 \\ &\lim_{r\to\infty}\int\int \delta v(r)\left(\nabla_r\frac{1}{|r-r'|}\right)\cdot\hat{n}\mathrm{d}a = \lim_{r\to\infty}\left[\frac{\langle\delta q\rangle}{r}\frac{\partial}{\partial r}\left(\frac{1}{r}\right)\right](4\pi r^2) = 0 \end{aligned} \quad (68)$$

Equation 68 shows that the surface terms vanish if $\delta v(r)$ falls off to zero at infinity at least as fast as r^{-1}. We simplify the volume integral in eq 67 using Poisson's equation for a point charge, obtaining

$$\begin{aligned} \int\int\int\left(\nabla_r{}^2\frac{1}{|r-r'|}\right)\delta v(r)\,\mathrm{d}r &= \int\int\int[-4\pi\delta(r-r')]\delta v(r)\,\mathrm{d}r \\ &= -4\pi\delta v(r') \end{aligned} \quad (69)$$

Substitute eqs 68 and 69 into eq 67; then, substitute that result into eq 65 to obtain

$$V_{nn}[v+\delta v] - V_{nn}[v] = \frac{-1}{4\pi}[\int\delta v(r')\,\nabla_{r'}{}^2 v(r')\,\mathrm{d}r'] \quad (70)$$

The functional derivative is then

$$\begin{aligned} \frac{\delta V_{nn}[v]}{\delta v(r)} &= \frac{-1}{4\pi}\nabla_r{}^2 v(r) \\ &= -\rho_{\mathrm{nuc}}(r) \end{aligned} \quad (71)$$

where $-\rho_{nuc}(r) = -\sum_\alpha Z_\alpha\delta(r-R_\alpha)$ is -1 times the nuclear charge density.

It is now easy to see how the electrostatic potential arises as a reactivity indicator in density-functional theory. Consider the change in total energy due to adding a small point charge, q, at the point r_p:

$$\begin{aligned} \Delta(E+V_{nn}) &= \int\left[\left(\frac{\delta E}{\delta v(r)}\right)_N + \frac{\delta V_{nn}}{\delta v(r)}\right]\frac{q}{|r-r_p|}\,\mathrm{d}r \\ &= q\int\frac{\rho(r) - \sum_\alpha Z_\alpha\delta(r-R_\alpha)}{|r-r_p|}\,\mathrm{d}r \\ &= -q\Phi(r_p) \end{aligned} \quad (72)$$

Appendix B. Error Analysis for the Reactive Site Interaction Model

This appendix performs an error analysis for the single reactive site interaction approximation to eq 21. The errors

incurred by replacing the integrals in eq 21 with the point charge representations in eq 27 are conveniently summarized using the following expressions:

$$-\epsilon_{(i)}\Phi_{\text{nucleophile}}(r_p) \equiv \int\left[\sum_{\beta\in\text{electrophile}} Z_\beta\delta(r-R_\beta) - \rho_{\text{electrophile}}(r) - q^{(0)}_{\text{electrophile}}\delta(r-r_p)\right]\Phi_{\text{nucleophile}}(r)\,dr \quad (73)$$

$$\epsilon_{(ii)}\Phi_{\text{nucleophile}}(r_p) \equiv [f^{+}_{\text{electrophile}}(r) - f^{(+)}_{\text{electrophile}}\delta(r-r_p)]\Phi_{\text{nucleophile}}(r)\,dr \quad (74)$$

$$-\epsilon_{(iii)}\upsilon^{f^-}_{\text{nucleophile}}(r_p) \equiv \int\left[\sum_{\beta\in\text{electrophile}} Z_\beta\delta(r-R_\beta) - \rho_{\text{electrophile}}(r) - q^{(0)}_{\text{electrophile}}\delta(r-r_p)\right]\upsilon^{f^-}_{\text{nucleophile}}(r)\,dr \quad (75)$$

$$\epsilon_{(iv)}\upsilon^{f^-}_{\text{nucleophile}}(r_p) \equiv \int\left[f^{+}_{\text{electrophile}}(r) - f^{(+)}_{\text{electrophile}}\delta(r-r_p)\right]\upsilon^{f^-}_{\text{nucleophile}}(r)\,dr \quad (76)$$

We now rationalize these forms. It is important to recognize that each of these integrals is really a Coulomb integral associated with the attractive forces that electrons feel toward nuclei and the repulsive forces they feel toward other electrons. For example, eq 74 could be rewritten as

$$\epsilon_{(ii)}\Phi_{\text{nucleophile}}(r_p) \equiv \iint\frac{[f^{+}_{\text{electrophile}}(r) - f^{(+)}_{\text{electrophile}}\delta(r-r_p)][\sum_{\alpha\in\text{nucleophile}} Z_\alpha\delta(r'-R_\alpha) - \rho_{\text{nucleophile}}(r')]}{|r-r'|}\,dr\,dr' \quad (77)$$

In an expression like this, it seems reasonable to assume that the largest error is associated with the approximation to the electron−electron repulsion-type term

$$-\iint\frac{f^{+}_{\text{electrophile}}(r)\,\rho_{\text{nucleophile}}(r')}{|r-r'|}\,dr\,dr' \approx -\iint\frac{f^{+}_{\text{electrophile}}\,\delta(r-r_p)\,\rho_{\text{nucleophile}}(r')}{|r-r'|}\,dr\,dr' \quad (78)$$

instead of with the approximation to the electron−nuclear attraction-type term

$$\iint\frac{f^{+}_{\text{electrophile}}(r)[\sum_{\alpha\in\text{nucleophile}} Z_\alpha\delta(r'-R_\alpha)]}{|r-r'|}\,dr\,dr' \approx \iint\frac{f^{+}_{\text{electrophile}}\delta(r-r_p)[\sum_{\alpha\in\text{nucleophile}} Z_\alpha\delta(r'-R_\alpha)]}{|r-r'|}\,dr\,dr' \quad (79)$$

Because the integrands in eq 78 are very large when r and r' are close together, the point charge approximation does not provide a good representation for the interaction between the asymptotic tails of $\rho_{\text{nucleophile}}(r)$ and $f^{+}_{\text{electrophile}}(r)$. This suggests that (1) the left-hand side of eq 78 is less than the right-hand side and (2) the error in eq 78 is significantly bigger than the error in eq 79. Extending this argument to the other integrals in eqs 73−76, we see the following:

(a) In general, the left-hand side in eqs 73 and 76 should be positive. This implies that, in general, $\epsilon_{(i)} > 0$ *and* $\epsilon_{(iv)} > 0$.

(b) In general, the left-hand side in eqs 74 and 75 should be negative. This implies that, in general, $\epsilon_{(ii)} > 0$ *and* $\epsilon_{(iii)} > 0$.

The sign convention in eqs 73−76 is based on these observations. Recall that the electrostatic potential of a nucleophile is usually negative near a reactive site [$\Phi_{\text{nucleophile}}(r_p) < 0$] while the potential due to the Fukui function is usually positive near a reactive site [$\upsilon^{f^-}_{\text{nucleophile}}(r_p) > 0$]. Referring back to a and b, it is apparent that the sign convention in eqs 73−76 implies that $\epsilon_{(i)}$, $\epsilon_{(ii)}$, $\epsilon_{(iii)}$, and $\epsilon_{(iv)}$ are usually positive constants. We will now characterize the size of these constants.

When eqs 73 and 74 are examined, it seems clear that the size of the error should be related to the magnitude of $\Phi_{\text{nucleophile}}(r)$ in the region where the electrophile is attacking. Similarly, in eqs 75 and 76, the error incurred by the point charge approximation should be small when $\upsilon^{f^-}_{\text{nucleophile}}(r)$ is small in the region where the electrophile is attacking. On the other hand, if $\upsilon^{f^-}_{\text{nucleophile}}(r)$ is large at the reactive site, this suggests that $f^{-}_{\text{nucleophile}}(r)$ is also large near the reactive site, which suggests that the point charge approximation to the integrals in eq 75 will be poor.

Hence, (c) the error in the left-hand sides of eqs 73 and 74 is roughly proportional to the magnitude of the electrostatic potential at the position where the electrophile attacks, $\Phi_{nucleophile}(r_p)$.

(d) The error in the left-hand sides of eqs 75 and 76 is roughly proportional to the magnitude of the Fukui potential at the position where the electrophile attacks, $\upsilon^{f^-}_{nucleophile}(r_p)$.

By including an appropriate dependence on $\Phi_{\text{nucleophile}}(r_p)$ or $\upsilon^{f^-}_{\text{nucleophile}}(r_p)$ in the right-hand side of the defining eqs 73−76, we ensure that $\epsilon_{(i)}$−$\epsilon_{(iv)}$ depend only weakly on the relative magnitude of the electrostatic and Fukui potentials at the reactive sites. Our choice of definition, then, ensures that $\epsilon_{(i)}$−$\epsilon_{(iv)}$ measure the *intrinsic error* in the point charge approximations.

In the discussion surrounding eq 78, we pointed out that the primary error in the single reactive site approximation was associated with the interaction between the tails of electronic distributions centered on the electrophile and the nucleophile. This error is expected to be most severe in eq 76, because the Fukui functions (1) have slow asymptotic decays and (2) are concentrated on the "frontiers" of the reagents.

(e) We expect $\epsilon_{(iv)}$ *to be relatively large. The error due to the point charge approximation can be rather large in this case because the interaction between the asymptotic tails of the Fukui functions might be significantly underestimated by the result from the point charge representation.*

By contrast, the error in eq 73 might be relatively small, because the electron density is concentrated near the atomic nuclei.

(f) We expect $\epsilon_{(i)}$ *to be relatively small, because the interaction between the asymptotic tails of the electron*

densities should be only a small part of the total interaction between the electronic distributions.

The errors in eqs 74 and 75 should be intermediate.

(g) We expect $\epsilon_{(ii)}$ and $\epsilon_{(iii)}$ to be relatively large compared to $\epsilon_{(i)}$ because the Fukui function in these integrals is concentrated on the frontiers of the molecule. However, we expect $\epsilon_{(ii)}$ and $\epsilon_{(iii)}$ to be smaller than $\epsilon_{(iv)}$, because point charge approximation is the least accurate in eq 76, where both of the Fukui functions are involved.

Using results from a, b, e, f, and g, we have the following ordering of relative errors.

$$0 < \epsilon_{(i)} < \epsilon_{(ii)} \approx \epsilon_{(iii)} < \epsilon_{(iv)} \tag{80}$$

It should be stressed that the preceding analysis is strictly qualitative. Exceptions to the ordering in eq 80 will occur. Henceforth, we will never rely upon the details of this analysis; the approximate ordering of errors in eq 80 is sufficient to establish our results.

As discussed in sections III.A and III.B, the key parameters in the reactivity model are the number of electrons donated by the nucleophile to the electrophile, $\Delta N \leq 0$, and the quantity

$$\tilde{\kappa} = q^{(0)}_{\text{electrophile}} + \Delta N f^{(+)}_{\text{electrophile}} \tag{81}$$

Inserting the expression for $\tilde{\kappa}$ into eq 27 gave the oversimplistic regioselectivity indicator in eq 30.

$\tilde{\kappa}$ measures the relative importance of electrostatic and electron-transfer effects. When $\tilde{\kappa}$ is significantly positive, the electrophile is highly charged and electron transfer is minimal; such reactions are expected to be electrostatically controlled. When $\tilde{\kappa}$ is significantly negative, then electron transfer to the electrophilic site is important, but the electrophile is not especially highly charged; such reactions are expected to be electron-transfer-controlled.

As long as $\tilde{\kappa}$ is very different from zero, it is reasonable to neglect the errors due to the point-charge approximations to the integrals, because they should be small compared to $\tilde{\kappa}$. To address the case where $\tilde{\kappa} \approx 0$, insert the expressions for $\epsilon_{(i)}$, $\epsilon_{(ii)}$, $\epsilon_{(iii)}$, and $\epsilon_{(iv)}$ into eq 21. This gives an "error-corrected" version of the reactive-site interaction model in eq 30, namely,

$$\tilde{\Xi}^{\tilde{\kappa},\epsilon_{(i)}-\epsilon_{(iv)}}_{\Delta N \leq 0}(r_p) \equiv [\tilde{\kappa} + (-\epsilon_{(i)} + \epsilon_{(ii)}\Delta N)]\, \Phi_{\text{nucleophile}}(r_p) - \Delta N[\tilde{\kappa} + (-\epsilon_{(iii)} + \epsilon_{(iv)}\Delta N)]\, v^{f^-}_{\text{nucleophile}}(r_p) \tag{82}$$

Because $\epsilon_{(i)} < \epsilon_{(iii)}$, $\epsilon_{(ii)} < \epsilon_{(iv)}$, and $\Delta N \leq 0$,

$$-\epsilon_{(iii)} + \epsilon_{(iv)}\Delta N < -\epsilon_{(i)} + \epsilon_{(ii)}\Delta N \tag{83}$$

Comparing eqs 31 and 82 allows us to make the identification

$$\epsilon_\rho = -\epsilon_{(i)} + \epsilon_{(ii)}\Delta N \tag{84}$$

$$\epsilon_f = -\epsilon_{(iii)} + \epsilon_{(iv)}\Delta N \tag{85}$$

Equation 83 then establishes that $\epsilon_\rho > \epsilon_f$, as stated in eq 32. The general-purpose reactivity indicator then follows from the analysis in section III.B. In particular, because

$$\tilde{\kappa} - \epsilon_{(i)} + \epsilon_{(ii)}\Delta N > \tilde{\kappa} - \epsilon_{(iii)} + \epsilon_{(iv)}\Delta N \tag{86}$$

the *qualitative* structure of the interaction energy model in eq 82 is recaptured by the simple expression

$$\Xi^{\kappa}_{\Delta N \leq 0}(r_p) \equiv (\kappa + 1)\Phi_{\text{nucleophile}}(r_p) - \Delta N(\kappa - 1) v^{f^-}_{\text{nucleophile}}(r_p) \tag{87}$$

The new parameter, κ, is linearly related to the more fundamental $\tilde{\kappa}$ via the equation

$$\kappa = 2\left(\frac{\tilde{\kappa} - \epsilon_{(i)} + \epsilon_{(ii)}\Delta N}{\epsilon_{(iii)} - \epsilon_{(i)} + (\epsilon_{(ii)} - \epsilon_{iv})\Delta N}\right) - 1 \tag{88}$$

The approximate proportionality reported in eq 38 is clearly accurate whenever $\tilde{\kappa}$ is large compared to the error terms. The motivation for this transformation of variables is that it gives a simple appealing qualitative picture, with $\kappa = +1$ and $\kappa = -1$ corresponding to pure electrostatic control and pure electron-transfer control, respectively.

References

(1) Albright, T. A.; Burdett, J. K.; Whangbo, M. H. *Orbital Interactions in Chemistry*; Wiley-Interscience: New York, 1985.

(2) Fukui, K. Role of Frontier Orbitals in Chemical Reactions. *Science* **1987**, *218*, 747−754.

(3) Fukui, K.; Yonezawa, T.; Nagata, C. A Free-Electron Model for Discussing Reactivity in Unsaturated Hydrocarbons. *J. Chem. Phys.* **1953**, *21* (1), 174−176.

(4) Fukui, K.; Yonezawa, T.; Shingu, H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J. Chem. Phys.* **1952**, *20* (4), 722−725.

(5) Fukui, K.; Yonezawa, T.; Nagata, C. Theory of Substitution in Conjugated Molecules. *Bull. Chem. Soc. Jpn.* **1954**, *27* (7), 423−427.

(6) Woodward, R. B.; Hoffmann, R. Conservation of Orbital Symmetry. *Angew. Chem., Int. Ed. Engl.* **1969**, *8* (11), 781−853.

(7) Hoffmann, R.; Woodward, R. B. Conservation of Orbital Symmetry. *Acc. Chem. Res.* **1968**, *1* (1), 17−22.

(8) Hoffmann, R.; Woodward, R. B. Selection Rules for Concerted Cycloaddition Reactions. *J. Am. Chem. Soc.* **1965**, *87* (9), 2046−2048.

(9) Woodward, R. B.; Hoffmann, R. Selection Rules for Sigmatropic Reactions. *J. Am. Chem. Soc.* **1965**, *87* (11), 2511−2513.

(10) Woodward, R. B.; Hoffmann, R. Stereochemistry of Electrocyclic Reactions. *J. Am. Chem. Soc.* **1965**, *87* (2), 395−397.

(11) Parr, R. G.; Yang, W. Density Functional Approach to the Frontier-Electron Theory of Chemical Reactivity. *J. Am. Chem. Soc.* **1984**, *106* (14), 4049−4050.

(12) Yang, W.; Parr, R. G.; Pucci, R. Electron Density, Kohn−Sham Frontier Orbitals, and Fukui Functions. *J. Chem. Phys.* **1984**, *81* (6), 2862−2863.

(13) Ayers, P. W.; Levy, M. Perspective on "Density Functional Approach to the Frontier-Electron Theory of Chemical Reactivity" by Parr RG, Yang W (1984). *Theor. Chem. Acc.* **2000**, *103* (3−4), 353−360.

(14) Ayers, P. W.; Parr, R. G. Variational Principles for Describing Chemical Reactions: The Fukui Function and Chemical Hardness Revisited. *J. Am. Chem. Soc.* **2000**, *122* (9), 2010−2018.

(15) Melin, J.; Ayers, P. W.; Ortiz, J. V. The Electron-Propagator Approach to Conceptual Density-Functional Theory. *J. Chem. Sci.* **2005**, *117* (5), 387−400.

(16) Bartolotti, L. J.; Ayers, P. W. An Example where Orbital Relaxation Is an Important Contribution to the Fukui Function. *J. Phys. Chem. A* **2005**, *109* (6), 1146−1151.

(17) Cerjan, C. J.; Miller, W. H. On Finding Transition-States. *J. Chem. Phys.* **1981**, *75* (6), 2800−2806.

(18) Simons, J.; Jorgensen, P.; Taylor, H.; Ozment, J. Walking on Potential-Energy Surfaces. *J. Phys. Chem.* **1983**, *87* (15), 2745−2753.

(19) Dewar, M. J. S. A Critique of Frontier Orbital Theory. *THEOCHEM* **1989**, *59*, 301−23.

(20) Anderson, J. S. M.; Melin, J.; Ayers, P. W. Conceptual Density-Functional Theory for General Chemical Reactions, Including Those That Are Neither Charge- nor Frontier-Orbital-Controlled. 2. Application to Molecules Where Frontier Molecular Orbital Theory Fails. *J. Chem. Theory Comput.* **2007**, *3*, 375−389.

(21) Li, Y.; Evans, J. N. S. The Fukui Function − A Key Concept Linking Frontier Molecular-Orbital Theory and the Hard−Soft−Acid−Base Principle. *J. Am. Chem. Soc.* **1995**, *117* (29), 7756−7759.

(22) Melin, J.; Aparicio, F.; Subramanian, V.; Galvan, M.; Chattaraj, P. K. Is the Fukui Function a Right Descriptor of Hard−Hard Interactions? *J. Phys. Chem. A* **2004**, *108* (13), 2487−2491.

(23) Coulson, C. A.; Longuet-Higgins, H. C. The Electronic Structure of Conjugated Systems. I. General Theory. *Proc. R. Soc. London, Ser. A* **1947**, *192*, 39−60.

(24) Coulson, C. A.; Longuet-Higgins, H. C. The Electronic Structure of Conjugated Systems. II. Unsaturated Hydrocarbons and Their Hetero-Derivatives. *Proc. R. Soc. London, Ser. A* **1947**, *192*, 16−32.

(25) Coulson, C. A.; Longuet-Higgins, H. C. The Electronic Structure of Conjugated Systems. III. Bond Orders in Unsaturated Molecules. *Proc. R. Soc. London, Ser. A* **1948**, *193*, 447−456.

(26) Coulson, C. A.; Longuet-Higgins, H. C. The Electronic Structure of Conjugated Systems. IV. Force Constants and Interaction Constants in Unsaturated Hydrocarbons. *Proc. R. Soc. London, Ser. A* **1948**, *193*, 456−464.

(27) Wolfsberg, M.; Helmolz, L. The Spectra and Electronic Structure of Tetrahedral Ions MnO_4^-, CrO_4^{-2} and ClO_4^-. *J. Chem. Phys.* **1952**, *20* (5), 837−843.

(28) Nalewajski, R. F.; Parr, R. G. Legendre Transforms and Maxwell Relations in Density Functional Theory. *J. Chem. Phys.* **1982**, *77* (1), 399−407.

(29) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L., Jr. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49* (23), 1691−1694.

(30) Yang, W.; Zhang, Y.; Ayers, P. W. Degenerate Ground States and Fractional Number of Electrons in Density and Reduced Density Matrix Functional Theory. *Phys. Rev. Lett.* **2000**, *84* (22), 5172−5175.

(31) Zhang, Y.; Yang, W. Perspective on "Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy". *Theor. Chem. Acc.* **2000**, *103* (3−4), 346−348.

(32) Cohen, M. H.; Ganduglia-Pirovano, M. V. Electronic and Nuclear Chemical Reactivity. *J. Chem. Phys.* **1994**, *101* (10), 8988−97.

(33) Ayers, P. W. Strategies for Computing Chemical Reactivity Indices. *Theor. Chem. Acc.* **2001**, *106* (4), 271−279.

(34) The contribution to molecular interaction energies from induced dipoles is normally less than 1 kcal/mol.

(35) Ayers, P. W.; Parr, R. G. Variational Principles for Describing Chemical Reactions. Reactivity Indices Based on the External Potential. *J. Am. Chem. Soc.* **2001**, *123* (9), 2007−2017.

(36) Ayers, P. W. Atoms in Molecules, an Axiomatic Approach. I. Maximum Transferability. *J. Chem. Phys.* **2000**, *113* (24), 10886−10898.

(37) Berkowitz, M. Density Functional-Approach to Frontier Controlled Reactions. *J. Am. Chem. Soc.* **1987**, *109* (16), 4823−4825.

(38) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103* (5), 1793−1873.

(39) Toro-Labbe, A.; Jaque, P.; Murray, J. S.; Politzer, P. Connection between the Average Local Ionization Energy and the Fukui Function. *Chem. Phys. Lett.* **2005**, *407* (1−3), 143−146.

(40) Toro-Labbe, A.; Gutierrez-Oliva, S.; Concha, M. C.; Murray, J. S.; Politzer, P. Analysis of Two Intramolecular Proton-Transfer Processes in Terms of the Reaction Force. *J. Chem. Phys.* **2004**, *121* (10), 4570−4576.

(41) Ayers, P. W.; Anderson, J. S. M.; Bartolotti, L. J. Perturbative Perspectives on the Chemical Reaction Prediction Problem. *Int. J. Quantum Chem.* **2005**, *101* (5), 520−534.

(42) Yang, W.; Mortier, W. J. The Use of Global and Local Molecular Parameters for the Analysis of the Gas-Phase Basicity of Amines. *J. Am. Chem. Soc.* **1986**, *108* (19), 5708−11.

(43) Ayers, P. W.; Morrison, R. C.; Roy, R. K. Variational Principles for Describing Chemical Reactions: Condensed Reactivity Indices. *J. Chem. Phys.* **2002**, *116* (20), 8731−8744.

(44) (a) Ayers, P. W.; Parr, R. G.; Pearson, R. G. Elucidating the Hard/Soft Acid/Base Principle: A Perspective Based on Half-Reactions. *J. Chem. Phys.* **2006**, *124* (19), 194107. (b) Ayers, P. W. The Physical Basis of the Global and Local Hard/Soft Acid/Base Principles. *Faraday Discuss.* **2007**, *135*, 161−190. (c) Ayers, P. W. An Elementary Derivation of the Hard/Soft Acid/Base Principle. *J. Chem. Phys.* **2005**, *122*, 141102.

(45) Ayers, P. W.; Anderson, J. S. M.; Rodriguez, J. I.; Jawed, Z. Indices for Predicting the Quality of Leaving Groups. *Phys. Chem. Chem. Phys.* **2005**, *7* (9), 1918−1925.

(46) Liu, S. B.; Parr, R. G. Additional Functional Relations in the Density Functional Theory of Finite Interacting Electronic Systems. *Chem. Phys. Lett.* **1997**, *278* (4−6), 341−344.

(47) De Proft, F.; Liu, S. B.; Parr, R. G. Chemical Potential, Hardness, Hardness and Softness Kernel and Local Hardness in the Isomorphic Ensemble of Density Functional Theory. *J. Chem. Phys.* **1997**, *107* (8), 3000−3006.

(48) Parr, R. G.; Liu, S. B. Some Functional Relations in the Density Functional Theory of Finite Interacting Electronic Systems. *Chem. Phys. Lett.* **1997**, *276* (1−2), 164−166.

(49) Senet, P. Kohn−Sham Orbital Formulation of the Chemical Electronic Responses, Including the Hardness. *J. Chem. Phys.* **1997**, *107* (7), 2516−2524.

(50) Senet, P. Nonlinear Electronic Responses, Fukui Functions and Hardnesses as Functionals of the Ground-State Electronic Density. *J. Chem. Phys.* **1996**, *105* (15), 6471−6489.

(51) Roy, R. K.; Pal, S.; Hirao, K. On Non-Negativity of Fukui Function Indices. *J. Chem. Phys.* **1999**, *110* (17), 8236−8245.

(52) Roy, R. K.; Hirao, K.; Pal, S. On Non-Negativity of Fukui Function Indices. II. *J. Chem. Phys.* **2000**, *113* (4), 1372−1379.

(53) Bultinck, P.; Carbo-Dorca, R. Negative and Infinite Fukui Functions: The Role of Diagonal Dominance in the Hardness Matrix. *J. Math. Chem.* **2003**, *34* (1−2), 67−74.

(54) Bultinck, P.; Carbo-Dorca, R.; Langenaeker, W. Negative Fukui Functions: New Insights Based on Electronegativity Equalization. *J. Chem. Phys.* **2003**, *118* (10), 4349−4356.

(55) Ayers, P. W. Can One Oxidize an Atom by Reducing the Molecule that Contains It? *Phys. Chem. Chem. Phys.* **2006**, *8*, 3387−3390.

(56) Langenaeker, W.; Deproft, F.; Geerlings, P. Development of Local Hardness Related Reactivity Indexes − Their Application in A Study of the Se at Monosubstituted Benzenes Within the Hsab Context. *J. Phys. Chem.* **1995**, *99* (17), 6424−6431.

(57) Klopman, G. Chemical Reactivity and the Concept of Charge and Frontier-Controlled Reactions. *J. Am. Chem. Soc.* **1968**, *90*, 223−234.

(58) Perez, P.; Simon-Manso, Y.; Aizman, A.; Fuentealba, P.; Contreras, R. Empirical Energy−Density Relationships for the Analysis of Substituent Effects in Chemical Reactivity. *J. Am. Chem. Soc.* **2000**, *122* (19), 4756−4762.

(59) Pal, S.; Chandrakumar, K. R. S. Critical Study of Local Reactivity Descriptors for Weak Interactions: Qualitative and Quantitative Analysis of Adsorption of Molecules in the Zeolite Lattice. *J. Am. Chem. Soc.* **2000**, *122* (17), 4145−4153.

(60) Nguyen, L. T.; Le, T. N.; De Proft, F.; Chandra, A. K.; Langenaeker, W.; Nguyen, M. T.; Geerlings, P. Mechanism of [2+1] Cycloadditions of Hydrogen Isocyanide to Alkynes: Molecular Orbital and Density Functional Theory Study. *J. Am. Chem. Soc.* **1999**, *121* (25), 5992−6001.

(61) Berkowitz, M.; Parr, R. G. Molecular Hardness and Softness, Local Hardness and Softness, Hardness and Softness Kernels, and Relations among these Quantities. *J. Chem. Phys.* **1988**, *88* (4), 2554−2557.

(62) Gross, E. K. U.; Kohn, W. Local Density-Functional Theory of Frequency-Dependent Linear Response. *Phys. Rev. Lett.* **1985**, *55* (26), 2850−2852.

(63) By some measures, the proton has infinite chemical hardness!

(64) Gazquez, J. L.; Mendez, F. The Hard and Soft Acids and Bases Principle − An Atoms in Molecules Viewpoint. *J. Phys. Chem.* **1994**, *98* (17), 4591−4593.

(65) Mendez, F.; Gazquez, J. L. Chemical-Reactivity of Enolate Ions − The Local Hard and Soft Acids and Bases Principle Viewpoint. *J. Am. Chem. Soc.* **1994**, *116* (20), 9298−9301.

(66) Chattaraj, P. K. Chemical Reactivity and Selectivity: Local HSAB Principle versus Frontier Orbital Theory. *J. Phys. Chem. A* **2001**, *105* (2), 511−513.

(67) Geerlings, P.; De Proft, F. HSAB Principle: Applications of Its Global and Local Forms in Organic Chemistry. *Int. J. Quantum Chem.* **2000**, *80* (2), 227−235.

(68) Parr, R. G.; Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105* (26), 7512−7516.

CT600164J

JCTC Journal of Chemical Theory and Computation

Conceptual Density-Functional Theory for General Chemical Reactions, Including Those That Are Neither Charge- nor Frontier-Orbital-Controlled. 2. Application to Molecules Where Frontier Molecular Orbital Theory Fails

James S. M. Anderson,[†] Junia Melin,[‡] and Paul W. Ayers*,[†]

Department of Chemistry, McMaster University, Hamilton, Ontario, Canada L8S 4M1, and Department of Chemistry, Kansas State University, Manhattan, Kansas 66506-3701

Received May 10, 2006

Abstract: This paper examines cases where frontier molecular orbital theory is known to fail, specifically electrophilic aromatic substitution reactions on isoquinoline and borazarophenanthrenes. While we are able to explain the experimental regioselectivity preferences for isoquinoline without too much difficulty, describing the regioselectivity of the borazarophenanthrenes is much more challenging. This is attributed to the fact that these molecules lie between the electrostatic (or charge) control and electron-transfer (or frontier-orbital) control paradigms. These molecules can, however, be described using the general-purpose reactivity indicator introduced in the first paper of this series. The variation of the general-purpose reactivity indicator with respect to the parameters is readily summed up using what we term "reactivity transition tables", which provide a compact summary of which products form under different reaction conditions. For the otherwise problematic molecules considered here, the new reactivity indicator performs better than either the Fukui function or the electrostatic potential alone.

I. Introduction

In the first paper in this series,[1] the authors derived a general-purpose reactivity indicator that is capable of describing not only electrostatically (or charge) controlled reactions and electron-transfer- (or frontier-orbital-) controlled reactions[2] but also reactions that lie between these two extremes. This paper will apply this reactivity indicator to a particularly challenging set of molecules, where ordinary reactivity predictors have been observed to fail.

Before applying the reactivity indicator, we briefly summarize the results from the first paper in this series. The goal of the first paper was to derive a reactivity indicator that could truly be called a "general-purpose" reactivity indicator. That is, we sought a reactivity indicator that describes the full spectrum of chemical reactivity, from strong electrostatic control (minimum Fukui function is good), to joint electrostatic and electron-transfer control (maximum Fukui function is good), to strong electron-transfer control [maximum Fukui function is good, but maximum (for nucleophiles) or minimum (for electrophiles) electrostatic potential is also good]. To achieve this goal, we used a perturbative expansion about the separated reagent limit to derive an expression for the interaction energy between an electrophile and a nucleophile (see eq 21 in part 1).

To derive a reactivity indicator, we introduced a single reactive-site interaction model for electrophiles and nucleophiles. In this model, the reactive site of the attacking electrophile/nucleophile is modeled with a point charge and a condensed Fukui function. Inserting this model in the expression for the interaction energy and performing a careful error analysis led to the desired indicators. One of our indicators is appropriate for predicting where an electrophile will attack a nucleophile

* Corresponding author fax: (905) 522-2509; e-mail: ayers@mcmaster.ca.

[†] McMaster University.

[‡] Kansas State University.

10.1021/ct6001658 CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/16/2007

$$\Xi^{\kappa}_{\Delta N\le 0}(r) \equiv (\kappa+1)\,\Phi_{\text{nucleophile}}(r) - \Delta N(\kappa-1)\,\upsilon^{f^-}_{\text{nucleophile}}(r) \quad (1)$$

The other indicator is appropriate for predicting where a nucleophile will attack an electrophile

$$\Xi^{\kappa}_{\Delta N\ge 0}(r) \equiv -(\kappa+1)\,\Phi_{\text{electrophile}}(r) + \Delta N(\kappa-1)\,\upsilon^{f+}_{\text{electrophile}}(r) \quad (2)$$

Because these indicators model the interaction energy of the molecule, the molecule is most reactive in the places where $\Xi^{\kappa}_{\Delta N}(r)$ is the most negative.

By simply varying the value of κ, the entire spectrum of chemical reactivity can be described, ranging from strong electrostatic control ($\kappa > 1$), to pure electrostatic control ($\kappa = 1$), to joint control by electrostatics and electron-transfer effects ($-1 < \kappa < 1$), to pure electron-transfer control ($\kappa = -1$), to strong electron-transfer control ($\kappa < -1$). The value of κ that is appropriate for a particular reaction can be estimated using the approximate proportionalities:

$$\kappa \overset{\propto}{\approx} q^{(0)}_{\text{electrophile}} + \Delta N f^{(+)}_{\text{electrophile}} \qquad \text{for nucleophiles}$$

$$\kappa \overset{\propto}{\approx} -q^{(0)}_{\text{nucleophile}} - \Delta N f^{(-)}_{\text{nucleophile}} \qquad \text{for electrophiles} \quad (3)$$

The constant of proportionality is positive and, on the basis of our experience, has order of magnitude 1.

The reactivity indicators in eqs 1 and 2 also depend on the extent of electron transfer. The amount of electron transfer could be computed by minimizing the expression for the interaction energy directly, but the simple formula proposed by Parr and Pearson should be adequate for qualitative purposes[3]

$$\Delta N_{\text{electrophile}} \approx \frac{\mu_{\text{nucleophile}} - \mu_{\text{electrophile}}}{\eta_{\text{nucleophile}} + \eta_{\text{electrophile}}} \approx \frac{(I_{\text{nucleophile}} - I_{\text{electrophile}}) + (A_{\text{nucleophile}} - A_{\text{electrophile}})}{2(I_{\text{nucleophile}} + I_{\text{electrophile}} - A_{\text{nucleophile}} - A_{\text{electrophile}})} \quad (4)$$

In this equation, μ denotes the electronic chemical potential[4] and η denotes the chemical hardness.[3] The second line of this equation approximates the chemical potential and chemical hardness using the vertical ionization potential (I) and vertical electron affinity (A) of the reagents. Because nucleophiles transfer electrons to electrophiles, we have chosen a sign convention where $\Delta N \le 0$ for nucleophiles and $\Delta N \ge 0$ for electrophiles. The molecules we are studying in this paper are nucleophiles, so it is the $\Delta N \le 0$ case that is of greatest interest here.

The general-purpose reactivity indicators are seen to have a dependence on the Fukui potential, $\upsilon^{f\pm}(r)$, and the molecular electrostatic potential, $\Phi(r)$. The electrostatic potential is essential for describing reactions that are electrostatically controlled ($\kappa = 1$); the Fukui potential is essential for describing reactions that are electron-transfer-controlled ($\kappa = -1$). In all other cases, both the electrostatic potential and the Fukui potential play a role in determining a molecule's regioselectivity preferences.

It is useful to approximate the electrostatic potential and the Fukui functions using atomic charges[5,6]

$$\Phi^{(0)}(r) \sim \sum_{\alpha}\frac{q^{(0)}_{\alpha}}{|r-R_{\alpha}|} \quad (5)$$

$$\upsilon^{f^-}(r) \sim \sum_{\alpha}\frac{f^-_{\alpha}}{|r-R_{\alpha}|}$$

$$\upsilon^{f^+}(r) \sim \sum_{\alpha}\frac{f^+_{\alpha}}{|r-R_{\alpha}|} \quad (6)$$

where $q^{(0)}_{\alpha}$ denotes the atomic charges on the reagent. f^+_{α} and f^-_{α} are the condensed Fukui functions. Using these results, one can derive a condensed version of the indicators in eqs 1 and 2

$$\Xi^{\kappa}_{\Delta N\le 0}(r_p) \sim \sum_{\alpha\in\text{nucleophile}}\frac{\Xi^{\kappa}_{\Delta N\le 0,\alpha}}{|r-R_{\alpha}|}$$

$$\Xi^{\kappa}_{\Delta N\le 0,\alpha} = (\kappa+1)q^{(0)}_{\text{nucleophile},\alpha} - \Delta N(\kappa-1)f^-_{\text{nucleophile},\alpha} \quad (7)$$

$$\Xi^{\kappa}_{\Delta N\ge 0}(r_p) \sim \sum_{\alpha\in\text{electrophile}}\frac{\Xi^{\kappa}_{\Delta N\ge 0,\alpha}}{|r-R_{\alpha}|}$$

$$\Xi^{\kappa}_{\Delta N\ge 0,\alpha} = -(\kappa+1)q^{(0)}_{\text{electrophile},\alpha} + \Delta N(\kappa-1)f^+_{\text{electrophile},\alpha} \quad (8)$$

For special cases, this reactivity indicator recovers the reactivity patterns that would be predicted on the basis of the Fukui function ($\kappa = -1$) or the electrostatic potential ($\kappa = 1$) alone. The new indicator has no value, then, unless it supersedes the description of chemical reactivity that is possible using these reactivity indicators in isolation. This suggests that the new indicator might be useful for studying molecules where frontier molecular orbital theory (FMO) would be expected to work but has been observed to fail. The regioselectivity of electrophilic aromatic substitution reactions is usually well-described using both FMO and density functional theory analogues to frontier molecular orbital theory like the Fukui function.[7−9] This is not always true, however: Dewar showed that FMO fails to describe electrophilic aromatic substitution in isoquinoline, 10-hydroxy-10,9-borazarophenanthrene, and 10-methyl-10,9-borazarophenanthrene. The question arises: does the Fukui function,[10−12] which extends the frontier molecular orbital theory but is nonetheless motivated by FMO ideas, fail in the same way? In section III, it is observed that, while the Fukui function does seem to work better than FMO for these molecules, it still fails to adequately describe their reactivity. (That is, the Fukui function fails "in the same way" as FMO, but not as badly.)

The Fukui function is intimately linked to the idea of electron transfer, so it is an appropriate indicator for "electron-transfer-controlled" (also called frontier-orbital-

controlled and Fukui-function-controlled) reactions. When hard reagents interact, electron transfer is either limited or occurs late in the chemical reaction profile; such reactions are usually called "electrostatically controlled" or "charge-controlled".[2] The electrostatic potential is a more appropriate indicator than the Fukui function in these cases.[13] Because the Fukui function fails to adequately describe the reactivity of these molecules, perhaps the electrostatic potential will suffice. The results in section III show that the electrostatic potential does not describe the reactivity of these molecules either.

These results suggest that the electrophilic substitution on borazarophenanthrenes represents a difficult, but otherwise suitable, test for the general-purpose reactivity indicator, $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$, that we derived and discussed in the first paper of this series. Indeed, the theoretical developments in the first paper were motivated by our inability to describe these molecules using ordinary reactivity indicators. Section III contains the main results; we observe that $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$ does an excellent job of describing a variety of electrophilic aromatic substitution reactions on borazarophenanthrenes. Before presenting our results, however, we need to state our computational methods.

II. Computational Methods

To demonstrate the power of these indicators, we decided to analyze molecules where Dewar found contradictions to frontier molecular orbital effects.[14] In what follows, all calculations were conducted using Gaussian 03[15] and the B3LYP[16−18] functional with the 6-31++G* basis set.[19] Figures were generated using GaussView 3.0. The atomic charges used to compute condensed reactivity indicators were obtained from four different methods: the Mulliken population analysis[20−23] and natural population analysis (NPA)[24−26] approaches to partitioning the density matrix and the Merz−Singh−Kollman[27,28] (MSK) and CHelpG[29] (CHG) methods for fitting the electrostatic potential.

III. Results and Discussion

A. Overview. We will explore three of the molecules (isoquinoline, 10-hydroxy-10,9-borazarophenanthrene, and 10-methyl-10,9-borazarophenanthrene) Dewar gave as examples where FMO fails to adequately describe regioselectivity.[14] (Dewar gave a fourth example, nitrobenzene. This molecule has been extensively studied using reactivity indicators associated with conceptual DFT and will not be revisited here. More information on nitrobenzene can be found, for example, in the paper of Langenaeker et al.[7])

For each molecule, we will first present the experimentally observed reactivity preferences. Then, we will present the reactivity preferences predicted by FMO (which predicts that electrophilic attack occurs where the magnitude of the highest occupied molecular orbital is largest), electrostatic considerations (which predicts that electrophilic attack occurs where the electrostatic potential is most negative), and the Fukui function (which predicts that electrophilic attack occurs where $f^{-}(r)$ is the largest). When no single method can predict the observed reactivity, we will examine the more general index, $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$, which combines information from the electrostatic potential and the Fukui function.

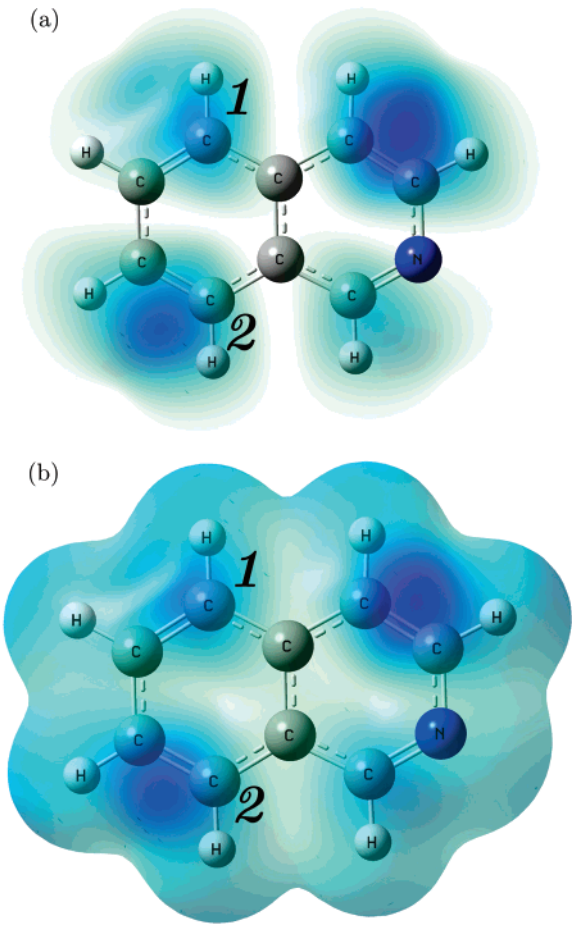
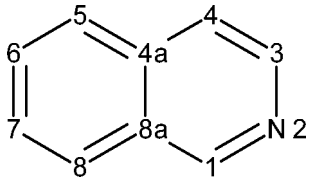


Figure 1. (a) Square magnitude of the HOMO orbital, $|\phi_{HOMO}(r)|^2$, (b) the Fukui function from below, $f^{-}(r)$, is plotted on the 0.0004 isodensity surface of the isoquinoline molecule. The molecule should be most reactive where these functions are the largest. The numbers denote the experimentally observed reactivity preferences.[30]

B. Isoquinoline. Experimental studies on isoquinoline[30]



have shown that the most reactive site of this molecule is carbon 5, with secondary reactivity at carbon 8. Products from electrophilic substitution on carbon 4 were not found, so this site is believed to be unreactive. Figure 1a reports the value of the highest occupied molecular orbital density, $|\phi_{HOMO}(r)|^2$, on the van der Waals surface of the molecule. We model the van der Waals surface with the $\rho(r) = 0.0004$ isodensity surface; this models the reactive surface of the molecule and is an appropriate indicator of site selectivity.[31−33] In accord with Dewar's FMO analysis, an electrophilic attack is predicted to occur on the double bonds between carbons 3 and 4, then carbons 7 and 8, and finally carbons 5 and 6. This reactivity order is reversed relative the experimental

products: carbon 5 is the primary reaction site; carbon 8 is the secondary reaction site, and carbon 4 is unreactive.

Sometimes, orbital relaxation effects are important for describing chemical reactivity, and so the Fukui function (which includes orbital relaxation effects)[11] sometimes performs better than $|\phi_{HOMO}(r)|^2$ for predicting sites of electrophilic attack. Orbital relaxation effects have been shown to be important in electrophilic aromatic substitution[7] and electrophilic attack on double bonds (in organic molecules)[31] and multiple bonds (in inorganic complexes with metal−metal bonds).[32] It seems plausible, then, that the Fukui function will locate the appropriate sites for electrophilic attack on isoquinoline. To investigate this possibility, we plotted the value of the Fukui function on the molecular van der Waals surface (see Figure 1b). While the Fukui function and $|\phi_{HOMO}(r)|^2$ are quantitatively different, they are qualitatively similar: the Fukui function at carbon 4 is slightly larger than that at carbon 8, which is significantly larger than the Fukui function at the dominant reaction site (carbon 5). It is interesting that this trend is altered somewhat if condensed Fukui functions are used. With condensed Fukui functions and ChelpG charges, we have $f_{C5} = 0.229$, $f_{C4} = 0.193$, and $f_{C8} = 0.165$. While carbon 4 is still predicted to be reactive, at least the first-choice reaction site is now identified correctly. To the best of our knowledge, this is the first molecule where condensed Fukui functions perform decisively better than Fukui functions plotted on a reactive surface. In this case, the Fukui function on carbon 5 is concentrated near the nucleus, and thus, while the Fukui potential (which depends on the total size of the Fukui function in the vicinity of the atom) will be large in the region of carbon 5, the amplitude of the Fukui function [which decays exponentially, rather than as $(1/r)$] has decreased almost to zero on the reactive surface. Thus chastened, we will henceforth focus on condensed reactivity indicators.

For reactions between hard molecules, the appropriate indicator of site selectivity is the electrostatic potential. This is clear already from the works of Klopman and Berkowitz[2,34] but is also a topic of recent emphasis in the conceptual DFT literature.[13,35,36] It follows very clearly from our analysis also, because the extent of electron transfer

$$\Delta N_{\text{nucleophile}} \approx \frac{\mu_{\text{electrophile}} - \mu_{\text{nucleophile}}}{\eta_{\text{electrophile}} + \eta_{\text{nucleophile}}} \tag{9}$$

will be small when the hardness of the reagents is large. When $|\Delta N|$ is small, the dominant contribution to both $\Xi^{\kappa}_{\Delta N \le 0}(r_p)$ and $\Xi^{\kappa}_{\Delta N \ge 0}(r_p)$ is from the molecular electrostatic potential. While isoquinoline is not especially hard, insofar as the Fukui function has failed to successfully describe its reactivity, it seems desirable to explore the electrostatic potential. This is done by computing the atomic charges, which represent the "condensed" electrostatic potential (cf. eq 5).[37] The charges from several different population analysis schemes are plotted in Figure 2a. The Mulliken charges are manifestly unreliable, as is to be expected for a basis set including diffuse functions. Henceforth, we will not report the results from the Mulliken population analysis. The other charge schemes reported in Figure 2a are more reasonable, with the two methods of electrostatic potential fitting (MSK and CHG) giving similar results. The NPA scheme is less similar, which may also be due to this method's stronger dependence on the basis set or due to the fact that NPA, unlike MSK and CHG, is based on a population analysis of the density matrix, and not on fitting the electrostatic potential.

Recall that the reactive site of strong electrophiles is usually positively charged. Electrophilic attack thus tends to occur at the most negatively charged sites of the nucleophile. Looking at Figure 2a, it is clear that the most negative sites in the molecule are the nitrogen atom and carbon 4. Carbon 4, however, is unreactive. The charges on carbon 5 (the most reactive position) and carbon 8 (the second most reactive position) are not especially small.

Sometimes, one argues that one should add the charges of hydrogen atoms into the charges of the atoms they are bonded to. This seems especially useful in electrostatic fitting procedures: because the carbon−hydrogen bond is short, it can be difficult to determine how to partition a reactive carbon's charge between the carbon atom and the adjacent hydrogen. More generally, the regioselectivity of a molecule is usually determined by interactions that occur when the molecular substrate and the attacking reagent are in van der Waals contact. Because the separation between the molecules is much larger than the length of a carbon−hydrogen bond, from the perspective of the attacking reagent, $-CH_n$ groups appear as a single point charge. On the basis of this reasoning, it is preferable to consider "functional group" charges that are computed by adding the charges of hydrogen atoms to the charges of the adjacent "heavy" atom. These data are reported in Figure 2b. Unfortunately, this does not alter the fact that carbon 4 is predicted to be reactive, while the molecular sites that actually are reactive are predicted to be relatively unreactive.

Further thought about the chemistry of isoquinoline rectifies these unsuccessful predictions. Isoquinoline is relatively basic, $pK_b = 8.6$. (Unsurprisingly, this is similar to the pK_b of pyridine, which is 8.7.) Experimentally, it is difficult to perform an electrophilic substitution reaction on pyridine: electrophiles are Lewis acids, and so the solutions used for electrophilic aromatic substitutions are also acidic.[38] The experimental studies of Dewar were carried out in a mixture of nitric and sulfuric acids![30] Even when less extreme conditions are used, the pH of those solutions is almost always less than 5.3, and in that environment, pyridine is protonated. Clearly, performing an electrophilic attack on a protonated molecule will be difficult! Isoquinoline, which is marginally less basic than pyridine, is also expected to be in its protonated state when it undergoes electrophilic aromatic substitution. On the basis of this logic, we performed calculations on protonated isoquinoline.[39] The atomic charges are reported in Figure 3a (raw charges) and 3b (summed with adjacent hydrogens). The condensed Fukui functions are reported in Figure 4a (computed from atomic charges) and 4b (computed from atomic charges summed with adjacent hydrogens). On the basis of the charges, we would predict that the reaction occurs at carbon 5, with carbon 8 and carbon 4 having similar reactivity. On the basis

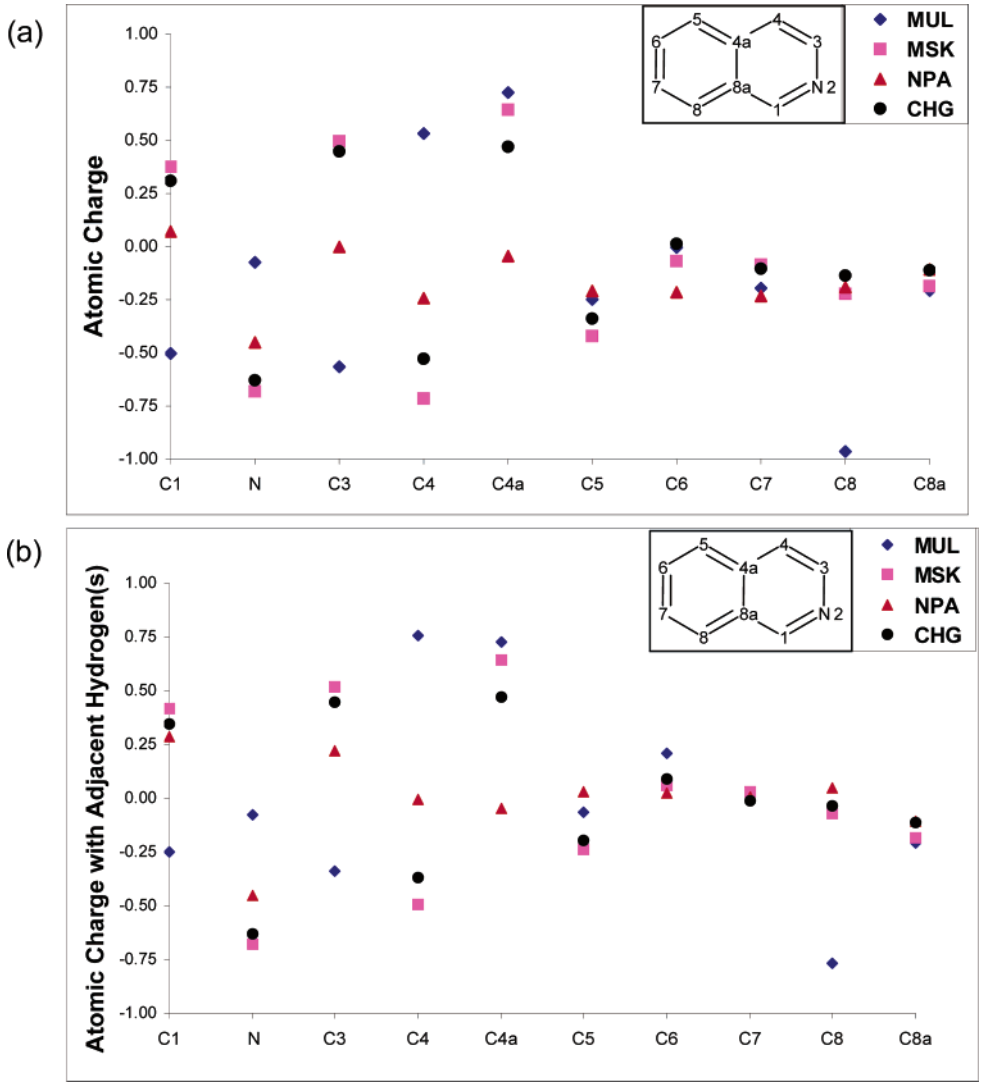


Figure 2. Atomic charges (condensed electrostatic potential) of isoquinoline. (a) The atomic charges on the indicated atoms. (b) The atomic charges on the indicated atoms plus the charges of hydrogen atoms bonded to those atoms. The atomic numbering scheme is included as an inset. Experimentally, C5 is the most reactive, followed by C8. C4 is unreactive.

of the Fukui function, we would predict that carbon 5 and carbon 8 are both highly reactive (carbon 5 perhaps slightly more so) and that carbon 4 is not very reactive. This agrees with the experiment: the substantial negative charge on carbon 5, coupled with its significant Fukui function, makes this site highly susceptible to electrophilic attack. Carbon 8 is also favorable electrostatically and on the basis of its Fukui function, but its less negative charge is associated with slightly diminished reactivity. Carbon 4 is not significantly reactive compared to carbon 5 or carbon 8.

Perhaps protonating isoquinoline is too strong an assumption. One could, after all, carry out an electrophilic substitution in an aprotic solvent. To explore this possibility, we considered the complex of pyridine with the sodium cation, which is an extremely weak Lewis acid (see Figure 5). In this case, carbon 4 remains the most negatively charged reactive site. However, the Fukui function is highly concentrated on carbon 5 and, to a lesser extent, carbon 8. Because isoquinoline is not especially hard, it seems reasonable to infer that, under any reasonable set of experimental conditions, the "ion-paired isoquinoline" that is subject to chemical reaction will react first at carbon 5, with a secondary product associated with reaction at carbon 8.

We could apply our general-purpose reactivity indicator to isoquinoline, but the picture that emerges is quite boring because the electrostatically favored and electron-transfer-favored sites on protonated isoquinoline are the same. Instead, we will apply our reactivity indicator to a much more challenging problem: electrophilic aromatic substitution on 10,9-borazarophenanthrenes.

C. 10-R-10,9-Borazarophenanthrene. 1. Summary of Experimental Observations. In addition to isoquinoline, Dewar also pointed out two, more challenging, molecules where experimental results were not in accord with frontier molecular orbital theory. Experimental studies of 10-R-10,9-borazarophenanthrene (R = OH or CH_3) indicate that carbon 8 and carbon 6 are both susceptible to electrophilic attack.[40−42] Chlorination of 10-methyl-10,9-borazarophenanthrene favors carbon 8 over carbon 6,[42] while chlorination of 10-hydroxy-10,9-borazarophenanthrene gives the disubstituted product corresponding to reaction at both carbon 6 and carbon 8.[41] Nitration produces a mixture of the products associated with

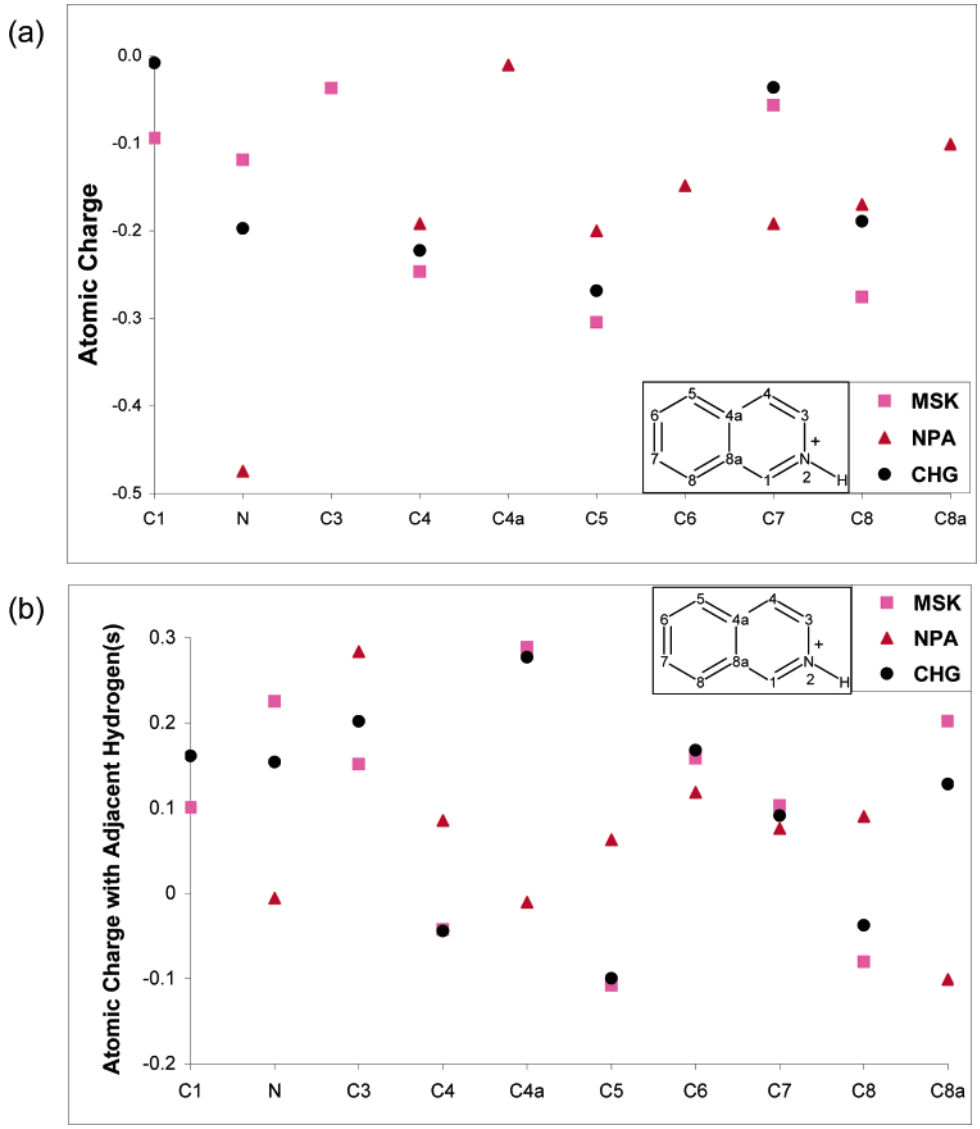


Figure 3. Atomic charges (condensed electrostatic potential) of *protonated* isoquinoline.

reaction at carbon 6 and carbon 8.[42] It seems reasonable to infer that carbon 6 and carbon 8 are the most reactive positions, with carbon 8 being slightly more reactive, at least for some electrophiles. If one increases the temperature and the amount of reagent, then one can add another chlorine to 10-hydroxy-10,9-borazarophenanthrene; forming the 2,6,8-trisubstituted product.[41] Carbon 2 is essentially unreactive with respect to nitration and bromination, however.[41,42] We infer, then, that the experimental reactivity profile can be summarized as

$$C8 \gtrsim C6 \gg C2 \qquad (10)$$

Theoretical electronic-structure treatments suggest that carbon 4 and carbon 2 should have similar reactivities. It is known, however, that steric congestion significantly reduces the rate of reaction at carbon 4.[43]

One might expect that, like isoquinoline, 10-R-10,9-borazarophenanthrene would be protonated. This is not the case. First of all, note that the nitrogen atom in 10,9-borazarophenanthrene already has four bonds, which reduces its susceptibility to protonation. Second, the presence of the adjacent boron atom (which accepts electrons from the p orbital of the sp^2 hybridized nitrogen atom) reduces the basicity of the nitrogen. Protonation does occur, but only when the compound is heated in concentrated sulfuric acid.[40,44] (Upon protonation in that environment, the central ring breaks and boron is lost.) Under any reasonable set of experimental conditions, then, 10-R-10,9-borazarophenanthrene is not protonated.

2. Frontier Molecular Orbitals, Condensed Fukui Functions, and Atomic Charges. Frontier molecular orbital theory does not predict the reactivity in these molecules. According to FMO, carbon 2 is slightly more reactive than carbon 4, which is slightly more reactive than the bond between carbon 6 and carbon 7.[14] (Both carbon 6 and carbon 7 have substantial contributions from the HOMO orbital.) There is very little frontier molecular orbital density on carbon 8.

The condensed Fukui functions for 10-hydroxy-10,9-borazarophenanthrene (R = OH) are reported in Figures 6b and 7b. The condensed Fukui function gives results that are more in line with the experiment than frontier molecular

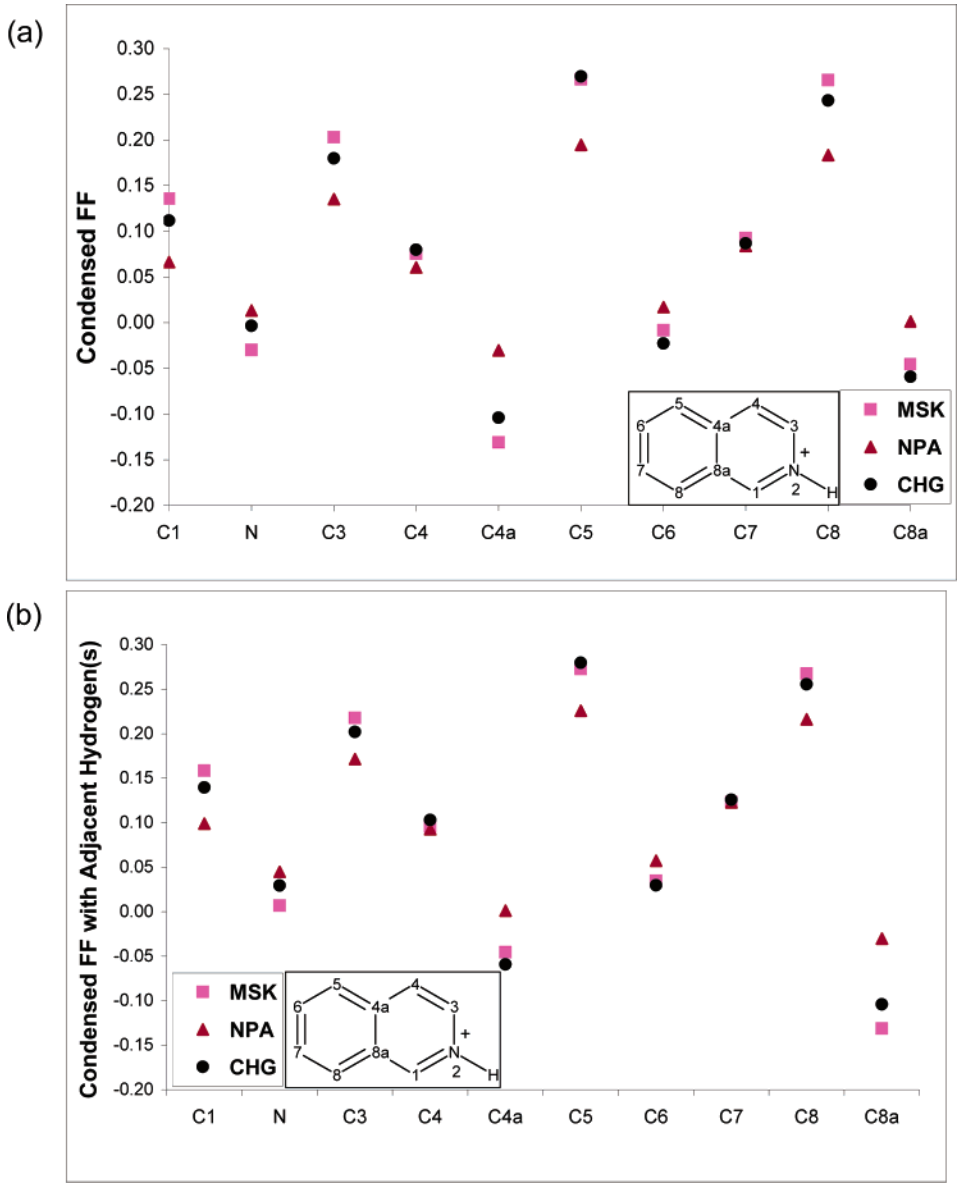
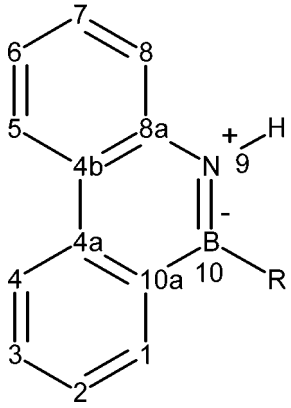


Figure 4. Condensed Fukui function of protonated isoquinoline. In a, we compute the Fukui function from the difference of atomic charges. In b, we add to each indicated atom the condensed Fukui functions of the hydrogen atoms bonded to it. The atomic numbering scheme is included as an inset. Experimentally, C5 is the most reactive, followed by C8. C4 is unreactive.



orbital theory. In particular, the Fukui function predicts that carbon 6 is the most reactive position in the molecule. Carbon 2 and carbon 4b are the next most reactive positions, followed by carbon 4. The Fukui function predicts that carbon 8 is essentially unreactive. The predicted reactivity at carbon 4b demonstrates a recurrent feature in electrophilic polyaromatic substitution reactions. Qualitative reactivity indicators often predict ipso addition of electrophiles to polyaromatic compounds, but in many cases, ipso addition is a mechanistic dead end: although it is sometimes the most favorable orientation for the reactants, the barrier separating the ipso reactive intermediate from stable product molecules is very high. For example, because there are no hydrogen atoms at carbon 4b, electrophilic aromatic substitution cannot occur. Consequently, addition at carbon 4b requires a loss of aromaticity.

The atomic charges are reported in Figures 6a and 7a. Though carbon 10a, nitrogen, and oxygen are all negatively charged, these sites are not susceptible to electrophilic aromatic substitution. While a positively charged electrophile might form an "ion-pair" association complex with one of these sites, no further reaction at these sites is possible under ordinary conditions. For carbon 10a, there is no hydrogen

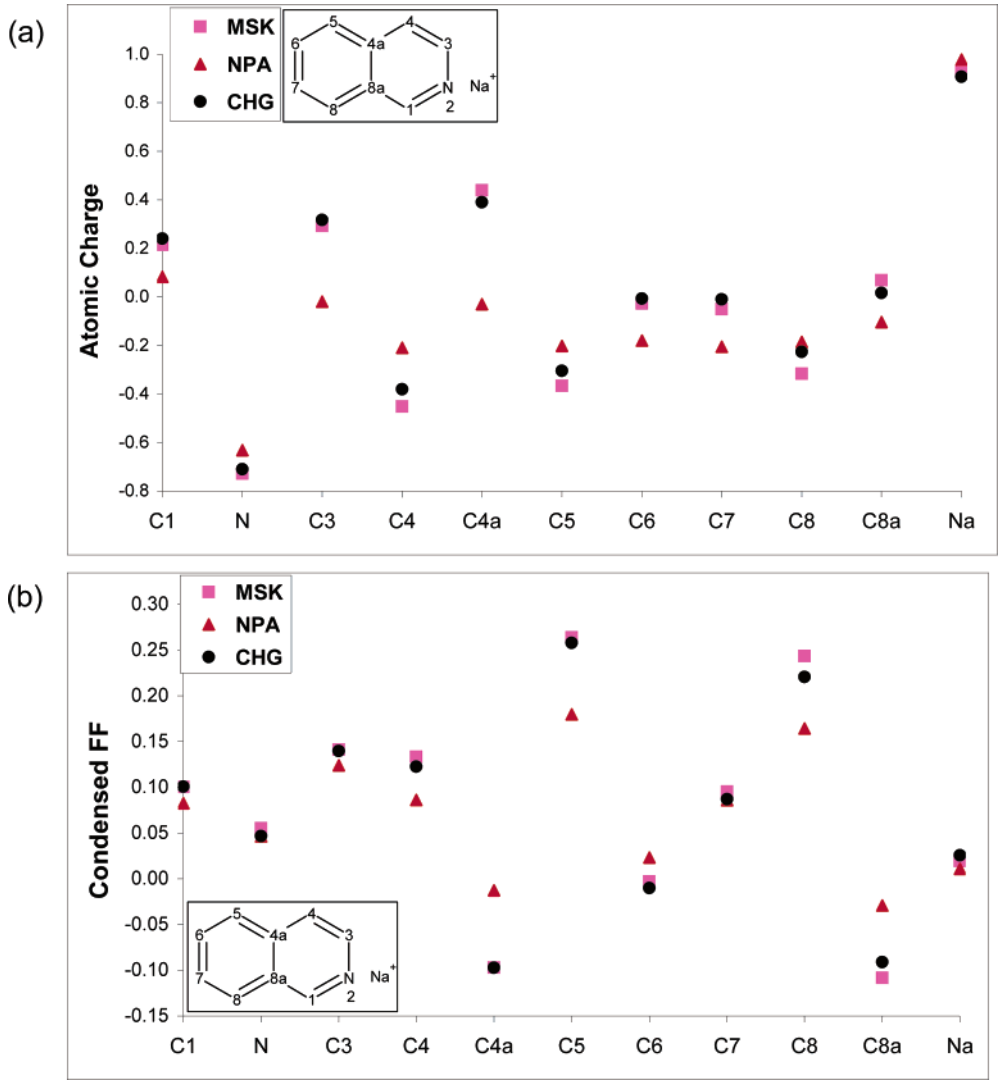


Figure 5. (a) Atomic charges and (b) condensed Fukui functions of isoquinoline with a sodium "spectator cation." Qualitatively, the plot with hydrogens summed into adjacent carbons is very similar.

atom to serve as a leaving group. While there is a hydrogen bound to the nitrogen atom, bonds between hydrogen and electron-deficient sp^2 hybridized nitrogen atoms are very strong, so this hydrogen atom is a poor leaving group.[45] Oxygen−hydrogen bonds are generally stronger than carbon−hydrogen bonds, so the hydrogen atom bonded to oxygen is also a poor leaving group. Consequently, we expect the chemistry at C10a, nitrogen, and oxygen is limited to the formation and dissociation of weak "ion-pairing"-type interactions.

Among sites that are susceptible to electrophilic aromatic substitution, carbon 8 is the most negatively charged, with carbon 6 and carbon 4 having somewhat smaller, but still substantial, negative charges. Carbon 2 is also negatively charged.

We predict, then, that carbon 8 and carbon 6 are the most highly reactive sites on the basis of electrostatic effects and electron-transfer effects, respectively. Carbon 2 and carbon 4 represent tradeoffs: these molecular sites have reasonably large Fukui functions and reasonably negative charges. If the reactivity of 10-hydroxy-10,9-borazarophenanthrene was strongly electrostatically controlled, then carbon 6 should be less reactive because large values for the Fukui function are unfavorable in that situation. Similarly, if the reactivity was strongly electron-transfer-controlled, then carbon 8 should be less reactive because small values of the electrostatic potential are unfavorable in that situation. Both carbon 8 and carbon 6 are observed to be reactive, so the reactivity must be jointly controlled by electrostatic and electron-transfer effects.

Because MSK and CHG charges are both designed to reproduce the electrostatic potential, these two choices of charges should resemble one another and, hopefully, also the charges derived from NPA. Examining Figure 7, it seems that the agreement between the different population analysis schemes is slightly better if the charges on the heavy atoms are combined with the charges on their adjacent hydrogen atoms. We will base the rest of our analysis on the plots with the hydrogen atoms summed in; because of the similarity between the results in Figures 6 and 7, this will not affect our results very much. Indeed, the main trends in the values of the charges and the condensed Fukui functions seem to be reproduced no matter which population analysis scheme is being employed. This is reassuring, because there

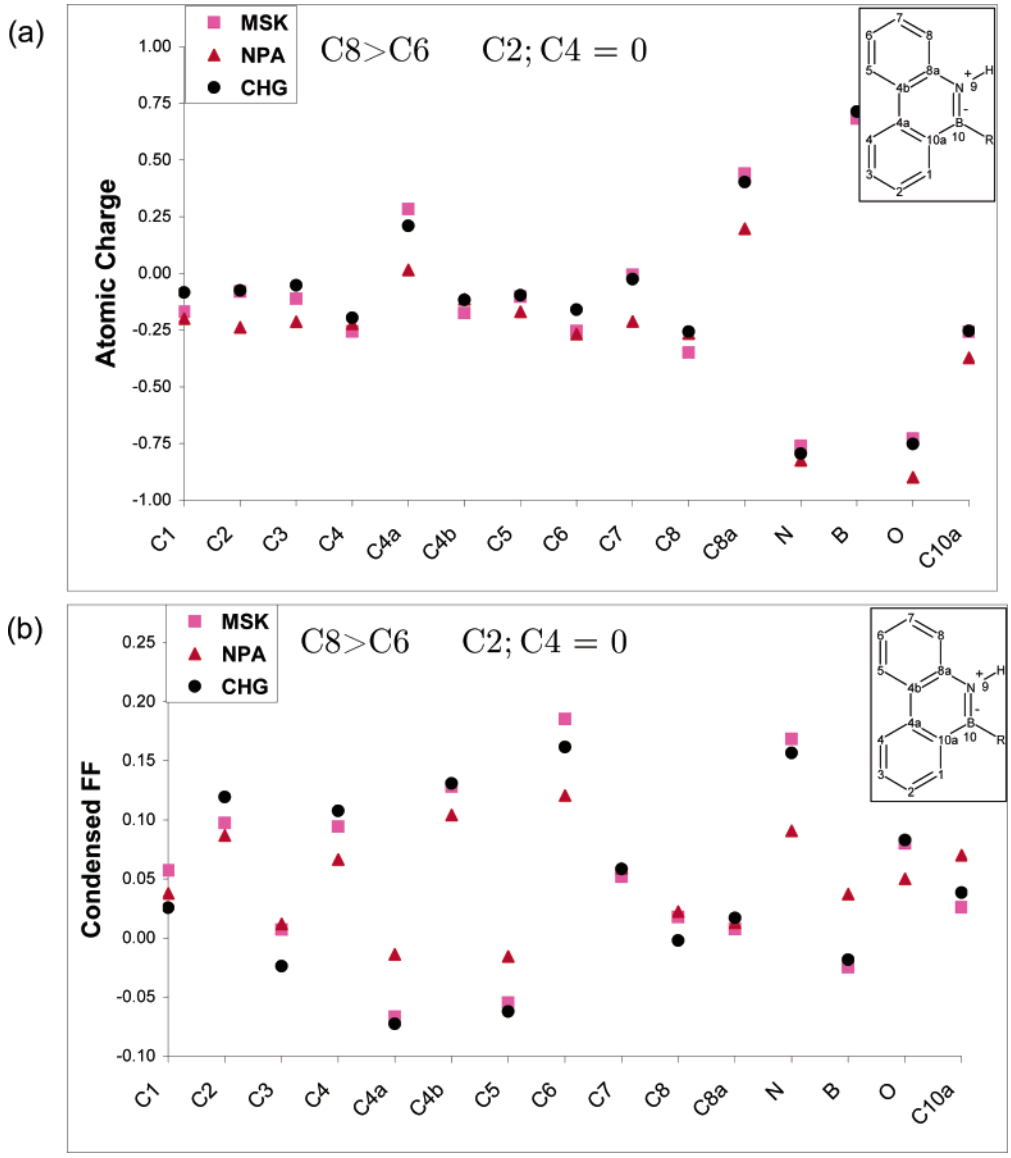


Figure 6. (a) Atomic charges and (b) condensed Fukui functions of 10-hydroxy-10,9-borazarophenanthrene. The atomic numbering scheme is included as an inset. Experimentally, C8 is the most reactive, followed by C6 and then C2. Other sites are unreactive.

is no a priori reason to assert the superiority of any one of the population analysis schemes.[46]

3. Application of the Condensed General-Purpose Reactivity Indicator. Because 10-hydroxy-10,9-borazarophenanthrene is a case of joint electrostatic and electron-transfer control, it seems to be a suitable test for the general-purpose reactivity indicator developed in the previous paper.[1] It is difficult to distill the immense amount of information in this indicator

$$\Xi^{\kappa}_{\Delta N\leq 0,\alpha} = (\kappa+1)q^{(0)}_{\text{nucleophile},\alpha} - \Delta N(\kappa-1)f^{-}_{\text{nucleophile},\alpha} \quad (11)$$

into an easily digestible form, however. Recall that $\Xi^{\kappa}_{\Delta N,\alpha}$ depends on two parameters: κ, which measures the relative importance of electrostatic control ($\kappa \approx 1$) and electron-transfer control ($\kappa \approx -1$), and ΔN, which measures the amount of electron transfer. $\Xi^{\kappa}_{\Delta N,\alpha}$ is a bivariate function for every atom in the system! What we really want to know is which site is most reactive and how (and whether) the choice of reaction site changes as a result of changing electrophilic reagents and reaction conditions. We have found that the key information about the reactivity of the molecule can be summarized using what we term "reactivity transition tables" (see Tables 1−4). To make a reactivity transition table, one starts by computing the value of $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$ for every atom in the molecule and for the entire chemically relevant range of choices for the amount of electron transfer ($-1 \leq \Delta N \leq 0$) and the extent of electrostatic/electron-transfer control (κ). As established in the previous section, electrophilic attack on 10-hydroxy-10,9-borazarophenanthrene is jointly controlled by electrostatic and electron-transfer effects, so we consider only $-1 \leq \kappa \leq 1$. In constructing the reactivity transition tables, we restrict ourselves to carbons that are susceptible to electrophilic aromatic substitution.

One determines which atom is the most reactive by locating the atom with the smallest value of $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$. We insert this value in the "first choice" reactivity transition table and then color-code the cell so that it is clear which carbon

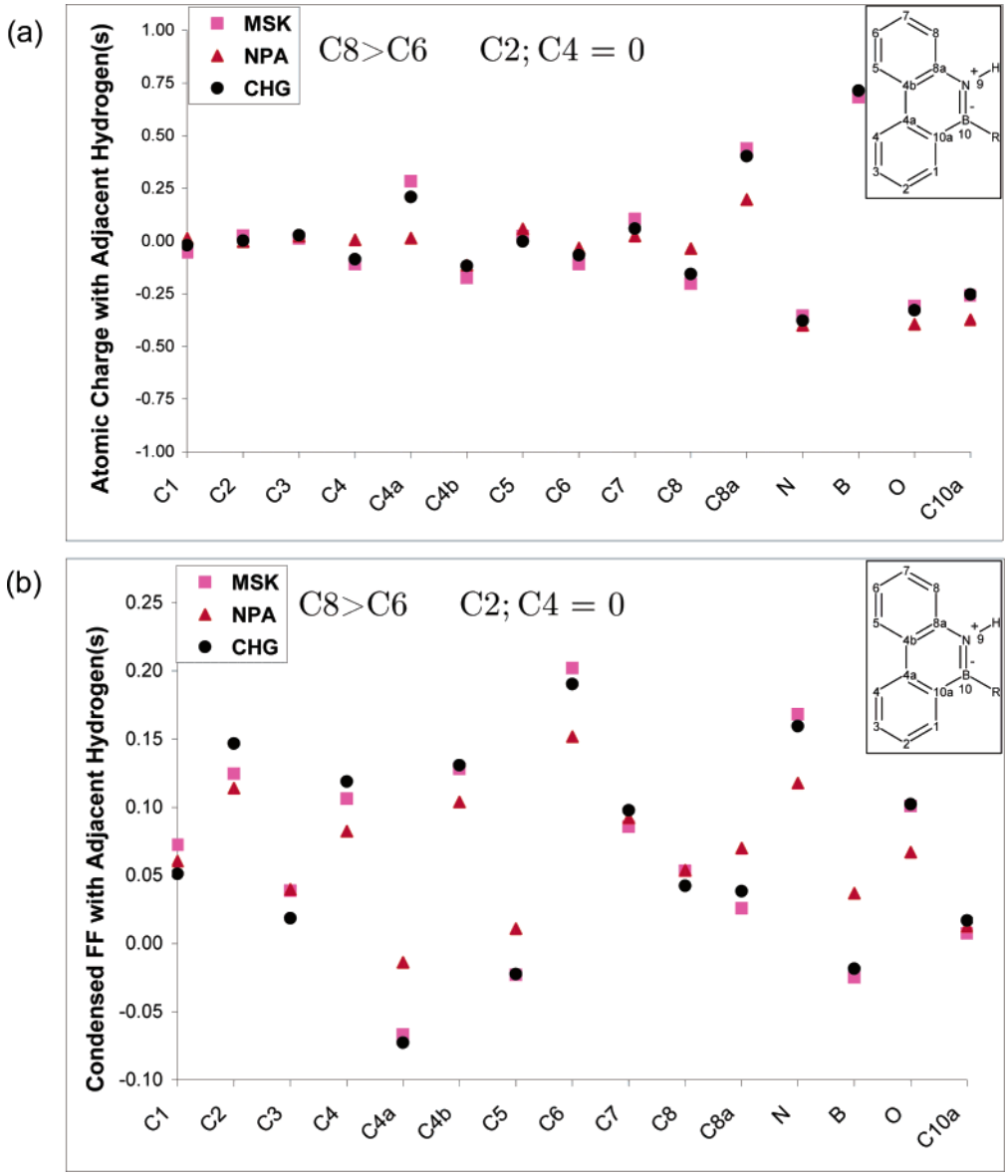


Figure 7. (a) "Hydrogen-summed" atomic charges and (b) "hydrogen-summed" condensed Fukui functions of 10-hydroxy-10,9-borazarophenanthrene. Unlike Figure 6, the reactivity indicators for hydrogen atoms have been added to the reactivity indicators of the heavy atoms to which they are bonded.

is reacting. The second most reactive atom is the one with the second smallest value for $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$. We insert this value in the "second choice" reactivity transition table and color-code the cells to indicate the second most reactive carbon. If the difference between the values of $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$ for the first and second choices is relatively small, then, one expects that both possible product molecules will form. If the difference between the "first choice" and "second choice" values of $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$ is relatively large, vigorous reaction conditions may be required to form the secondary product.

Reactivity transition tables contain both qualitative and quantitative data on reactivity. At a quantitative level, the "first choice" and "second choice" reactivity transition tables give information about the relative favorability of the primary and secondary products. At a qualitative level, the reactivity transition table can be read as a "phase diagram" for chemical reactivity. Examine Table 1a. When the nucleophile reacts with a very hard electrophile (so that the reaction is mostly electrostatically controlled and $\kappa \approx 1$), reactions occur at carbon 8. As the electrophile becomes softer and the extent of electron transfer increases, carbon 6 becomes the preferred site for reactivity. In the "transition region" between carbon 8 and carbon 6, one would expect a mixture of products. In this way, a reactivity transition table contains information about how to choose the electrophile so that a desired product is formed.

Reactivity transition tables for different types of population analyses are provided in Table 1 (NPA), Table 2 (MSK), and Table 3 (CHelpG). In each case, carbon 8 is the most favorable site when electrostatic effects are dominant ($\kappa \approx 1$ and $\Delta N \approx 0$) and carbon 6 is the most favorable site when electron-transfer effects are dominant ($\kappa \approx -1$ and $\Delta N \approx -1$). Experimental results indicate that the chlorination and nitration of 10-hydroxy-10,9-borazarophenanthrene occur on carbon 8 and carbon 6, with a small preference for carbon 8. In these reactions, the electrophile is reasonably hard, and

Table 1. Reactivity Transition Tables for 10-Hydroxy-10,9-borazarophenanthrene Using Natural Population Analysis for the Charges and Fukui Functions on Hydrogens Summed into the Adjacent Heavy Atoms[a]

(a)

	κ										
N	1.0	0.8	0.6	0.4	0.2	0	-0.2	-0.4	-0.6	-0.8	-1.0
-1.0	-0.072	-0.090	-0.113	-0.137	-0.161	-0.185	-0.208	-0.232	-0.256	-0.280	-0.303
-0.9	-0.072	-0.086	-0.107	-0.128	-0.149	-0.169	-0.190	-0.211	-0.232	-0.252	-0.273
-0.8	-0.072	-0.083	-0.101	-0.119	-0.137	-0.154	-0.172	-0.190	-0.207	-0.225	-0.243
-0.7	-0.072	-0.080	-0.095	-0.110	-0.124	-0.139	-0.154	-0.168	-0.183	-0.198	-0.212
-0.6	-0.072	-0.077	-0.089	-0.101	-0.112	-0.124	-0.136	-0.147	-0.159	-0.170	-0.182
-0.5	-0.072	-0.074	-0.083	-0.092	-0.100	-0.109	-0.117	-0.126	-0.135	-0.143	-0.152
-0.4	-0.072	-0.071	-0.077	-0.082	-0.088	-0.094	-0.099	-0.105	-0.110	-0.116	-0.121
-0.3	-0.072	-0.068	-0.071	-0.073	-0.076	-0.078	-0.081	-0.083	-0.086	-0.088	-0.091
-0.2	-0.072	-0.067	-0.065	-0.064	-0.064	-0.063	-0.063	-0.062	-0.062	-0.061	-0.061
-0.1	-0.072	-0.066	-0.060	-0.055	-0.052	-0.048	-0.045	-0.041	-0.037	-0.034	-0.030
0.0	-0.072	-0.065	-0.058	-0.051	-0.043	-0.036	-0.029	-0.022	-0.014	-0.007	
	Carbon 2		Carbon 4			Carbon 6			Carbon 8		

(b)

	κ										
N	1.0	0.8	0.6	0.4	0.2	0	-0.2	-0.4	-0.6	-0.8	-1.0
-1.0	-0.066	-0.076	-0.079	-0.083	-0.095	-0.117	-0.139	-0.161	-0.184	-0.206	-0.228
-0.9	-0.066	-0.075	-0.077	-0.080	-0.086	-0.106	-0.125	-0.145	-0.165	-0.185	-0.205
-0.8	-0.066	-0.074	-0.075	-0.076	-0.078	-0.094	-0.112	-0.129	-0.147	-0.165	-0.182
-0.7	-0.066	-0.073	-0.073	-0.073	-0.073	-0.083	-0.098	-0.113	-0.129	-0.144	-0.160
-0.6	-0.066	-0.071	-0.071	-0.070	-0.069	-0.071	-0.084	-0.098	-0.111	-0.124	-0.137
-0.5	-0.066	-0.070	-0.069	-0.067	-0.065	-0.063	-0.071	-0.082	-0.092	-0.103	-0.114
-0.4	-0.066	-0.069	-0.066	-0.063	-0.061	-0.058	-0.057	-0.066	-0.074	-0.083	-0.091
-0.3	-0.066	-0.068	-0.064	-0.060	-0.056	-0.052	-0.048	-0.050	-0.056	-0.062	-0.068
-0.2	-0.066	-0.065	-0.062	-0.057	-0.052	-0.047	-0.042	-0.037	-0.038	-0.042	-0.046
-0.1	-0.066	-0.062	-0.059	-0.054	-0.048	-0.041	-0.035	-0.029	-0.023	-0.021	-0.023
0.0	-0.066	-0.059	-0.053	-0.046	-0.039	-0.033	-0.026	-0.020	-0.013	-0.007	
	Carbon 2		Carbon 4			Carbon 6			Carbon 8		

[a] (Part a) First choice: the minimum values of $\Xi^{\kappa}_{\Delta N \leq 0}$ denote where the molecule is most reactive. (Part b) Second choice: the second-smallest values for $\Xi^{\kappa}_{\Delta N \leq 0}$, denoting the second most reactive carbon.

Table 2. Reactivity Transition Tables for 10-Hydroxy-10,9-borazarophenanthrene Using CHelpG Population Analysis with the Charges and Fukui Functions on Hydrogens Summed into the Adjacent Heavy Atoms[a]

(a)

	κ										
N	1.0	0.8	0.6	0.4	0.2	0	-0.2	-0.4	-0.6	-0.8	-1.0
-1.0	-0.312	-0.289	-0.267	-0.244	-0.233	-0.257	-0.282	-0.307	-0.331	-0.356	-0.381
-0.9	-0.312	-0.289	-0.265	-0.241	-0.218	-0.238	-0.259	-0.280	-0.301	-0.322	-0.343
-0.8	-0.312	-0.288	-0.263	-0.239	-0.214	-0.219	-0.236	-0.253	-0.270	-0.288	-0.305
-0.7	-0.312	-0.287	-0.262	-0.236	-0.211	-0.200	-0.214	-0.227	-0.240	-0.253	-0.266
-0.6	-0.312	-0.286	-0.260	-0.234	-0.208	-0.181	-0.191	-0.200	-0.210	-0.219	-0.228
-0.5	-0.312	-0.285	-0.258	-0.231	-0.204	-0.177	-0.168	-0.173	-0.179	-0.185	-0.190
-0.4	-0.312	-0.284	-0.257	-0.229	-0.201	-0.173	-0.145	-0.147	-0.149	-0.150	-0.152
-0.3	-0.312	-0.284	-0.255	-0.226	-0.197	-0.169	-0.140	-0.120	-0.118	-0.116	-0.114
-0.2	-0.312	-0.283	-0.253	-0.224	-0.194	-0.165	-0.135	-0.105	-0.088	-0.082	-0.076
-0.1	-0.312	-0.282	-0.251	-0.221	-0.191	-0.160	-0.130	-0.100	-0.069	-0.048	-0.038
0.0	-0.312	-0.281	-0.250	-0.219	-0.187	-0.156	-0.125	-0.094	-0.062	-0.031	
	Carbon 2		Carbon 4			Carbon 6			Carbon 8		

(b)

	κ										
N	1.0	0.8	0.6	0.4	0.2	0	-0.2	-0.4	-0.6	-0.8	-1.0
-1.0	-0.174	-0.180	-0.187	-0.208	-0.221	-0.206	-0.212	-0.219	-0.234	-0.264	-0.294
-0.9	-0.174	-0.178	-0.182	-0.197	-0.218	-0.194	-0.198	-0.202	-0.211	-0.237	-0.264
-0.8	-0.174	-0.176	-0.177	-0.185	-0.202	-0.190	-0.184	-0.185	-0.187	-0.211	-0.235
-0.7	-0.174	-0.173	-0.173	-0.174	-0.187	-0.186	-0.170	-0.169	-0.168	-0.185	-0.206
-0.6	-0.174	-0.171	-0.168	-0.165	-0.172	-0.181	-0.155	-0.152	-0.149	-0.158	-0.176
-0.5	-0.174	-0.169	-0.163	-0.158	-0.157	-0.162	-0.150	-0.135	-0.130	-0.132	-0.147
-0.4	-0.174	-0.166	-0.158	-0.150	-0.143	-0.143	-0.145	-0.119	-0.111	-0.105	-0.117
-0.3	-0.174	-0.164	-0.154	-0.143	-0.133	-0.124	-0.122	-0.111	-0.092	-0.082	-0.088
-0.2	-0.174	-0.161	-0.149	-0.136	-0.123	-0.111	-0.099	-0.094	-0.076	-0.060	-0.059
-0.1	-0.174	-0.159	-0.144	-0.129	-0.114	-0.099	-0.084	-0.069	-0.057	-0.039	-0.029
0.0	-0.174	-0.157	-0.139	-0.122	-0.104	-0.087	-0.070	-0.052	-0.035	-0.017	
	Carbon 2		Carbon 4			Carbon 6			Carbon 8		

[a] (Part a) First choice: the minimum values of $\Xi^{\kappa}_{\Delta N \leq 0}$ denote where the molecule is most reactive. (Part b) Second choice: the second-smallest values for $\Xi^{\kappa}_{\Delta N \leq 0}$, denoting the second most reactive carbon.

so the reaction is probably more electrostatically controlled than it is electron-transfer-controlled ($0 < \kappa < 1$). Electron transfer to the electrophile will be important, but incomplete ($\Delta N \approx -0.5$). In Tables 2a and 3a (on the basis of CHelpG and MSK population analysis, with the hydrogenic contributions summed into the adjacent heavy atoms), these experimental conditions place one near the transition region between carbon 8 and carbon 6, but in regions where carbon 8 would be predicted to be slightly more reactive. In Table 1a (on the basis of natural population analysis), this places

Table 3. Reactivity Transition Tables for 10-Hydroxy-10,9-borazarophenanthrene Using Merz−Singh−Kollman Population Analysis with the Charges and Fukui Functions on Hydrogens Summed into the Adjacent Heavy Atoms[a]

(a)

	κ										
N	1.0	0.8	0.6	0.4	0.2	0	-0.2	-0.4	-0.6	-0.8	-1.0
-1.0	-0.697	-0.631	-0.565	-0.499	-0.453	-0.439	-0.426	-0.412	-0.398	-0.385	-0.371
-0.9	-0.697	-0.631	-0.564	-0.498	-0.438	-0.421	-0.404	-0.386	-0.369	-0.351	-0.334
-0.8	-0.697	-0.630	-0.563	-0.497	-0.430	-0.402	-0.381	-0.360	-0.339	-0.318	-0.297
-0.7	-0.697	-0.630	-0.563	-0.496	-0.428	-0.384	-0.359	-0.334	-0.309	-0.284	-0.260
-0.6	-0.697	-0.630	-0.562	-0.494	-0.427	-0.365	-0.337	-0.308	-0.280	-0.251	-0.223
-0.5	-0.697	-0.629	-0.561	-0.493	-0.425	-0.358	-0.314	-0.282	-0.250	-0.218	-0.185
-0.4	-0.697	-0.629	-0.561	-0.492	-0.424	-0.356	-0.292	-0.256	-0.220	-0.184	-0.148
-0.3	-0.697	-0.629	-0.560	-0.491	-0.423	-0.354	-0.285	-0.230	-0.191	-0.151	-0.111
-0.2	-0.697	-0.628	-0.559	-0.490	-0.421	-0.352	-0.283	-0.214	-0.161	-0.118	-0.074
-0.1	-0.697	-0.628	-0.558	-0.489	-0.420	-0.350	-0.281	-0.212	-0.142	-0.084	-0.037
0.0	-0.697	-0.627	-0.558	-0.488	-0.418	-0.349	-0.279	-0.209	-0.139	-0.070	
	Carbon 2		Carbon 4			Carbon 6			Carbon 8		

(b)

	κ										
N	1.0	0.8	0.6	0.4	0.2	0	-0.2	-0.4	-0.6	-0.8	-1.0
-1.0	-0.513	-0.494	-0.481	-0.467	-0.433	-0.366	-0.318	-0.286	-0.254	-0.221	-0.194
-0.9	-0.513	-0.491	-0.473	-0.456	-0.431	-0.365	-0.307	-0.273	-0.238	-0.204	-0.175
-0.8	-0.513	-0.487	-0.466	-0.445	-0.424	-0.363	-0.296	-0.260	-0.223	-0.187	-0.156
-0.7	-0.513	-0.483	-0.458	-0.433	-0.409	-0.361	-0.294	-0.246	-0.208	-0.170	-0.136
-0.6	-0.513	-0.479	-0.451	-0.422	-0.394	-0.359	-0.292	-0.233	-0.193	-0.153	-0.117
-0.5	-0.513	-0.476	-0.443	-0.411	-0.379	-0.347	-0.290	-0.222	-0.178	-0.136	-0.097
-0.4	-0.513	-0.472	-0.436	-0.400	-0.364	-0.328	-0.287	-0.219	-0.163	-0.119	-0.078
-0.3	-0.513	-0.468	-0.429	-0.389	-0.349	-0.310	-0.270	-0.217	-0.148	-0.102	-0.058
-0.2	-0.513	-0.466	-0.421	-0.378	-0.334	-0.291	-0.248	-0.204	-0.145	-0.085	-0.039
-0.1	-0.513	-0.464	-0.414	-0.367	-0.320	-0.273	-0.225	-0.178	-0.131	-0.073	-0.019
0.0	-0.513	-0.462	-0.411	-0.359	-0.308	-0.257	-0.205	-0.154	-0.103	-0.051	
	Carbon 2		Carbon 4			Carbon 6			Carbon 8		

[a] (Part a) First choice: the minimum values of $\Xi^{\kappa}_{\Delta N \leq 0}$ denote where the molecule is most reactive. (Part b) Second choice: the second smallest values for $\Xi^{\kappa}_{\Delta N \leq 0}$, denoting the second most reactive carbon.

Table 4. Reactivity Transition Tables for 10-Methyl-10,9-borazarophenanthrene Using Natural Population Analysis with the Charges and Fukui Functions on Hydrogens Summed into the Adjacent Heavy Atoms[a]

(a)

	κ										
N	1.0	0.8	0.6	0.4	0.2	0	-0.2	-0.4	-0.6	-0.8	-1.0
-1.0	-0.067	-0.066	-0.074	-0.094	-0.115	-0.135	-0.159	-0.183	-0.207	-0.232	-0.256
-0.9	-0.067	-0.066	-0.069	-0.087	-0.105	-0.123	-0.143	-0.165	-0.187	-0.209	-0.231
-0.8	-0.067	-0.065	-0.065	-0.080	-0.096	-0.112	-0.128	-0.147	-0.166	-0.186	-0.205
-0.7	-0.067	-0.064	-0.062	-0.073	-0.086	-0.100	-0.113	-0.129	-0.146	-0.163	-0.179
-0.6	-0.067	-0.064	-0.061	-0.066	-0.077	-0.088	-0.099	-0.111	-0.125	-0.140	-0.154
-0.5	-0.067	-0.063	-0.059	-0.059	-0.067	-0.076	-0.084	-0.093	-0.105	-0.117	-0.128
-0.4	-0.067	-0.063	-0.058	-0.054	-0.058	-0.064	-0.070	-0.076	-0.084	-0.093	-0.102
-0.3	-0.067	-0.062	-0.057	-0.052	-0.048	-0.052	-0.056	-0.060	-0.064	-0.070	-0.077
-0.2	-0.067	-0.061	-0.056	-0.050	-0.045	-0.040	-0.042	-0.043	-0.045	-0.047	-0.051
-0.1	-0.067	-0.061	-0.055	-0.049	-0.043	-0.036	-0.030	-0.027	-0.026	-0.025	-0.026
0.0	-0.067	-0.060	-0.054	-0.047	-0.040	-0.033	-0.027	-0.020	-0.013	-0.007	
	Carbon 2		Carbon 4			Carbon 6			Carbon 8		

(b)

	κ										
N	1.0	0.8	0.6	0.4	0.2	0	-0.2	-0.4	-0.6	-0.8	-1.0
-1.0	-0.033	-0.054	-0.065	-0.086	-0.110	-0.134	-0.156	-0.176	-0.196	-0.217	-0.237
-0.9	-0.033	-0.051	-0.064	-0.078	-0.100	-0.122	-0.141	-0.159	-0.177	-0.195	-0.213
-0.8	-0.033	-0.049	-0.063	-0.070	-0.089	-0.109	-0.127	-0.143	-0.158	-0.174	-0.190
-0.7	-0.033	-0.047	-0.060	-0.063	-0.079	-0.096	-0.113	-0.126	-0.140	-0.153	-0.166
-0.6	-0.033	-0.044	-0.055	-0.057	-0.069	-0.083	-0.097	-0.110	-0.121	-0.131	-0.142
-0.5	-0.033	-0.042	-0.050	-0.056	-0.059	-0.070	-0.082	-0.093	-0.102	-0.110	-0.119
-0.4	-0.033	-0.039	-0.046	-0.052	-0.050	-0.057	-0.066	-0.075	-0.083	-0.089	-0.095
-0.3	-0.033	-0.037	-0.041	-0.045	-0.047	-0.045	-0.051	-0.058	-0.064	-0.067	-0.071
-0.2	-0.033	-0.035	-0.036	-0.038	-0.039	-0.039	-0.036	-0.040	-0.043	-0.046	-0.047
-0.1	-0.033	-0.032	-0.031	-0.030	-0.029	-0.029	-0.028	-0.024	-0.023	-0.024	-0.024
0.0	-0.033	-0.030	-0.027	-0.023	-0.020	-0.017	-0.013	-0.010	-0.007	-0.003	
	Carbon 2		Carbon 4			Carbon 6			Carbon 8		

[a] (Part a) First choice: the minimum values of $\Xi^{\kappa}_{\Delta N \leq 0}$ denote where the molecule is most reactive. (Part b) Second choice: the second smallest values for $\Xi^{\kappa}_{\Delta N \leq 0}$, denoting the second most reactive carbon.

one in the region where carbon 6 is the most favorable, though one is still reasonably close to the region where carbon 8 would be favored.

Examining the second-choice reactivity transition diagrams in Tables 1b (NPA), 2b (ChelpG), and 3b (MSK), one finds that carbon 4 and carbon 2 are also predicted to be reactive.

Carbon 4 would be predicted to be more reactive than carbon 2 except when electron-transfer effects are dominant ($\kappa \approx -1$). Experimentally, carbon 4 is much less reactive than one would expect on the basis of electronic-structure considerations alone. Dewar noted this effect as early as 1956 and suggested that the abnormally low reactivity of carbon 4 is due to steric effects, rather than electronic structure considerations.[43] Our indicator does not include any information about steric hindrance, so it is not surprising that it also overestimates the reactivity of carbon 4. Examining the values of $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$ where carbon 4 emerges as the "second choice" reactivity site, we observe that there is a large gap between the predicted reactivity of carbon 4 and the reactivity of the most reactive site.[47] This indicates that, whenever carbon 4 is the "second choice" reactive site, the "first choice" reactive site is much more reactive. When the second-choice reactive site is much less favorable than the first-choice reactive site, one expects that the secondary product will be a very small percentage of the total yield and, thus, difficult to isolate and characterize.

The overall picture that emerges is fairly convincing: carbon 6 and carbon 8 are the most reactive, with the population analysis schemes based on electrostatic fitting successfully predicting that carbon 8 should be slightly more reactive. Carbon 2 is predicted to be much less reactive, which agrees with the experimental observation that reactions at carbon 2 only occur under very special conditions: chlorination at high temperatures with excess chlorine.[41]

Our analysis also indicates that, under the appropriate conditions, carbon 6 should be more reactive than carbon 8. Indeed, Friedel−Crafts acetylation of 10-hydroxy-10,9-borazarophenanthrene occurs predominately at carbon 6, with the diacetylated product corresponding to reaction at both carbon 6 and carbon 8 as a secondary product.[44] The monoacetylated product corresponding to reaction at carbon 8 is not observed. Dewar explained this by hypothesizing that the nitrogen atom is complexed by the $AlCl_3$ catalyst, which might make carbon 8 sterically inaccessible.[44,48] Our analysis suggests another possibility, however. The mechanism of Friedel−Crafts acetylation involves the addition of the resonance-stabilized acetyl carbocation (CH_3CO^+) to the aromatic ring. Because this is a carbocation, we expect that $\Delta N \approx -1$. Moreover, this carbocation is resonance-stabilized and, additionally, might be somewhat stabilized by complexation of the catalyst. As such, the charge on the electrophilic carbon atom is relatively small, and moreover, we expect that the acetyl carbocation is relatively soft. Accordingly, we expect that the reaction will be mostly electron-transfer-controlled, $-1 < \kappa < 0$. Referring to Tables 1a−3a, one finds that, under these conditions, carbon 6 is more reactive than any of the other molecular sites.

To this point, we have focused on hydroxylated 10,9-borazarophenanthrene (R = OH) rather than on the methylated compound. For the hydroxylated compound, the various population analysis schemes gave qualitatively similar results. This was not true for $R = CH_3$, and summing the charges of the hydrogen atoms into the adjacent heavy atoms did not substantially improve the agreement between the various population analysis schemes. Indeed, even though the MSK and CHelpG charges are both based on fitting the electrostatic potential, the results from these population analysis schemes were significantly different for $R = CH_3$. This can be contrasted with the favorable results in Figure 7a, where the different electrostatic potential fitting methods gave substantially similar results. Given the unreliability of our charges when $R = CH_3$, we have chosen to focus our discussion on the hydroxylated compound. Nonetheless, the main results for $R = CH_3$ are broadly similar: carbon 6 and carbon 8 are ordinarily the most reactive, with carbon 4 and carbon 2 being less reactive. In many cases, carbon 4 is more reactive than carbon 2, and steric effects need to be invoked to describe the lack of reactivity at this site.[43] An exception occurs for the natural population analysis scheme, which is reported in Table 4a and b. In that case, carbon 8, carbon 6, and carbon 2 are all reasonably reactive, though carbon 8 and carbon 6 are likely to be the most reactive sites for chlorination and nitration.[49]

IV. Summary

To explore the validity of our methods, we studied three molecules where frontier molecular orbital theory fails to predict the correct reactivity. This analysis underscores, among other things, the importance of chemical reasoning when applying reactivity indicators. For example, while our results for isoquinoline were in stark disagreement with the experiment, isoquinoline is protonated under the experimental conditions. Adding a proton to isoquinoline brought our predictions into agreement with the experimental results. Similarly, when exploring electrophilic aromatic substitution on 10-R-10,9-borazarophenanthrenes, it was important to remember that "ipso" attack—attack on carbon atoms that do not have any adjacent hydrogen atoms—is a mechanistic dead end because there is no leaving group that the electrophile can "substitute" for.

The complex reactivity of 10-R-10,9-borazarophenanthrenes provided an ideal situation for testing the general-purpose reactivity indicator derived in the first paper of this series. Experimentally, chlorination, bromination, and nitration of 10-R-10,9-borazarophenanthrene occurs primarily on carbons 6 and 8 (with carbon 8 slightly favored). At higher temperatures, chlorination of 10-hydroxy-10,9-borazarophenanthrenes also gives some of the trichlorinated product, with the additional reaction occurring at carbon 2. Friedel−Crafts acetylation of 10-R-10,9-borazarophenanthrenes occurs primarily on carbon 6. Using the condensed version of our general-purpose reactivity indicator, $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$, we were able to explain these results: carbon 6 and carbon 8 are the most reactive sites, with carbon 8 favored for hard electrophiles; carbon 2 is significantly less reactive than carbon 6 and carbon 8 but is predicted to be more reactive than any other site except the sterically hindered carbon 4; because the acetyl carbocation is resonance-stabilized and a very good electron acceptor, the most reactive site should be carbon 6. To obtain these results, we used reactivity transition tables, which list the value of $\Xi^{\kappa}_{\Delta N\leq 0,\alpha}$ at the most reactive (Tables 1a−4a) and second most-reactive (Tables 1b−4b) sites. The entries in the table are then color-coded according to the identity of the most reactive site. For molecules with multiple reactive

sites, reactivity transition tables provide a useful way of predicting how the regioselectivity of the molecule depends on the characteristics of the attacking reagent.

We are encouraged by these results. The new indicator, $\Xi^{\kappa}_{\Delta N}$, coincides with conventional conceptual DFT reactivity indicators whenever they work but also provides theoretical insight and computational results that can be used to clarify situations where conventional reactivity indicators fail. Our future efforts in this area will focus on extending these results to other types of chemical reactions and refining the present indicator to account for the polarization of reactive sites.

Acknowledgment. Helpful discussions with Dr. David C. Thompson and Dr. Cherif Matta are acknowledged. NSERC, the Canada Research Chairs, and PREA provided funding for the Canadian authors. This research was performed when the second author visited McMaster University in the winter of 2005, and she wishes to thank the chemistry department at McMaster University for their hospitality.

References

(1) Anderson, J. S. M.; Melin, J.; Ayers, P. W. Conceptual Density-Functional Theory for General Chemical Reactions, Including Those That Are Neither Charge- nor Frontier-Orbital-Controlled. 1. Theory and Derivation of a General-Purpose Reactivity Indicator. *J. Chem. Theory Comput.* **2007**, *3*, 358−374.

(2) Klopman, G. Chemical Reactivity and the Concept of Charge and Frontier-Controlled Reactions. *J. Am. Chem. Soc.* **1968**, *90*, 223−234.

(3) Parr, R. G.; Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105* (26), 7512−7516.

(4) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. Electronegativity: The Density Functional Viewpoint. *J. Chem. Phys.* **1978**, *68* (8), 3801−3807.

(5) Yang, W.; Mortier, W. J. The Use of Global and Local Molecular Parameters for the Analysis of the Gas-Phase Basicity of Amines. *J. Am. Chem. Soc.* **1986**, *108* (19), 5708−11.

(6) Ayers, P. W.; Morrison, R. C.; Roy, R. K. Variational Principles for Describing Chemical Reactions: Condensed Reactivity Indices. *J. Chem. Phys.* **2002**, *116* (20), 8731−8744.

(7) Langenaeker, W.; Demel, K.; Geerlings, P. Quantum-Chemical Study of the Fukui Function As a Reactivity Index. 2. Electrophilic Substitution on Mono-Substituted Benzenes. *THEOCHEM* **1991**, *80*, 329−342.

(8) Meneses, L.; Tiznado, W.; Contreras, R.; Fuentealba, P. A Proposal for a New Local Hardness as Selectivity Index. *Chem. Phys. Lett.* **2004**, *383* (1−2), 181−187.

(9) Martinez, A.; Vazquez, M. V.; Carreon-Macedo, J. L.; Sansores, L. E.; Salcedo, R. Benzene Fused Five-Membered Heterocycles. A Theoretical Approach. *Tetrahedron* **2003**, *59* (34), 6415−6422.

(10) Ayers, P. W.; Levy, M. Perspective on "Density Functional Approach to the Frontier-Electron Theory of Chemical Reactivity" by Parr RG, Yang W (1984). *Theor. Chem. Acc.* **2000**, *103* (3−4), 353−360.

(11) Yang, W.; Parr, R. G.; Pucci, R. Electron Density, Kohn−Sham Frontier Orbitals, and Fukui Functions. *J. Chem. Phys.* **1984**, *81* (6), 2862−2863.

(12) Parr, R. G.; Yang, W. Density Functional Approach to the Frontier-Electron Theory of Chemical Reactivity. *J. Am. Chem. Soc.* **1984**, *106* (14), 4049−4050.

(13) Melin, J.; Aparicio, F.; Subramanian, V.; Galvan, M.; Chattaraj, P. K. Is the Fukui Function a Right Descriptor of Hard−Hard Interactions? *J. Phys. Chem. A* **2004**, *108* (13), 2487−2491.

(14) Dewar, M. J. S. A Critique of Frontier Orbital Theory. *THEOCHEM* **1989**, *59*, 301−23.

(15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Peersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvetti, O.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian Inc.: Wallingford, CT, 2004.

(16) Becke, A. D. Density-Functional Exchange-Energy Approximation With Correct Asymptotic-Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38* (6), 3098−3100.

(17) Becke, A. D. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648−5652.

(18) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37* (2), 785−789.

(19) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. 20. Basis Set for Correlated Wave-Functions. *J. Chem. Phys.* **1980**, *72* (1), 650−654.

(20) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions 1. *J. Chem. Phys.* **1955**, *23*, 1833.

(21) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions 2. *J. Chem. Phys.* **1955**, *23*, 1841.

(22) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions 4. *J. Chem. Phys.* **1955**, *23*, 2343.

(23) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions 3. *J. Chem. Phys.* **1955**, *23*, 2338.

(24) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular Interactions from a Natural Bond Orbital, Donor−Acceptor Viewpoint. *Chem. Rev.* **1988**, *88* (6), 899−926.

(25) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural-Population Analysis. *J. Chem. Phys.* **1985**, *83* (2), 735−746.

(26) Reed, A. E.; Weinhold, F. Natural Bond Orbital Analysis of Near-Hartree−Fock Water Dimer. *J. Chem. Phys.* **1983**, *78* (6), 4066−4073.

(27) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11* (4), 431−439.

(28) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5* (2), 129−145.

(29) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials − The Need for High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.* **1990**, *11* (3), 361−373.

(30) Dewar, M. J. S.; Maitlis, P. M. Electrophilic Substitution. Part XI. Nitration of Some Six-Membered Nitrogen-Heterocyclic Compounds in Sulfuric Acid. *J. Chem. Soc.* **1957**, 2521−2528.

(31) Flurchick, K.; Bartolotti, L. Visualizing Properties of Atomic and Molecular-Systems. *J. Mol. Graphics* **1995**, *13* (1), 10−13.

(32) Bartolotti, L. J.; Ayers, P. W. An Example where Orbital Relaxation Is an Important Contribution to the Fukui Function. *J. Phys. Chem. A* **2005**, *109* (6), 1146−1151.

(33) We have confirmed that the qualitative features of these plots do not alter when the isodensity surface is moved closer to the molecule. To the eye, plots generated using the 0.001 and 0.002 isodensity surfaces look the same as the ones in Figure 1.

(34) Berkowitz, M. Density Functional-Approach to Frontier Controlled Reactions. *J. Am. Chem. Soc.* **1987**, *109* (16), 4823−4825.

(35) Chattaraj, P. K. Chemical Reactivity and Selectivity: Local HSAB Principle versus Frontier Orbital Theory. *J. Phys. Chem. A* **2001**, *105* (2), 511−513.

(36) Hocquet, A.; Toro-Labbe, A.; Chermette, H. Intramolecular Interactions along the Reaction Path of Keto−Enol Tautomerism: Fukui Functions as Local Softnesses and Charges as Local Hardnesses. *THEOCHEM* **2004**, *686* (1−3), 213−218.

(37) Because of the slow asymptotic decay of the electrostatic potential and the Fukui potential, plots such as that in Figure 1 do not provide useful information.

(38) Miller, B. *Advanced Organic Chemistry: Reactions and Mechanisms*; Prentice Hall: Upper Saddle River, NJ, 1998.

(39) The reader might notice that, with regard to atomic charges, the nitrogen atom and carbon 4 have similar negative charges in the electrostatic charge-fitting schemes (Figure 2a). Note, however, that, once one considers the hydrogen-summed charges (Figure 2b), protonation on the nitrogen is clearly preferred. The protonation of carbon 4 would be associated with the electrophilic aromatic substitution of hydrogen for hydrogen. So even if carbon 4 were protonated, the concentration of the tetrahedral reactive intermediate would be very, very low. It follows that the case where the molecule is protonated at carbon 4 is unlikely to contribute to the overall chemical reactivity in any significant way. The protonation of carbon 4 could be studied by deuterium exchange experiments, but we are unaware of experimental studies of this type.

(40) Dewar, M. J. S. Heteroaromatic Boron Compounds. *Prog. Boron Chem.* **1964**, *1*, 235−263.

(41) Dewar, M. J. S.; Kubba, V. P. New Heteroaromatic Compounds. VII. Chloro and Bromo Derivatives of 10-Hydroxy-10,9-borazarophenanthrene. *J. Org. Chem.* **1960**, *25*, 1722−1724.

(42) Dewar, M. J. S.; Kubra, V. P. New Heteroaromatic Compounds IV. The Nitration and Chlorination of 10-Methyl-10,9-borazarophenanthrene. *Tetrahedron* **1959**, *7*, 213−222.

(43) Dewar, M. J. S.; Warford, E. W. T. Electrophilic Substitution. III. The Nitration of Phenanthrene. *J. Chem. Soc.* **1956**, 3570−3572.

(44) Dewar, M. J. S.; Kubba, V. P. New Heteroaromatic Compounds. XI. Friedel−Crafts Accetylation of 10-Methyl- and 10-Hydroxyl-10,9-borazarophenanthrene. *J. Am. Chem. Soc.* **1961**, *83*, 1757−1760.

(45) To the extent that electrophilic attack on the nitrogen atom does occur, it will be associated with molecular decomposition (just like protonation on this site). Though Dewar's experiments often produced the featured products in high yield, some of the impurities might be due to the decomposition of the starting reagent induced by electrophilic attack on the nitrogen atom. It is impossible to quantify the importance of this process, so we will focus on the reaction pathways for which there is incontrovertible experimental evidence.

(46) One might argue that, because our model is based on changes in the interaction potential, charges from electrostatic potential fitting are the more natural choice. While we share this view, it is also true that, if one is far enough from the molecule, the charges based on natual population analysis (or any other density-matrix population analysis method) reproduce the electrostatic potential.

(47) This is especially true in Tables 2 and 3 but not as true for Table 1.

(48) We attempted to add a sodium cation near the nitrogen atom to mimic the complexation. If there is a potential energy well in this area, it seems to be very shallow: in our calculations, the cation always either dissociated or migrated to the top of one of the aromatic rings.

(49) Again, naive application of the reactivity indicator sometimes indicates that ipso attack is favored. The carbon of the methyl group is also predicted to be rather reactive, but electrophilic substitution at this carbon cannot occur because of the absence of the electrofuge. In the text and in the tables, we focus our treatment on the ring carbons that are subject to substitution, carbons 1−8.

CT6001658

JCTC Journal of Chemical Theory and Computation

QM/MM Minimum Free-Energy Path: Methodology and Application to Triosephosphate Isomerase

Hao Hu, Zhenyu Lu, and Weitao Yang*

Department of Chemistry, Duke University, Durham, North Carolina 27708

Received July 21, 2006

Abstract: Structural and energetic changes are two important characteristic properties of a chemical reaction process. In the condensed phase, studying these two properties is very challenging because of the great computational cost associated with the quantum mechanical calculations and phase space sampling. Although the combined quantum mechanics/molecular mechanics (QM/MM) approach significantly reduces the amount of the quantum mechanical calculations and facilitates the simulation of solution-phase and enzyme-catalyzed reactions, the required quantum mechanical calculations remain quite expensive and extensive sampling can be achieved routinely only with semiempirical quantum mechanical methods. QM/MM simulations with ab initio QM methods, therefore, are often restricted to narrow regions of the potential energy surface such as the reactant, product and transition state, or the minimum-energy path. Such ab initio QM/MM calculations have previously been performed with the QM/MM-free energy (QM/MM-FE) method of Zhang et al. (*J. Chem. Phys.* **2000**, *112*, 3483−3492) to generate the free-energy profile along the reaction coordinate using free-energy perturbation calculations at fixed structures of the QM subsystems. Results obtained with the QM/MM-FE method depend on the determination of the minimum-energy reaction path, which is based on local conformations of the protein/solvent environment and can be difficult to obtain in practice. To overcome the difficulties associated with the QM/MM-FE method and to further enhance the sampling of the MM environment conformations, we develop here a new method to determine the QM/MM minimum free-energy path (QM/MM-MFEP) for chemical-reaction processes in solution and in enzymes. Within the QM/MM framework, we express the free energy of the system as a function of the QM conformation, thus leading to a simplified potential of mean force (PMF) description for the thermodynamics of the system. The free-energy difference between two QM conformations is evaluated by the QM/MM free-energy perturbation method. The free-energy gradients with respect to the QM degrees of freedom are calculated from molecular dynamics simulations at given QM conformations. With the free energy and free-energy gradients in hand, we further implement chain-of-conformation optimization algorithms in the search for the reaction path on the free-energy surface without specifying a reaction coordinate. This method thus efficiently provides a unique minimum free-energy path for solution and enzyme reactions, with structural and energetic properties being determined simultaneously. To further incorporate the dynamic contributions of the QM subsystem into the simulations, we develop the reaction path potential of Lu, et al. (*J. Chem. Phys.* **2004**, *121*, 89−100) for the minimum free-energy path. The combination of the methods developed here presents a comprehensive and accurate treatment for the simulation of reaction processes in solution and in enzymes with ab initio QM/MM methods. The method has been demonstrated on the first step of the reaction of the enzyme triosephosphate isomerase with good agreement with previous studies.

Introduction

With the overwhelming details of real-time atomic motions, computer simulations have provided unprecedented insight into the puzzle of chemical reaction processes.[1−5] Complementary to experimental studies, simulation methods can provide information that is often not easily accessed by conventional experimental approaches. For example, simulations can determine the transition-state structure of a reaction process, which is difficult to obtain from experimental

* To whom correspondence should be addressed. E-mail: weitao.yang@duke.edu.

10.1021/ct600240y CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/06/2007

methods. Simulations can also reveal the site-specific interactions influencing an enzymatic reaction process for which experimental studies are more costly or are often derived from indirect evidence.[6−13] However, challenging problems persist in simulating reaction processes in solution and in enzymes because of many technical limitations such as the accuracy of the force field and the convergence of the statistical sampling.

For reactions in solution or enzymes, quantum mechanical (QM) treatment of the whole molecular system is computationally prohibitive in general because there are too many electronic degrees of freedom. On the other hand, classical molecular mechanics (MM) is incapable of describing the electron redistribution during the bond breaking/forming events in reactive processes. To address this difficulty, a method was proposed to combine quantum mechanics and molecular mechanics in the simulations.[14] In this hybrid approach, only a small portion of the molecular system that is important to the reaction is treated by QM, while the rest of the system is simulated by simplified MM force fields. Presumably, the QM treatment accurately captures the most important changes at the site of the chemical reaction, while the MM treatment takes into account the contributions of the environment as an approximate, yet computationally economic, solution.[14,15]

Since there is no restriction on the choice of the QM level of theory in the QM/MM approach, many different QM methods have been used in QM/MM simulations of reaction processes. The QM methods vary from the semiempirical level empirical valence bond (EVB),[3] MNDO,[16] AM1,[17] PM3,[18,19] and self-consistent charge density functional tight binding (SCC-DFTB)[20,21] to ab initio Hartree−Fock, MP2, and density functional theory (DFT)[22−24] methods. Compared with the plentiful choices of QM methods, the MM subsystems vary much less between simulations and often assume the form of common classical MM force fields which are composed of empirical covalent terms (i.e., bond, bond angle, dihedral, and improper dihedral terms) and nonbonded van der Waals and electrostatic interactions. With such flexible combinations of QM and MM methods, simulation studies of many reaction processes have demonstrated that within current computational capabilities, the QM/MM method has become the most effective way for simulating condensed phase reactions.[5, 9−13, 25−28]

Different QM methods usually possess correspondingly different levels of accuracy and computational cost, both of which may vary significantly for a given simulation. Consequently, dependent on the implementation of a specific QM level of theory, specific applications of QM/MM methods have followed different routes to maximize the effectiveness of the QM/MM method. Each route essentially evolves into a distinct QM/MM method with each method focusing on different aspects of the reaction process. As each method possesses different advantages and disadvantages, a brief review of these QM/MM methods will be given to set the stage for presenting our new development.

In general, when semiempirical methods such as EVB, MNDO, AM1, PM3, and SCCDFTB are used as the QM model, the computational cost is low with the currently available computational capacity. As a result, classical approaches such as umbrella sampling,[29] PMF calculations, and thermodynamic integration (TI) can be applied in a straightforward manner to the calculation of the structural and energetic properties of a reaction. Usually, a reaction coordinate is chosen as a combination of geometric and energetic terms. The conformations are then sampled, and the free-energy changes of the molecular system along this reaction coordinate are computed. Because broad sampling of phase space is attainable in this situation, the convergence of the results is satisfactory. However, the results may be less accurate and reliable than those of the ab initio QM calculations because of the well-known deficiencies in semiempirical QM methods, a problem which originates in the approximations made both in the theory and in the parametrization process. It is thus highly desirable to use high-level accurate QM methods. This demand has in fact motivated the development of several QM/MM free-energy simulation techniques based on ab initio quantum mechanics.

Jorgensen's group developed the quantum mechanical free energy (QM-FE) approach.[30−34] In this approach, the reaction path is optimized for a model reaction system in the gas phase. To compute the free energy of the reaction process, free-energy perturbation (FEP) is applied along the preoptimized gas-phase reaction path with the inclusion of the QM/MM interactions. The interactions between the QM and MM subsystems are treated classically by either taking them directly from MM force fields or by fitting them to a QM calculation for a molecular cluster in gas phase. This approach has been successfully applied to many solution reactions and some enzymatic reactions.[35] However, since the reaction path is predetermined in the gas phase, it is not clear whether the solution reaction will follow the same path. The results would most likely be more reliable if the path were determined in situ.

Warshel's group pioneered in the development and application of QM/MM methods and has recently constructed several approaches to calculate accurate ab initio QM/MM free energies.[25,26] A strategy for combining those approaches, namely, the hybrid ab initio quantum mechanics/molecular mechanics (QM(ai)/MM) method,[36−40] was developed for the study of condensed phase reaction processes. The first step and the essence of this method is to build an EVB potential that approximates the potential-energy surface from ab initio QM methods. In the second step, the EVB potential is used to perform long time-scale dynamic sampling of the whole molecular system which ensures the convergence of statistical sampling in phase space. The free energies are also calculated on this EVB surface using standard sampling approaches. The final step of this method is to recover the free-energy difference between the EVB surface and the ab initio surface. Free-energy perturbation combined with the linear response approximation (LRA) is used to evaluate this term. The application of this method to a wide range of problems has demonstrated its success. There are, however, several technical concerns. The first is that the construction of the EVB potential is a nontrivial problem and often requires clear and clever understanding of the chemical process. The second relates to how well the EVB potential approximates the ab

initio surface. The third concern is that this method does not provide direct determination of the transition-state structure. Nevertheless, the idea of using a reference potential has proven to be successful and applicable in general, from which several more recent approaches share a similar spirit.[41]

Our group has developed a combined quantum mechanics/molecular mechanics free-energy perturbation (QM/MM-FE) method for the simulation of enzymatic reaction processes.[1,10,42,43] Our approach consists of three major components: the pseudobond ab initio QM/MM method, which provides a smooth interface between the QM and MM subsystem and thus a well-defined potential energy surface, the efficient iterative optimization procedure, which determines the reaction paths within a realistic enzyme environment, and free-energy calculations, which take into account the fluctuation of the enzyme system. To calculate the free energy of activation (or the potential of mean force) in the QM/MM-FE method, we made two assumptions: (1) the dynamics of the QM and MM subsystems are independent of each other, and (2) the QM subsystem fluctuations are harmonic. We then calculated the contribution to the free energy from the fluctuations in the MM subsystem with the FEP method and approximated the contribution to the free energy from the fluctuations in the QM subsystem with harmonic frequency calculations. In addition, we approximate the QM/MM electrostatic interactions by the interactions between point charges of MM atoms and point charges of QM atoms; the latter being fitted to the QM electrostatic potential (ESP). This approximation leads to pairwise interactions that are separable and cost much less in the calculations. Therefore, the reaction path for the entire enzyme system can be iteratively optimized by this simplified QM/MM description.[1,43] To further include the dynamical contribution of the QM subsystem, the reaction path potential (RPP) method was developed.[2] Applications of this QM/MM-FE method have been shown to be successful.[11,12,44,45] The accuracy of the QM/MM-FE method has been tested against full free-energy simulations in two other laboratories[41,46] and has been shown to be excellent. Further efforts have also been made, employing the idea of reference potentials, to improve the accuracy in the construction of the thermodynamic cycle and FEP simulations. The main limitation of the QM/MM-FE method is, however, that the optimization of the reaction path depends on the choice of the initial conformation,[47] although it is debatable how much this dependence can bias the results in the simulations of enzymatic reactions (Cisneros and Yang, unpublished work). In solution reactions, nevertheless, this dependence becomes the main obstacle in the application of the static, iterative optimization to the calculation of reaction paths[1,43] because of the disorder and rapid change of the positions of solvent molecules.

Extension of the capability of the QM/MM-FE approach thus requires elimination of the conformational dependence of the reaction path. To do so, a naïve approach would be to carry out a set of several QM/MM-FE simulations with different starting conformations and then average over the individual simulation results. Although logically sound, this idea becomes practically intractable for two reasons. First, when the MM environments undergo significant conformational change during the reaction process, there is no guarantee that a converged reaction path can be obtained for given starting conformations. Second, averaging over results of a set of simulations is not trivial because proper weighting is required for each simulation. To make the result meaningful, a rigorous theory is required to guide the selection of starting structures and the averaging of results. This theory is currently lacking.

To solve this problem, the conformational dependence of the QM/MM-FE method must be traced back to its origin. Like many other ab initio QM/MM methods, the potential-energy surface is sampled first in the QM/MM-FE method, and the free-energy profile is built afterward for selected points on that energy surface. Consequently, the results depend on the selection of the initial conformation whose energy surface is used for reaction path optimization. To remove this dependence, the reaction path must be determined on a free-energy surface in which the contributions of the MM conformations are appropriately included by ensemble averaging. The choice of an initial conformation thus becomes irrelevant in the construction of the reaction path. Optimization of the reaction path on the free-energy surface also possesses another advantage. That is, the interpretations of experimental studies are primarily based on classical transition state theory which relates the reaction rate constant with the free-energy change. Thus, a theoretical free-energy description of the reaction process correlates naturally with experimental study. In other words, the reaction path is computed on the free-energy surface so that there is no need to map between different surfaces; the structures of the reactant state, transition state, and product state determined in this manner genuinely match the definitions in the classical transition state theory. For this purpose, methods for sampling the free-energy surface of the reaction system must be developed.

Simulation methods have been proposed for the exploration of the free-energy (as opposed to the potential energy) reaction path for small molecular reaction systems in gas phase, since the calculation of the free energy is achievable for those systems even with high-level QM methods.[48−50] In this case, the full system is treated quantum mechanically, and the free energy is simulated with the "brute-force" approach (i.e., direct sampling of the phase space of the entire QM/MM system). The free-energy path is determined with the assistance of a predefined reaction coordinate. It is obvious that such methods cannot be applied to reactions in condensed phases for which the direct-phase space sampling of the QM system is prohibitive. For those complicated reaction systems, methods have been proposed to explore the free-energy surface of the system under the QM/MM framework, mainly, with semiempirical QM methods.[51−55] Applications have been reported but are limited mostly to the optimization of the transition-state structure of aqueous reactions. Thus, a complete, practical, and effective treatment based on ab initio quantum mechanics is still lacking.

We develop here an ab initio QM/MM minimum free-energy path (QM/MM-MFEP) method to achieve dual goals in a single simulation (i.e., optimization of the reaction path

on a free-energy surface and calculation of the free-energy profile of the reaction process). The essence of this method is to calculate the reaction path on the PMF surface of the QM/MM system. To accomplish these goals, we develop formulas to compute both the free energies and the free-energy gradients of the system so that we are able to use all the degrees of freedom of the QM subsystem to define the reaction coordinate and hence optimize the reaction path and calculate the associated free-energy changes. The relative free energies between different QM conformations are computed by the QM/MM-FE approach previously developed in our laboratory; the free-energy gradients of the QM subsystem are computed through molecular dynamics sampling of the MM environment. The reaction path is then optimized by means of a chain-of-conformation approach. Compared with other approaches, this method has several distinct features: it drastically minimizes the computational need of the QM calculations which then allows the use of a very accurate high-level QM method; it uses all of the QM degrees of freedom to define the reaction coordinate which relieves the bias from an improper choice of the reaction coordinate; it naturally generates a one-dimensional reaction profile without the need of converting from a high-dimensional potential-energy surface; and most importantly, the reaction path is optimized on the free energy surface, and thus there is no longer a dependence on the choice of the initial molecular conformations.

The paper is organized as follows. We will first review the QM/MM approach with approximate QM/MM electrostatic interactions. We will then show how to approximate the free energy as a function of the degrees of freedom of a selected subset and how to calculate the free-energy gradients which allows conformation/path minimization on a free-energy surface. With the QM/MM free-energy perturbation approach, we then show how to optimize the free-energy reaction path with available chain-of-conformation methods such as the nudged elastic band (NEB) method[56] and the Ayala−Schlegel second-order minimum-energy path (MEP) method,[57] both having been adapted and extended by our laboratory for application to QM/MM simulations of large and complex systems.[10,58] We can also use the most recent and very efficient quadratic string method (QSM)[59] for reaction path determination. We further introduce an RPP method[2] using the mean-field approximation for the contribution of the effects of the MM environment. Finally, we present an application of this method to the first step of the reaction catalyzed by triosephosphate isomerase (TIM).

Theory

QM/MM Hamiltonian with ESP Charge Simplification. For simulating complex reaction processes in the condensed phase, we describe a reaction system by the combined QM/MM approach: we select a small structural part of the system to be described by the QM method and the rest by an MM force field. The total energy of the whole system is then

$$E(r_{QM}, r_{MM}) = E_{QM}(r_{QM}) + E_{QM/MM}(r_{QM}, r_{MM}) + E_{MM}(r_{MM}) \quad (1)$$

where r_{QM} and r_{MM} represent the coordinates of the QM and MM subsystems, respectively. The three terms on the right-hand side are the quantum mechanical energy of the QM subsystem, the interaction between the QM and MM subsystems, and the molecular mechanical energy of the MM subsystem, respectively. The interaction between the QM and MM parts is composed of electrostatic, van der Waals, and covalent QM/MM link terms. The bonds connecting the QM and MM parts, which are usually present in an enzymatic reaction system, are described by the pseudobond method.[42] Thus, one can decompose the total energy of the system into the sum of different components as

$$E(r_{QM}, r_{MM}) = E_{QM}(r_{QM}) + E_{QM/MM,ele}(r_{QM}, r_{MM}) + E_{QM/MM,vdw}(r_{QM}, r_{MM}) + E_{QM/MM,cova}(r_{QM}, r_{MM}) + E_{MM}(r_{MM}) \quad (2)$$

The first two terms on the right-hand side, the QM energy and the electrostatic interactions between the QM and MM parts, are obtained together via a combined QM/MM Hamiltonian in a self-consistent electronic structure calculations. To calculate the free energy, we follow the QM/MM-FE approach developed previously[1] and make an approximation here that the QM/MM interaction can be further decomposed as

$$\begin{aligned} E_{QM}(r_{QM}) + E_{QM/MM,ele}(r_{QM}, r_{MM}) &\simeq E'_{QM}(r_{QM}) + \sum_{j\in MM}\sum_{i\in QM}\frac{q_j Q_i}{|r_i - r_j|} \\ &= E'_{QM}(r_{QM}) + E'_{QM/MM,ele}(r_{QM}, r_{MM}) \end{aligned} \quad (3)$$

where q and Q are the point charges of the MM and QM atoms, respectively. The former is taken from the MM force field, and the latter must be determined by fitting the ESP from QM calculations[60,61] in a proper MM environment. The underlying assumption here is that the fluctuation of the electrostatic potential of the QM subsystem caused by the fluctuation of the MM conformations can be neglected because it is small compared to the magnitude of the QM electrostatic potential. It should be noted that the QM electrostatic potential has already been polarized by the MM environment at a well-defined conformational state. This assumption can be further improved as discussed later. When the geometry of the QM subsystem is frozen and the MM subsystem is fluctuating, as in the case of the QM/MM-FE approach, it is not efficient to calculate the ESP charges for every MM conformation. On the other hand, QM ESP charges determined in a single MM environment are strongly influenced by the particular MM conformation, so the use of this set of QM ESP charges in the calculation of the electrostatic interactions with other MM conformations will thus be biased.

To reduce this bias in the original QM/MM-FE approach, we develop here the following strategy: we will use the mean field of the electrostatic potential from the MM subsystem to generate the QM ESP charges. That is, given multiple MM conformations, the effective way to determine the best set of ESP charges would be to use the mean electrostatic field of the MM conformations. The QM Hamiltonian is then

$$H = H_{QM}(r_{QM}) + \frac{1}{N}\sum_{n=1}^{N}\sum_{j\in MM}\sum_{i\in QM}\frac{q_j}{|r_i - r_{n,j}|} \quad (4)$$

where N is number of MM conformations recorded in the trajectory, $r_{n,j}$ represents the coordinates of MM atom j of the nth MM conformation. Within this averaged electrostatic field, the wave function of the QM system is solved, and ESP charges are obtained. For this scheme, the best way to determine the QM ESP charges will be an iterative self-consistent approach: starting from a set of ESP charges (from a single MM conformation perhaps), sample the MM conformations by molecular dynamics (MD) or Monte Carlo (MC) simulation and record the trajectories, calculate the ESP charges using the mean electrostatic field from trajectories as in eq 4, and repeat the MM sampling and recalculate the ESP charges until they are converged. In practice, we have found that the ESP charges from a short MD simulation of the MM subsystem are usually acceptable, and we therefore did not pursue this self-consistent scheme any further. This simplified QM/MM approach with ESP charges solved within the mean-field approximation now provides us a feasible force field for efficient simulation of the reaction system.

Potential of Mean Force of the QM Coordinates. For the calculation of thermodynamic properties, an accurate force field and a converged phase space sampling are both necessary. For a complex system of interest (e.g., a chemical reaction in solution or an enzyme), the enormous number of degrees of freedom makes it inefficient and often impossible to treat the whole system with high accuracy. An effective scheme would involve focusing on a small number of degrees of freedom thought to be the most important to the reaction process and modeling the remaining contributions by a simplified description. Therefore, we would expect to reach a balance between accessible precision and affordable complexity by combining fine- and coarse-grained methods together. This idea of combining theories of different levels has been proposed and applied in many ways in simulation studies. For instance, the hybrid QM/MM method is an implementation of this idea in the construction of the simulation force field and its effectiveness is well appreciated. Likewise, one would also like to seek a simplified thermodynamic description of the system, based only on a small number of important degrees of freedom. This is toward the potential of mean force description of a reaction system in which the contribution of a large number of less important degrees of freedom have been ensemble-averaged out in the free-energy expression of the whole system. It is obvious that such a PMF description of the reaction system bears two advantages compared with conventional approaches based on the exploration of the potential-energy surface: the state of the system is consistently defined in a much simpler way without the complications from the environment, and the thermodynamic properties are genuinely obtained without extra effort. There have been applications reported for the sampling of PMF surfaces, either for small molecular systems in gas phase,[48−50] or for systems in the condensed phase but with simplified QM approaches.[51−55]

Here, we focus on the calculation of such a PMF with ab initio QM/MM methods. We derive below a PMF description of a reaction system within the QM/MM framework, although the theory is not dependent on the QM/MM method. The partition function of a QM/MM system is

$$Z_0 = \int \exp(-\beta E(r_{QM}, r_{MM}))dr_{QM}^{M}dr_{MM}^{M'} \quad (5)$$

in which E is the total energy of the system as a function of both r_{QM} and r_{MM} and M and M' represent the number of degree of freedom of the QM and MM subsystems, respectively. The free energy of the system is then

$$A_0 = -\frac{1}{\beta}\ln(Z_0) \quad (6)$$

If we focus on the conformation of a selected subset of the system (e.g., the QM subsystem), we have a free-energy expression of r_{QM} which is also regarded as a potential of mean force of r_{QM}, that is

$$Z_0(r_{QM}) = \int \exp(-\beta E(r_{QM}, r_{MM}))dr_{MM}^{M'}$$
$$A_0(r_{QM}) = -\frac{1}{\beta}\ln(Z_0(r_{QM})) \quad (7)$$

The integration of the potential of mean force in the r_{QM} space recovers the complete free energy of the system

$$A_0 = -\frac{1}{\beta}\ln[\int \exp(-\beta A_0(r_{QM}))dr_{QM}^{M}] \quad (8)$$

With the MM contributions averaged out, the conformational space of the whole reaction system has been reduced to the potential of mean force surface of the QM subsystem. The problem of determining the reaction path and the activation free energy in a very complicated phase space of the whole system becomes a greatly simplified problem of exploring the PMF surface of the QM degrees of freedom.

Before proceeding, we would like to make two comments on this PMF expression. First, with the assumption of the ergodicity of the MD (or MC) simulation, this PMF surface takes into account the complete thermodynamic contribution of the MM environment, while the direct dynamic (thermal) contribution of the QM part is not considered. For the latter term, the only rigorous way to compute it is direct ab initio QM sampling, which is obviously too expensive to accomplish in the near future. However, with the equations derived below, one may either estimate the contribution of the dynamics of the QM part by computing the (thermodynamic) frequencies of the QM subsystem or by employing the reaction-path potential method to carry out synergized dynamics sampling between the QM and MM parts. Second, in this expression, the reaction coordinate is defined in Cartesian space. Of course, one can use more general geometrical coordinates to describe the reaction process and derive correspondingly the PMF expression for those coordinates. However, as we will discuss in later sections, caution needs to be taken to handle the mass-metric term which requires the evaluation of the inertial forces.

To sample the PMF surface of the QM conformation, the gradients for the PMF surface (i.e., the derivatives of the

PMF with respect to the coordinates of the QM subsystem) must be calculated. The free-energy gradient acting on QM atom i is computed as

$$\frac{\partial A_0(r_{QM})}{\partial r_{QM,i}} = \frac{\partial\left[-\frac{1}{\beta}\ln(Z_0(r_{QM}))\right]}{\partial r_{QM,i}} = \frac{\int \frac{\partial E(r_{QM}, r_{MM})}{\partial r_{QM,i}} \exp(-\beta E(r_{QM}, r_{MM}))\mathrm{d}r_{MM}^{M'}}{Z_0(r_{QM})} = \left\langle \frac{\partial E(r_{QM}, r_{MM})}{\partial r_{QM,i}} \right\rangle_{r_{MM}} \quad (9)$$

where the bracket represents ensemble averaging and the subscript r_{MM} represents an ensemble of MM conformations. Therefore, the gradient of the PMF is in fact an ensemble average of the gradients of the QM atoms, which must be evaluated by sampling the phase space of the MM subsystem with the QM conformation frozen. Similarly, when the second derivatives (i.e., Hessian) of the QM atoms are desired for computation of the entropic contributions, for example, the formula is

$$\frac{\partial^2 A_0(r_{QM})}{\partial r_{QM,i}\partial r_{QM}j} = \left\langle \frac{\partial^2 E}{\partial r_{QM,i}\partial r_{QM,j}} \right\rangle_{r_{MM}} - \beta\left[\left\langle \frac{\partial E}{\partial r_{QM,i}} \frac{\partial E}{\partial r_{QM,j}} \right\rangle_{r_{MM}} - \left\langle \frac{\partial E}{\partial r_{QM,i}} \right\rangle_{r_{MM}} \left\langle \frac{\partial E}{\partial r_{QM,j}} \right\rangle_{r_{MM}}\right] \quad (10)$$

To implement the calculation of the PMF force, we take the expression of the QM/MM energy as defined in eq 2 and also assume the QM density is frozen as defined in eq 3. The force is then

$$\frac{\partial A_0(r_{QM})}{\partial r^0_{QM,i}} = \frac{\partial E'_{QM}(r^0_{QM})}{\partial r^0_{QM,i}} + \left\langle \frac{\partial E'_{QM/MM,ele}(r^0_{QM}, r_{MM})}{\partial r^0_{QM,i}} + \frac{\partial E_{QM/MM,vdw}(r^0_{QM}, r_{MM})}{\partial r^0_{QM,i}} + \frac{\partial E_{QM/MM,cova}(r^0_{QM}, r_{MM})}{\partial r^0_{QM,i}} \right\rangle_{r_{MM}} \quad (11)$$

Accordingly, the second derivatives are written as

$$\frac{\partial^2 A_0(r_{QM})}{\partial r_{QM,i}\partial r_{QM,j}} = \frac{\partial^2 E_{QM}}{\partial r_{QM,i}\partial r_{QM,j}} + \left\langle \frac{\partial^2 E_{QM/MM}}{\partial r_{QM,i}\partial r_{QM,j}} \right\rangle_{r_{MM}} - \beta\left[\left\langle \frac{\partial E_{QM/MM}}{\partial r_{QM,i}} \frac{\partial E_{QM/MM}}{\partial r_{QM,j}} \right\rangle_{r_{MM}} - \left\langle \frac{\partial E_{QM/MM}}{\partial r_{QM,i}} \right\rangle_{r_{MM}} \left\langle \frac{\partial E_{QM/MM}}{\partial r_{QM,j}} \right\rangle_{r_{MM}}\right] \quad (12)$$

The availability of the first and second derivatives allows us to explore the phase space of the QM subsystem much more efficiently by conventional methods such as energy minimization and MD.

Reaction Path Minimization. Equation 7 essentially represents a multidimensional PMF: additional procedures are required to convert it into the quantity (activation free energy) directly measured in experiments, which consists of one and only one canonical order parameter to characterize the reaction progress. Therefore, to employ this equation in the QM/MM free-energy simulation of the reaction process, a proper set of (Cartesian or internal) geometric or energetic coordinates is usually identified such that their combination closely mimics the canonical order parameter. This set of coordinates is commonly known as the reaction coordinate. In practice, the determination of the reaction coordinate is nontrivial, especially for many complicated reactions catalyzed by enzymes. Instead of choosing a reaction coordinate more or less arbitrarily and always bearing the risk of it being incomplete or inappropriate,[62] we allow the reaction coordinate to be described by the coordinates of the entire QM subsystem, which eliminates the risk of it being improperly defined to the largest extent possible. Moreover, to efficiently employ the QM coordinates as the reaction coordinate, we select a series of discrete conformations parallel to the reaction process. The distance vectors between two adjacent conformations are used as a local reaction coordinate, while many local reaction coordinates are pieced together to constitute the global reaction coordinate. The chain of conformations is then optimized by means of well-established methods[43,56−59] with the free-energy profile of the reaction process determined simultaneously via FEP.

To carry out the minimization for a chain of conformations efficiently, the free-energy gradients, and maybe even the second derivatives, must be computed for the individual QM conformations. The relative free energies between adjacent QM conformations must also be computed. The former step is achieved by eqs 11 and 12, while the latter may be performed by using previously developed QM/MM-FE methods.[1] In this method, the free energy difference, ΔA, between two adjacent QM conformations is computed as

$$\Delta A_{i\to j} = -kT \ln\langle \exp(-\Delta E_{i\to j}/kT)\rangle_i \quad (13)$$

where the subscripts i and j are the indices of two adjacent conformations, the brackets represent an ensemble average, and the potential energy difference is

$$\Delta E_{i\to j} = E_j - E_i \quad (14)$$

We now outline our QM/MM-MFEP method as follows. (1) An initial set of discrete conformations for the QM subsystem connecting the reactant state to the product state using any interpolation scheme are generated, and the MM environment for each QM conformation is relaxed, if necessary. These intermediates states, plus the reactant and product states, form a chain of conformations. (2) For each conformation, the QM energy, ESP point charges, and forces for the QM subsystem are computed using a standard QM/MM scheme. (3) An MD simulation on the MM subsystems of each image is performed with the QM conformations frozen. The QM/MM interactions employed in the MD simulation are described by eq 3. During each MD step, the QM/MM forces on the QM atoms and the energy differences between a given QM conformation and its adjacent QM conformation(s) are computed and recorded for computing the free energy difference. (4) After a period of MD simulation, the free-energy difference between two adjacent conformations, as well as the free-energy forces, is calculated.

Then one step of coupled optimization, with a method such as NEB, Ayala−Schlegel MEP, or QSM, is applied to optimize all conformations on the chain. (5) The optimization process is exited if the path is converged by the predefined criteria. Otherwise, the procedure is repeated from step 2.

As discussed in previous sections, to obtain the highest quality ESP charges, the best approach would be to repeat steps 2 and 3 in an iterative, self-consistent manner. However, in practice, we have found that the ESP charges obtained after one MD simulation are accurate enough for future calculations. To reduce the computational time required, we have also made the approximation of using the MM conformations sampled during the MD simulation of the previous minimization step as the electrostatic background for the QM calculation, instead of self-consistently calculating them during the current minimization step. In practical simulations where the step size of the optimization can be controlled, we have found that this approximation works effectively (data not shown).

Since one minimization cycle of the QM/MM-MFEP method consists of only one QM energy and force calculation, a certain length of MD simulation, and a single optimization step, the computational need for the QM calculations has been significantly reduced. The time spent on the QM calculations becomes minor compared to that of the MD simulations. For a typical QM/MM simulation of an enzyme reaction, the number of QM atoms is usually on the order of 10^1-10^2, while the number of MM atoms varies drastically between 10^4 and 10^6. In such a case, one ordinary hybrid DFT, such as B3LYP,[63,64] calculation with a double-ζ basis set for the QM subsystem takes only minutes to a few hours, while the MD simulation of 10^2 ps will take several tens of CPU hours. Because the MD simulation competes as the computational bottleneck, and the time required for the QM calculation is relatively small, very high-level ab initio QM methods may be used to improve the accuracy. On the other hand, the MD simulations may be carried out with well-established parallel algorithms to speed up the MD sampling.

To facilitate the convergence of the reactant-path optimization, it appears to be practically more efficient if the reactant and product states are optimized prior to the iterative MD sampling/optimization procedure and their QM conformations frozen during the subsequent path optimization process.

Reaction Path Potential with a Mean MM Field. In the QM/MM-MFEP method, the QM subsystem is frozen to reduce the computational cost of the QM calculations. This treatment essentially removes the dynamics of the QM part and decouples any possible dynamic correlation between the active site and the conformational change of the enzyme environment. However, it should be noted that the coupling between the chemical reaction process and the large amplitude conformational motion of the enzyme can be seamlessly interfaced with the QM/MM-MFEP method (Hu et al., unpublished work). Therefore, the only approximate term is the harmonic approximation of the QM subsystem. Since the details of the dynamics of the QM active site associate directly to the question of whether or not the enzyme achieves its catalytic function through coupling between the dynamics of the active site and other structural components, it is very important to simulate the dynamics of the QM subsystem.

To achieve the goal of QM dynamic sampling with a computationally affordable approach, we extend the previously developed RPP method[2] in the framework of a mean-field approach. The RPP originates from the idea of the reaction path Hamiltonian,[65] which expands the exact quantum mechanical Hamiltonian of a reaction system to different orders of perturbations, thus allowing the inexpensive yet accurate calculation of energetics for the region of phase space around the reaction path. In our method, the QM subsystem is spatially embedded in a buffer of atoms sampled during MD simulation. That is, the QM subsystem experiences the mean field of the MM environment. For each QM atom, α, the external electrostatic field is

$$v_{\mathrm{MM}}(r_\alpha) = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M_n}\frac{q_m}{r_{\alpha m}} \tag{15}$$

where r_a represents the coordinates of QM atom α, N is the number of snapshots recorded during MD simulation, M_N is the number of MM atoms recorded in the Nth snapshot, q_m is the atomic charge of MM atom m, and $r_{\alpha m}$ is the distance between MM atom m and QM atom α. Because the changes of the ESP-fitted charges of the QM atoms can be represented by the response to the perturbations by truncating the response to the first order of perturbations, we can calculate the change of the ESP charges with respect to the changes in the geometries and external electrostatic potentials. To do so, we compute two response kernels for the QM atoms, namely, the changes of the QM ESP charges in response to the change of the external electrostatic potential

$$\chi_{\alpha\beta} = \left(\frac{\partial Q_\alpha}{\partial v_\beta}\right)_N \tag{16}$$

and the changes of the QM ESP charges in response to changes in the QM geometries

$$\kappa_{\alpha\beta} = \left(\frac{\partial Q_\alpha}{\partial \mathrm{r}_\beta}\right)_N \tag{17}$$

where Q_α is the ESP-fitted change of QM atom α, v_β is the mean electrostatic potential of QM atom β defined in eq 15, and the subscript N defines the constraint that the number of electrons must remain a constant N. The first response kernel, $\chi_\alpha\beta$, was introduced by Kato and co-workers,[66] and the second kernel, $\kappa_{\alpha\beta}$, was introduced by our group.[2] With these response properties, the polarized charge of QM atom α is computed as

$$Q_\alpha(r, v_{\mathrm{MM}}) = Q_\alpha^{\mathrm{ref}} + \sum_{\beta\in\mathrm{QM}}\chi_{\alpha\beta}[v_{\mathrm{MM}}(r_\beta^{\mathrm{ref}}) - v_{\mathrm{MM}}^{\mathrm{ref}}(r_\beta^{\mathrm{ref}})] + \sum_{\beta\in\mathrm{QM}}\kappa_{\alpha\beta}[r_\beta - r_\beta^{\mathrm{ref}}] \tag{18}$$

The superscript ref designates the reference state in the absence of the perturbation. The electrostatic interaction

between the QM and MM subsystems is the simple Coulombic interaction between point charges

$$E^{\mathrm{ESP}}_{\mathrm{QM/MM,ele}} = \sum_{m\in\mathrm{MM}}\sum_{\alpha\in\mathrm{QM}} \frac{q_m Q_\alpha(r, \upsilon_{\mathrm{MM}})}{|r_\alpha - r_m|} = \sum_{\alpha\in\mathrm{QM}} Q_\alpha(r, \upsilon_{\mathrm{MM}})\upsilon_{\mathrm{MM}}(r_\alpha) \quad (19)$$

The internal energy of the QM subsystem is defined as

$$E_1(r_{\mathrm{QM}}, \upsilon_{\mathrm{MM}}) = \langle\Psi|H_{\mathrm{eff}}|\Psi\rangle - E^{\mathrm{ESP}}_{\mathrm{QM/MM,ele}} \quad (20)$$

which can be computed as[2]

$$E_1(r_{\mathrm{QM}}, \upsilon_{\mathrm{MM}}) = E_1(r^{\mathrm{min}}_{\mathrm{QM}}, \upsilon^{\mathrm{ref}}_{\mathrm{MM}}) + \sum_{\alpha\in\mathrm{QM}}\left(\frac{\partial\langle\Psi|H_{\mathrm{eff}}|\Psi\rangle}{\partial r_\alpha}\Big|_{r^{\mathrm{min}}_{\mathrm{QM}}, \upsilon^{\mathrm{ref}}_{\mathrm{MM}}} - \sum_{\beta\in\mathrm{QM}}\kappa_{\beta\alpha}\upsilon_{\mathrm{MM}}{}^{\mathrm{ref}}(r^{\mathrm{min}}_\beta) - Q^{\mathrm{min}}_\alpha\frac{\partial\upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\alpha)}{\partial r^{\mathrm{min}}_\alpha}\right)\Delta r_\alpha + \frac{1}{2}\sum_{\alpha,\beta\in\mathrm{QM}}\Delta r_\alpha\left(\frac{\partial^2\langle\Psi|H_{\mathrm{eff}}|\Psi\rangle}{\partial r_\alpha\partial r_\beta}\Big|_{r^{\mathrm{min}}_{\mathrm{QM}}, \upsilon^{\mathrm{ref}}_{\mathrm{MM}}} - \kappa_{\alpha\beta}\frac{\partial\upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\alpha)}{\partial r^{\mathrm{min}}_\alpha} - \kappa_{\beta\alpha}\frac{\partial\upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\beta)}{\partial r^{\mathrm{min}}_\beta} - Q^{\mathrm{min}}_\alpha\frac{\partial^2\upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\alpha)}{\partial r^{\mathrm{min}}_\alpha\partial r^{\mathrm{min}}_\beta}\delta_{\alpha\beta}\right)\Delta r_\beta - \sum_{\alpha\in\mathrm{QM}}\sum_{\beta\in\mathrm{QM}}\chi_{\alpha\beta}\upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\alpha)[\upsilon_{\mathrm{MM}}(r^{\mathrm{min}}_\beta) - \upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\beta)] - \frac{1}{2}\sum_{\alpha\in\mathrm{QM}}\sum_{\beta\in\mathrm{QM}}[\upsilon_{\mathrm{MM}}(r^{\mathrm{min}}_\alpha) - \upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\alpha)]\chi_{\alpha\beta}[\upsilon_{\mathrm{MM}}(r^{\mathrm{min}}_\beta) - \upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\beta)] - \sum_{\alpha\in\mathrm{QM}}\sum_{\beta\in\mathrm{QM}}(r_\alpha - r^{\mathrm{min}}_\alpha)\frac{\partial\upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\alpha)}{\partial r^{\mathrm{min}}_\alpha}\times\chi_{\alpha\beta}[\upsilon_{\mathrm{MM}}(r^{\mathrm{min}}_\beta) - \upsilon^{\mathrm{ref}}_{\mathrm{MM}}(r^{\mathrm{min}}_\beta)] \quad (21)$$

That is, the energy is expanded to second order of perturbations around the initial conformation (denoted as $r^{\mathrm{min}}_{\mathrm{QM}}$) and the initial electrostatic potential it bears (denoted as $\upsilon^{\mathrm{ref}}_{\mathrm{MM}}$). With the definitions of eqs 18 and 19, the total energy of the system is then defined in a similar manner to eq 2 as

$$E_{\mathrm{RPP}} = E_1(r_{\mathrm{QM}}, \upsilon_{\mathrm{MM}}) + E^{\mathrm{ESP}}_{\mathrm{QM/MM,ele}}(r_{\mathrm{QM}}, r_{\mathrm{MM}}) + E_{\mathrm{QM/MM,vdw}}(r_{\mathrm{QM}}, r_{\mathrm{MM}}) + E_{\mathrm{QM/MM,cov}}(r_{\mathrm{QM}}, r_{\mathrm{MM}}) + E_{\mathrm{MM}}(r_{\mathrm{MM}}) \quad (22)$$

The derivation of the RPP allows dynamic sampling of the QM subsystem without the need for expensive QM calculations at every step and also allows the direct simulation of the free-energy difference between different conformational states through well-established classical or quantum free-energy simulation techniques.[44,45]

Simulation Details. To examine the validity and effectiveness of the minimum free-energy path method proposed here, we have studied the first reaction step catalyzed by the enzyme TIM. As shown in Figure 1, the reaction involves a proton transfer from the substrate to the side-chain carboxylate group of Glu164. This reaction step has been studied many times previously.[1,2,7] Since our intention here is to examine the applicability of the QM/MM-MFEP method, we have carried out the simulation under a setup similar to studies previously reported.

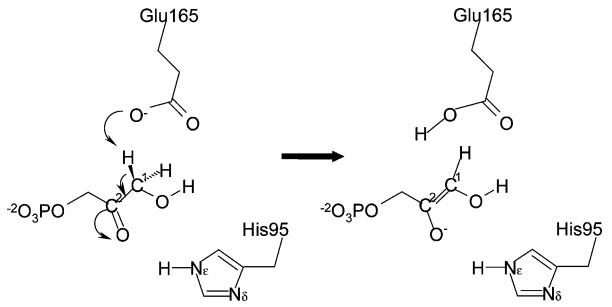


Figure 1. Active site and the reaction catalyzed by the enzyme triosephosphate isomerase (TIM).

For this system, all protein atoms were included in the simulation, and a solvation sphere of 21 Å was created around the C1 atom of the substrate molecule. Any residues or water molecules in which all atoms are greater than 16 Å away from all atoms of the substrate were restrained by a harmonic force of 10 kcal mol^{-1} Å^{-2} with respect to their initial minimized positions. A flat-bottomed restraint was added to all free water molecules to prevent them from crossing the restrained shell of atoms and escaping into the vacuum space. This restraint potential takes no effect until the distance between a specific water molecule and the C1 atom of the substrate is greater than 16.5 Å. The final system was composed of 6618 atoms, of which 3795 belong to the protein and the substrate molecule and 2823 belong to the 941 water molecules.

The parameters of the AMBER force field[67] incorporated into the TINKER program[68] were used to model the classical MM interactions. In all simulations, a dual cutoff of 9 and 15 Å was used to separate the short- and medium-range interactions. The nonbonded pair lists were updated every 16 fs. The multi-timestep method was used for integration,[69,70] with time steps of 1 and 4 fs for the short- and medium-range forces, respectively. The temperature of the system was kept at 300 K by a Berendsen thermostat.[71] The medium-range QM/MM electrostatic interactions were modeled as pure classical interactions between the ESP point charges on the QM atoms and the point charges on the MM atoms; only the short-range neighboring MM atoms were included in the quantum mechanical calculation for the energy, gradient, and ESP charges. In other words, only the polarization effects from the short-range MM atoms were considered for the QM atoms.

The MD simulations and minimizations were carried out with the program Sigma,[72−74] which was interfaced with Gaussian 03[34] to perform the QM calculations. For the QM calculations, the HF/3-21G method was used, while the QM/MM hybrid bonds were modeled with the pseudobond method.[42] These choices were made for ease of comparison with previous studies, with no attention paid to the effect of the level of theory and basis set on the actual reaction mechanism.

For this model reaction system, the initial structures of the reactant and product states were obtained from previous

studies.[1,2] The reaction process was modeled by linear interpolation between the reactant and product geometries to yield 13 intermediate conformational states. These 15 structures were used for all optimization and simulation studies. As described in the methods section, the QM part was frozen during the dynamics sampling of the MM part (i.e., their positions did not change and their velocities are set to zero). The SHAKE algorithm[75] was completely turned off for the protein molecule, which means the interactions of all protein bonds were explicitly computed, including the crossing bonds between the QM and MM atoms.

From the interpolated structures, the reaction path was determined by a three-stage optimization: stand-alone free-energy minimization of the reactant and product states, NEB optimization,[56] and Ayala−Schlegel MEP optimization[57] on the QM/MM free-energy surface.[11,43] After the first minimization step, the QM structures of the reactant and product states were kept frozen in the subsequent NEB and Ayala−Schlegel MEP optimizations.

All optimizations were carried out with the algorithm outlined in the previous section. In general, each optimization cycle consists of one QM calculation which yields QM energies, gradients, and ESP charges, followed by MD simulations for calculating the free-energy gradients of the QM subsystem and the free-energy differences between adjacent conformations. After that, an optimization step is made on the basis of the free energies and the free-energy gradients of the chain of conformations. The length of MD simulation in each optimization step has a vital impact on the quality and efficiency of the global optimization process. Longer simulations yield better convergence of the energies and gradients. Because there is no easy and automated tool to determine the required length of the MD simulations, the convergence properties of the free energies and free-energy gradients of each QM atom must be determined from trial simulations.

Nevertheless, to speed up the calculation, MD simulations of different lengths can be carried out in the NEB optimization process. In the current system, each NEB optimization step in the beginning stage consisted of one QM calculation followed by 40 ps of MD simulation for the computation of the free energies and free-energy gradients. After 20 steps of optimization, the MD simulation time was extended to 80 ps for better convergence. An additional 40 steps of NEB optimization were then carried out and yielded a reaction path which was used as the input for the Ayala−Schlegel MEP optimization in the next stage. We observed slow convergence in the later stage of the NEB optimization, consistent with the observations of many others.

The Ayala-Schlegel MEP optimization was performed, starting from the last NEB path, to determine the exact reaction path and, most importantly, the structure of the transition state. In all MEP optimizations, the MD simulation time was 128 ps to ensure good convergence of the free energies and free-energy gradients. The MEP convergence criteria were loosened to 1×10^{-3} hartree for energy and 1×10^{-3} hartree bohr^{-1} for the gradients. Different sets of optimizations were carried out with different maximum step sizes, but all optimizations converged to nearly identical transition-state structures and activation free energies.

After we obtained the exact MEP path, reaction-path potential calculations with the mean-field approximation were carried out for the reactant state and the transition state, respectively, and each yielded an RPP function that allowed us to simulate the dynamics of the QM subsystem as the same as to carry out ordinary MM simulations. The free-energy difference between the transition state and the reactant state was then simulated by the slow-growth free-energy simulation method.[76−78]

Results

All the optimization algorithms implemented here depend on the evaluation of the relative QM/MM free energies between different conformational states and the free-energy gradients defined in eq 9. The convergences of both quantities are crucial for the effectiveness of the optimization process.[79] To demonstrate how the free-energy gradients converge in a simulation, we plotted the evolution of the free-energy gradient components of all QM atoms in the course of a simulation of 128 ps (Figure 2), with the first 12 ps disregarded as an equilibration process. It was clear that most components achieved good convergence within 30 ps. After 80 ps, all components reached a stable plateau. The free-energy perturbations between different conformational states converged, in general faster, than the gradients. One can in fact examine the correctness of the free-energy gradients by carrying out free-energy perturbation with numerical differentiation of the QM coordinates.

To illustrate the effectiveness of our approach, Figure 3 plots the final free-energy profile for the NEB/FEP optimization. The final optimized activation free energies were 22.0 and 8.4 kcal mol^{-1} for the forward and backward reactions, respectively. Not surprisingly, these numbers are larger than those of previously reported similar studies.[1,2] Since NEB converges very slowly when the path gets close to the exact path, it is very difficult to locate the exact reaction path and transition state. For this reason, we did not pursue other complicated NEB schemes to improve the accuracy of the results.

Figure 4 illustrates the free-energy profile determined in the MEP/FEP optimizations. The activation free energies were 21.6 and 4.8 kcal mol^{-1} for the forward and backward reactions, respectively. These values are close to what has been reported previously at the same QM level of theory. The structure of the transition state is also close to the one previously reported.[1] At the transition state, the difference between the C1−H1 and H1−O2 bond lengths was 0.37 Å. In previous studies, this distance difference was used as the reaction coordinate and was determined to be 0.4 Å at the transition state.

The free-energy difference between the reactant state and the transition state was determined by the slow growth method with the reaction-path potentials. The latter was computed from the reaction path optimized in the MEP/FEP simulations. The activation free energy was determined to be 19.8 kcal/mol in this way. This value is slightly lower than those obtained previously. Such a difference may be

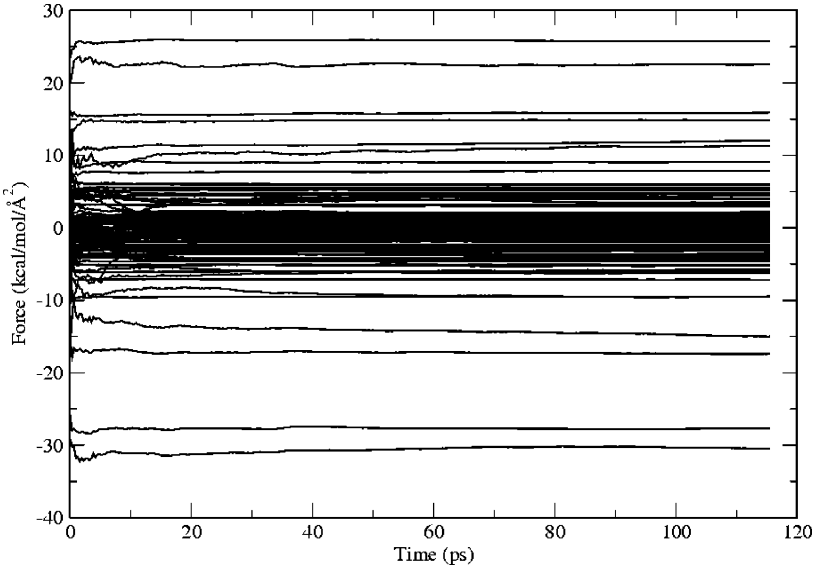


Figure 2. Convergences of the free-energy gradients of all QM atoms in the TIM system. The *x* axis is the simulation time, while the *y* axis is the free-energy gradient defined in eq 9.

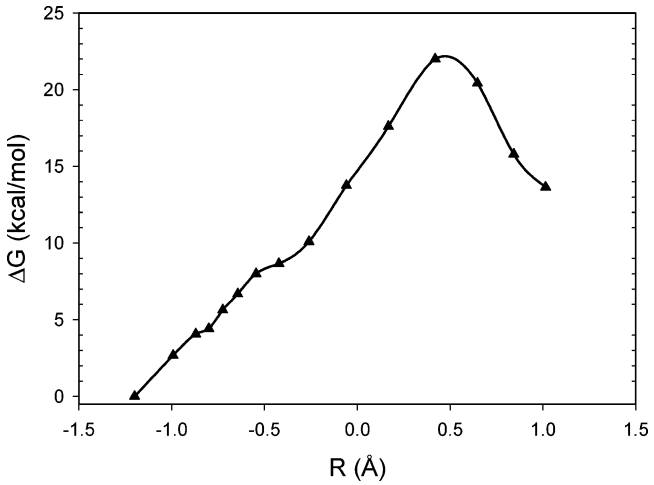


Figure 3. Free-energy profile from NEB optimization by the QM/MM-MFEP method for the proton-transfer reaction in TIM. To make the results comparable to previous studies, the *x* axis, *R*, is defined as the distance difference $R_{C-H} - R_{O-H}$, instead of the indices of the conformations used in the simulations.

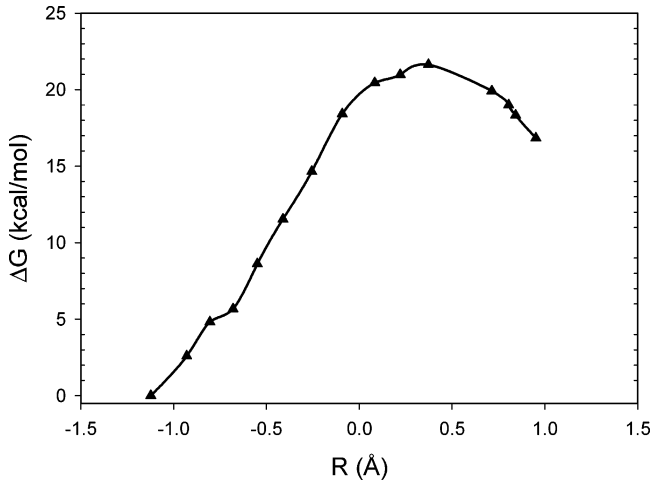


Figure 4. Free-energy profile from Ayala−Schlegel MEP optimizations by the QM/MM-MFEP method for the proton-transfer reaction in TIM. The *x* axis is defined in the same way as in Figure 2.

attributed to the subtle difference between the reaction paths determined by different methods or to the longer MD relaxation of the MM environment in the current study.

Discussion

Comparison with Related Studies on TIM. In the current study, there is an apparent resemblance between the activation free energies obtained from NEB/FEP and MEP/FEP optimizations. It should be noted, however, that such a resemblance has no implications on the effectiveness of each method. Instead, it is well-known that for complicated systems, the NEB method converges slowly and often fails to locate the exact position of the transition state. In addition to the slow convergence, the errors from two other known problems of the NEB method, straddling and corner-cutting the true transition state, tend to cancel out. As a result, the similar free-energy barriers in both calculations come from the fact that the NEB optimization does not locate the true transition state; otherwise, a higher barrier would be expected for the NEB method. It would be interesting to use the recently developed QSM method[59] which displays superlinear convergence and has been shown to be significantly more efficient than NEB.

For the model reaction step of TIM we have investigated here, the results of the present study are in good agreement with previous studies employing the same QM level of theory.[1,2] The structures of the reactant state, transition state, and product state are very similar to those previously obtained. The only notable difference is that the activation free energy of the forward reaction process obtained here is ~2 kcal mol^{-1} lower than previous results. Several factors may contribute to this difference, including: extensive MD sampling of the enzyme MM subsystem in this study, minimization on the free-energy surface rather than on the potential-energy surface. The fact that the structural differ-

ences between the current and previous studies are small tells us that the results from the current approach are consistent with previous work.

Compared with many other QM/MM simulation methods, the current QM/MM-MFEP method uses all of the QM degrees of freedom as a reaction coordinate, and thus does not require the explicit choice of the reaction coordinate prior to the simulation process. This considerably reduces the risk of incorrectly choosing the reaction coordinate because it has been shown that improper choice of the reaction coordinate will bias the simulation and slow down the convergence.[62] The problem of the inappropriate choice of the reaction coordinate is more severe in simulations using coordinate-driving types of techniques in which the choice of the reaction coordinate not only strongly influences the ability to sample phase space correctly but also causes technical difficulties when the changes of specific geometrical properties are stepwise or nonlinearly correlated. With the coordinates of the whole QM subsystem naturally utilized as the reaction coordinate, one no longer bears those problems.

Of course this method of constructing a reaction coordinate is an advantage but not a unique feature for the QM/MM-MFEP method. It is obvious that there is no need to explicitly specify the geometrical reaction coordinate in any QM/MM simulation of a reaction, as long as the global reaction coordinate is pieced together from a chain of conformations such as those implemented in the NEB, Ayala−Schlegel MEP, and QSM algorithms. On the other hand, the QM/MM-MFEP approach does not exclude the use of a well-defined reaction coordinate. If a reasonable reaction coordinate is available, one could by all means use this reaction coordinate in the QM/MM-MFEP simulation. In this situation, the known reaction coordinate will speed up the convergence of the free-energy simulation, simplify the definition of the free-energy gradients, and subsequently allow the use of other free-energy simulation techniques.

The most important improvement of the QM/MM-MFEP approach compared with the original QM/MM-FE approach is that the results of the simulation no longer depend on the choice of the initial conformations, thus eliminating the bias from the initial structure. In many simulation methods previously proposed, including the QM/MM-FE method, the reaction path is usually determined with a static enzyme structure (i.e., on a single zero-temperature potential-energy surface. Even though the enzyme structure is energetically minimized during the process of determining the reaction path, the configuration of the protein molecule is nearly immobile structurally (i.e., large-scale conformational changes such as domain motions or even the transition of side-chain rotamer states are almost completely prohibited). This dilemma causes two problems in the simulation of enzymatic reactions. On one hand, when specific MM groups undergo a significant conformational change during the reaction process, it is technically difficult to obtain a converged reaction path. On the other hand, since in the native state an enzyme molecule can access an enormous number of conformational minima, the choice of a particular minimum state will bias the results. According to the rigorous statistical mechanics principle, it is incorrect to use a reaction path obtained within one particular enzyme conformation to represent the thermodynamic behavior of the enzyme. Moreover, a reaction path determined on the zero-temperature potential-energy surface is not same as the thermodynamic path measured in experiments. Those problems were overcome by the QM/MM-MFEP method in a theoretically sounded way.

It may appear that the agreement between the results of the current and previous study is at odds with the idea that the results of QM/MM-FE method were strongly dependent on the choice of the initial structure model. It has been shown that this dependence can lead to some uncertainties in the simulation studies.[47] Since the dependence was removed in current approach, one is expected to see a clear difference between the current and previous results. To resolve this discrepancy, we note that the setup of the current model system does have a significant impact on the results. The numerous positional restraints applied to those atoms outside the active sphere of atoms to a large extent limits the sampling of phase space of the protein molecule and, subsequently, the phase space of the QM atoms of the active site. In spite of these structural restraints, we still see a significant difference between the computed activation free energies of the potential energy and minimum free-energy paths.

Connection to Previous Studies of Sampling the Free Energy Surface. The idea of representing the structural and thermodynamic properties of a molecular system in terms of the PMF surface of a few variables is a quite general idea in statistical mechanics.[76,80] Past simulation study often focused on building the PMF surface in simulations through various techniques such as umbrella sampling,[29] free-energy perturbation,[81] and thermodynamic integration.[76] Only recently, it was proposed to explore the phase space of the system by directly walking the PMF surface.[52−55] Several groups have extended the theory and the simulation techniques and have also reported several example applications to different reaction systems.[48−50,79] The work reported here should still be regarded as an important further improvement and enrichment of the idea because of several key differences between this work and those reported previously.

The QM/MM-MFEP method was built within the ab initio QM/MM framework and thus is appropriate for simulating reaction processes in enzymes and in solution. In our method, the contribution of the MM environment has been explicitly considered. In contrast, for other previously reported applications,[48−50,79] either the reaction process took place in the gas phase or semiempirical QM methods were used. Therefore, the QM/MM-MFEP is better suited for simulating complicated reaction systems with higher accuracy.

The difference in the means by which the reaction coordinate is chosen also results in different implementations of the free-energy simulation approaches in the QM/MM-MFEP and other methods. In many other methods, the calculation of the free energy relies on the determination of a reaction coordinate which is usually defined as a set of geometrical variables. The free energy is computed by umbrella sampling or by computing the effective forces

acting on the reaction coordinate, and then the work and the free energies along the reaction path are computed. Even though those approaches are still applicable in the QM/MM-MFEP method, our QM/MM model also allows the free energy to be computed through direct free-energy perturbation on a chain of conformations. This provides additional flexibility when the reaction coordinate is not so easily defined by geometrical terms.

One more important improvement in the QM/MM-MFEP method is the application of the mean-field approximation. This approximation was used in modeling the electrostatic interactions from the MM environment to the QM system in both the optimization of the reaction path and the generation of the reaction path potential. The use of a mean MM field has improved the numerical stability and quality of the ESP charges, thus speeding up the convergence of the QM/MM-MFEP calculations. The reaction path potential has been also derived within the mean field of the MM atoms. This allows the dynamics of the QM system to be simulated without using the straightforward yet expensive full QM-MD methods.

The PMF surface of the QM/MM-MFEP method is built by means of molecular dynamic simulation with frozen QM atoms. The freezing of the QM subsystem makes the current method resemble some other methods such as the "blue moon sampling" which was developed with the constrained dynamics sampling approach.[82−84] It is clear that all these methods share the same origin from thermodynamic integration.[76,85] Interestingly however, the use of Cartesian coordinates of the QM subsystem as the reaction coordinate of the QM/MM-MFEP method again leads to the difference in the method for carrying out simulation and analyzing the results. When the reaction coordinate is chosen as a combination of general geometric variables such as bonds, bond angles, and dihedral angles, the PMF expression of those variables is still valid and takes a similar form to eq 7.[82,84] However, the calculation of the free-energy gradients with respect to those variables is not as simple as we derived here in eq 9. In fact, it was shown that an additional term has to be included to count the metric tensor effect, caused by the overconstraint on the momentum space.[82,84,86−88] Previous work also indicated the importance of this term.[2,79] On the contrary, when Cartesian coordinates are used as the reaction coordinate, the metric matrix becomes a unit matrix, thus makes no contribution to the evaluation of the free-energy gradients.[84]

Like many other methods for locating the reaction path, the QM/MM-MFEP method developed here also has limitations. Apparently, the solution of the QM/MM-MFEP method depends on the effectiveness of the optimization algorithms. As the NEB, Ayala−Schlegel MEP, and QSM methods implemented here are all developed for locating local minima, the QM/MM-MFEP will also be a localized minimal path. This feature often does not become a serious problem because in many systems good chemical intuitions often exist for the reaction mechanism. Abundant structural and biochemical information is usually available for enzymatic reactions from extensive biochemical experiments which will guide the simulation study. In such a case, the computed path will be (at least) very close to global minimal path, and the essence of the chemistry will be captured to large extent with caution. Yet another factor that will impact the accuracy of the results is the issue of time scale. Because we derived the free-energy force as an ensemble-average of the forces bore in MD simulation (eq 9), we take the well-known ergodicity assumption of MD sampling and also assume the results converge within routine simulation period. As shown in Figure 2, this assumption may hold well in the case of TIM. There are, however, many examples in the enzyme-catalyzed reactions that long-time conformational dynamics play significant roles in the reaction process.[89,90] For those molecules, apparently even nanosecond MD simulations will not be enough to characterize the slow conformational dynamics. To combat this problem, new methods which combine the advantages of the QM/MM-MFEP method and other enhanced sampling approaches must be sought.

Further Improvements and Extensions to the QM/MM-MFEP Method. Even though we have shown here the effectiveness of the QM/MM-MFEP method, the precision of the method can still be improved by better modeling of the QM/MM interactions. Specifically, the abilities and accuracy of this method could be enhanced in at least two ways.

(*a) Improvement of the Electrostatic Representation of the QM Atoms in the MD Simulation.* Consistent with the QM/MM-FE approach, a simple ESP-fitted charge is used for each QM atom in the MD simulations. This is equivalent to the case where the whole QM electron density is approximated by electrostatic monopoles at the atomic positions. It is obvious that the addition of higher-order electrostatic multipoles such as point dipoles to each QM atom will greatly improve the quality of the ESP fitting, thus improving the precision of the results. For enzymes, the improvement from atomic dipole moments may not be significant because the large structural environment of the MM subsystem will overwhelmingly modulate the motion of the small QM subsystem. For reactions in solution, however, dipoles become more important since the structure of the solute−solvent cluster is strongly influenced by the how correctly the local electrostatic interactions between the QM solute and MM solvents are described.

(b) Improvement of the Accuracy of the QM/MM Calculation by Surmounting the Frozen-Density Assumption. For a given QM conformation, the electron density is assumed to be frozen during the subsequent MD simulation process in the current method. We have reported that the effect on the energetics is small for frozen electron densities. Still, the description of the QM electron density may be improved by allowing it to fluctuate in response to changes in the MM environment. Without any loss in efficiency, the linear response-polarizable charge model developed in the RPP method may be employed. Since the QM atoms are frozen, the charges in eq 18 are simplified to

$$Q_\alpha(r, \upsilon_{MM}) = Q_\alpha^{min} + \sum_{\beta \in QM} \chi_{\alpha\beta}[\upsilon_{MM}(r_\beta^{min}) - \upsilon_{MM}^{ref}(r_\beta^{min})] \quad (23)$$

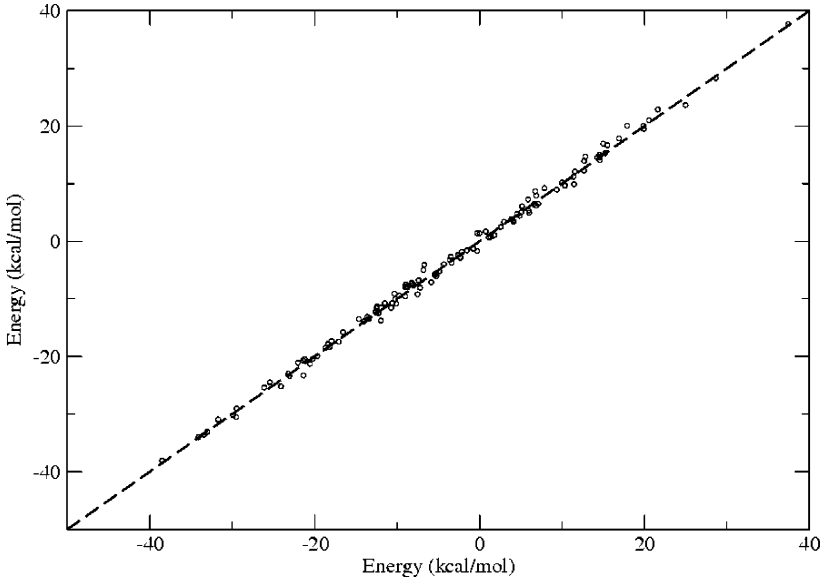


Figure 5. Comparison between the energies calculated for different MM environments with the same QM conformation. The *x* axis is the energy calculated by direct QM/MM method in which the exact QM electron density is solved for each QM/MM conformation; the *y* axis is the energy calculated by the RPP method with polarizable QM charges. The RPP is constructed using the mean-field of all MM charges as the external electrostatic potential for the QM part. For better display, the energies on both axes have been shifted by the same amount.

For higher quality reference charges and the response kernel, the mean-field approach should be used.

To show the validity of this polarizable charge model of the QM atoms, we compared the QM/MM energies from two sources of a frozen QM structure with a set of different MM environments in Figure 5. One is computed from direct QM/MM calculation for each MM environment, and the other from the classical electrostatic interaction between the MM atoms and the polarized QM charges defined in eq 23. The response kernel here was computed by employing the mean field of the collection of all MM atoms. The results clearly indicated that the polarizable QM charge model can accurately reproduce the fluctuating QM density and thus the QM/MM interactions. The excellent agreement between the energies calculated from both methods also supports the use of the MM mean field in the QM/MM-MFEP method.

Significance of the QM/MM-MFEP Method. The conventional method for the determination of the chemical reaction path relies on the techniques of exploring the energy surface of a small, isolated molecular system. For those systems, the conformational states are often very well defined and the number of conformational states is small enough that it is possible to carry out exhaustive calculations for every state. When the reaction process occurs in the condensed phase (e.g., solution reactions and enzymatic reactions), the many degrees of freedom make it impossible to evaluate all conformational states one by one. Furthermore, the dynamics are hierarchically organized in proteins, making it difficult to define the microconformational states with the canonical reaction process for complicated enzymatic reactions.[91,92] For this purpose, a PMF description of the reaction system with a small number of degrees of freedom will reduce much of the complexity of the phase space and simplify the simulation study of reaction processes.

The QM/MM-MFEP method allows the determination of reaction paths on the free-energy surface, rather than the potential energy surface. It has been shown that dynamics can play an important role in the selection of the reaction path. The free-energy path is one and only one path that corresponds exactly to the macroscopic thermodynamic reaction process. Therefore, it is crucial to develop simulation methods with the capability of sampling the free -energy surface, as in the QM/MM-MFEP approach.

Conclusions

The accuracy of the simulations of condensed phase reaction processes depends on two factors: the ability of the method to faithfully describe the reactivity of the system under study and the convergence of the statistical sampling of phase space. The former is solved by introducing quantum mechanics into the classical force field, while the latter can only be achieved by long time scale MD simulation (or MC). For reaction processes in the condensed phase, the complexity of the problem makes it difficult if not impossible to achieve the quantum mechanical treatment for the entire system. For this reason, the hybrid QM/MM method has been developed to reduce the computational cost. Even so, the excessive expense of high accuracy ab initio QM calculation still remains the bottleneck of the simulation and prevents the widespread application of the technique.

To reduce the cost of the QM calculations and bridge the gap between expensive ab initio QM calculations and long time scale MD sampling, we have developed the QM/MM-MFEP method to simulate reaction processes in the condensed phase, the essence of which to sample the dynamic free-energy surface rather than the static potential-energy surface. Distinct features of this method include the expression of the free-energy profile of the system as a PMF surface along a chain of conformations described by hybrid ab initio

quantum mechanics and classical molecular mechanics, with the contribution of the MM environment properly simulated by classical MD simulation and the relative free energies between distinct conformational states by computed QM/MM FEP. The method can be applied to numerous research projects, from the optimization of a reaction path by the NEB or second-order Ayala−Schlegel MEP method to the simulation of the full energetics including the correlation between the dynamics of the QM active site and the MM environment by the reaction-path potential method. Therefore, it provides a complete stepwise recipe for simulating reactions under the framework of the QM/MM approach. The example application of the QM/MM-MFEP method on the TIM system illustrates its validity and effectiveness.

Acknowledgment. Financial support from the National Institute of Health is gratefully appreciated. We also thank Steven K. Burger and Jerry M. Parks for very helpful discussions. We thank the reviewers for prompt information on a related paper published in the *Journal of Chemical Physics* (Vol. 125, 024106) at the same time of the submission of this manuscript.

Appendix. Potential of Mean Force as a Mean Field Approximation to the Free Energy

We give here a different derivation of the PMF expression through a mean field approximation to the free energy, providing an interesting interpretation of the PMF and also the minimum on the PMF surface.

The partition function Z_0 of a system (with momentum part factoring out) is written as

$$Z_0 = \int \exp(-\beta E)\mathrm{d}r^N \tag{24}$$

in which E is the energy of the system and r is the coordinate. It is noted that E is a function of r. The equilibrium configuration density/distribution is then defined as

$$\rho_0 = \frac{\exp(-\beta E)}{Z_0} \tag{25}$$

which obviously is normalized as

$$\int \rho_0 \mathrm{d}r^N = 1 \tag{26}$$

The free energy of the system is

$$A_0 = \frac{-1}{\beta}\ln(Z_0) = \int \rho_0\left[E + \frac{1}{\beta}\ln(\rho_0)\right]\mathrm{d}r^N \tag{27}$$

We can define a free-energy functional of an arbitrary distribution ρ

$$A[\rho] = \int \rho\left[E + \frac{1}{\beta}\ln(\rho)\right]\mathrm{d}r^N \tag{28}$$

It is clear that, if $\rho = \rho_0$, $A[\rho]$ approaches its minimal value as $A(\rho_0) = A_0$. This equation, known in the textbook of statistical mechanics, then serves as the variational principle for determining the equilibrium distribution.

Suppose we have a system that is described by QM/MM force field. The energy of the system is written as $E(R, r)$, again with R representing the coordinates of the QM subsystem and r being the coordinates of the MM subsystem. By taking the mean-field approximation, we have

$$\rho_{\mathrm{MF}}(R, r) = \rho_R(R)\rho_r(r) \tag{29}$$

If we further assume that the QM atoms are frozen in space $\rho_R(R) = \delta(R - R_0)$, then we have

$$\begin{aligned} A[\rho] &= \int \rho_{\mathrm{MF}}(R, r)\left[E(R, r) + \frac{1}{\beta}\ln(\rho_{\mathrm{MF}}(R, r))\right]\mathrm{d}R^M\mathrm{d}r^{M'} \\ &= \int \rho_r(r)E(R_0, r)\mathrm{d}r + \frac{1}{\beta}\int \rho_r(r)\ln(\rho_r(r))\mathrm{d}r + \text{const} \end{aligned} \tag{30}$$

Omitting the constant term, which is independent of the choice of R_0, we now seek the condition for the minimization of $A[\rho]$ from all possible $\rho_r(r)$, which gives

$$\rho_r^0(r) = \frac{\exp(-\beta E(R_0, r))}{\int \exp(-\beta E(R_0, r))\mathrm{d}r} \tag{31}$$

Using $\rho_r^0(r)$ in eq 30, we obtain the free energy as a function of R_0, the QM fixed coordinates in the frozen QM approximation

$$A(R_0) = -\frac{1}{\beta}\ln\int \exp(-\beta E(R_0, r))\mathrm{d}r + \text{const} \tag{32}$$

which is just the potential of mean force (PMF) in the QM coordinate (eq 7 of the text). Therefore, the PMF $A(R_0)$ is the approximation to the free energy under the mean field and the frozen QM approximation. The minimum of the PMF $A(R_0)$ thus provides the best estimate to the free energy of the QM/MM system under the same approximations with optimal QM frozen coordinates. This appears to be an interesting interpretation of PMF and the minimum on the PMF.

References

(1) Zhang, Y.; Liu, H.; Yang, W. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* **2000**, *112*, 3483−3492.

(2) Lu, Z.; Yang, W. Reaction path potential for complex systems derived from combined ab initio quantum mechanical and molecular mechanical calculations. *J. Chem. Phys.* **2004**, *121*, 89−100.

(3) Warshel, A. *Computer modeling of chemical reactions in enzymes and solutions*; John Wiley & Sons: New York, 1991.

(4) Fersht, A. *Structure and mechanism in protein science;* W. H. Freeman and Company: New York, 1998.

(5) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. How enzymes work: Analysis by modern rate theory and computer simulations. *Science* **2004**, *303*, 186−195.

(6) Shurki, A.; Warshel, A. Why does the Ras switch "break" by oncogenic mutations? *Proteins: Struct., Funct., Genet.* **2004**, *55*, 1−10.

(7) Cui, Q.; Karplus, M. Triosephosphate isomerase: A theoretical comparison of alternative pathways. *J. Am. Chem. Soc.* **2001**, *123*, 2284−2290.

(8) Cui, Q.; Elstner, M.; Karplus, M. A Theoretical Analysis of the Proton and Hydride Transfer in Liver Alcohol Dehydrogenase (LADH). *J. Phys. Chem. B* **2002**, *106*, 2721−2740.

(9) Liu, H. Y.; Zhang, Y. K.; Yang, W. T. How is the active site of enolase organized to catalyze two different reaction steps? *J. Am. Chem. Soc.* **2000**, *122*, 6560−6570.

(10) Cisneros, G. A.; Liu, H. Y.; Lu, Z. Y.; Yang, W. T. Reaction path determination for quantum mechanical/molecular mechanical modeling of enzyme reactions by combining first order and second order "chain-of-replicas" methods. *J. Chem. Phys.* **2005**, 122.

(11) Cisneros, G. A.; Liu, H. Y.; Zhang, Y. K.; Yang, W. T. Ab initio QM/MM study shows there is no general acid in the reaction catalyzed by 4-oxalocrotonate tautornerase. *J. Am. Chem. Soc.* **2003**, *125*, 10384−10393.

(12) Cisneros, G. A.; Wang, M.; Silinski, P.; Fitzgerald, M. C.; Yang, W. The protein backbone makes important contributions to 4-oxalocrotonate tautomerase enzyme catalysis: Understanding from theory and experiment. *Biochemistry* **2004**, *43*, 6885−6892.

(13) Cisneros, G. A.; Wang, M.; Silinski, P.; Fitzgerald, M. C.; Yang, W. T. Theoretical and experimental determination on two substrates turned over by 4-oxalocrotonate tautomerase. *J. Phys. Chem. A* **2006**, *110*, 700−708.

(14) Warshel, A.; Levitt, M. Theoretic studies of enzymic reactions: Dielectric electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(15) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.

(16) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4507.

(17) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(18) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209−220.

(19) Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* **1989**, *10*, 221−264.

(20) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260−7268.

(21) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. A self-consistent charge density-functional based tight-binding scheme for large biomolecules. *Phys. Status Solidi B* **2000**, *217*, 357−376.

(22) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1994.

(23) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864−B871.

(24) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(25) Shurki, A.; Warshel, A. Structure/function correlations of proteins using MM, QMcharacterMM, and related approaches: Methods, concepts, pitfalls, and current progress. *Adv. Protein Chem.* **2003**, *66*, 249−313.

(26) Warshel, A. Computer simulations of enzyme catalysis: Methods, progress, and insights. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.

(27) Gao, J.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.

(28) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* **2006**, *110*, 6458−6469.

(29) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(30) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. Theoretical examination of the S_N2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.* **1985**, *107*, 154−163.

(31) Jorgensen, W. L. Free energy calculations: A breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* **1989**, 22.

(32) Kuhn, B.; Kollman, P. A. QM-FE and molecular dynamics calculations on catechol O-methyltransferase: Free energy of activation in the enzyme and in aqueous solution and regioselectivity of the enzyme-catalyzed reaction. *J. Am. Chem. Soc.* **2000**, *122*, 2586−2596.

(33) Singh, U. C.; Kollmann, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $CH_3Cl + Cl^-$ exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* **1986**, *7*, 718−730.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(35) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689−1700.

(36) Wesolowski, T.; Warshel, A. Ab initio free energy perturbation calculations of solvation free energy using the frozen density functional approach. *J. Phys. Chem.* **1994**, *98*, 5183−5187.

(37) Wesolowski, T.; Muller, R. P.; Warshel, A. Ab initio frozen density functional calculations of proton transfer reactions in solution. *J. Phys. Chem.* **1996**, *100*, 15444−15449.

(38) Muller, R. P.; Warshel, A. Ab initio calculations of free energy barriers for chemical reactions in solution. *J. Phys. Chem.* **1995**, *99*, 17516−17524.

(39) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: The nucleophilic attack in subtilisin. *J. Phys. Chem. B* **1998**, *102*, 2293−2301.

(40) Strajbl, M.; Hong, G.; Warshel, A. Ab initio QM/MM simulation with proper sampling: "First principle" calculations of the free energy of the autodissociation of water in aqueous solution. *J. Phys. Chem. B* **2002**, *106*, 13333−13343.

(41) Rod, T. H.; Ryde, U. Accurate QM/MM free energy calculations of enzyme reactions: Methylation by catechol O-methyltransferase. *J. Chem. Theory. Comput.* **2005**, *1*, 1240−1251.

(42) Zhang, Y.; Lee, T.-S.; Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **1999**, *110*, 46−54.

(43) Liu, H.; Lu, Z.; Cisneros, G. A.; Yang, W. Parallel iterative reaction path optimization in ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions. *J. Chem. Phys.* **2004**, *121*, 697−706.

(44) Wang, M.; Lu, Z.; Yang, W. Transmission coefficient calculation for proton transfer in triosephosphate isomerase based on the reaction path potential method. *J. Chem. Phys.* **2004**, *121*, 101−107.

(45) Wang, M.; Lu, Z.; Yang, W. Nuclear quantum effects on an enzyme-catalyzed reaction with reaction path potential: Proton transfer in triosephosphate isomerase. *J. Chem. Phys.* **2006**, *124*, 124516.

(46) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: Application to an enzymatic reaction. *J. Chem. Theory Comput.* **2006**, *2*, 452−461.

(47) Klahn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions. *J. Phys. Chem. B* **2005**, *109*, 15645−15650.

(48) Schenter, G. K.; Garrett, B. C.; Truhlar, D. G. Generalized transition state theory in terms of the potential of mean force. *J. Chem. Phys.* **2003**, *119*, 5828−5833.

(49) Fleurat-Lessard, P.; Ziegler, T. Tracing the minimum-energy path on the free-energy surface. *J. Chem. Phys.* **2005**, *123*, 084101.

(50) Yang, S.-Y.; Hristov, I.; Fleurat-Lessard, P.; Ziegler, T. Optimizing the structures of minimum and transition state on the free energy surface. *J. Phys. Chem. A* **2005**, *109*, 197−204.

(51) Hirao, H.; Nagae, Y.; Nagaoka, M. Transition-state optimization by the free energy gradient method: Application to aqueous-phase Menshutkin reaction between ammonia and methyl chloride. *Chem. Phys. Lett.* **2001**, *348*, 350−356.

(52) Okuyama-Yoshida, N.; Nagaoka, M.; Yamabe, T. Potential energy function for intramolecular proton transfer reaction of glycine in aqueous solution. *J. Phys. Chem. A* **1998**, *102*, 285−292.

(53) Okuyama-Yoshida, N.; Kataoka, K.; Nagaoka, M.; Yamabe, T. Structure optimization via free energy gradient method: Application to glycine zwitterion in aqueous solution. *J. Chem. Phys.* **2000**, *113*, 3519−3524.

(54) Nagaoka, M.; Okuyama-Yoshida, N.; Yamabe, T. Origin of the transition state on the free energy surface: Intramolecular proton transfer reaction of glycine in aqueous solution. *J. Phys. Chem. A* **1998**, *102*, 8202−8208.

(55) Okuyama-Yoshida, N.; Nagaoka, M.; Yamabe, T. Transition-state optimization on free energy surface: Toward solution chemical reaction ergodography. *Int. J. Quantum Chem.* **1998**, *70*, 95−103.

(56) Jónsson, H.; Mills, G.; Jacobsen, K. W. *Nudged elastic band method for finding minimum energy paths of transitions*; World Scientific: Singapore, 1998.

(57) Ayala, P. Y.; Schlegel, H. B. A combined method for determining reaction paths, minima, and transition state geometries. *J. Chem. Phys.* **1997**, *107*, 375−384.

(58) Xie, L.; Liu, H.; Yang, W. Adapting the nudged elastic band method for determining minimum-energy paths of chemical reactions in enzymes. *J. Chem. Phys.* **2004**, *120*, 8039−8052.

(59) Burger, S. K.; Yang, W. Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *J. Chem. Phys.* **2006**, *124*, 054109.

(60) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.

(61) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(62) Bolhuis, P. G.; Dellago, C.; Chandler, D. Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5877−5882.

(63) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(64) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785.

(65) Miller, W. H.; Handy, N. C.; Adams, J. E. Reaction path Hamiltonian for polyatomic molecules. *J. Chem. Phys.* **1980**, *72*, 99−112.

(66) Morita, A.; Kato, S. Ab initio molecular orbital theory on intramolecular charge polarization: Effect of hydrogen abstraction on the charge sensitivity of aromatic and nonaromatic species. *J. Am. Chem. Soc.* **1997**, *119*, 4021−4032.

(67) Cornell, W. D.; Cieplak, P.; Bayly, C.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(68) Ponder, J. W. *TINKER, Software Tools for Molecular Design*, version 3.8; Washington University: St. Louis, MO, 2000.

(69) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990−2001.

(70) Schlick, T.; Skeel, R. D.; Brünger, A. T.; Kalé, L. V.; Board, J. A.; Hermans, J.; Schulten, K. Algorithmic challenges in computational molecular biophysics. *J. Comput. Phys.* **1999**, *151*, 9−48.

(71) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(72) Mann, G.; Yun, R. H.; Nyland, L.; Prins, J.; Board, J.; Hermans, J. The Sigma MD program and a *generic* interface applicable to multi-functional programs with complex, hierarchical command structure. In *Computational Methods for Macromolecules: Challenges and Applications*; Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modelling; Schlick, T., Gan, H. H., Eds.; Springer-Verlag: Berlin, 2002; pp 129−145.

(73) Hu, H.; Elstner, M.; Hermans, J. Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine dipeptides (Ace−Ala−Nme and Ace−Gly−Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 451−463.

(74) Hu, H.; Yang, W. Dual-topology/dual-coordinate free-energy simulation using QM/MM force field. *J. Chem. Phys.* **2005**, *123*, 041102.

(75) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(76) Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulatin: A primer. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Eds. ESCOM: Leiden, The Netherlands, 1989; Vol. 1, pp 1−26.

(77) Hermans, J. A simple analysis of noise and hysteresis in free energy simulations. *J. Phys. Chem.* **1991**, *95*, 9029−9032.

(78) Hu, H.; Yun, R. H.; Hermans, J. Reversibility of free energy simulations: Slow growth may have a unique advantage. (With a note on use of Ewald summation). *Mol. Simul.* **2002**, *28*, 67−80.

(79) Li, G.; Cui, Q. Direct determination of reaction paths and stationary points on potential of mean force surfaces. *J. Mol. Graphics Modell.* **2005**, *24*, 82−93.

(80) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1986.

(81) Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(82) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **1989**, *156*, 472−477.

(83) Ryckaert, J. P.; Ciccotti, G. Introduction of Andersen's demon in the molecular dynamics of systems with constraints. *J. Chem. Phys.* **1983**, *78*, 7368−7374.

(84) Sprik, M.; Ciccotti, G. Free energy from constrained molecular dynamics. *J. Chem. Phys.* **1998**, *109*, 7737−7744.

(85) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F. Statistical mechanics and molecular dynamics: The calculation of free energy. In *Molecular Dynamics and Protein Structure*; Hermans, J., Ed. Polycrystal Book Service: Western Springs, IL, 1985; pp 43−46.

(86) Straatsma, T. P.; Zacharias, M.; McCammon, J. A. Holonomic constraint contributions to energy differences from thermodynamic integration molecular dynamics simulations. *Chem. Phys. Lett.* **1992**, *196*, 297−302.

(87) de Otter, W. K.; Briels, W. J. The calculation of free-energy differences by constrained molecular-dynamics simulations. *J. Chem. Phys.* **1998**, *109*, 4139−4146.

(88) Schlitter, J.; Klähn, M. A new concise expression for the free energy of a reaction coordinate. *J. Chem. Phys.* **2003**, *118*, 2057−2060.

(89) Antikainen, N. M.; Smiley, R. D.; Benkovic, S. J.; Hammes, G. G. Conformation coupled enzyme catalysis: Single-molecule and transient kinetics investigation of dihydrofolate reductase. *Biochemistry* **2005**, *44*, 16835−16843.

(90) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. Enzyme dynamics during catalysis. *Science* **2002**, *295*, 1520−1523.

(91) Amadei, A.; de Groot, B. L.; Ceruso, M. A.; Paci, M.; Di Nola, A.; Berendsen, H. J. C. A kinetic model for the internal motions of proteins: Diffusion between multiple harmonic wells. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 283−292.

(92) Kitao, A.; Hayward, S.; Go, N. Energy landscape of a native protein: Jumping-among-minima model. *Proteins: Struct., Funct., Genet.* **1999**, *33*, 496−517.

CT600240Y

JCTC Journal of Chemical Theory and Computation

Critical Assessment of the Performance of Density Functional Methods for Several Atomic and Molecular Properties

Kevin E. Riley,† Bryan T. Op't Holt,† and Kenneth M. Merz, Jr.*

Department of Chemistry, and Quantum Theory Project, University of Florida, P.O. Box 118435, Gainesville, Florida 32611-8435

Received June 1, 2006

Abstract: The reliable prediction of molecular properties is a vital task of computational chemistry. In recent years, density functional theory (DFT) has become a popular method for calculating molecular properties for a vast array of systems varying in size from small organic molecules to large biological compounds such as proteins. In this work, we assess the ability of many DFT methods to accurately determine atomic and molecular properties for small molecules containing elements commonly found in proteins, DNA, and RNA. These properties include bond lengths, bond angles, ground-state vibrational frequencies, electron affinities, ionization potentials, heats of formation, hydrogen-bond interaction energies, conformational energies, and reaction barrier heights. Calculations are carried out with the 3-21G*, 6-31G*, 3-21+G*, 6-31+G*, 6-31++G*, cc-pVxZ, and aug-cc-pVxZ (x = D and T) basis sets, while bond-distance and bond-angle calculations are also done using the cc-pVQZ and aug-cc-pVQZ basis sets. Members of the popular functional classes, namely, local spin density approximation, generalized gradient approximation (GGA), meta-GGA, hybrid-GGA, and hybrid-meta-GGA, are considered in this work. For the purpose of comparison, Hartree−Fock and second-order many-body perturbation methods are also assessed in terms of their ability to determine these physical properties. Ultimately, it is observed that the split valence bases of the 6-31G variety provide accuracies similar to those of the more computationally expensive Dunning-type basis sets. Another conclusion from this survey is that the hybrid-meta-GGA functionals are typically among the most accurate functionals for all of the properties examined in this work.

1. Introduction

In the past several years, it has become clear that, with the availability of increasingly powerful computers and improvements in computational methods, it will soon be possible to perform high-level ab initio calculations on large biological systems. Indeed, many DFT studies on biological systems have been carried out,[1−25] and some preliminary large-scale biological studies, treating entire proteins, have already been performed using ab initio methods.[26−29] Density functional theory (DFT) methods scale favorably with molecular size compared to Hartree−Fock and post-Hartree−Fock methods and have the further advantage over the Hartree−Fock method of describing electron correlation effects.[30,31] For these reasons, DFT is the ab initio technique most well-suited for studies on large compounds and promises to be at the forefront of ab initio investigations on biologically relevant systems such as proteins and DNA.

In this study, 37 DFT methods along with Hartree−Fock (HF) and second-order many-body perturbation (MP2) are assessed for their ability to accurately calculate nine molecular properties. These properties include bond lengths, bond angles, ground-state vibrational frequencies, electron affinities (EA), ionization potentials (IP), heats of formation (HOF), hydrogen-bond interaction energies, conformational

* Corresponding author e-mail: merz@qtp.ufl.edu.
† These authors contributed equally to this work.

10.1021/ct600185a CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/20/2007

Table 1. The 37 DFT Methods and Two Wave-Function Methods Used in This Work, with Appropriate References

method	reference	method	reference
HF	90	**hybrid-GGA**	
MP2	91	B1LYP	92−94
LSDA		B3LYP	93−96
SVWN5	98, 99	PBE1PBE	97
SPL	98, 101	B3P86	93, 100
c-SVWN5(0.3)	53, 98, 99	B3PW91	93, 101−103
GGA		B98	104
BLYP	93, 94	**meta-GGA**	
BPW91	93, 101, 102	VSXC	105
PBELYP	94, 97	BB95	93, 106
PBEP86	97, 100	MPWB95	106, 107
PBEPW91	97, 101, 102	TPSS	108, 109
PBEPBE	97	MPWKCIS	106, 110−112
PW91LYP	94, 101, 102	PBEKCIS	97, 110−112
PW91P86	100−102	TPSSKCIS	108−112
PW91PW91	101, 102	**hybrid-meta-GGA**	
MPWLYP	94, 100, 107	BB1K	93, 106, 113
MPWP86	100, 107	B1B95	93, 106
MPWPW91	101, 102, 107	TPSS1KCIS	108−112, 114
MPWPBE	97, 107	PBE1KCIS	73, 97, 110−112
G96LYP	94, 115	MPW1KCIS	78, 92, 106, 110−112
G96P86	100, 115		
HCTH	116		

energies, and reaction barrier heights. Each property is associated with its own test set. For consistency, each density functional method, wave function method, and basis set considered in this survey was used to calculate each of the nine properties (except for EA, for which nondiffuse basis functions were omitted).

DFT is not a single method but a family of methods because the exact density functional is unknown. Density functional methods can be divided into several classes according to the types of functional dependencies that they possess; examples of the five commonly used DFT categories are included in this study. The simplest type of DFT is the local spin density approximation (LSDA), which depends only on electron density. Generalized gradient approximation (GGA) functionals depend on the electron density and its reduced gradient, while meta-GGA functionals also depend on the kinetic energy density. Hybrid and meta-hybrid functionals are combinations of GGA and meta-GGA functionals with Hartree−Fock exchange. Table 1 lists all of the functionals investigated in this work and indicates the category to which each functional belongs. For the purpose of comparison to other popular molecular electronic structure methods, both the HF and MP2 methods are also included in this study.

Perdew and Schmidt's "Jacob's ladder" approach for the systematic improvement of density functional approximations contains five rungs, with each possessing more accurate approximations than the one below it.[32] The five classes of density functionals investigated in this work contain elements of the first four rungs, with the hybrid-meta-GGA being the most complex. Functionals residing on the highest step in this scheme would include an exact exchange term and "exact partial correlation".[32,33] There have been a limited number of "fifth-rung" functionals developed in the past few years, but these types of functionals are not yet widely used. An example of a "fifth-rung" functional is that of Perdew and co-workers that combines exact exchange and second-order correlation with a gradient-corrected density functional.[34,35] Since the use of functionals in this work is confined to more widely used DFT methods, functionals from the fifth rung of "Jacob's ladder" are not included here.

In this study, we employ a wide variety of basis sets ranging in size from the small, 3-21G*, basis to the very large, aug-cc-pVQZ, basis. The Pople-type split valence basis sets, such as 3-21G, 6-31G, and so forth, are used extensively throughout chemistry and are very well-validated; in this study, we utilized five of these basis sets, namely, 3-21G*, 6-31G*, 3-21+G*, 6-31+G*, and 6-31++G*.[36−39] The correlation-consistent basis sets of Dunning are also very widely used and are typically used in conjunction with high-level, post-Hartree−Fock techniques such as the configuration interaction and coupled cluster methods.[40] These basis sets incorporate functions with high angular momentum (d,f,g,...) and have been optimized to describe correlation effects in atoms, but they also describe correlation effects in molecules quite well.[41] It has been shown that, when used with DFT methods, the correlation-consistent basis sets yield optimized geometries that agree with experimental results very well. However, these basis sets have the disadvantage of being computationally expensive compared to the Pople-type basis sets. In this work, we employ the cc-pVxZ and aug-cc-pVxZ basis sets with x = D, T, and Q for geometry calculations, while all others are performed at the x = D and T levels. MP2 geometry optimization calculations at the cc-pVQZ and aug-cc-pVQZ levels are not performed because of their prohibitive cost. We realize that the most appropriate basis sets for each individual property may not have been included in our survey. However, after careful consideration, we decided to limit the number of basis sets we employed because of computational constraints. We feel that the basis sets tested here are widely implemented and are familiar to most computational chemists.

This article contains a quantitative assessment of a wide variety of quantum chemical methods available in a widely used quantum chemistry software package. We have had to base our study on a limited number of functionals and basis sets because of time considerations. New functionals are constantly being developed, and it is impossible for a work such as this one to keep up with these new developments. Examples of more recent functionals are M05[42] and M05-2X[43] of Truhlar and co-workers; however, these are not yet readily available for use. While we have included some analysis of the information garnered by this research, it is beyond the scope of these articles to discuss in depth the mathematical composition of functionals. Scuseria and Staroverov[44] have composed a review that provides an overview of the development of most classes of density functional methods. Readers interested in the development of these types of quantum methods should refer to this and other similar reviews.

One of our primary interests is to predict the performance of density functional methods for calculations on biological

systems. With this in mind, the test sets used in this work represent a collection of systems containing atoms commonly found in biomolecules such as proteins and DNA, namely, C, H, N, O, S, and P. For most test sets, only systems whose physical properties have been determined experimentally have been included. All of the test sets employed in this survey are given in the Supporting Information.

In order to estimate the performance of DFT methods in calculating accurate molecular geometries in proteins, we have carried out geometry optimizations on a test set of 44 molecules whose structures are well-characterized experimentally. The test set comprises 71 bond lengths and 34 bond angles that can be compared directly with experimental results. One would expect that the functional/basis set combination that is chosen to optimize these structures would have a great impact on the accuracy with which the bond lengths and angles can be computed; for this reason, geometry optimizations have been carried out using a large number of density functional methods along with several basis sets. In a similar fashion, we have constructed a vibrational frequency test set of 35 molecules whose harmonic vibrational frequencies are experimentally well-characterized. A total of 145 vibrational frequencies are contained within this test set.

Several studies have been carried out to assess the performance of DFT methods for calculating molecular geometries and vibrational frequencies.[41,45−52] Johnson et al. conducted studies using the S (Slater), B (Becke88), SVWN5, BVWNV, SLYP, and BLYP functionals (as well as the HF, MP2, and QCISD methods) along with the 6-31G* basis set to asses the accuracy with which these methods predict bond lengths, bond angles, and vibrational frequencies, as well as atomic energies, dipole moments, and atomization energies.[52] For geometric and vibrational properties, a test set composed of 32 small molecules from the G2 set was employed (44 bond lengths, 18 bond angles, and 110 frequencies). Raymond and Wheeler assessed the performance of the B3LYP method paired with the aug-cc-pVxZ (x = D, T, and Q) basis sets in calculating the molecular geometries, energies, and harmonic vibrational frequencies of a set of 19 small inorganic molecules.[41] More recently, Wang and Wilson carried out studies using a test set of 17 molecules to determine the accuracy of molecular geometries and several other molecular properties calculated using the B3LYP, B3PW91, B3P86, BLYP, and BP86 functionals paired with the cc-pVxZ and aug-cc-pVxZ (x = D, T, Q, and 5) basis sets.[45] Riley et al. assessed the accuracy of the BLYP, B3LYP, SVWN5, c-SVWN5, and S (Slater) functionals, along with the 3-21G*, 3-21+G*, 6-311G**, and 6-311+G** basis sets, for calculating molecular geometries and vibrational frequencies along with several other atomic and molecular properties.[53]

Next, we consider three more physical properties: electron affinity, ionization potential, and heat of formation. The electron affinity is the energy gained by a neutral system upon attachment of an additional electron, thus forming an anion. This is a very challenging property to compute because of the difficulty associated with the treatment of anions and their, often loosely bound, extra electron. The ionization potential, the energy required to remove an electron from a bound state to infinite separation, has been known for some time to be an important property of atoms and molecules. The ability to predict ionization potentials accurately has deep implications in the field of photoelectron spectroscopy. The heat of formation is the change in enthalpy that occurs when a molecule is formed from its constituent elements in their most stable states. This physical parameter is used to assess the stability of a molecule, to estimate the amount of energy released in a reaction, and to calculate other thermodynamic functions.[54]

The test sets used for the IP, EA, and HOF calculations are derived from the Gaussian G2/97 test set.[55,56] In an effort to increase the contribution of phosphorus-containing systems in our test sets, a few non-G2 additions have been made. The electron affinity test set contains 25 atoms and molecules; 24 of these come from the G2/97 test set while one of them, PO_2, does not. The ionization potential test set contains 37 atoms and molecules. All but one of these systems, PO_2, comes from the G2/97 test set. The heat of formation test set contains 156 molecules. Three of these molecules are not from the G2/97 set: PH, PO_2, and CH_3-PH_2.

Several studies have been conducted to assess the performance of DFT and wave-function-based methods in computing electron affinities, ionization potentials, and heats of formation.[54−63] Curtiss et al. carried out a study to evaluate density functional methods as well as the Gaussian-2 method for computing electron affinities and ionization potentials using a test set including 58 electron affinities and 88 ionization potentials.[56] In this study, the B3LYP, B3PW91, B3P86, BLYP, BP86, and SVWN functionals were used along with the 6-311+G(3*df*,2*p*) basis set. Ernzerhof and Scuseria evaluate the performance of the SVWN, SVWN5, BLYP, B3LYP, VSXC, PBEPBE, and PBE1PBE functionals for calculating atomization energies, ionization potentials, electron affinities, and bond lengths; this study was also carried out using the 6-311+G(3*df*,2*p*) basis set.[63] Curtiss et al. assess the SVWN, BLYP, and B3LYP methods along with several Gaussian-3 methods for the calculation of heats of formation, ionization potentials, electron affinities, and proton affinities using the 6-311+G(3*df*,2*p*) basis set.[57] Brothers and Merz carried out a study in which several LSDA, GGA, and hybrid-GGA functionals were evaluated for the computation of heats of formation with small basis sets, namely 3-21G*, 3-21+G*, and MIDI!.[61]

Another goal of this work is to assess the performance of DFT for describing hydrogen-bonding interaction energies, conformational energies, and reaction barrier heights. Hydrogen-bonding plays a critical role in many physical phenomena. Among other things, these interactions are important in the formation of clusters; the stability of large molecules such as proteins, DNA, and polysaccharides; and the formation of protein−ligand complexes. Conformational energies are an important measure of a computational method's ability to accurately predict both the geometry and the electronic energy of a molecular system. The reaction barrier height is a measure of the energy necessary to drive

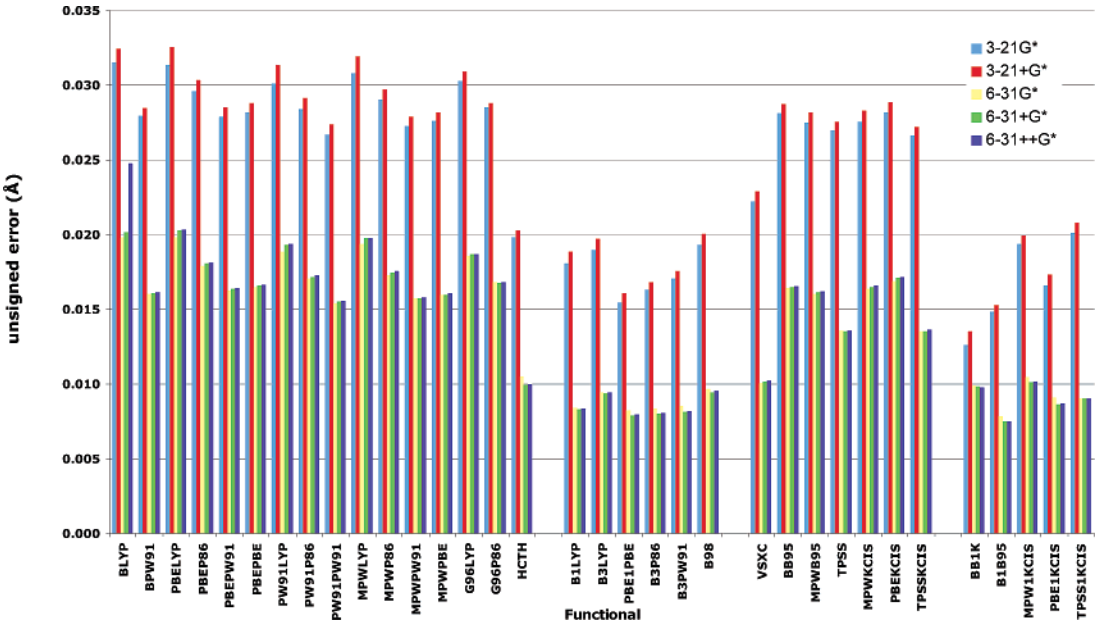


Figure 1. Average unsigned bond length errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople-type basis sets.

a stable system into a metastable transition state; this property has important implications in chemical reactions and kinetics.

The hydrogen-bonding test set contains 10 hydrogen-bound systems whose interaction energies have been well-characterized by high-level theoretical techniques. The conformational energy test comprises 10 small molecules whose experimental conformational energies are known. There are two different barrier height test sets; the first of these contains 23 barrier heights for hydrogen-transfer reactions of small molecules with radical (nonsinglet) transition states, while the second contains six barrier heights for reactions of larger molecules with singlet transition states. Henceforth, in this work, the former of these test sets will be referred to as the small radical barrier height set (SRBH) and the latter will be referred to as the large singlet barrier height set (LSBH).

Several studies have been carried out to assess the accuracy of density functional theory in predicting hydrogen-bonding interaction energies.[64−73] Tsuzuki and Lüthi evaluated the BLYP, B3LYP, and PW91PW91 functionals, as well as the MP2 and HF methods, for the prediction of hydrogen-bond interaction energies. These studies were carried out using the Dunning-type basis sets, cc-pVxZ (x = D, T, Q, and 5). Zhao and Truhlar carried out studies to determine the accuracy of DFT methods for several types of nonbonding interactions. These are hydrogen bonding, charge transfer, dipole interaction, and the weak (dispersion) interaction. These studies were done using a very large number of functionals along with the 6-31+G(*d,p*), MG3S[74] [a modified version of 6-31+G(3*d*2*f*,2*df*,2*p*)], and aug-cc-pVTZ basis sets. To our knowledge, there have been only a limited number of studies concerned with the accuracy with which DFT methods predict conformational energies.[74−76] Truhlar et al. evaluate the conformational energies of several conformer pairs of 1,2-ethandiol and butadiene. These studies were done using a number of functionals based on the MPW correlation functional along with several basis sets. There have been a number of studies carried out to evaluate the accuracy with which DFT methods describe reaction barrier heights.[77−81] In a recent study, Truhlar et al. test the accuracy of a large number of functionals along with several different basis sets for calculating the barrier heights of 38 hydrogen-transfer and 38 non-hydrogen-transfer reactions.[78]

2. Methods

All of the calculations carried out in this study were performed using Gaussian 03, version C.01.[82] The authors were informed that MPWKCIS, MPW1KCIS, PBEKCIS, and PBE1KCIS functionals were not correctly coded in Gaussian 03, version C.01. Therefore, calculations involving these functionals were rerun using Gaussian 03, version D.01. For DFT calculations, the default numerical grid in Gaussian 03 was used to evaluate the density functional theory integrals. Geometry optimizations were carried out for each of the functional/basis set combinations using the default optimization algorithm in Gaussian 03.

Previous studies have used molecular geometries obtained at high levels of theory for the calculation of molecular properties at lower levels of theory. Because information based on high levels of theory is not available for large biomolecules, we feel that it is more appropriate to optimize the molecular geometry using the same basis set and density functional that is being evaluated for some given molecular property. In this work, most physical properties are evaluated at the geometric minimum obtained using the same method/basis set combination. Unfortunately, because of computer time restrictions, it was not possible to optimize the molecules from the heat of formation test set at the aug-cc-pVTZ level. For these calculations, single-point calculations at the TPSS1KCIS/aug-cc-pVDZ optimized geometry were used to estimate the heat of formation.

Ionization potentials and electron affinities were calculated adiabatically. Only basis sets containing diffuse functions

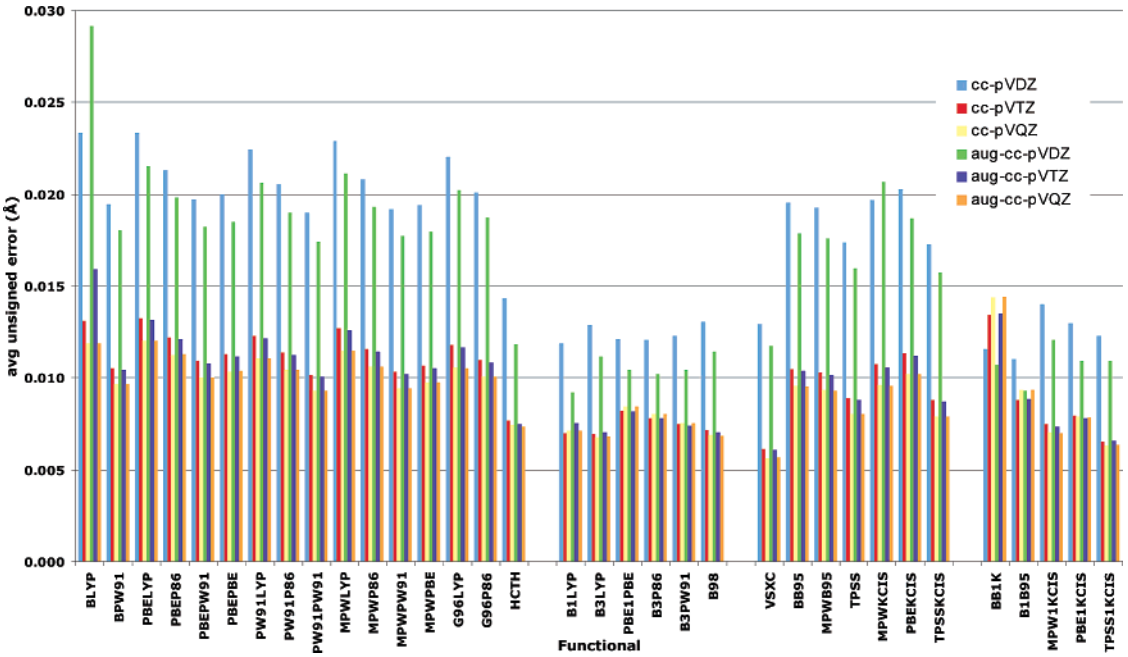


Figure 2. Average unsigned bond length errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning-type basis sets.

Table 2. Bond Lengths, Bond Angles, and Vibrational Frequencies for the HF, MP2, and LSDA Methods

method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	cc-pVQZ	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVQZ
					(a) Bond Lengths (Å)						
HF	0.015	0.016	0.015	0.014	0.014	0.013	0.018	0.019	0.012	0.017	0.019
MP2	0.024	0.027	0.011	0.012	0.012	0.014	0.009		0.015	0.009	
SVWN5	0.024	0.024	0.018	0.017	0.017	0.021	0.014	0.014	0.019	0.014	0.014
SPL	0.025	0.025	0.018	0.017	0.017	0.022	0.014	0.014	0.019	0.014	0.014
c-SVWN5	0.034	0.035	0.025	0.025	0.025	0.029	0.019	0.018	0.027	0.019	0.018
					(b) Bond Angles (degrees)						
HF	1.85	2.45	1.45	1.55	1.54	1.40	1.50	1.51	1.48	1.53	1.53
MP2	1.55	2.02	1.31	1.39	1.37	1.65	1.22		1.26	1.17	
SVWN5	1.80	2.58	1.36	1.27	1.27	1.60	1.25	1.22	1.28	1.24	1.23
SPL	1.80	2.55	1.37	1.27	1.27	1.60	1.25	1.22	1.27	1.23	1.22
c-SVWN5	1.69	2.31	1.52	1.28	1.28	1.75	1.24	1.20	1.25	1.19	1.17
					(c) Vibrational Frequencies (cm^{-1})						
HF	210	203	236	235	234	211	209	203	207		
MP2	151	147	149	141	141	126	122	113	117		
SVWN5	83	87	51	50	50	58	51	59	52		
SPL	82	87	52	50	50	57	52	58	51		
c-SVWN5	93	100	58	57	58	75	63	71	65		

were considered for electron affinity calculations. This was done because is it well-known that diffuse functions are necessary in order to properly treat anions. Heats of formation were calculated using the method specified in the "Thermochemistry in Gaussian" white paper available at http://www.Gaussian-.com/g_whitepap/thermo.htm.[83] Values for ionization potentials and electron affinities were calculated adiabatically. Most experimental data for heats of formation, ionization potentials, and electron affinities were obtained from the G2/97 and G3 test set papers; additional experimental data for phosphorus compounds were obtained from the NIST chemistry WebBook at http://www.nist.gov.[84]

For the investigations involving the small radical barrier height test set, because of substantial problems associated with transition-state convergence, we have used the geometries determined by Lynch and Truhlar at the QCISD/MG3 level for all calculations of this property (see http://comp.chem.umn.edu/truhlar/).[77]

For the hydrogen-bonding interaction energy calculations, the counterpoise method of Boys and Bernardi is employed in order to account for the basis set superposition error.[85] For each level of theory considered in this work, the geometries of the hydrogen-bound dimers are fully optimized on the counterpoise hypersurface, and the constituent monomers are also fully optimized. Because of the difficulties associated with the extraction of the zero-point-exclusive binding energies from experimental data, we have used binding energies obtained at a very high level of theory as reference values. These reference values were determined at the CCSD(T) basis set limit by Tsuzuki and Lüthi.[72]

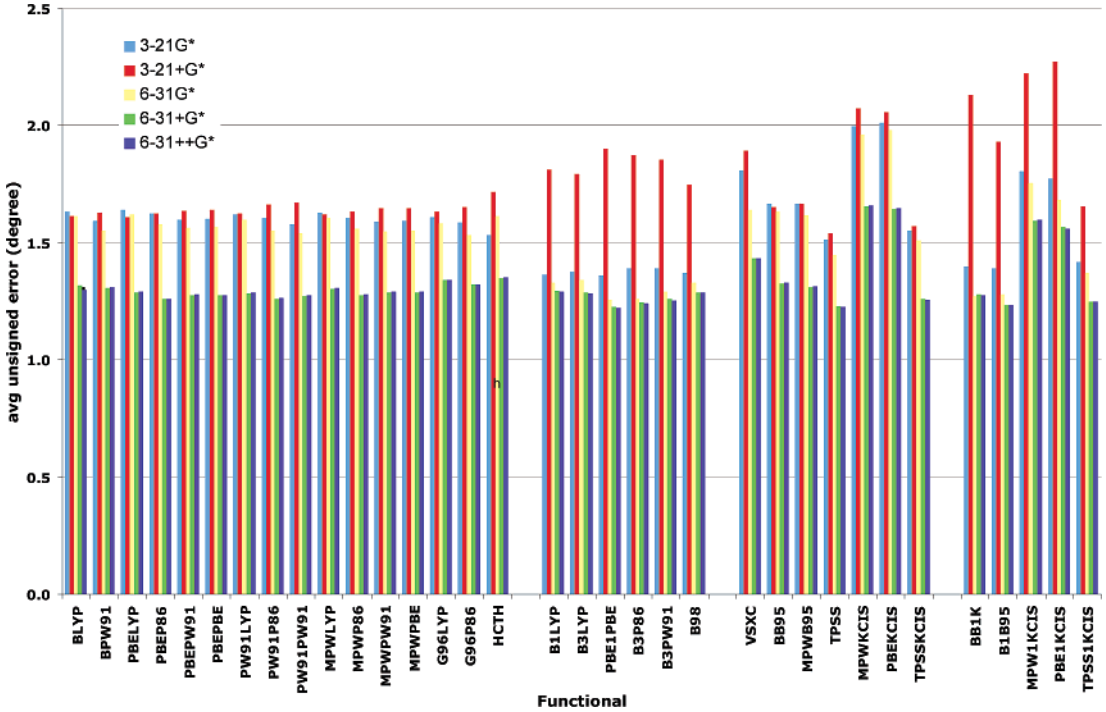


Figure 3. Average unsigned bond angle errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople-type basis sets.

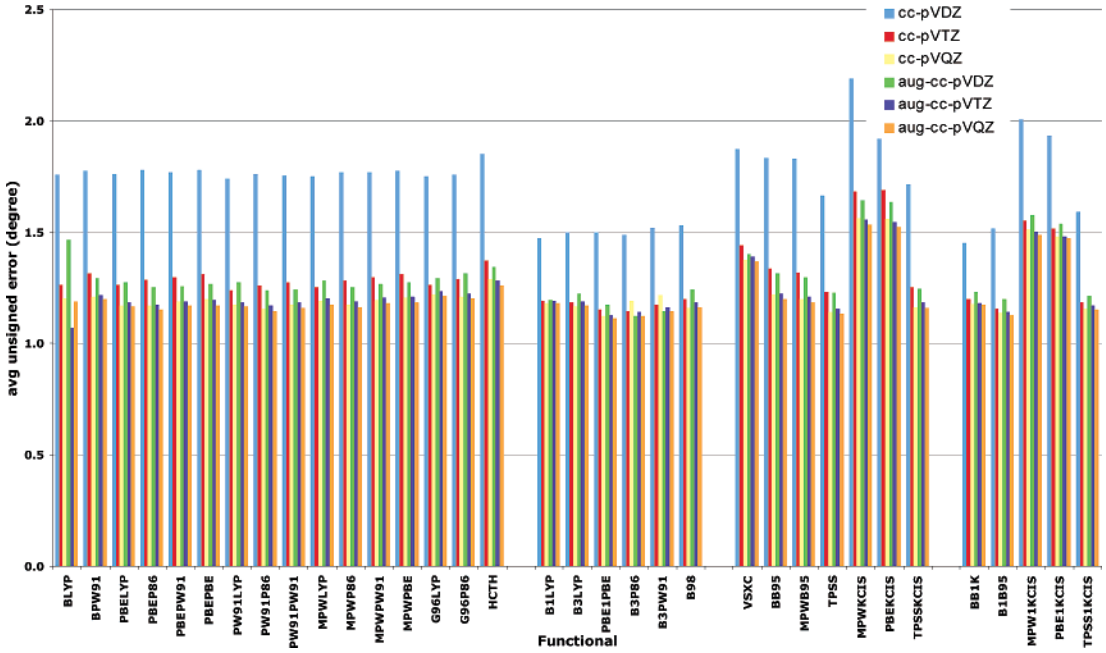


Figure 4. Average unsigned bond angle errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning-type basis sets.

Experimental conformational energy reference data were all obtained from ref 86. These quantities are calculated as potential energy differences, that is, the difference in electronic energy between the most stable conformer and the least stable one. This is the method employed in several other studies.[87−89]

The barrier height reference data forSRBH are in the form of zero-point exclusive, Born−Oppenheimer barrier heights. These barrier heights are simply calculated as the difference in electronic energy between the transition state and the reactants. For LSBH, the data are all directly from experimental results, and so it is necessary to include vibrational effects into the calculation of the barrier heights. These barrier heights are calculated as the difference in the thermally corrected total enthalpy between the transition state and the ground state of the reactant(s). For all six reactions in the singlet set, initial coordinates of the transition states were constructed to have sensible transition-state-like geometries. The transition state optimization method in Gaussian 03 was implemented using the *ts, calcall,* and *noeigentest*

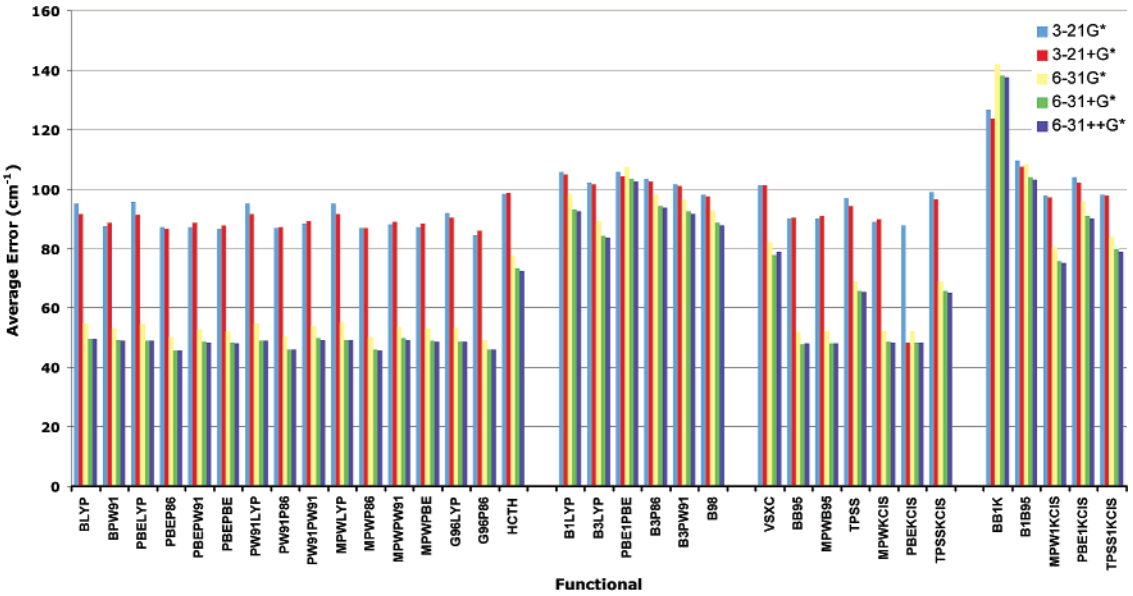


Figure 5. Average unsigned vibrational frequency errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Pople-type basis sets.

keywords. Reaction barrier heights were obtained from optimized reactant and transition-state structures for all functional and basis set combinations. MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ LSBH reaction calculations were not performed because of their considerable computational expense.

3. Discussion

3.1. Bond Lengths. The average unsigned bond length errors for the gradient-corrected, GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals, with Pople-type basis sets, are given in Figure 1. HF, MP2, and LSDA results are given in Table 2. Inspection of these figures reveals that the hybrid-GGA and hybrid-meta-GGA methods generally yield the lowest bond length errors when paired with these basis sets. The GGA and meta-GGA functionals both yield similar results with errors that are slightly greater than those of the hybrid and hybrid-meta functionals. Among the LSDA functionals, SVWN5 and SPL obtain errors that are comparable to those of GGA and meta-GGA, while c-SVWN5 yields errors that are higher than those of the other LSDA methods. Of the GGA functionals, from the meta-GGA class, the VSXC functional yields the lowest unsigned bond length errors.

Figure 2 gives average unsigned bond length errors for the gradient-corrected methods along with the Dunning-type basis sets. It can be seen that, overall, the hybrid-GGA functionals yield the lowest errors. Meta-GGA functionals and hybrid-meta-GGA functionals perform nearly as well, followed by the GGA functionals. LSDA methods yield errors that are generally higher than those of GGA functionals. It should be noted that, for the smallest of the Dunning-type basis sets, cc-pVDZ and aug-cc-pVDZ, the hybrid-GGA and hybrid-meta-GGA methods generally outperform all other methods by a significant margin. For the larger, cc-pVTZ, cc-pVQZ, aug-cc-pVTZ, and aug-cc-pVQZ, basis sets, the hybrid-GGA, hybrid-meta-GGA, and meta-GGA functionals all yield, on average, comparable results. As in the case of the Pople-type basis sets, VSXC yields the lowest bond length errors.

Generally, increasing the basis set size results in smaller bond length errors; the most dramatic improvements can be seen when the small 3-21G* and 3-21+G* basis sets are compared to the 6-31G* and 6-31+G* basis sets. The addition of diffuse functions to hydrogen atoms in the 6-31++G* makes little difference in the magnitudes of bond length errors. For the Dunning-type bases, the addition of diffuse functions improves accuracy only slightly. It is interesting to note that, for all of the functionals considered here, the bond length errors obtained using the 6-31G* and 6-31+G* basis sets are lower than those obtained using the much more computationally expensive cc-pVDZ and aug-cc-pVDZ basis sets.

One of the key features of the bond length data is that there is little variation within a particular functional class for a given basis set. For example, the average unsigned bond length errors calculated with the hybrid-GGA functionals along with the 6-31G* basis set are all within 0.002 Å of one another. This trend is especially evident for the GGA, hybrid-GGA, and meta-GGA functionals, whereas there is quite a bit more variation within the LSDA and hybrid-meta-GGA class of functionals. Another aspect of this trend is that, typically, the amount of variation within a particular functional class decreases as the basis set size increases. Individual functionals that deviate significantly from the other functionals within their class are c-SVWN, HCTH, B98, VSXC, and BB1K.

For singlets, the hybrid-GGA and hybrid-meta-GGA methods yield the lowest errors when paired with most basis sets. However, for the TZ bases, the hybrid-GGA, meta-GGA, and hybrid-meta-GGA methods yield the best singlet results. For the small Pople-type basis sets, 3-21G* and 3-21+G*, the LSDA methods yield results that are slightly better than those of the GGA functionals. At higher basis

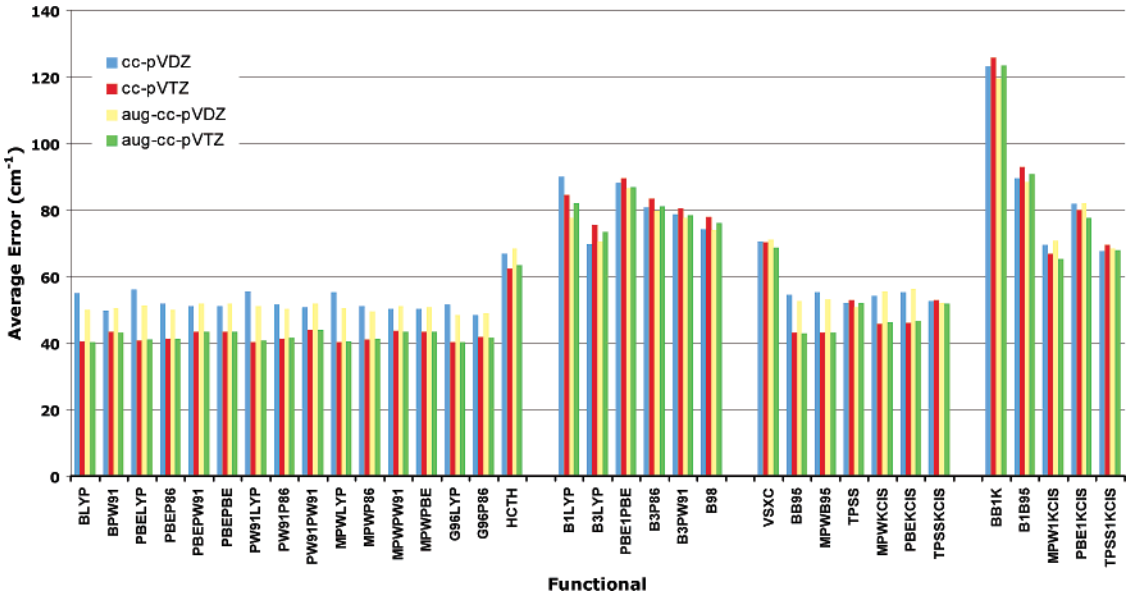


Figure 6. Average unsigned vibrational frequency errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with the Dunning-type basis sets.

Table 3. Average Unsigned Electron Affinity, Ionization Potential, and Heat of Formation Errors for the HF, MP2, LSDA, and B3P86 Methods[a,b]

Electron Affinity

method	3-21+G*	6-31+G*	6-31++G*	aug-cc-pVDZ	aug-cc-pVTZ
HF	26.40	29.05	28.95	28.90	29.63
MP2	11.14	10.73	10.62	5.81	4.70
SVWN5	9.37	6.25	6.36	7.58	7.53
SPL	9.92	6.79	6.90	8.14	8.04
c-SVWN5	12.04	14.80	14.65	13.60	13.68
B3P86	16.78	13.02	13.08	13.99	13.53

Ionization Potential

method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	26.53	23.00	25.70	24.74	24.76	26.33	25.74	25.19	25.40
MP2	15.59	9.77	10.45	10.42	5.84	11.44	9.78	6.74	5.53
SVWN5	10.74	10.29	7.57	8.52	8.48	7.68	8.33	8.59	8.75
SPL	10.50	10.37	7.49	8.62	8.57	7.46	8.42	8.70	8.83
c-SVWN5	24.93	17.31	22.57	19.08	19.05	23.37	19.85	19.06	18.69
B3P86	15.05	19.01	14.25	15.87	15.86	13.83	15.22	15.84	15.68

Heat of Formation

method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	289.94	294.01	252.94	256.28	258.10	261.27	248.30	259.40	248.49
MP2	58.23	55.95	83.74	84.53	83.27	96.10			
SVWN5	125.83	127.57	135.97	128.06	121.78	124.24	135.06	122.98	133.85
SPL	128.66	117.21	143.48	133.70	127.46	129.90	144.22	128.62	139.49
c-SVWN5	13.45	9.44	19.47	26.46	27.14	19.48	29.44	18.95	18.10

[a] All values in kcal/mol. [b] B3P86 for EA and IP only.

sets, all other density functional classes outperform the LSDA methods. In terms of radical systems, hybrid-GGA and hybrid-meta-GGA are the density functional methods that yield the lowest bond length errors for all basis sets, since the best results overall are obtained by MP2. For all basis sets, the meta-GGA methods obtain errors that are higher than those of the hybrid-GGA and hybrid-meta-GGA methods.

For all methods except MP2, single and triple bonds are calculated much more accurately than double bonds. For all basis sets, the LSDA, GGA, and meta-GGA methods tend to yield lower unsigned errors for triple bonds than for double bonds. Hybrid-GGA and hybrid-meta-GGA also obtain lower errors for triple bonds (compared to single bonds) when paired with Pople-type and cc-pVDZ/aug-cc-pVDZ basis sets, but they give higher errors for triple bonds with the larger Dunning-type basis sets. Hartree−Fock and MP2 generally obtain lower bond length errors for single bonds than triple bonds.

3.2. Bond Angles. The average unsigned bond angle errors are shown in Figures 3 and 4 and in Table 2. Generally, the hybrid-GGA and hybrid-meta-GGA methods produce the lowest errors. The best results among the small basis sets are generally obtained with the hybrid-GGA functionals along with the 3-21G* basis set. The hybrid-meta-GGA functionals also yield very good results, although there is some variation within this class. The LSDA methods produce bond angle errors that are higher than most of the gradient-corrected functionals. The addition of diffuse functions to the basis set results in a significant increase in the average bond angle errors for these functionals. The SVWN5 and SPL functionals both yield very similar errors, while the c-SVWN5 functional obtains results that are only slightly better than its LSDA counterparts when paired with the small bases. Overall, the best small basis set results are obtained by the B1LYP/3-21G* and PBE1PBE/3-21G* methods, which both calculate the average unsigned bond angle error to be 1.36°.

The larger Pople basis sets, 6-31G*, 6-31+G*, and 6-31++G*, generally produce better bond angle results for the basis sets containing diffuse functions. The wave-function-based methods yield lower errors when paired with the 6-31G* basis set. It should also be noted that the addition of diffuse functions to hydrogen atoms in the 6-31++G* basis set does not result in any significant improvement over the 6-31+G* basis. Once again, the hybrid-GGA and hybrid-meta-GGA methods generally give the lowest unsigned bond angle errors. As in the case of the small Pople basis sets, there is a great deal of variation in the meta-GGA class of functionals. LSDA methods all yield similar results when paired with basis sets containing diffuse functions. However, the c-SVWN5 functional produces errors that are significantly higher than those of SVWN5 and SPL when paired with the 6-31G* basis set. For 6-31G*, HF generally outperforms the GGA and meta-GGA methods and yields errors that are higher than those of all hybrid-GGA and hybrid-meta-GGA methods.

One aspect of the data for the Dunning-type basis sets is that the errors obtained with the cc-pVDZ basis set are much higher than those of all other Dunning basis sets for all of the methods considered except Hartree−Fock. In fact, cc-pVDZ is generally outperformed by all other basis sets, with the exception of 3-21+G*, for all functional methods. The hybrid-GGA and hybrid-meta-GGA functionals generally yield the lowest unsigned bond angle errors for the Dunning-type basis sets. Generally speaking, increasing the basis set size results in lower unsigned errors; this trend is especially evident for the GGA and meta-GGA functionals paired with nondiffuse basis sets. When diffuse functions are added to the cc-pVTZ basis set, there is a significant increase of accuracy for GGA and meta-GGA functionals. For most functionals, there is a small decrease in bond angle error upon the addition of diffuse functions to cc-pVQZ. Overall, the lowest error of 1.11° is obtained by the hybrid-GGA PBE1PBE/aug-cc-pVQZ method. PBEP86 and MPWP86 are the GGA functionals that typically yield the lowest unsigned bond angle errors, while HCTH is the least accurate in this class. It should be noted that BLYP exhibits trends that are quite different than those of other GGA functionals. BLYP/aug-cc-pVDZ yields errors that are much higher than those of all other GGA/aug-cc-pVDZ combinations, while BLYP/aug-cc-pVTZ produces errors that are significantly lower than those of all other GGA/aug-cc-pVTZ combinations. LSDA methods all yield similar results, with the exception of c-SVWN/cc-pVDZ. HF performs better when paired with the nondiffuse basis sets, and larger Dunning-type bases generally yield larger errors than the smaller Dunning bases. The unsigned bond angle errors obtained with MP2 improve with increasing basis set size and with the addition of diffuse functions to the basis set.

The HF, MP2, LSDA, GGA, and meta-GGA methods obtain lower bond angle errors for singlet states for most Dunning- and large Pople-type basis sets. By and large, the hybrid-GGA and hybrid-meta-GGA methods yield lower errors for radicals than for singlets. Most methods produce better results for radical species when paired with the 3-21+G* basis set. For singlets, the LSDA functionals produce the lowest unsigned bond angle errors among all functional-based methods when paired with the Dunning-type basis sets except cc-pVDZ and with the 6-31+G* and 6-31++G* Pople-type basis sets. Hybrid-GGA methods give the lowest errors among DFT methods for 3-21G* and 6-31G*. In terms of radical species, the hybrid-GGA methods give the lowest unsigned bond angle errors for all basis sets considered in this work. Hybrid-meta-GGA methods generally produce errors that are slightly larger than those of hybrid-GGA functionals.

3.3. Vibrational Frequencies. Figures 5 and 6 show the average unsigned vibrational frequency errors for gradient-corrected DFT methods. Vibrational errors for the LSDA methods as well as HF and MP2 are given in Table 2. It has been observed previously that frequency errors for MP2 and HF methods are larger than those of most DFT methods. It is observed in this study that the hybrid-GGA and hybrid-meta-GGA functionals, which include a Hartree−Fock-like exchange term, are less accurate than other classes of DFT functionals in predicting harmonic vibrational frequencies. As expected, HF and MP2 errors are far worse than errors obtained with DFT calculations. Compared with the GGA, meta-GGA, and LSDA classes of functionals, MP2 errors are higher by a factor of about 2 or 3, while HF errors are typically about 3−5 times higher. The lowest vibrational frequency error obtained by the MP2 method is 113 cm^{-1} for MP2/aug-cc-pVDZ, while its highest error is 151 cm^{-1} for MP2/3-21G*. HF errors for vibrational frequencies range from 203 cm^{-1}, for HF/3-21+G*, to 236 cm^{-1}, for HF/6-31G*.

The inclusion of diffuse functions in the basis set does not greatly affect the ability of the functionals to predict vibrational frequencies. As shown in the figures, 3-21G* and 3-21+G* are nearly identical in performance. There is only a small improvement, on the order of 5 cm^{-1}, with the use of the 6-31+G* and 6-31++G* basis sets as opposed to 6-31G*. Augmented correlation-consistent basis sets do not perform markedly better than their nondiffuse counterparts for the double-ζ and triple-ζ basis sets.

The three LSDA functionals yield lower average errors than hybrid-meta-GGA and hybrid-GGA functionals, but

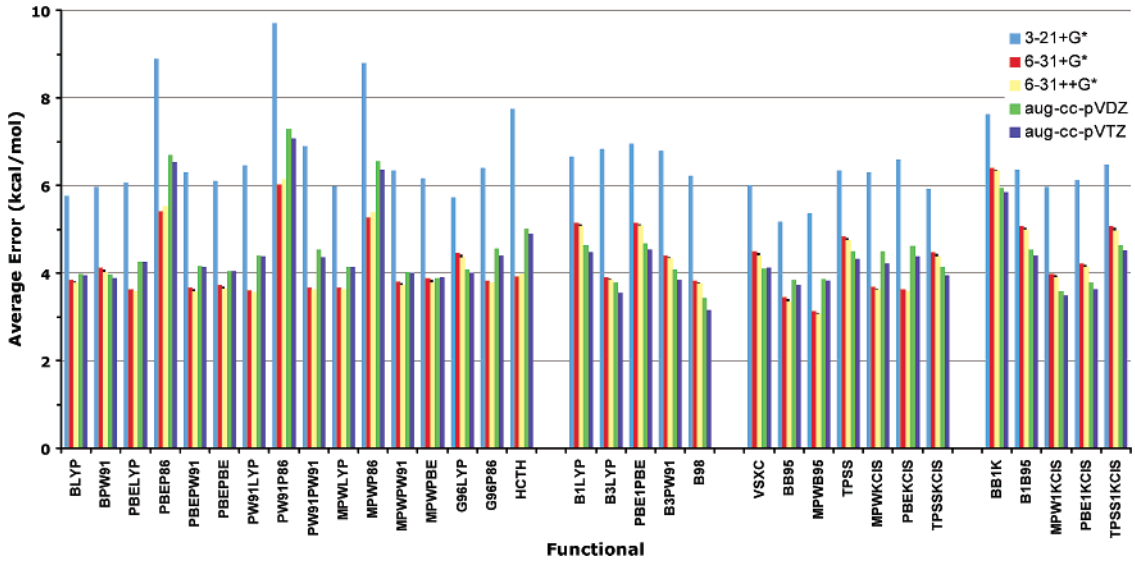


Figure 7. Average unsigned electron affinity errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals.

they are generally higher than the errors of meta-GGA and GGA functionals. The SPL and SVWN5 functionals perform slightly better than c-SVWN5, with SPL performing the best of the three. The 6-31+G* and 6-31++G* basis sets yield the lowest errors for the SPL functional, around 50 cm^{-1}.

Within the GGA class, there is little variation of performance between functionals with the exception of HCTH, which yields vibrational frequency errors that are significantly higher than any of the other functionals. In this class, the Dunning-style triple-ζ basis sets give the lowest unsigned vibrational frequency errors. Both the diffuse and nondiffuse variants of this basis set produce unsigned errors of 40−44 cm^{-1}. It is interesting to note that the 6-31+G* and 6-31++G* basis sets, which are much less computationally expensive than the cc-pVTZ and aug-cc-pVTZ basis sets, yield errors that are only 5−8 cm^{-1} higher than these Dunning-type basis sets. 6-31+G* and 6-31++G* also give results that typically outperform the cc-pVDZ and aug-cc-pVTZ basis sets. Of the 16 functionals in the GGA family, we find that MPWLYP and MPWP86 perform the best, but their advantage over most of the other functionals is only slight.

B3LYP is the most accurate of the hybrid-GGA class for calculating vibrational frequencies. B3LYP/cc-pVDZ yields the lowest error in this class at 70 cm^{-1}. PBE1PBE does not perform as well as the other members of this class. The meta-GGA class also shows some variance between functionals. VSXC, which is among the best meta-GGA methods for calculating bond lengths, performs poorly for vibrational frequencies. BB95, MPWB95, MPWKCIS, and PBEKCIS are similar in performance throughout. BB95/aug-cc-pVTZ and MPWB95/aug-cc-pVTZ yield the lowest error of the class at 43 cm^{-1}. Finally, the hybrid-meta-GGA group does not perform as well as most other classes because of its inclusion of the HF exchange. TPSS1KCIS and MPW1KCIS are the best functionals in this functional class; the cc-pVTZ and aug-cc-pVTZ basis sets typically give the lowest average errors within this group.

3.4. Electron Affinities. The electron affinity average unsigned errors are shown in Table 3 and in Figure 7; the first of these gives the results for the HF, MP2, B3P86, and LSDA functional methods; the second gives all GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functional method results except for the hybrid-GGA functional B3P86, which yields very poor results. The best overall result for electron affinities is obtained using the meta-GGA functional MPWB95 along with the 6-31++G* basis set, yielding an average unsigned error of 3.08 kcal/mol. The worst average unsigned error among the DFT methods is 16.78 kcal/mol and is obtained with the hybrid-GGA B3P86 functional combined with the 3-21+G* basis set. Among all methods studied in this work, the HF/aug-cc-pVTZ method yields the worst result with an average error of 29.63 kcal/mol.

Among all of the functional types considered here, the LSDA methods produce the highest errors; this is an expected result as these are the least sophisticated functionals and they lack gradient-dependent terms. It is also interesting to note that results obtained with methods that incorporate the P86 correlation functional, with the exception of G96P86, are significantly worse than those obtained using the other functionals within a given class.

Of the three LSDA functionals, the best result of 6.25 kcal/mol is obtained with SVWN5/6-31+G*. Among the GGA functionals, the PW91LYP/6-31++G* method yields the lowest average unsigned error of 3.56 kcal/mol; the PW91LYP/6-31+G*, PBEPW91/6-31++G*, and PBELYP/6-31++G* functional/basis set combinations all obtain errors lower than 3.60 kcal/mol. The B98/aug-cc-pVTZ method gives the smallest average error among the hybrid-GGA functionals of 3.15 kcal/mol. Among the meta-GGA functionals, the MPWB95/6-31++G* method yields the lowest error with a value of 3.08 kcal/mol; it is also noteworthy that the MPWB95/6-31+G* functional gives extremely good results with an error value of 3.12 kcal/mol. The MPW1KCIS/aug-cc-pVTZ method yields an error value of 3.48 kcal/mol, the lowest error among all hybrid-meta-GGA methods.

The Hartree−Fock method performs very poorly in describing electron affinities. This can be explained by the fact that, since an anion has one electron more than its neutral counterpart, correlation effects have a stronger effect on the

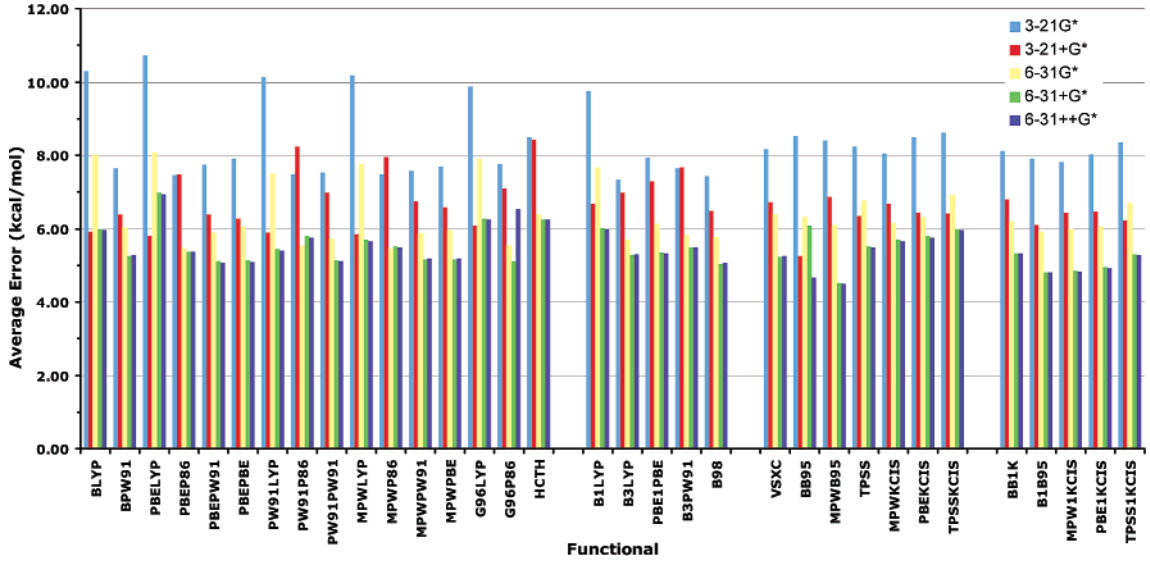


Figure 8. Average unsigned ionization potential errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals with Pople-type basis sets.

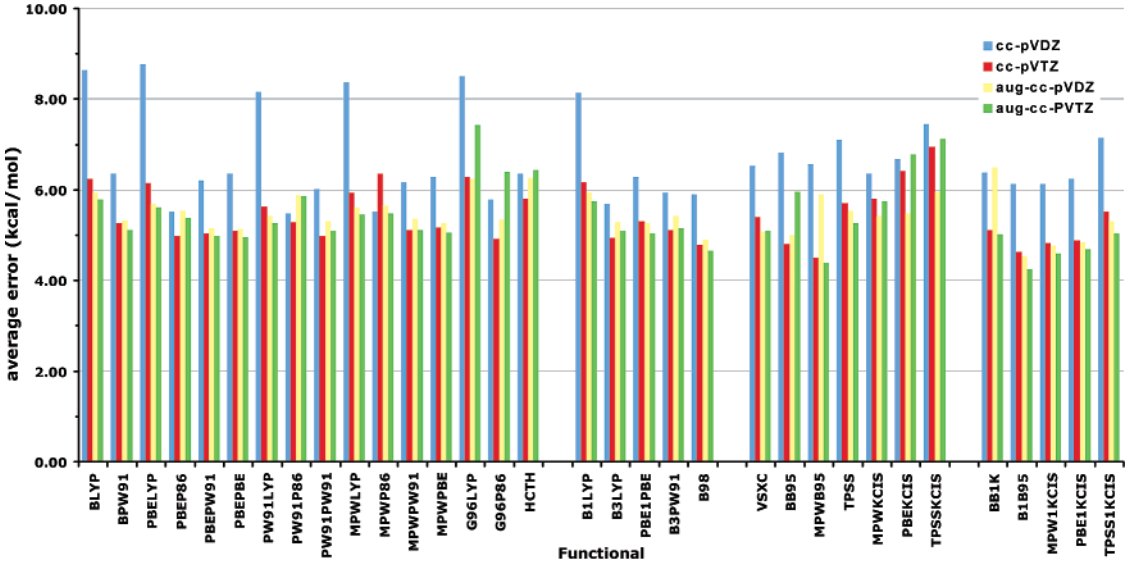


Figure 9. Average unsigned ionization potential errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals with Dunning-type basis sets.

negatively charged ion than on the neutral system. Because of the neglect of correlation effects in the Hartree−Fock technique, there is a pronounced discrepancy in its description of neutral and anionic species.

One salient aspect of these data is that, not surprisingly, the 3-21+G* basis set performs very poorly compared to the larger basis sets for all functionals. This basis set does, however, outperform all other basis sets when combined with the Hartree−Fock method and gives results that are only slightly worse than those obtained with the larger, 6-31+G* and 6-31++G*, basis sets when combined with the MP2 method. The lowest unsigned error for the very small (and inexpensive) 3-21+G* basis set is 5.17 kcal/mol as calculated with the BB95 functional.

It is also surprising that, generally, the 6-31+G* and 6-31++G* basis sets obtain results that are comparable to, or in many cases superior to, the aug-cc-pVDZ and aug-cc-pVTZ results. As can be seen in Figure 7 and Table 3, the average unsigned errors for the 6-31+G* and 6-31++G* basis sets are generally lower than those for the aug-cc-pVDZ and aug-cc-pVTZ basis sets for the LSDA, GGA, and meta-GGA functional classes. This trend is especially pronounced for the LSDA and GGA functionals; two of the three LSDA functionals obtain better results when combined with the smaller, Pople-6-31G-type, basis sets compared to the results they obtained when used in conjunction with the larger, Dunning-cc-pVXZ-type, basis sets. The Pople-type basis sets outperform the Dunning-type basis sets for 13 of the 16 GGA functionals. For the hybrid-GGA and hybrid-meta-GGA functionals, the Dunning-type basis sets typically outperform the Pople-type basis sets by a small margin (≤0.5 kcal/mol). The meta-GGA functionals represent a “mixed bag” in terms of this trend; here, the smaller basis sets outperform the larger

ones (not including the 3-21+G* basis set) for four of the seven functionals studied in this work.

It should be noted that the addition of diffuse functions to hydrogen atoms in the 6-31++G* basis set does not lead to results that are significantly different than those obtained with the 6-31+G* basis set. It is also interesting to note that the aug-cc-pVTZ basis set generally yields results that are only slightly better than the aug-cc-pVDZ basis set results.

3.5. Ionization Potentials. The average ionization potential unsigned errors for each functional/basis set combination are given in Figures 8 and 9 and in Table 3. Table 3 gives results of HF, MP2, LSDA functionals, and B3P86 (which yields high errors) for all basis sets. The best result for ionization potentials is obtained with the hybrid-meta-GGA functional B1B95 combined with the aug-cc-pVTZ basis set, yielding an error of 4.25 kcal/mol. The worst average unsigned error among the density functional methods is 15.05 kcal/mol and is obtained using the hybrid-GGA functional B3P86 along with the 3-21+G* basis set. Overall, the largest unsigned error of 26.53 kcal/mol is obtained with the HF/3-21G* method.

Of the three LSDA functionals, the best ionization potential unsigned error of 7.46 kcal/mol is obtained using the SPL/cc-pVDZ method. Among the GGA functionals, G96P86/cc-pVTZ yields the lowest unsigned error of 4.93 kcal/mol; it should also be noted that the PW91PW91/cc-pVTZ, PBEP86/cc-pVTZ, PBEPW91/aug-cc-pVTZ, and PBEPBE/aug-cc-pVTZ methods all yield errors lower than 5.00 kcal/mol. The B98/aug-cc-pVTZ method yields an average unsigned error of 4.65 kcal/mol, the lowest unsigned error among all hybrid-GGA methods. The functional/basis set combination yielding the best result in the meta-GGA class of functionals is MPWB95/aug-cc-pVTZ with a calculated value of 4.38 kcal/mol. Among the hybrid-meta-GGA functionals, the lowest average unsigned error is given by the B1B95/aug-cc-pVTZ method with a value of 4.25 kcal/mol.

As in the case of electron affinities, the Hartree−Fock method does a very poor job in predicting ionization potentials; this can be explained in the same way as above. The cation has fewer electrons than the neutral systems and, thus, exhibits less correlation effects. The Hartree−Fock method's inability to describe electron correlation will lead to a more accurate prediction for the electronic energy of the cation as compared to the neutral species.

As one might expect, the small, 3-21G* and 3-21+G*, basis sets typically perform very poorly in predicting ionization potentials compared to the larger basis sets. Generally, the 3-21+G* basis set predicts average errors that are substantially lower than those of the 3-21G* basis set. The best ionization potential result for the 3-21G* basis set is obtained with the hybrid-GGA B3LYP functional with an average error of 7.33 kcal/mol, while the 3-21+G* basis set has the lowest average unsigned error of 5.26 kcal/mol when used in conjunction with the meta-GGA BB95 functional.

Inspection of the average unsigned errors for individual functionals in Figures 8 and 9 and in Table 3 reveals that the cc-pVTZ, 6-31+G*, 6-31++G*, aug-cc-pVDZ, and aug-cc-pVTZ basis sets all yield fairly similar results that are typically superior to the results obtained with the 3-21G*, 3-21+G*, 6-31G*, and cc-pVDZ basis sets. An exception to this is the LSDA functionals, for which the 6-31G* and cc-pVDZ basis sets yield the lowest average unsigned errors. As the Pople-type 6-31+G* and 6-31++G* basis sets are computationally much less expensive to use compared to the larger Dunning correlation-consistent basis sets, it is very promising, in terms of biological applications, that such high-quality results can be obtained using the smaller basis sets. It should be noted that the MPWB95/(6-31+G*, 6-31++G*) methods (4.53 and 4.50 kcal/mol) outperform all other meta-GGA methods with the exception of the MPWB95/(cc-pVTZ, aug-cc-pVTZ) methods (4.49 and 4.38 kcal/mol). Similarly, the B1B95/(6-31+G*, 6-31++G*) methods (4.81 and 4.80 kcal/mol) yield better results than all other hybrid-meta-GGA methods with the exception of B1B95/(aug-cc-pVDZ, cc-pVTZ, aug-cc-pVTZ) methods (4.64, 4.53, and 4.25 kcal/mol). For the 6-31++G* basis set, the best GGA result of 5.07 kcal/mol is obtained with the PBEPW91 functional. The best hybrid-GGA result for Pople-type basis sets is 5.05 kcal/mol and is obtained with the B98/6-31+G* method.

As with the electron affinities, the addition of diffuse functions to hydrogen atoms in the 6-31++G* basis set seems to have a negligible effect on the calculation of ionization potentials compared to the 6-31+G* basis set. There is also only a small difference between the results obtained with the cc-pVTZ and aug-cc-pVTZ basis sets. There is, however, a marked difference in the quality of the cc-pVDZ and aug-cc-pVDZ basis set results.

3.6. Heats of Formation. Unsigned errors for the HOF test set are listed in Figures 10 and 11 and in Table 3. Overall, the combination that gives the lowest unsigned error is PBE1KCIS/aug-cc-pVTZ at 3.64 kcal/mol. Neglecting errors from the LSDA, MP2, and HF methods, the overall least accurate combination is PW91P86/6-31G* with an average unsigned error of 51.4 kcal/mol. The MPWLYP/3-21G* method yields an unsigned error of 5.66 kcal/mol, which is the lowest error for the relatively inexpensive 3-21G* and 3-21+G* basis sets.

For the Pople basis sets, the accuracy of HOF calculations is dependent on the size of the basis set for the hybrid-GGA and meta-hybrid GGA classes of functionals. As shown in Figure 10, all of the meta-hybrid GGA functionals and all but one of the hybrid-GGA functionals yield much higher errors for the 3-21G* and 3-21+G* basis sets than for the 6-31G*, 6-31+G*, and 6-31++G* basis sets, while the other functional classes show no such dependency. The use of diffuse basis sets with GGA and meta-GGA functionals appears to increase the accuracy of the methods as 3-21+G*, 6-31+G*, and 6-31++G* produce typically lower HOF errors than their nondiffuse counterparts. The opposite effect is observed when diffuse bases are paired with hybrid-GGA or hybrid-meta-GGA functionals. The most accurate functional/basis combination within the set of Pople bases is TPSSKCIS/6-31+G*, yielding an average unsigned error of 4.76 kcal/mol.

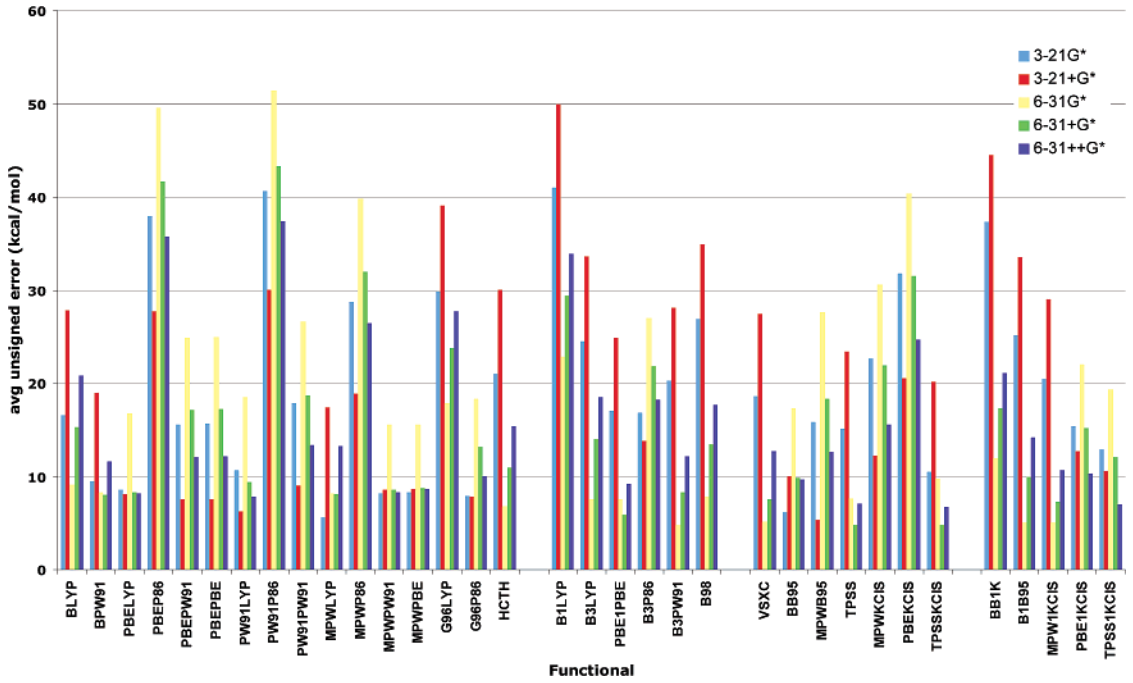


Figure 10. Average unsigned heat of formation errors for the five Pople-style basis sets employed in this study.

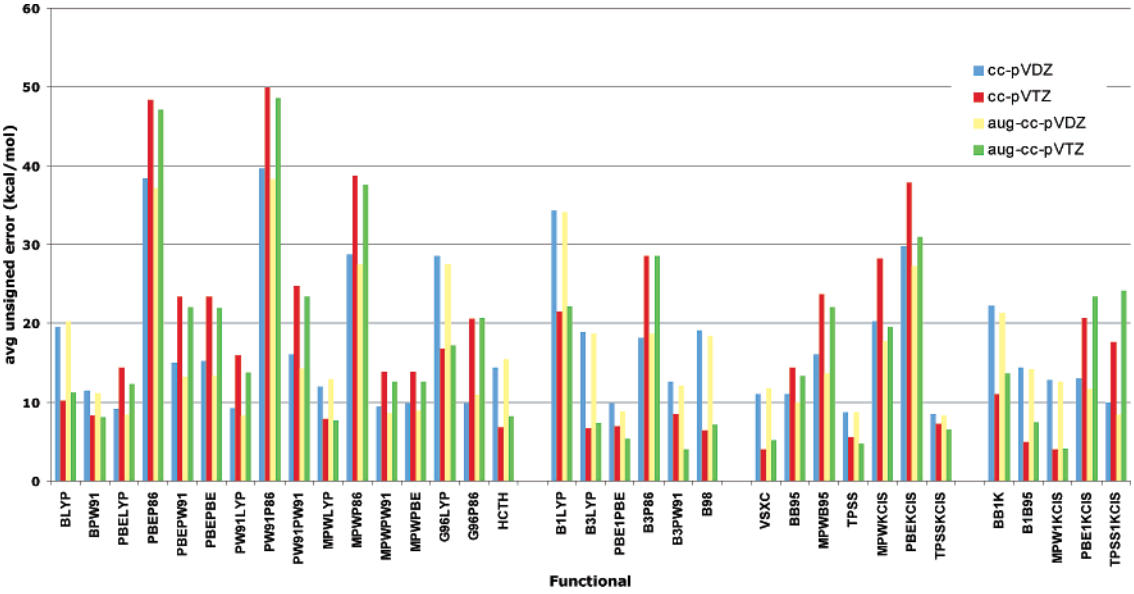


Figure 11. Average unsigned heat of formation errors for the Dunning-type correlation-consistent basis functions used in this work.

Generally, the hybrid-meta-GGA class of functionals produces the most accurate HOF calculations. Within this class B1B95/6-31G* yields the most accurate results with a 5.03 kcal/mol average unsigned error. For all five functionals in this class, the 6-31G* basis is the most accurate of the Pople-type bases. The meta-GGA class of functionals yields errors slightly larger than those of the hybrid-meta-GGA class, on the whole. Functionals employing the TPSS exchange perform the best in this class. BB95 also performs well. TPSSTPSS/6-31+G* and TPSSKCIS/6-31+G* are the most accurate combinations within this functional category, producing average errors of 4.76 kcal/mol. The 6-31+G* basis yields the lowest error for each of the functionals in this class. Within the hybrid-GGA class, no functional's performance is particularly good. Within the GGA family, the P86 correlation term should be avoided as these data suggest that it is generally ill-suited for HOF calculations. Four functionals, PBELYP, MPWLYP, MPWPW91, and MPWPBE, perform very well. With the exception of MPWP86, all functionals containing the MPW exchange perform well. MPWLYP/3-21G* gives the most accurate results of the entire class with an average unsigned error of 5.66 kcal/mol, which is remarkable for such an inexpensive method. The accuracy of this method surpasses that of many of the more expensive methods included in this study. As seen in Figure 10, the 6-31G* basis generally produces the highest average errors with a few exceptions.

As seen in Figure 11, the Dunning-style correlation-consistent basis sets yield a smaller range of errors than the Pople-style bases. This may be due to the fact that the cc-

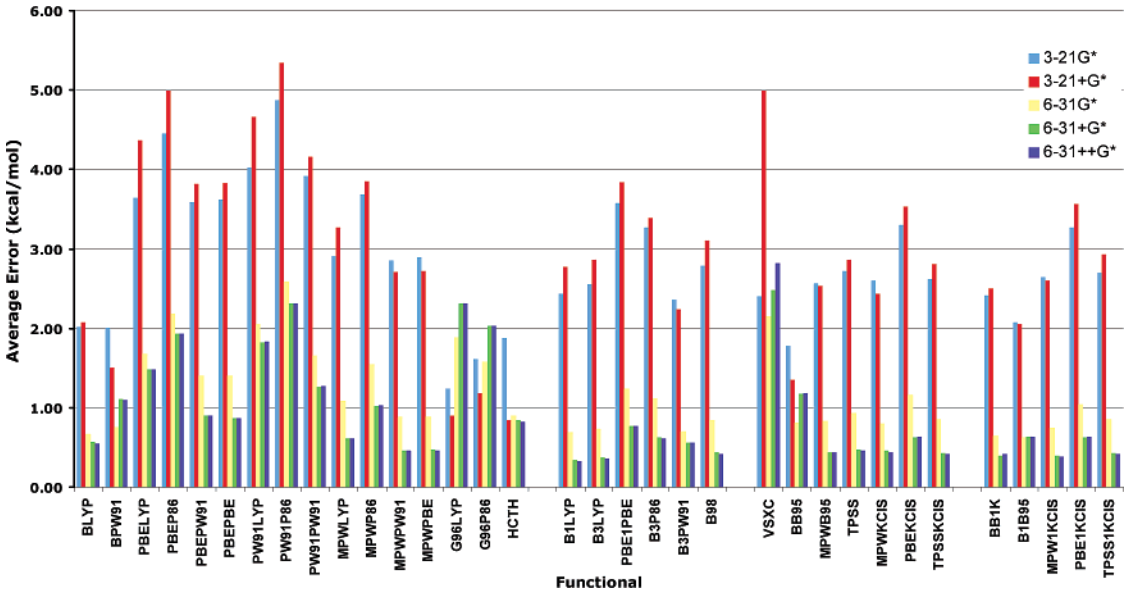


Figure 12. Average unsigned hydrogen-bond interaction energy errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople-type basis sets.

pVDZ basis, the smallest Dunning-style basis used, is still quite large and is more accurate than the 3-21G* and 3-21+G* bases. As discussed previously, enthalpy calculations were preformed at the same functional/basis combination as the geometry optimization of each molecule, the exception being aug-cc-pVTZ, for which single-point enthalpy calculations were performed at the TPSS1KCIS/aug-cc-pVDZ geometries. The lowest average error for all Dunning-style bases is obtained with B3PW91/aug-cc-pVTZ at 3.95 kcal/mol. Within each class of functional, there is a mixture of accurate and inaccurate methods.

Within the hybrid-meta-GGA class, MPW1KCIS is the most accurate, with the MPW1KCIS/cc-pVTZ method producing an average error of 3.97 kcal/mol. Augmentation of the bases with diffuse functions tends to reduce the accuracy of methods in this functional class. Once again, the meta-GGA functionals prove to be the second-most accurate DFT methods for HOF calculation. On the whole, the TPSSKCIS and TPSS methods produce the best results among meta-GGA functionals. However, VSXC/aug-cc-pVDZ is the most accurate combination in the class with an average unsigned error of 3.98 kcal/mol. The use of DZ versus TZ bases does not seem significant within this class, as the TZ bases produce the largest errors in three of the seven functionals. The hybrid-GGA class of functionals reveals the same trends with the Dunning-type bases as with the Pople bases. PBE1PBE is the most accurate functional, and B3PW91/aug-cc-pVTZ produces the lowest error at 3.95 kcal/mol. Within this class, the expansion of the basis set from DZ to TZ does not enhance the accuracy of the HOF calculations as the TZ bases produce nearly equivalent errors for all functionals in the set. The behavior of the GGA class of functionals with the Dunning-type basis sets is again similar to that of the Pople bases. Methods containing the P86 correlation term are again very poor at predicting heats of formation, while those containing the MPW exchange term are more accurate. BPW91 also performs well compared to the rest of the functionals in this group. The most accurate method within this class is HCTH/aug-cc-pVDZ, which produces an average error of 6.83 kcal/mol. A total of 12 of the 16 functionals in the class show a decrease in accuracy with the addition of diffuse functions. Again, there is little difference in values obtained with DZ bases as opposed to TZ methods, as most functionals show only a slight increase in accuracy when using the TZ bases instead of the DZ sets.

Of the LSDA methods, c-SVWN5 performs notably well at predicting heats of formation. c-SVWN5/3-21+G* yields an average unsigned error of 9.44 kcal/mol, the best value in this group. In terms of heat of formation, c-SVWN5 is 4−13 times more accurate than other LSDA methods. Other LSDA methods do not accurately predict HOF.

3.7. Hydrogen-Bonding Interaction Energies. Figures 12 and 13 give the average hydrogen-bonding interaction energy unsigned errors for gradient-corrected density functional methods along with Pople- and Dunning-type basis sets, respectively. Table 4 gives the average unsigned hydrogen-bond interaction errors for HF, MP2, and LSDA functional methods. Overall, the best result is obtained with MP2/aug-cc-pVDZ with an average error of 0.25 kcal/mol. The best result among density functional methods is 0.31 kcal/mol as obtained by MPWLYP/aug-cc-pVTZ. The largest overall error of 10.26 kcal/mol is obtained by the c-SVWN5/aug-cc-pVDZ method.

Not surprisingly, the small Pople-type basis sets, 3-21G* and 3-21+G*, generally yield poor results in terms of hydrogen bonding. For most functionals, the errors obtained with these small bases are greater than 2.00 kcal/mol. Some notable examples of small basis methods that perform fairly well are HCTH/3-21+G* (0.84 kcal/mol), G96LYP/3-21+G* (0.90 kcal/mol), and MP2/3-21G* (0.85 kcal/mol). The best result for these small basis sets combined with one of the LSDA methods, which are very computationally inexpensive, is 7.99 kcal/mol as calculated using c-SVWN5/3-21G*.

As one might expect, the 6-31+G* and 6-31++G* basis sets, which contain diffuse functions, generally outperform

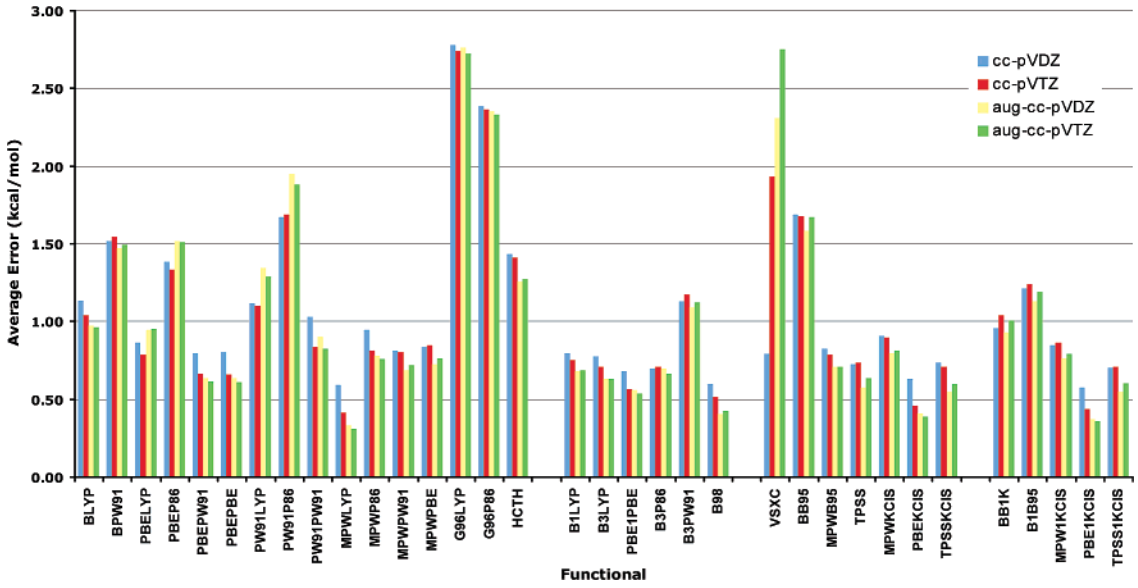


Figure 13. Average unsigned hydrogen-bond interaction energy errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning-type basis sets.

Table 4. Average Unsigned Hydrogen-Bond Interaction Energy, Conformational Energy, and Reaction Barrier Height Errors for the HF, MP2, and LSDA Methods[a]

Hydrogen-Bond Interaction Energy

method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	1.13	1.60	0.91	1.07	1.08	1.56	1.77	1.68	1.73
MP2	0.85	1.14	0.49	0.29	0.28	1.29	0.42	0.25	0.30
SVWN5	9.68	10.24	6.48	6.20	6.21	5.49	5.76	6.04	5.97
SPL	9.66	10.24	6.48	6.21	6.21	5.48	5.76	6.04	5.97
c-SVWN5	7.99	8.66	5.13	4.85	4.85	9.90	10.06	10.26	10.18

Conformational Energy

method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	27.6	38.7	24.2	22.1	22.2	27.7	20.3	22.0	21.0
MP2	30.8	21.7	18.8	16.8	15.5	19.8	7.4	6.8	8.9
SVWN5	51.9	39.1	19.5	15.6	15.4	17.1	9.9	8.8	11.6
SPL	52.1	39.3	19.7	15.6	15.5	17.4	10.1	8.9	11.6
c-SVWN5	51.2	34.4	17.4	17.0	17.0	17.8	10.2	10.0	15.6

Small Reaction Radical Barrier Height

method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	10.79	10.78	12.49	12.84	12.78	11.49	12.51	12.08	13.10
MP2	6.70	5.66	6.53	6.63	6.45	3.46	3.35	2.98	3.14
SVWN5	21.95	19.83	17.73	16.70	16.73	19.31	17.65	18.16	17.12
SPL	21.94	19.75	17.71	16.75	16.79	19.26	17.61	18.20	17.26
c-SVWN5	18.82	16.38	14.18	14.18	13.23	16.05	14.18	14.70	13.46

Large Reaction Singlet Barrier Height

method	3-21G*	3-21+G*	6-31G*	6-31+G*	6-31++G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ	aug-cc-pVTZ
HF	7.79	8.69	13.93	13.89	13.89	13.73	14.32	13.39	14.39
MP2	5.68	5.41	5.18	5.28	7.07	7.71	8.80		
SVWN5	19.86	17.81	12.61	12.04	12.08	12.01	11.83	12.54	12.21
SPL	19.82	17.78	12.59	12.02	12.05	11.99	11.75	12.49	11.83
c-SVWN5	14.11	13.25	13.01	11.31	11.35	11.22	11.93	11.74	12.33

[a] All values in kcal/mol.

6-31G* in terms of hydrogen bonding; there are seven methods for which this is not the case; these are HF, BPW91, G96LYP, G96P86, VSXC, BB95, and B1B95. Somewhat surprisingly, there is typically only a small advantage to using

the 6-31++G* basis set, which incorporates diffuse functions for hydrogen atoms, as compared to the 6-31+G* basis set.

For the large Pople-type basis sets, the MP2 method performs very well with average unsigned binding energies of 0.28 and 0.29 kcal/mol with 6-31++G* and 6-31+G*, respectively (these values represent the second- and third-best overall results). Hartree−Fock performs fairly well with these basis sets with a best value of 0.91 kcal/mol when combined with 6-31G*.

The LSDA functionals perform poorly for hydrogen bonding when combined with the large Pople-type basis sets. The SVWN5 and SPL functionals both yield errors greater than 6.00 kcal/mol with these bases. The c-SVWN5 functional, which gives results that are substantially better than those of the other two LSDA methods, still only yields a best result of 4.85 kcal/mol (with both the 6-31+G* and 6-31++G* basis sets).

There is a great deal of variation in the hydrogen-bonding results obtained with the GGA functionals. The lowest interaction energy error of 0.46 kcal/mol is obtained with the MPWPW91 functional combined with both the 6-31+G* and 6-31++G* basis sets. The highest error of 2.59 kcal/mol is given by PW91P86/6-31G*. Other noteworthy methods in this class are MPWPBE/(6-31+G*,6-31++G*) (0.47 kcal/mol) and BLYP/6-31++G* (0.55 kcal/mol). It is interesting to note that, generally, functionals containing the P86 correlation functional perform poorly while functionals containing the MPW exchange functional perform fairly well when used along with the large Pople-type basis sets. The MPWP86 functional performs moderately well with an average error of 1.03 kcal/mol for MPWP86/6-31++G*.

For the large Pople-type basis sets, the best result among hybrid-GGA methods is 0.33 kcal/mol as calculated with the B1LYP/6-31++G* method; it should also be noted that this is the best overall result for these basis sets among density functional methods. B1LYP/6-31+G* gives a slightly higher average unsigned interaction energy of 0.34 kcal/mol, while B3LYP also performs well with average errors of 0.36 and 0.38 kcal/mol with 6-31+G* and 6-31++G*, respectively.

Among the meta-GGA methods, the lowest interaction energy error of 0.42 kcal/mol is obtained with the TPSS1KCIS/6-31++G* method. The VSXC functional performs very poorly compared to the other meta-GGA functionals (indeed, it performs poorly compared to most gradient-corrected functionals). It is interesting to note that four of the seven functionals in this class obtain errors lower than 0.50 kcal/mol when combined with the 6-31+G* and 6-31++G* basis sets; these functionals are MPWB95, TPSS, MPWKCIS, and TPSSKCIS.

Each of the five hybrid-meta-GGA functionals performs quite well for hydrogen-bond interaction energies when paired with the 6-31+G* and 6-31++G* basis sets, with no method obtaining average errors larger than 1.00 kcal/mol. The best result in this class is 0.38 kcal/mol and is given by the MPW1KCIS/6-31++G* method. Other noteworthy methods are BB1K/6-31+G* (0.40 kcal/mol), BB1K/6-31++G* (0.41 kcal/mol), MPW1KCIS/6-31++G* (0.42 kcal/mol), and TPSS1KCIS/6-31+G* (0.42 kcal/mol).

As in the case of the Pople-type basis sets, the Dunning-type basis sets that contain diffuse functions, aug-cc-pVDZ and aug-cc-pVTZ, yield better hydrogen-bond interaction energies than the ones that do not, cc-pVDZ and cc-pVTZ, for a majority of the functionals considered in this work. Generally speaking, the cc-pVTZ functional outperforms the smaller cc-pVDZ basis set for hydrogen bonding; it should be noted that this is not the case for Hartree−Fock or any of the LSDA functionals. The aug-cc-pVTZ basis set typically outperforms the aug-cc-pVDZ basis set for LSDA, GGA, and hybrid-GGA functionals, while the smaller basis, aug-cc-pVDZ, yields better results when combined with the meta-GGA and hybrid-meta-GGA functionals.

The Hartree−Fock method yields fairly large errors with the Dunning-type basis sets, with the lowest unsigned error being 1.65 kcal/mol for HF/cc-pVDZ and the highest being 1.77 kcal/mol for HF/cc-pVTZ. These values are significantly higher than those obtained with the large Pople-type basis sets. The MP2 method performs very well with most Dunning-type basis sets; MP2/aug-cc-pVDZ produces an average unsigned error of 0.25 kcal/mol, which is the best value obtained for hydrogen-bond interaction energies obtained in this work.

Once again, the LSDA functionals perform very poorly compared to the other DFT methods. Among these methods, the SPL and SVWN5 functionals generally yield results that are almost identical for all Dunning-type basis sets. The c-SVWN5 method yields unsigned errors that are significantly higher than those of SPL and SVWN5. The best LSDA result of 5.48 kcal/mol is obtained with SPL/cc-pVDZ. The worst LSDA result is given by c-SVWN5/aug-cc-pVDZ with a value of 10.26 kcal/mol.

Among the GGA methods, the Dunning-type basis sets outperform the large Pople-type bases for 9 of the 16 functionals. The MPWLYP functional performs significantly better than all other functionals in this class, with the best value of 0.31 kcal/mol given by MPWLYP/aug-cc-pVTZ. The highest unsigned error of 2.78 kcal/mol is obtained with the G96LYP/cc-pVDZ method. Both functionals containing the G96 correlation functional, G96LYP and G96P86, give very poor results with errors that are greater than 2.00 kcal/mol for all Dunning-type basis sets.

For the hybrid-GGA methods, four of the six functionals, B1LYP, B3LYP, PBE1PBE, and B3P86, obtain errors that are between 0.50 and 0.80 kcal/mol; the remaining functional, B3PW91, does not perform as well, producing errors that are above 1.00 kcal/mol for all basis sets. For all Dunning-type basis sets, the B98 functional produces the best hydrogen-bonding results. The best result within this class is given by the B98/aug-cc-pVDZ method, with a value of 0.40 kcal/mol.

Among the meta-GGA methods, PBEKCIS stands out as being notably better than all other functionals. Values of 0.32 and 0.34 kcal/mol are obtained when PBEKCIS is combined with the aug-cc-pVDZ and aug-cc-pVTZ basis sets, respectively; these are the lowest unsigned errors within this class. The VSXC and BB95 functionals both perform poorly.

Within the hybrid-meta-GGA class of functionals, PBE1KCIS yields the lowest errors for hydrogen-bonding

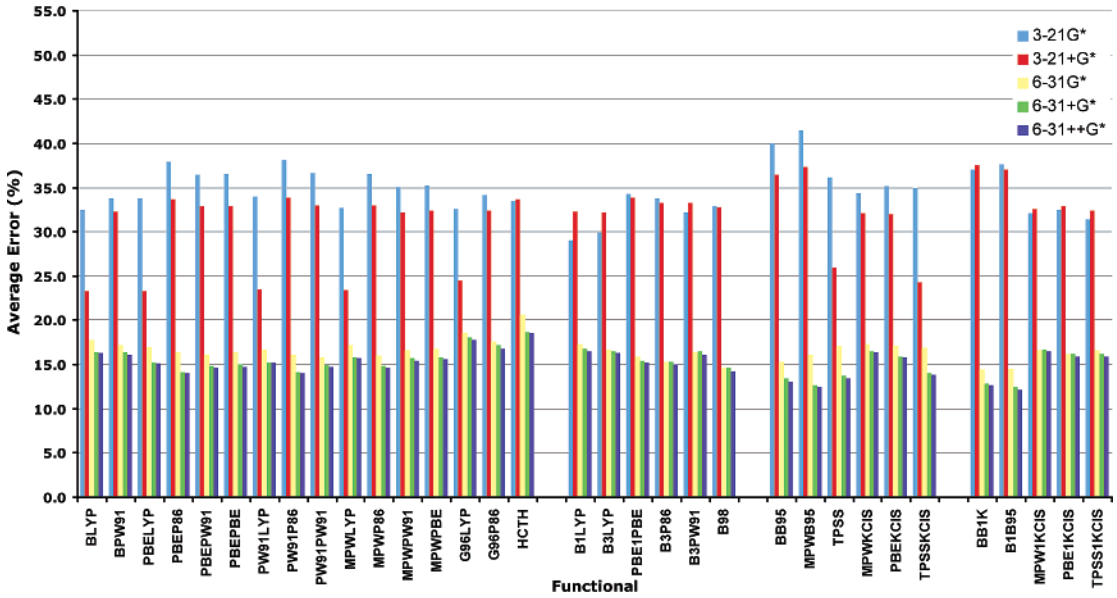


Figure 14. Average unsigned conformational energy errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople-type basis sets.

interaction energies for all Dunning-type basis sets. The best result in this class is obtained with PBE1KCIS/aug-cc-pVDZ with a value of 0.36 kcal/mol. The next-best functional in this class is TPSS1KCIS, whose lowest error is 0.55 kcal/mol at the TPSS1KCIS/aug-cc-pVDZ level.

3.8. Conformational Energies. The average unsigned conformational energy errors are given in Figures 14 and 15 and in Table 4. There are great differences in the conformational energies of the systems considered here; for example, the experimental difference in energy between the orthogonal and planar conformers of ethylene is 65.0 kcal/mol, whereas the experimental value for the conformational energy for the anti and eclipsed forms of methanol is 1.1 kcal/mol. For this reason, the conformational energies are reported in percent error, that is

$$\text{error}_{\%} = \frac{\Delta E_{\text{exp}} - \Delta E_{\text{theory}}}{\Delta E_{\text{exp}}} \times 100$$

Overall, the best result of 6.8% is obtained with the MP2/aug-cc-pVDZ method. The best result among density functional methods is 7.9% as calculated using MPWB95/cc-pVTZ. The worst conformational energy error is that of VSXC/3-21G* with a value of 81.9%.

As seen in Figure 14 and Table 4, the small Pople-type basis sets, 3-21G* and 3-21+G*, give conformational energy errors that are typically much greater than those of the larger Pople-type basis sets, 6-31G*, 6-31+G*, and 6-31++G*. Generally, 3-21+G* outperforms 3-21G*; there are several exceptions to this rule in the hybrid-GGA and hybrid-meta-GGA classes of functionals; also, 3-21G* yields slightly lower errors than 3-21+G* for the GGA functional HCTH. For these small basis sets, the LSDA method produces conformational energies that are significantly worse than those of the gradient-corrected density functional methods. The lowest unsigned error for small Pople-type basis sets is obtained with the MP2/3-21+G* method with a value of 21.7%; for DFT methods, the best value of 23.3% is obtained with the PBELYP/3-21+G* method.

For the Pople-type basis sets, the best conformational energy results can be found within the meta-GGA and hybrid-meta-GGA functional classes. The best overall result of 12.2% is obtained with the hybrid-meta-GGA B1B95/6-31++G* method. It should be noted that, although the BB1K and B1B95 methods perform very well, the remaining three functionals in the hybrid-meta-GGA class, MPW1KCIS, PBE1KCIS, and TPSS1KCIS, yield errors that are about 2−4% higher. Within the meta-GGA group of functionals, BB95, MPWB95, and TPSS all yield very low conformational energy errors. The lowest unsigned error in this class is produced by the MPWB95/6-31++G* method with a value of 12.4%. Among the hybrid-GGA functionals, B98 obtains errors that are about 1% lower than those of the next-best functional, B3P86. The lowest error in this class is obtained at the B98/6-31++G* level with an average unsigned error of 14.2%. Two GGA functionals, PBEP86 and PW91P86, produce the best results within their class; both yield an error value of 14.0% when paired with the 6-31++G* basis set. Among the LSDA functionals, SPL and SVNWV both yield the same error values of 15.6% and 15.5% when paired with the 6-31+G* and 6-31++G* bases, respectively. Hartree−Fock generates errors that are significantly higher than those obtained by most DFT methods; the best error value of 22.1% is obtained with 6-31+G*. The MP2 method obtains errors of 15.5% when paired with 6-31++G* and 16.8% with the 6-31+G* basis set.

The basis sets that include diffuse functions, 6-31+G* and 6-31++G*, generally give unsigned errors that are substantially lower than those obtained using the 6-31G* basis function. It is also interesting to note that 6-31++G* outperforms 6-31+G* for most of the functionals considered here.

Figure 15 and Table 4 give the conformational energy unsigned errors for the Dunning-type basis sets. Here, it can

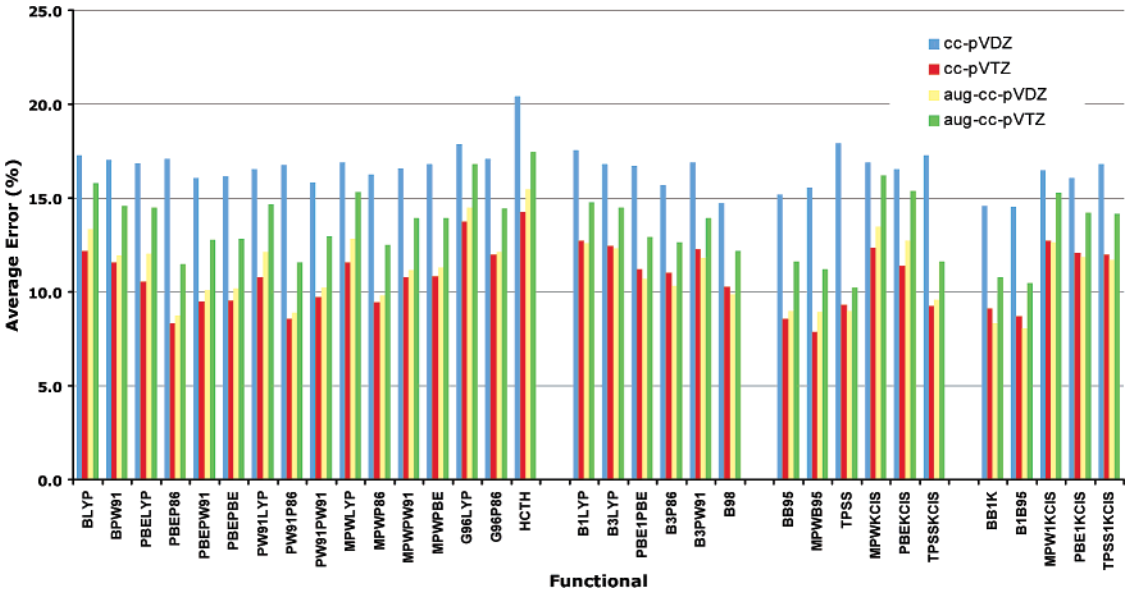


Figure 15. Average unsigned conformational energy errors for GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning-type basis sets.

be seen that there is no class of functional that stands out as being substantially more accurate than another. Some of the lowest unsigned errors are obtained with BB1K and B1B95 (hybrid-meta-GGA), BB95 and MPWB95 (meta-GGA), and PBEP86 and PW91P86 (GGA). The B98 functional produces the best hybrid-GGA results, which are not quite as good as the best results obtained by other DFT methods. It is also interesting to note that each of the LSDA methods studied here yields results that are competitive with many of those obtained with the more sophisticated gradient-corrected techniques. Among all density functional methods considered in this work, the lowest unsigned error obtained for this property is 7.9% as calculated using the MPWB95/cc-pVTZ method. Once again, the VSXC functional (meta-GGA) performs very poorly for describing conformational energies. The MP2 method also yields very good results for all Dunning-type basis sets except for cc-pVDZ. The best overall conformational energy result obtained in this study is 6.8% and is given by the MP2/aug-cc-pVDZ method. Hartree−Fock produces errors that are significantly higher than those of most DFT techniques.

Among the Dunning-type basis sets, aug-cc-pVDZ and cc-pVTZ tend to yield the lowest errors. The aug-cc-pVDZ basis set gives the best results for all of the hybrid-GGA functionals, all of the LSDA functionals, and all of the hybrid-meta-GGA functionals except MPW1KCIS. The cc-pVTZ basis set yields the lowest unsigned errors for all of the GGA functionals and for all of the meta-GGA functionals except VSXC and MPWKCIS. The cc-pVDZ basis set produces the largest errors among Dunning-type basis sets for each of the computational techniques employed in this study with the exception of VSXC.

3.9 Barrier Heights. a. Barrier Heights for Reactions of Small Systems with Radical Transition States (SRBH). Figures 16 and 17 give the average unsigned barrier height errors of the SRBH systems for gradient-corrected functionals along with the Pople- and Dunning-type basis sets, respectively. Table 4 gives the SRBH barrier height errors for the HF, MP2, and LSDA functional methods along with all basis sets considered in this work. Overall, the best result is obtained with the BB1K/aug-cc-pVTZ method with an average unsigned error of 1.05 kcal/mol. The highest error, 21.95 kcal/mol, is produced with the SVWN5/3-21G* functional/basis combination. Again, we would like to point out that these barrier heights are based on single-point calculations at geometries determined at the QCISD/MG3 level of theory.

Inspection of these data reveals that the DFT methods that include exact exchange, that is, the hybrid-GGA and hybrid-meta-GGA methods, generally yield the lowest barrier height errors. The LSDA methods, which are based solely on the electron density, produce the largest unsigned errors.

The LSDA, GGA, and meta-GGA methods perform poorly for SRBH barrier heights. Each of the LSDA methods produces errors larger than 12 kcal/mol for all basis sets. Of the GGA functionals, only HCTH yields errors smaller than 6 kcal/mol. The best result in this class is obtained with the HCTH/6-31++G* method with an average unsigned error of 4.86 kcal/mol. Among the meta-GGA methods, only the VSXC functional obtains errors smaller than 6 kcal/mol. The smallest error in this class is 4.24 kcal/mol and is given by the VSXC/6-31++G* method.

Among the hybrid-GGA functionals, B1LYP yields the smallest errors for all basis sets; this functional produces its lowest error of 3.11 kcal/mol when paired with the aug-cc-pVTZ basis set. It should be noted that B1LYP/6-31++G* gives a slightly higher error of 3.23 kcal/mol. In the hybrid-meta-GGA class, the BB1K functional stands out as clearly being the best performer; indeed, for each basis set, this functional produces the best results among all methods considered in this work. The lowest error in this class is obtained with the BB1K/aug-cc-pVTZ method with a value of 1.05 kcal/mol. The next-best functional for the calculation of these SRBH barrier heights is B1B95, which produces the second-best results among all methods studied here. The

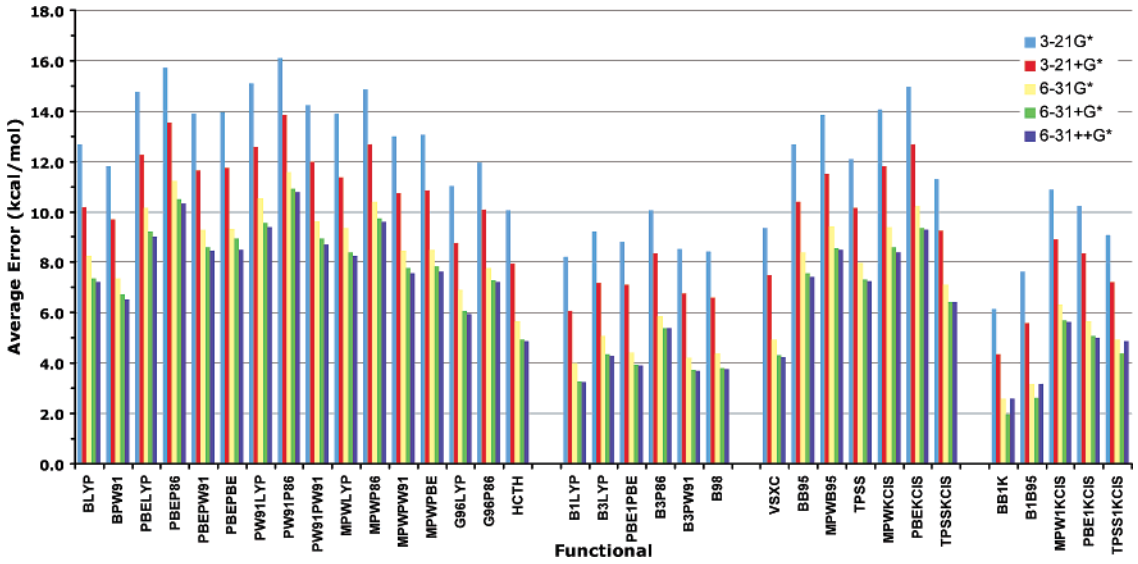


Figure 16. Average unsigned barrier height energy errors for small radical transition-state reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople-type basis sets.

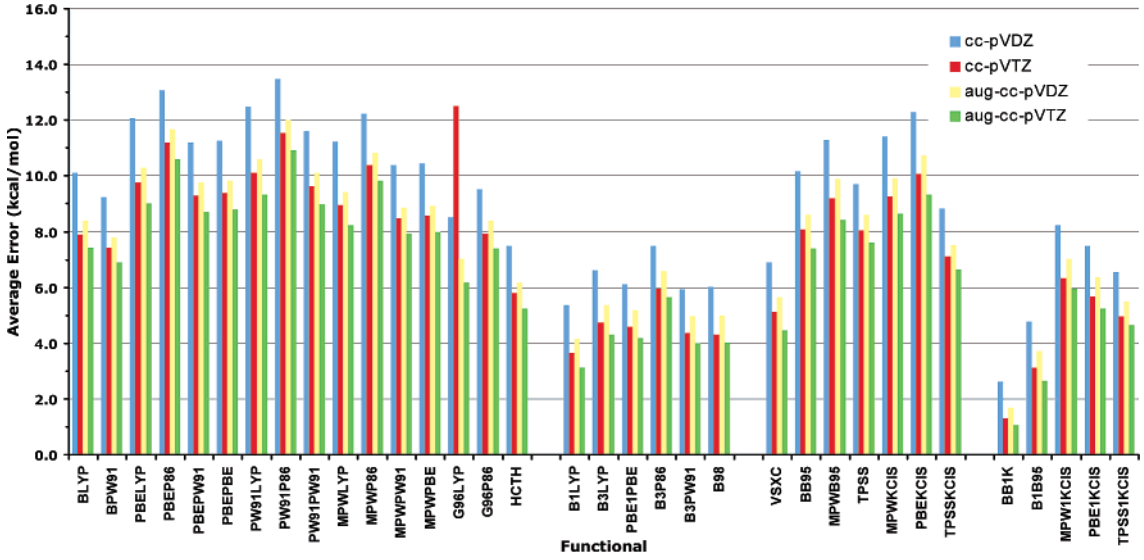


Figure 17. Average unsigned barrier height energy errors for small radical transition-state reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning-type basis sets.

lowest error given by this functional is 2.64 kcal/mol as calculated using the 6-31G* basis set.

The Hartree−Fock method performs very poorly in describing radical transition-state barrier heights; the lowest unsigned error attained with this technique is 10.78 kcal/mol, with the 3-21+G* basis set. MP2 yields fairly good results when paired with the Dunning-type basis sets but, when paired with the Pople-type basis sets, produces much larger errors. The lowest unsigned error attained with this method is 2.98 kcal/mol at the MP2/aug-cc-pVDZ level.

b. Barrier Heights for Reactions of Large Systems with Singlet Transition States (LSBH). Figures 18 and 19 and Table 4 show the reaction barrier heights for the six reactions listed in the LSBH test set. Transition-state barrier heights in this study are calculated as the difference between the temperature-corrected total enthalpy of the transition state and that of the reactants. All structures have been fully optimized at each functional/basis set combination. The values listed in the following tables are average values of the error in transition-state barrier height over all six reactions considered. These reactions include (1) the Diels−Alder reaction of butadiene and ethene forming cyclohexene, (2) the Cope rearrangement of 1,5 hexadiene, (3) the Claisen rearrangement of allyl vinyl ether to pentenal, (4) the electrocyclic rearrangement of cyclobutene to butadiene, (5) the 1,5-sigmatropic shift of 2,4 pentanedione, and (6) the 1,5-sigmatropic shift of 1,3-pentadiene.

Overall, the functional that provides the lowest average error over all six reactions for both Pople and Dunning basis sets is B1LYP. The average error for this functional is 2.63 kcal/mol for the 6-31++G* basis and 2.58 kcal/mol for the aug-cc-pVTZ basis. Generally, a marked improvement in accuracy is observed between basis sets for each functional. The 3-21G* and 3-21+G* basis sets are less accurate than the larger Pople-type bases by 3−4 kcal/mol, while the triple-ζ Dunning-style basis sets are more accurate than their

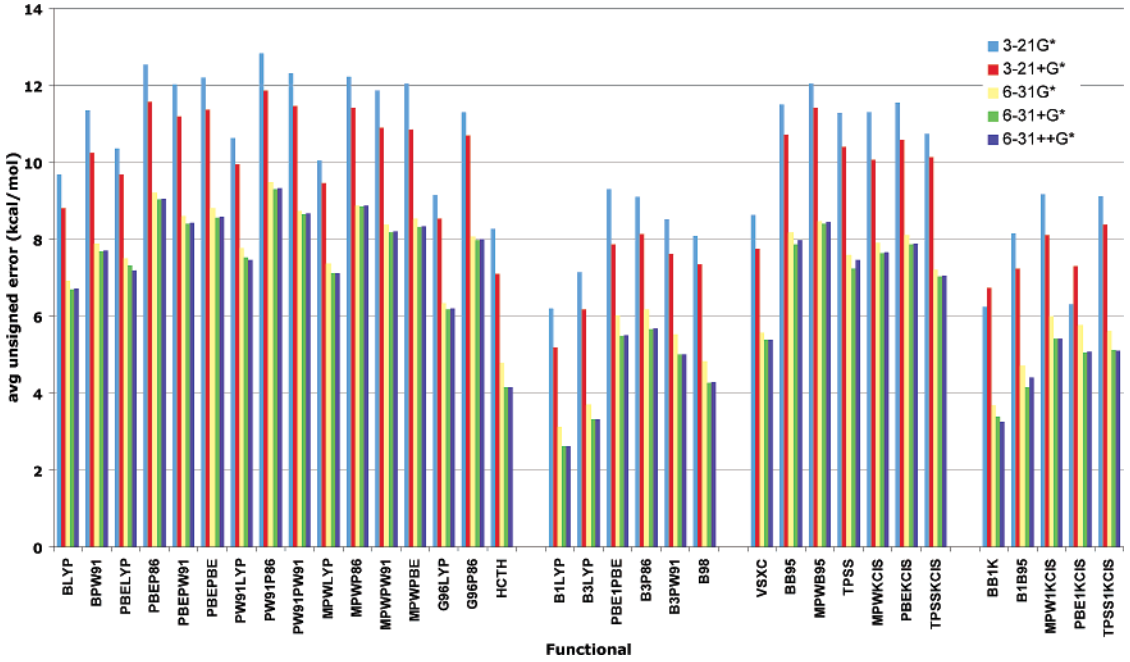


Figure 18. Average unsigned barrier height energy errors for large singlet transition-state reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Pople-type basis sets.

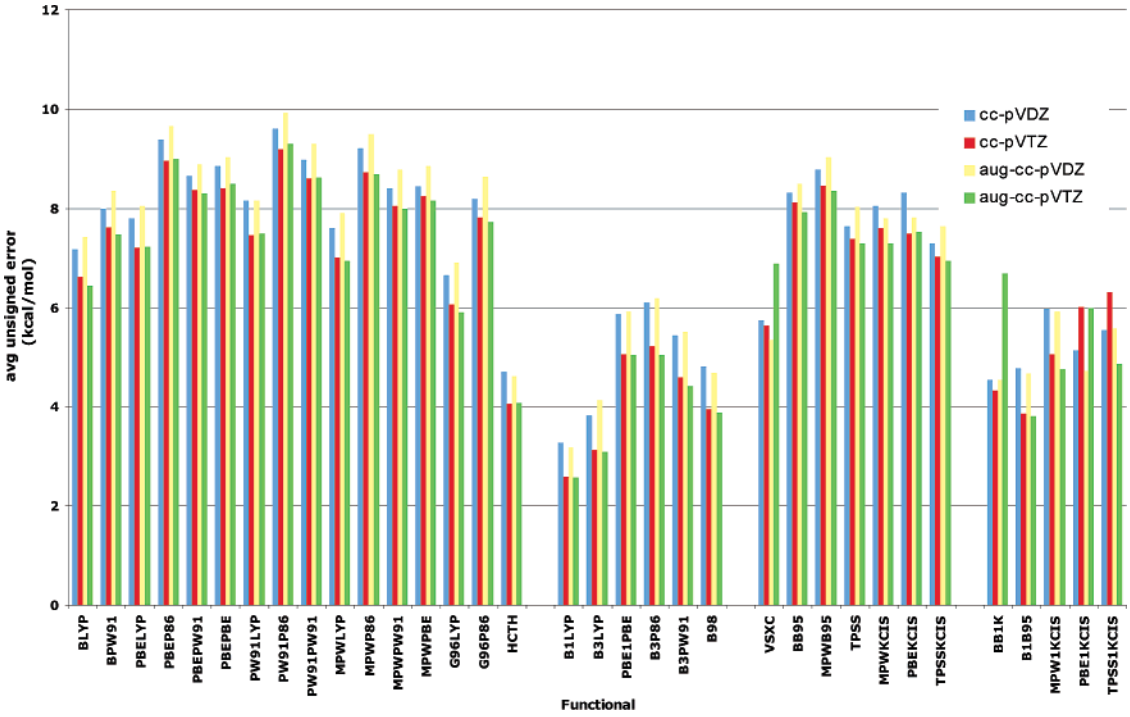


Figure 19. Average unsigned barrier height energy errors for large singlet transition-state reactions along with the GGA, hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals along with Dunning-type basis sets.

double-ζ counterparts by nearly 0.5 kcal/mol. The hybrid-GGA and hybrid-meta GGA functional classes perform markedly better for predicting barrier heights than the LSDA, GGA, and meta-GGA classes. This result indicates a trend that is the opposite of that observed for frequency calculations, for which functionals that include the Hartree−Fock exact exchange perform worse than those without a DFT "exact exchange" term. Since frequency calculations must be performed for transition-state optimizations, this result is somewhat surprising. Moreover, on its own, the HF method is more accurate than most DFT methods at predicting barrier heights when the 3-21G* and 3-21+G* basis sets are used. MP2 also performs well with the lower basis sets. In fact, for HF, basis sets larger than 3-21G* produce errors nearly twice as large as those given by the smallest bases.

Among the LSDA functionals, the c-SVWN5 functional gives the greatest accuracy, while SPL is slightly less accurate. The average barrier height error for c-SVWN5/cc-pVDZ is 11.22 kcal/mol. Typically, average errors within the LSDA class are near 12 kcal/mol except for the lower Pople-style basis sets, which returned errors from 18 to 20 kcal/mol for the SPL and SVWN5 functionals.

Within the GGA class, most functionals yield similar results, while the HCTH functional clearly returns the most accurate results. As mentioned, the accuracy of the barrier height calculations is highly basis-set-dependent for the Pople-type basis sets. HCTH/6-31+G* and HCTH/6-31++G* both yield an average error of 4.15 kcal/mol over all six reactions, while HCTH/cc-pVTZ produced an average error of 4.05 kcal/mol. For the class as a whole, errors for the 3-21G* and 3-21+G* basis sets average 9−13 kcal/mol, while errors for the larger Pople bases average 6−8 kcal/mol. The Dunning basis sets provide accuracy equivalent to the high-level Pople sets. Similar results are obtained by the meta-GGA class, in which VSXC is by far the most accurate. Once again, the lower Pople basis sets yield average errors of 8−11 kcal/mol, while the larger Pople sets and the correlation-consistent sets give average errors of 7−8 kcal/mol. The VSXC functional consistently yields lower errors than the other functionals in this class.

Hybrid-GGA methods perform better than either the GGA or meta-GGA methods, with B1LYP proving to be the most accurate functional tested. Again, a large dependence on the basis is observed with the Pople-type basis sets as the larger basis sets are much more accurate than 3-21G* and 3-21+G*. TZ Dunning-type sets are slightly more accurate than the DZ sets. B1LYP/6-31++G* and B1LYP/aug-cc-pVTZ are the most accurate functional/basis combinations in the entire test set, producing average errors of 2.63 and 2.58 kcal/mol, respectively. B3LYP also provides very accurate calculations for the barrier height test set. On the whole, the hybrid-GGA and hybrid-meta-GGA classes provide similar accuracy. Of the hybrid-meta-GGA class, BB1K yields the lowest average errors with the Pople-type bases, while B1B95 performs better when the Dunning-style sets are employed.

4. Conclusions

In terms of geometric parameters, hybrid-GGA and hybrid-meta-GGA generally yield the best results for both bond lengths and bond angles. The LSDA functionals generally do not perform as well as the more sophisticated functionals. The choice of basis set has a large impact on the quality of calculated geometric parameters. In terms of bond lengths, the large Pople-type basis sets, 6-31G*, 6-31+G*, and 6-31++G*, generally perform similarly to or better than the much larger (and more expensive) cc-pVDZ and aug-cc-pVDZ basis sets for all gradient-corrected functionals. For bond angles, the Dunning-type basis sets generally yield the best results. The largest of these bases, aug-cc-pVQZ, generally obtains the lowest bond angle errors for all DFT functional classes. The large Pople-type basis sets that incorporate diffuse functions typically yield bond angles that are only slightly less accurate than those obtained with the aug-cc-pVDZ and cc-pVTZ basis sets. For most functionals, 6-31++G* produces bond angle errors that are only 0.01−0.05° higher than those of aug-cc-pVDZ.

The methods that include DFT "exact exchange" perform very poorly for calculating the vibrational frequencies of molecules. For large Pople- and Dunning-type basis sets, these methods generally yield unsigned frequency errors that are 1.5−2 times larger than those obtained with methods that do not include exact exchange. For all basis sets, with the exception of 3-21G*, the GGA functionals produce the lowest average frequency errors. For LSDA and GGA functionals, the augmented Pople-type basis sets, 6-31+G* and 6-31++G*, typically produce errors that are slightly lower than those of aug-cc-pVDZ and slightly higher than those of aug-cc-pVTZ. For all functionals, the Pople-type basis sets yield errors that are comparable to the errors computed using all Dunning-type basis sets.

For electron affinities, there is no strong tendency for one functional class to significantly outperform another, with the exception of LSDA, which performs very poorly compared to all other functional groups. It is interesting to note that all functionals containing the P86 correlation functional (GGA and hybrid-GGA) perform very poorly. Functionals incorporating "exact exchange" tend to yield the smallest errors when combined with larger Dunning-type basis sets, while the other functional groups, LSDA, GGA, and meta-GGA, all obtain the most accurate results when used in conjunction with the 6-31+G* and 6-31++G* basis sets.

For ionization potentials, the best results are obtained with the hybrid-meta-GGA functionals. It is very promising, in terms of large-scale calculations, that the ionization potential results obtained with the 6-31+G* and 6-31++G* basis sets are comparable to those obtained using the much larger Dunning-type basis sets for most functionals. As one might expect, the inclusion of diffuse functions in the basis set greatly improves the results for this property.

For heats of formation, the meta-GGA and hybrid-meta-GGA classes of DFT functionals appear to be the most accurate. It is important to note that, in all classes except LSDA, one can find some functional/basis combination that performs well. Overall, the Dunning-style bases are more accurate than the Pople-type sets, with the cc-PVTZ and aug-cc-pVTZ bases yielding the lowest average unsigned errors for our 156-molecule heat of formation test set. However, it should be noted that one can achieve a very high level of accuracy with the MPWLYP/3-21G* method. This combination produces an average error of only 5.6 kcal/mol, which is only 2 kcal/mol less accurate than the best result obtained within the entire study. Within the GGA class of functionals, a wide range of accuracies is obtained.

Generally the hybrid-GGA, meta-GGA, and hybrid-meta-GGA functionals yield the best results for hydrogen-bond interaction energies. There is a large amount of variation among the GGA functionals, with some giving very good results and others performing very poorly. The MP2 method produces some of the lowest hydrogen-bonding interaction energy errors. For both the large Pople-type basis sets and the Dunning-type bases, the addition of diffuse functionals typically produces lower unsigned errors. The inclusion of diffuse functions on hydrogen atoms in the 6-31++G* basis does not generally increase the performance in terms of hydrogen-bonding interaction energies when compared to the 6-31+G* basis. For the large Pople- and Dunning-type bases that include diffuse functions, there is no clear tendency for one particular basis set to consistently produce the lowest errors within the GGA class of functionals; for all of the

Table 5. Rankings of Functional/Basis Set Combinations for All Physical Properties Considered in This Work[a]

rank	bond length		avg. unsigned error (Å)	HOF		(kcal/mol)
1		VSXC/cc-pVQZ	0.0056	1	B3PW91/aug-cc-pVTZ	3.95
2		VSXC/aug-cc-pVQZ	0.0057	2	MPW1kcis/cc-pVTZ	3.97
3		VSXC/aug-cc-pVTZ	0.0061	3	VSXC/cc-pVTZ	3.99
4		VSXC/cc-pVTZ	0.0061	4	MPW1KCIS/aug-cc-pVTZ	4.10
5		TPSS1KCIS/cc-pVTZ	0.0063	5	TPSSTPSS/aug-cc-pVTZ	4.73
1		B1B95/6-31+G*	0.0075	1	TPSSKCIS/6-31+G*	4.76
2		B1B95/6-31++G*	0.0075	2	TPSSTPSS/6-31+G*	4.77
3		B1B95/6-31G*	0.0078	3	B3PW91/6-31G*	4.79
rank	bond angle		(deg)	hydrogen-bond interaction energy		(kcal/mol)
1		BLYP/aug-cc-pVTZ	1.07	1	MPWLYP/aug-cc-pVTZ	0.31
2		PBE1PBE/aug-cc-pVQZ	1.11	2	B1LYP/6-31++G*	0.33
3		B3P86/aug-cc-pVQZ	1.12	3	MPWLYP/aug-cc-pVDZ	0.33
4		PBE1PBE/cc-pVQZ	1.12	4	B1LYP/6-31+G*	0.34
5		B3P86/aug-cc-pVDZ	1.12	5	PBE1KCIS/aug-cc-pVTZ	0.36
1		PBE1PBE/6-31++G*	1.22	1	B1LYP/6-31++G*	0.33
2		PBE1PBE/6-31+G*	1.23	2	B1LYP/6-31+G*	0.34
3		TPSSTPSS/6-31++G*	1.23	3	B3LYP/6-31++G*	0.36
rank	frequencies		(cm^{-1})	conformational energy		(% error)
1		G96LYP/aug-cc-pVTZ	40	1	MPWB95/cc-pVTZ	7.90
2		PW91LYP/cc-pVTZ	40	2	B1B95/aug-cc-pVDZ	8.10
3		BLYP/aug-cc-pVTZ	40	3	BB1K/aug-cc-pVDZ	8.30
4		G96LYP/cc-pVTZ	40	4	PBEP86/cc-pVTZ	8.30
5		MPWLYP/cc-pVTZ	40	5	BB95/cc-pVTZ	8.60
1		PBEP86/6-31+G*	46	1	B1B95/6-31++G*	12.20
2		PBEP86/6-31++G*	46	2	MPWB95/6-31++G*	12.40
3		MPWP86/6-31++G*	46	3	B1B95/6-31+G*	12.50
rank	EA		(kcal/mol)	SRBH		(kcal/mol)
1		MPWB95/6-31++G*	3.08	1	BB1K/aug-cc-pVTZ	1.05
2		MPWB95/6-31+G*	3.12	2	BB1K/cc-pVTZ	1.31
3		B98/aug-cc-pVTZ	3.15	3	BB1K/aug-cc-pVDZ	1.69
4		BB95/6-31++G*	3.35	4	BB1K/6-31+G*	1.95
5		B98/aug-cc-pVDZ	3.42	5	BB1K/6-31++G*	2.58
1		MPWB95/6-31++G*	3.08	1	BB1K/6-31+G*	1.95
2		MPWB95/6-31+G*	3.12	2	BB1K/6-31++G*	2.58
3		BB95/6-31++G*	3.35	3	BB1K/6-31G*	2.60
rank	IP		(kcal/mol)	LSBH		(kcal/mol)
1		B1B95/aug-cc-pVTZ	4.25	1	B1LYP/aug-cc-pVTZ	2.575
2		MPWB95/aug-cc-pVTZ	4.38	2	B1LYP/cc-pvTZ	2.591
3		MPWB95/cc-pVTZ	4.49	3	B1LYP/6-31++G*	2.631
4		MPWB95/6-31++G*	4.50	4	B1LYP/6-31+G*	2.637
5		MPWB95/6-31+G*	4.53	5	B3LYP/aug-cc-pVTZ	3.102
1		MPWB95/6-31++G*	4.50	1	B1LYP/6-31++G*	2.631
2		MPWB95/6-31+G*	4.53	2	B1LYP/6-31+G*	2.637
3		BB95/6-31++G*	4.67	3	B1LYP/6-31G*	3.123

[a] The first five functional/basis combinations include all basis sets, while the group of three functionals under each property shows the highest-ranking methods using the only Pople-type basis sets.

other functional classes, the 6-31+G* and 6-31++G* bases generally give the best results.

In terms of conformational energies, the meta-GGA and hybrid-meta-GGA functionals produce the lowest average errors. Not surprisingly, the large Pople-type basis sets, 6-31G*, 6-31+G*, and 6-31++G*, yield results that are typically about 10% better than those obtained using the smaller Pople-type bases, 3-21G* and 3-21+G*. For the large Pople-type basis sets, there is a slight improvement in the calculated conformational energies when diffuse functionals are employed. Overall, the basis sets that produce the lowest errors are the Dunning-type bases, aug-cc-pVDZ and cc-pVTZ.

Table 6. Average Functional Rankings and Standard Deviations for the "Top 15" Functionals along with 6-31+G* and aug-cc-pVDZ Basis Sets

	6-31+G*	avg. rank	std. dev.
1	B1B95	10.7	11.9
2	B98	11.9	7.5
3	TPSSKCIS	13.6	8.4
4	TPSSTPSS	13.7	8.4
5	PBE1PBE	13.8	10.9
6	B3LYP	13.9	9.0
6	MPWB95	13.9	11.2
8	TPSS1KCIS	14.0	8.2
9	B3PW91	14.2	9.2
9	BB1K	14.2	12.4
11	MPW1KCIS	14.8	10.5
12	MPWPW91	15.8	5.6
13	PBEPW91	16.6	8.5
14	PBE1KCIS	16.7	9.2
15	MPWPBE	16.8	5.3
	aug-cc-pVDZ	avg. rank	std. dev.
1	B98	10.1	8.8
2	B1B95	11.7	12.2
3	TPSS1KCIS	12.0	8.1
4	PBE1PBE	12.2	10.0
5	B3LYP	12.3	9.2
6	PBE1KCIS	12.8	10.9
7	TPSSTPSS	13.3	6.3
8	TPSSKCIS	13.6	7.6
9	B3PW91	13.8	9.0
10	MPWPW91	15.3	5.9
11	MPWPBE	15.7	6.7
12	MPW1KCIS	15.9	9.9
13	BB95	16.2	9.3
13	B1LYP	16.2	13.1
15	BB1K	16.7	13.8

One of the most salient aspects of the data concerning the barrier heights of small molecules with radical transition states (SRBH) is that functionals containing exact exchange terms generally produce the lowest average barrier height errors. The LSDA methods, which depend only on the electron density, produce errors that are significantly higher than those of all other methods considered here. In terms of basis sets, the inclusion of diffuse functions typically increases the accuracy with which the barrier heights of these reactions can be calculated. The lowest barrier height errors are generally produced with the 6-31+G*, 6-31++G*, and aug-cc-pVTZ bases.

As in the case of the SRBH reactions, the barrier heights of larger systems with singlet transition states (LSBH) are generally bette-described by functionals that contain exact exchange. The addition of diffuse functions to the 3-21G*, 6-31G*, and cc-pVTZ basis sets generally results in a lower unsigned average error; in the case of the cc-pVDZ basis set, however, the addition of diffuse functions typically increases the errors slightly. For the LSBH reactions, the 6-31+G*, 6-31++G*, and aug-cc-pVTZ basis sets generally produce the lowest errors for most methods studied in this work.

Here, we will attempt to summarize the results obtained in the entire study and draw some conclusions concerning the functionals that seem to offer the best compromise in terms of describing all of the physical properties investigated in this work. As we have generated a tremendous amount of data in this study, we will limit our discussion by considering only the results obtained by two popular basis sets, 6-31+G* and aug-cc-pVDZ.

One of the most interesting observations that can be made from the data presented here is that, for many physical properties, the large Pople-type basis sets (6-31G*, 6-31+G*, and 6-31++G*) produce results that are comparable to, or superior to, those given by the much larger and computationally expensive Dunning-type basis sets. For example, for the B1B95 functional, the 6-31+G* basis set outperforms the aug-cc-pVDZ basis set for bond distances, heats of formation, hydrogen-bond interaction energies, and reactions' barrier heights (both SRBH and LSBH); the average unsigned bond angle error obtained with the smaller basis set is only 0.034° higher than that of the larger basis, and the average unsigned ionization potential error for 6-31+G* is only 0.28 kcal/mol larger than that of aug-cc-pVDZ. The average unsigned electron affinity, vibrational frequency, and conformational energy errors are larger for 6-31+G* than for aug-cc-pVDZ.

Table 5 indicates the rankings of the top five functional/basis set combinations overall and the top three functional/basis set combinations among Pople-type basis sets for each property considered in this work. In Table 5, it can be seen that, for each physical property considered here, with the exception of conformational energies, the best results obtained with Pople-type basis sets are comparable to the best results produced by the larger Dunning-type bases.

One of the main goals of this survey is to get a rough estimate of a functional's performance in terms of its ability to describe all of the properties considered in this study. In order to accomplish this goal, we compare the average functional ranks and standard deviations for each of the functionals studied in this work. The average functional rank is given as the mean of a functional's rank for all of the properties considered here, and the standard deviation was also calculated.

Table 6 lists the average functional ranks and standard deviations of the 15 functionals with the lowest average ranks for the 6-31+G* and aug-cc-pVDZ basis sets. For both basis sets, there are five hybrid-meta-GGA and three meta-GGA functionals represented in the "top 15". The top 15 of the 6-31+G* basis also included four hybrid-GGA and three GGA functionals, while the top performers from the aug-cc-pVDZ set included five hybrid-GGA and two GGA functionals. In the aug-cc-pVDZ group, each of the "top five" functionals in terms of average functional rank contains "exact-exchange" terms. Whereas, only three of the "top five" of the 6-31+G* set contain an "exact-exchange" term. Also, for both basis sets, the only GGA functional to rank in the top 10 is MPWPW91/aug-cc-PVDZ.

Table 7 lists the 15 best functionals for the 6-31+G* basis set along with their unsigned errors for each of the properties considered in this work; for purposes of comparison, the lowest and highest unsigned errors for each property are given, as well as the mean unsigned error averaged over all of the functionals in this study. For the 6-31+G* basis set,

Table 7. Performance of the "Top 15" Functionals Along with the 6-31+G* Basis Set[a]

	HOF	IP	EA	H bond	freq	length	angle	conf E	SRBH	LSBH
B1B95	9.94	4.81	5.07	0.64	104.1	0.0074	1.23	12.47	2.64	4.16
B98	13.47	5.05	3.83	0.43	88.6	0.0094	1.29	14.59	3.81	4.28
TPSSKCIS	4.76	5.99	4.47	0.43	65.6	0.0135	1.26	14.07	6.41	7.04
TPSSTPSS	4.77	5.52	4.83	0.47	65.8	0.0135	1.23	13.73	7.33	7.25
PBE1PBE	5.94	5.34	5.15	0.77	103.4	0.0079	1.23	15.45	3.92	5.49
B3LYP	14.03	5.29	3.91	0.38	84.4	0.0093	1.29	16.54	4.36	3.32
MPWB95	18.31	4.53	3.12	0.44	48.0	0.0161	1.31	12.66	8.57	8.41
TPSS1KCIS	12.06	5.30	5.07	0.43	79.8	0.0090	1.25	16.23	4.37	5.12
B3PW91	8.32	5.48	4.39	0.56	92.5	0.0081	1.26	16.46	3.73	5.01
BB1K	17.34	5.33	6.40	0.40	138.2	0.0098	1.28	12.79	1.95	3.38
MPW1KCIS	7.28	4.86	3.98	0.40	75.9	0.0101	1.60	16.71	5.71	5.41
MPWPW91	8.57	5.16	3.80	0.46	49.7	0.0157	1.29	15.67	7.78	8.17
PBEPW91	17.18	5.10	3.67	0.89	48.8	0.0163	1.28	14.84	8.62	8.41
PBE1KCIS	15.20	4.95	4.22	0.63	91.0	0.0086	1.57	16.18	5.09	5.07
MPWPBE	8.76	5.16	3.87	0.47	48.9	0.0159	1.29	15.79	7.83	8.32
	TPSSKCIS	MPWB95	MPWB95	B1LYP	G96P86	B1B95	cSVWN5	B1B95	BB1K	B1LYP
lowest err. value	4.76	4.53	3.12	0.34	49	0.007	1.28	12.00	1.95	2.64
	SPL	cSVWN5	cSVWN5	SPL	BB1K	cSVWN5	VSXC	VSXC	SPL	SPL
highest err. value	133.7	19.08	14.8	6.21	142	0.025	1.56	44	16.75	12.04
avg. err.	15.82	5.78	4.54	0.93	66.01	0.014	1.33	16.24	6.74	6.62

[a] Errors given in the following units: bond length (Å), bond angle (degrees), frequency (cm^{-1}), ionization potential (kcal/mol), electron affinity (kcal/mol), heat of formation (kcal/mol), hydrogen-bond interaction energy (kcal/mol), conformational energy (percent error), and reaction barrier height (kcal/mol). Average errors include all 37 density functionals considered in this work.

Table 8. Performance of the "Top 15" Functionals along with the aug-cc-pVDZ Basis Set[a]

	HOF	IP	EA	H bond	freq	length	angle	conf E	SRBH	LSBH
B98	18.38	4.90	3.42	0.40	73.9	0.0114	1.24	9.85	4.98	4.68
B1B95	14.13	4.53	4.54	1.13	88.3	0.0093	1.20	8.06	3.72	4.67
TPSS1KCIS	8.35	5.30	4.64	0.55	68.5	0.0109	1.21	11.70	5.48	5.58
PBE1PBE	8.82	5.27	4.66	0.56	86.5	0.0104	1.17	10.71	5.17	5.92
B3LYP	18.66	5.28	3.78	0.63	70.5	0.0111	1.22	12.34	5.37	4.12
PBE1KCIS	11.63	4.84	3.79	0.37	82.2	0.0109	1.54	11.84	6.37	4.71
TPSSTPSS	8.72	5.53	4.49	0.57	50.9	0.0159	1.23	8.99	8.61	8.02
TPSSKCIS	8.31	5.96	4.13	0.55	52.1	0.0157	1.25	9.58	7.51	7.64
B3PW91	12.02	5.42	4.08	1.09	77.6	0.0104	1.15	11.80	4.96	5.52
MPWPW91	8.58	5.36	4.03	0.69	51.1	0.0177	1.27	11.17	8.85	8.77
MPWPBE	8.91	5.27	3.89	0.72	50.9	0.0180	1.27	11.29	8.92	8.85
MPW1KCIS	12.57	4.78	3.58	0.76	70.9	0.0121	1.58	12.64	7.00	5.92
BB95	9.94	5.00	3.85	1.58	52.6	0.0179	1.31	8.98	8.62	8.50
B1LYP	34.07	5.93	4.63	0.68	77.6	0.0092	1.20	12.61	4.15	3.19
BB1K	21.29	6.50	5.94	0.93	119.3	0.0107	1.23	8.31	1.69	4.54
	PW91LYP	B1B95	B98	PBEKCIS	G96LYP	B1LYP	cSVWN5	B1B95	BB1K	B1LYP
lowest err. value	8.24	4.53	3.43	0.32	48	0.009	1.25	8.1	1.69	3.19
	SPL	cSVWN5	B3P86	cSVWN5	BB1K	BLYP	BLYP	VSXC	SPL	SPL
highest err. value	128.62	19.06	13.99	10.26	119	0.029	1.6	51.5	18.2	12.54
avg. error	16.23	5.77	4.77	1.00	62.2	0.016	1.30	12.33	7.84	7.15

[a] Errors given in the following units: bond length (Å), bond angle (degrees), frequency (cm^{-1}), ionization potential (kcal/mol), electron affinity (kcal/mol), heat of formation (kcal/mol), hydrogen-bond interaction energy (kcal/mol), conformational energy (percent error), and reaction barrier height (kcal/mol). Average errors include all 37 density functionals considered in this work.

the B1B95 functional obtains the lowest average functional rank with a value of 10.7. However, the standard deviation of this functional is fairly high with a value of 11.9, since the method performs very well for some properties and

relatively poorly for others, as can be seen in Table 7. Other functionals that perform notably well are B98, TPSSTPSS, TPSS1KCIS, and PBE1PBE; each of these functionals gives reasonably good results for all of the physical properties here (with the possible exception of vibrational frequencies).

Table 8 lists the 15 best functionals for the aug-cc-pVDZ basis set along with their unsigned errors for each of the properties considered in this work in the same manner as was done for the 6-31+G* basis. For this basis set, there are a number of functionals that perform very well in terms of giving a good description of each of the physical properties in this work. The B98 functional has the lowest average functional ranking with a value of 10.1 (standard deviation = 8.8). When paired with aug-cc-pVDZ, B98 ranks in the top 11 functionals for all properties except HOF and vibrational frequency. B98's predicted heat of formation is in error by an average of 18.38 kcal/mol. TPSS1KCIS, which ranks as third best with the DZ basis, predicts HOF very well but is less accurate for electron affinity, conformational energy, and vibrational frequency. Other functionals of note are B1B95, PBE1PBE, and B3LYP.

Acknowledgment. We thank the NIH (GM066859) for providing the funding for this research and the Pittsburgh Supercomputing Center and the LIONXL cluster at Pennsylvania State University for supplying some of the computational resources for this project.

Supporting Information Available: Many additional tables including average signed errors and singlet and radical errors corresponding to our DFT survey as well as the test sets used for each of the nine properties calculated in this study. This information also includes functional type/basis-set-specific performance for all of the molecular properties addressed in this article. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Adesokan, A. A.; Fredj, E.; Brown, E. C.; Gerber, R. B. *Mol. Phys.* **2005**, *103*, 1505−1520.

(2) Basheerl, M. M.; Perles, C. E.; Volpe, P. L. O.; Airoldi, C. *J. Solution Chem.* **2006**, *35*, 625−637.

(3) Bassan, A.; Borowski, T.; Pelmenschikov, V.; Schofield, C.; Siegbahn, P. *Abstr. Pap.—Am. Chem. Soc.* **2005**, *229*, U800.

(4) Blomberg, M.; Siegbahn, P. *Biochim. Biophys. Acta* **2004**, *1658*, 145.

(5) Chen, B. G.; Zhang, M. Y.; Zhao, Y. Y.; Zhang, J.; Sun, C. C. *Gaodeng Xuexiao Huaxue Xuebao* **2006**, *27*, 1307−1310.

(6) Dannenberg, J. J. *Pept. Solv. H-Bonds* **2006**, *72*, 227.

(7) Dannenberg, J. J.; Asensio, A.; Kobko, N. *Abstr. Pap.—Am. Chem. Soc.* **2003**, *226*, U330.

(8) Dobes, P.; Otyepka, M.; Strnad, M.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 4297−4304.

(9) Dolker, N.; Maseras, F.; Lledos, A. *J. Phys. Chem. B* **2003**, *107*, 306−315.

(10) Dorcier, A.; Dyson, P. J.; Gossens, C.; Rothlisberger, U.; Scopelliti, R.; Tavernelli, I. *Organometallics* **2005**, *24*, 2114−2123.

(11) Ferretti, V.; Pretto, L.; Tabrizi, M. A.; Gilli, P. *Acta Crystallogr., Sect. B* **2006**, *62*, 634−641.

(12) Isaev, A.; Scheiner, S. *J. Phys. Chem. B* **2001**, *105*, 6420−6426.

(13) Otyepka, M.; Sklenovsky, P.; Horinek, D.; Kubar, T.; Hobza, P. *J. Phys. Chem. B* **2006**, *110*, 4423−4429.

(14) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 11139−11149.

(15) Reha, D.; Valdes, H.; Vondrasek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. *Chem.—Eur. J.* **2005**, *11*, 6803−6817.

(16) Rovira, C.; Parrinello, M. *Chem.—Eur. J.* **1999**, *5*, 250−262.

(17) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.

(18) Sulpizi, M.; Folkers, G.; Rothlisberger, U.; Carloni, P.; Scapozza, L. *Quant. Struct.-Act. Relat.* **2002**, *21*, 173−181.

(19) Szyczewski, A.; Pietrzak, J.; Mobius, K. *Acta Phys. Pol., A* **2005**, *108*, 119−126.

(20) Viswanathan, R.; Asensio, A.; Dannenberg, J. J. *J. Phys. Chem. A* **2004**, *108*, 9205−9212.

(21) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14198−14205.

(22) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2005**, *127*, 17216−17223.

(23) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2005**, *127*, 14534−14535.

(24) Xu, L. Z.; Huang, Y. W.; Yu, G. P.; Si, G. D.; Zhu, Q. *Struct. Chem.* **2006**, *17*, 235−239.

(25) Zhou, W. Q.; Yang, W.; Qiu, L. H. *THEOCHEM* **2005**, *130*, 133−141.

(26) Van Alsenoy, C.; Yu, C. H.; Peeters, A.; Martin, J. M. L.; Schafer, L. *J. Phys. Chem. A* **1998**, *102*, 2246−2251.

(27) Bour, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2002**, *65*, 45−59.

(28) Jalkanen, K. J.; Elstner, M.; Suhai, S. *THEOCHEM* **2004**, *675*, 61−77.

(29) Oldfield, E. *Philos. Trans. R. Soc. London, Ser. B* **2005**, *360*, 1347−1361.

(30) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(31) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864−B871.

(32) Perdew, J. P.; Schmidt, K. *AIP Conf. Proc.* **2001**, *577*, 1−20.

(33) Mattsson, A. *Science* **2002**, *298*, 759−760.

(34) Seidl, M.; Perdew, J. P.; Kurth, S. *Phys. Rev. Lett.* **2000**, *84*, 5070−5073.

(35) Seidl, M.; Perdew, J. P.; Kurth, S. *Phys. Rev. A* **2000**, *62*, 012502/012501−01250.2/012515.

(36) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939−947.

(37) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *J. Am. Chem. Soc.* **1982**, *104*, 2797−2803.

(38) Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Defrees, D. J.; Pople, J. A.; Binkley, J. S. *J. Am. Chem. Soc.* **1982**, *104*, 5039−5048.

(39) Hehre, W. J.; Ditchfie, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257−2261.

(40) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(41) Raymond, K. S.; Wheeler, R. A. *J. Comput. Chem.* **1999**, *20*, 207−216.

(42) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

(43) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364−382.

(44) Scuseria, G. E.; Staroverov, V. N. Progress in the Development of Exchange-Correlation Functionals. In *Theory and Applications of Computational Chemistry: The First 40 Years (A Volume of Technical and Historical Perspectives*); Dykstra, C. E., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 669−724.

(45) Wang, N. X.; Wilson, A. K. *J. Chem. Phys.* **2004**, *121*, 7632−7646.

(46) Wang, N. X.; Wilson, A. K. *Mol. Phys.* **2005**, *103*, 345−358.

(47) Bauschlicher, C. W. *Chem. Phys. Lett.* **1995**, *246*, 40−44.

(48) Neugebauer, A.; Häfelinger, G. *THEOCHEM* **2002**, *578*, 229.

(49) Neugebauer, A.; Häfelinger, G. *THEOCHEM* **2002**, *585*, 35−47.

(50) Scheiner, A. C.; Baker, J.; Andzelm, J. W. *J. Comput. Chem.* **1997**, *18*, 775−795.

(51) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1992**, *97*, 7846−7848.

(52) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612−5626.

(53) Riley, K. E.; Brothers, E. N.; Ayers, K. B.; Merz, K. M. *J. Chem. Theory Comput.* **2005**, *1*, 546−553.

(54) Mole, S. J.; Zhou, X.; Liu, R. *J. Phys. Chem.* **1996**, *100*, 14665−14671.

(55) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063−1079.

(56) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42−55.

(57) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374−7383.

(58) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107/124101−12410.7/124112.

(59) Curtiss, L. A.; Redfern, P. C.; Rossolov, V.; Kedziora, Z.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 9287−9295.

(60) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 8996−8999.

(61) Brothers, E. N.; Merz, K. M. *J. Phys. Chem. A* **2004**, *108*, 2904−2911.

(62) Rabuck, A. D.; Scuseria, G. E. *Chem. Phys. Lett.* **1999**, *309*, 450−456.

(63) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029−5036.

(64) Hobza, P.; Sponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315−1325.

(65) Paizs, B.; Suhai, S. *J. Comput. Chem.* **1998**, *19*, 575−584.

(66) Tuma, C.; Bosese, A. D.; Handy, N. C. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3939−3947.

(67) Rappe, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117−6128.

(68) Rabuck, A. D.; Scuseria, G. E. *Ther. Chem. Acc.* **2000**, *104*, 439−444.

(69) Ireta, J.; Neugebauer, J.; Scheffler, M. *J. Phys. Chem. A* **2004**, *108*, 5692−5698.

(70) Xu, X.; Goddard, W. A. *J. Chem. Phys. A* **2004**, *108*, 2305−2313.

(71) Helkier, A.; Klopper, W.; Helgaker, T.; Jorgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157.

(72) Tsuzuki, S.; Lüthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949−3957.

(73) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.

(74) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384−1388.

(75) Stamant, A.; Cornell, W. D.; Kollman, P. A.; Holgren, T. A. *J. Comput. Chem.* **1995**, *16*, 1483.

(76) Scheiner, P.; Fokin, A. A.; Pascal, R. A., Jr.; de Meijere, A. *Org. Lett.* **2006**, *8*, 3635−3638.

(77) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 842−846.

(78) Zhao, Y.; González-Garcia, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012−2018.

(79) Dickson, R. M.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 111101/111101−11110.1/111103.

(80) Dybala-Defratyka, A.; Paneth, P.; Pu, J. Z.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2475−2486.

(81) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem.* A **2000**, *104*, 4811−4815.

(82) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Chesseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Forseman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; AlLoham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 03*; version C.01; Gaussian Inc.: Pittsburgh, PA, 2003.

(83) Ochterski, J. W. Themochemistry in Gaussian. www.gaussian.com (accessed Sept 12, 2006).

(84) NIST. NIST Chemistry Webbook. http://webbook.nist.gov/chemistry/ (accessed Sept 12, 2006).

(85) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.

(86) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.

(87) Halgren, T. a. *J. Comput. Chem.* **1999**, *20*, 730−748.

(88) Freile, M. L.; Risso, S.; Curaqueo, A.; Zamora, M. A.; Enriz, R. D. *THEOCHEM* **2005**, *731*, 107−114.

(89) Zhong, H. Z.; Stewart, E. L.; Kontoyianni, M.; Bowen, J. P. *J. Chem. Theory Comput.* **2005**, *1*, 230−238.

(90) Roothan. *Rev. Mod. Phys.* **1951**, *23*, 69.

(91) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(92) Adamo, C.; Barone, V. J. *Chem. Phys. Lett.* **1997**, *274*, 242−250.

(93) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(94) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(95) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(96) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345−351.

(97) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(98) Slater, J. C. *Quantum Theory of Molecular and Solids*; McGraw-Hill: New York, 1974; Vol. 4.

(99) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.

(100) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(101) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244−13249.

(102) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671−6687.

(103) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(104) Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, *108*, 9624−9631.

(105) Voorhis, T. V.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400−410.

(106) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040−1046.

(107) Adamo, C.; Barone, V. J. *J. Chem. Phys.* **1998**, *108*, 664−675.

(108) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129−12137.

(109) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(110) Rey, J.; Savin, A. *Int. J. Quantum Chem.* **1998**, *69*, 581−590.

(111) Krieger, J. B.; Chen, J.; Iafrate, G. J.; Savin, A. In *Electron Correlations and Materials Properties*; Gonis, A., Kioussis, N., Eds.; Plenum: New York, 1999; pp 463.

(112) Toulouse, J.; Savin, A.; Adamo, C. *J. Chem. Phys.* **2002**, *117*, 10465−10473.

(113) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715−2719.

(114) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43−52.

(115) Gill, P. M. W. *Mol. Phys.* **1996**, *89*, 433−445.

(116) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264−6271.

CT600185A

JCTC Journal of Chemical Theory and Computation

Calculation of the $\mathscr{B}$ Term of Magnetic Circular Dichroism. A Time-Dependent Density Functional Theory Approach

Michael Seth,[†] Jochen Autschbach,[‡] and Tom Ziegler*,[†]

Department of Chemistry, University of Calgary, University Drive 2500, Calgary, AB T2N-1N4, Canada, and Department of Chemistry, State University of New York at Buffalo, 312 Natural Sciences Complex, Buffalo, New York 14260-3000

Received September 12, 2006

Abstract: A method for calculating the $\mathscr{B}$ term of magnetic circular dichroism utilizing time-dependent density functional theory is presented. The expression for the $\mathscr{B}$ term is formulated through the standard sum-over-states approach, and all necessary matrix elements and transition energies are provided by the time-dependent density functional theory calculation. Test calculations of the magnetic circular dichroism spectra of ethene, propene, furan and its heavier homologues, and pyrrole and two of its derivatives are presented. The discrepancy between theory and experiment previously observed for ethene is not resolved, but the experimental spectra of the aromatic compounds are very well reproduced by the theory.

1. Introduction

Magnetic circular dichroism (MCD) spectroscopy measures the difference in absorption of left and right circularly polarized light in the presence of a magnetic field. MCD spectroscopy can provide information about the geometry and magnetic properties of the system of interest as well as new insights into the assignment of the corresponding absorption spectrum. MCD spectra are usually analyzed through three types of spectral feature, namely, the $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$ terms.[1,2] $\mathscr{A}$ terms occur when the degeneracy of a state involved in the transition is broken by the applied magnetic field. $\mathscr{B}$ terms arise because of the mixing between states induced by the magnetic field. $\mathscr{C}$ terms appear when the degeneracy of the ground state is broken by the magnetic field, leading to unequal populations in the ground-state components. It is generally believed that the relative importance of contributions to a MCD spectrum follows the order $\mathscr{C} > \mathscr{A} > \mathscr{B}$. Often, when $\mathscr{A}$ and/or $\mathscr{C}$ terms are present in a MCD spectrum, the influence of $\mathscr{B}$ terms is assumed to be small for the purposes of spectral interpretation.

The temperature-dependent $\mathscr{C}$ terms can be present in an MCD spectrum only if the molecule being studied has a degenerate or near-degenerate ground state. For closed-shell molecules, only $\mathscr{A}$ and $\mathscr{B}$ terms are relevant. $\mathscr{A}$ terms can arise only if a molecule has degenerate states, that is, has a 3-fold or higher rotational or improper axis in its point group. The MCD spectrum of a lower-symmetry closed-shell molecule will be made up only of $\mathscr{B}$ terms. Thus, although $\mathscr{B}$ terms are expected to be smaller in magnitude, they provide the only contribution to the MCD spectra of many molecules. Further, some experiments have found that, even for spectra dominated by $\mathscr{A}$ or $\mathscr{C}$ terms, $\mathscr{B}$ terms must be considered to fully explain what is observed.[3,4]

Recently, we have explored ways to calculate MCD $\mathscr{A}$ and $\mathscr{C}$ terms through density functional theory (DFT) calculations.[5−7] The obvious next step is to implement the calculation of $\mathscr{B}$ terms for closed-shell molecules. Our efforts in this direction are detailed in this paper. Following our previous work, we utilize DFT to describe the ground state of the system at hand and describe the excited states with time-dependent DFT (TDDFT). This choice allows us to obtain results of reasonable accuracy while retaining the ability to treat quite large molecules.

A large number of calculations of MCD $\mathscr{B}$ terms exist in the literature. Most of these calculations have utilized

* Corresponding author fax: +1-403-289-9488; e-mail: ziegler@ucalgary.ca.
[†] University of Calgary.
[‡] University of New York at Buffalo.

10.1021/ct600283t CCC: $37.00 © 2007 American Chemical Society
Published on Web 12/24/2006

semiempirical methods.[8] The few ab initio calculations of MCD $\mathscr{B}$ terms include the quadratic response multiconfiguration self-consistent field study of Coriani and co-workers[9] and the generalized unrestricted Hartree−Fock sum-over states (SOS) and finite perturbation work of Honda et al.[10,11]

The outline of this paper is as follows. In the next section, we will review the equations defining the $\mathscr{B}$ term and show how TDDFT can be used in combination with these equations. The usefulness of our approach will then be illustrated through example calculations on the small molecules ethene and propene and a series of five-membered aromatic heterocyclic organic compounds.

2. Theory

Magnetic circular dichroism is defined as the difference in absorption (or absorption coefficients) of a substance with respect to left and right circularly polarized light induced by the presence of a magnetic field.

$$\Delta A_{\mathrm{MCD}}(A\rightarrow J) = \gamma(\langle|A|\mu_- |J\rangle|^2 - |\langle A|\mu_+|J\rangle|^2) \qquad (1)$$

where γ is a collection of constants, A and J are the initial and final states of a transition, and $\mu_\pm$ are the electric dipole moment operators corresponding to left and right circularly polarized light

$$\mu_- = \frac{1}{\sqrt{2}}(\mu_x - i\mu_y) \qquad (2)$$

$$\mu_+ = \frac{1}{\sqrt{2}}(\mu_x + i\mu_y) \qquad (3)$$

where i is $\sqrt{-1}$. As has already been noted, $\mathscr{B}$ terms in MCD spectroscopy arise because of the mixing between states induced by the applied magnetic field. For most magnetic fields of interest, the intensities of bands in the MCD spectrum corresponding to $\mathscr{B}$ terms vary linearly with the applied magnetic field. This MCD can be written as

$$\Delta A_{\mathrm{MCD},\mathscr{B}}(A\rightarrow J) = \gamma \mathscr{B} B f(\omega) \qquad (4)$$

where $\mathscr{B}$ is the $\mathscr{B}$ term, B is the applied magnetic field, and $f(\omega)$ is a band shape function. Throughout this work, we take $f(\omega)$ to be Gaussian in form

$$f_J(\omega) = \frac{1}{W\sqrt{\pi}} \mathrm{e}^{-(\omega-\omega_J)^2/W^2} \qquad (5)$$

for transition to state J centered at energy ω_J with width parameter W. A first-order perturbation treatment[2] gives the following expression for $\mathscr{B}$

$$\mathscr{B}(A\rightarrow J) = -\frac{2i}{3}\left[\sum_{\lambda K\kappa(K\neq A)} \frac{\langle K\kappa|\mathbf{L}|A\rangle}{\omega_K}(\langle A|\mu|J\lambda\rangle \times \langle J\lambda|\mu|K\kappa\rangle) + \sum_{\lambda K\kappa(K\neq J)} \frac{\langle J\lambda|\mathbf{L}|K\kappa\rangle}{\omega_K - \omega_J}(\langle A|\mu|J\lambda\rangle \times \langle K\kappa|\mu|A\lambda\rangle)\right] \qquad (6)$$

for nondegenerate ground state A and (possibly degenerate) excited states J and K with degenerate components λ and κ. $\mathbf{L}$ and μ are vectors composed of the Cartesian components of the angular momentum and electric dipole moment operators, respectively. ω_K is the energy of state K relative to the ground state. Equation 6 corresponds to a $\mathscr{B}$ term averaged over all possible orientations.

There is some arbitrariness with respect to sign and multiplicative constants in the definition of MCD parameters. We follow the conventions advocated by Piepho and Schatz.[2]

The first term in eq 6 describes contributions to the $\mathscr{B}$ term arising because of the magnetic-field-mediated mixing between the ground state and excited states. The second term describes contributions arising because of mixing between the excited state in the transition of interest and all other excited states.

Calculating a $\mathscr{B}$ term through eq 6 appears to be a rather formidable task. It is obviously impossible to include all excited states as required by the summation. In favorable cases, only a few states need to be included. Even if this is the case, if one wishes to make use of a wave-function-based approach to calculate $\mathscr{B}$, then one is forced to choose between methods that provide relatively low accuracy (configuration interaction singles) or those that are so prohibitively demanding in terms of computation resources [such as complete active space−self-consistent field (CASSCF) or multireference configuration interaction (MRCI)] that only relatively small molecules can be studied.

TDDFT provides a reasonable compromise between accuracy and computation expense. How can all the terms required to evaluate eq 6 be extracted from a TDDFT calculation? We consider the formulation of TDDFT utilized in the majority of implementations in quantum chemical codes, that is, that outlined by Casida.[12] The eigenvalue equation to be solved is

$$\Omega F = \omega^2 F \qquad (7)$$

where the Ω matrix is derived from the random-phase-approximation-like equations obtained from linear response TDDFT under the adiabatic approximation. The eigenvalues ω^2 are interpreted as the square of the transition energies. The eigenvectors, the "transition vector F", are closely related to the transition density from the reference state to the excited states

Most of the terms required to evaluate eq 6 can be obtained in a straightforward manner once eq 7 is solved. The ω_K terms are obviously taken directly from the eigenvalues of eq 7. Casida's original formulation included details of how to calculate $\langle A|\mu|K\rangle$ through the following formula[12]

$$\langle A|\mu|K\rangle = \sum_{ia} \sqrt{\frac{\epsilon_a - \epsilon_i}{\omega_K}}\mu_{ia}F_{ia}^K \qquad (8)$$

where ϵ_s is the energy of molecular orbital (MO) s, μ_{ia}^ξ are the matrix elements of the electric dipole moment operator with orbitals i and a, and F_{ia}^K is an element of the Kth F corresponding to the one-electron excitation from orbital i to orbital a. Here and throughout this paper, i, j, ... refer to occupied molecular orbitals; a, b, ... refer to unoccupied

orbitals; and *s*, *t*, ... refer to orbitals that could be occupied or unoccupied.

Similarly, it has been shown[13] that the terms corresponding to matrix elements of the angular momentum operator between the reference state and state *K*, $\langle A|\mathbf{L}|K\rangle$, have the form

$$\langle A|\mathbf{L}|K\rangle = \sum_{ia} \sqrt{\frac{\omega_K}{\epsilon_a - \epsilon_i}}\, \mathbf{L}_{ia} F_{ia}^K \tag{9}$$

where the slightly different form of eq 9 as compared to eq 8 arises because **L** is an antisymmetric operator while μ is a symmetric operator.

The remaining terms are matrix elements involving only excited states $\langle J|\mu|K\rangle$ and $\langle J|\mathbf{L}|K\rangle$. If $J = K$, we can use the expression given by Furche and Ahlrichs[14] utilizing the unrelaxed density to evaluate the difference between the values of a one-electron property for the reference and excited state

$$\langle J|\mathcal{O}|J\rangle - \langle A|\mathcal{O}|A\rangle = \frac{1}{2}\left[\sum_{iab} \frac{(\epsilon_a - \epsilon_i)(\epsilon_b - \epsilon_i) + \omega_J^2}{\omega_J\sqrt{(\epsilon_a - \epsilon_i)(\epsilon_b - \epsilon_i)}} F_{ia}^J F_{ib}^J \mathcal{O}_{ab} - \sum_{ija} \frac{(\epsilon_a - \epsilon_i)(\epsilon_a - \epsilon_j) + \omega_J^2}{\omega_J\sqrt{(\epsilon_a - \epsilon_i)(\epsilon_a - \epsilon_j)}} F_{ia}^J F_{ja}^J \mathcal{O}_{ij}\right] \tag{10}$$

where $\mathcal{O}_{st}$ is an element of the one-electron operator $\mathcal{O}$ over molecular orbitals *s* and *t*. $\langle A|\mathcal{O}|A\rangle$ is readily obtained from the reference state orbitals allowing $\langle J|\mathcal{O}|J\rangle$ to be calculated.

The off-diagonal elements between excited states are derived using a generalization of eq 10

$$\langle J|\mathcal{O}|K\rangle = \frac{1}{2}\left[\sum_{iab} \frac{(\epsilon_a - \epsilon_i)(\epsilon_b - \epsilon_i) + \omega_J\omega_K}{\sqrt{\omega_J\omega_K(\epsilon_a - \epsilon_i)(\epsilon_b - \epsilon_i)}} F_{ia}^J F_{ib}^K \mathcal{O}_{ab} - \sum_{ija} \frac{(\epsilon_a - \epsilon_i)(\epsilon_a - \epsilon_j) + \omega_J\omega_K}{\sqrt{\omega_J\omega_K(\epsilon_a - \epsilon_i)(\epsilon_a - \epsilon_j)}} F_{ia}^J F_{ja}^K \mathcal{O}_{ij}\right] \tag{11}$$

Equation 11 is obtained from eq 10 by relaxing the requirement that both transition densities be the same, giving $F_{ia}^J F_{ja}^K$ rather than $F_{ia}^J F_{ja}^J$ for instance. Thus, when the integrals of μ and **L** over MOs, the orbital eigenvalues, and transition energies and transition vectors available from a TDDFT calculation are inserted into eqs 8−11, all of the terms required to evaluate eq 6 are now available.

One issue remains that arises from the sum-over-states formulation of the $\mathcal{B}$ term applied here. In eq 6, the energy denominators often play an overriding role, determining the magnitude of the calculated $\mathcal{B}$ term. States that are close in energy with either the ground state or the excited state of interest generally will provide the dominant contribution to the $\mathcal{B}$ term of that transition because either ω_K or $\omega_K - \omega_J$ will be very small. This is a useful feature of the theory as it makes for simple interpretations of the observed spectra when states close in energy are present. The strong influence of the energy denominators puts very heavy demands on the theoretical method used to calculate the transition energies however. In a TDDFT calculation, the expected error of the calculated transition energies under favorable conditions (e.g., transitions well-described by one-electron excitations, a system with no transition metals) would be about 0.2 eV. Suppose two excited states (*J* and *K*) are separated by 0.15 eV. It would not be unusual if a TDDFT calculation predicted that these two states were separated by 0.05 or 0.25 eV. If the predicted energy difference was 0.05 eV, then the contribution to the $\mathcal{B}$ term of the transition $A \rightarrow J$ due to the mixing between states *K* and *J* would be too large by a factor of 3. If the predicted splitting was 0.25 eV, the $\mathcal{B}$-term contribution would be too small by a factor of 0.6. Thus, the contributions to the calculated $\mathcal{B}$ term that would be expected to be the most important are likely to be the least accurate.

It turns out that the problem caused by errors in the transition energies are less significant than they seem. From eq 6, it is apparent that, if state *K* mixes with state *J* to induce a $\mathcal{B}$ term at ω_J, then state *J* mixing with state *K* will induce a $\mathcal{B}$ term equal in magnitude but opposite in sign at energy ω_K. When two states are nearly degenerate, these two opposing $\mathcal{B}$ terms give rise to an overall band that has the derivative shaped characteristic of an $\mathcal{A}$ term.[2] It can be shown that, provided the difference in energy between the two states is significantly smaller than their absorption band widths, then the intensity of the resulting "pseudo-$\mathcal{A}$ term" is independent of that energy difference.[2] Thus, while errors in predicted transition energies may make it difficult to correctly reproduce the individual $\mathcal{B}$ terms of two closely lying states, the overall MCD spectrum is relatively insensitive to the exact energy difference, making it easier to predict the form of the spectrum.

3. Example Calculations

The approach to calculating MCD $\mathcal{B}$ terms described in the previous section was implemented into a development version of the Amsterdam Density Functional program.[15−19] In this section, we will describe several example calculations to illustrate the possibilities of the current method.

All of the calculations had several features in common. Molecular geometries were optimized in calculations utilizing valence triple-ζ basis sets with two added polarization functions (TZ2P) and the Becke-Perdew86 functional.[20−22]

Whenever the molecule of interest included the heavier elements Br, Se, or Te, relativistic effects were included by the spin-free version of the zeroth-order regularized approximation.[23−25]

It is expected that transitions to Rydberg states will be found in the energy range of interest of all of the molecules considered. A functional with correct asymptotic behavior and basis sets including diffuse functions therefore must be utilized in the TDDFT calculations. The SAOP functional[26,27] was chosen to satisfy the first requirement. For the smaller ethene and propene molecules, it was feasible to apply large, even-tempered basis sets with a double-ζ description of the core and a quadruple-ζ set describing the valence augmented with three polarization functions to all atoms. Sets of three

Table 1. Transition Energies and Oscillator Strengths of Ethene

transition	energy[a]	oscillator strength	other calculations		expt[d,e]
			MRCI[b]	STEOM-CCSD[c]	
$1^1A_g \rightarrow 1^1B_{3u}$ ($\pi \rightarrow$ 3s)	7.30	0.06	7.16	7.21 (0.08)	7.11
$1^1A_g \rightarrow 1^1B_{1g}$ ($\sigma \rightarrow \pi^*$)	7.65			8.49	
$1^1A_g \rightarrow 1^1B_{1u}$ ($\pi \rightarrow \pi^*$)	7.66	0.31	7.80	7.78 (0.37)	7.66
$1^1A_g \rightarrow 1^1B_{2g}$ ($\pi \rightarrow 3p_y$)	7.94		7.85	7.97	7.90
$1^1A_g \rightarrow 2^1B_{1g}$ ($\pi \rightarrow 3p_z$)	7.98		7.82	7.91	7.80

[a] Energies in eV. [b] Multireference configuration interaction results from ref 36. [c] Similarity transformed equation of motion coupled-cluster singles and doubles results from ref 37. Oscillator strengths in parentheses. [d] Ref 33. [e] Ref 32.

diffuse functions (3s3p3d for H and 3s3p3d3f for C) were then added to give the final QZ3P+3D basis set. For all other molecules, a similar basis set was chosen for all atoms forming part of the conjugated backbone of that molecule with the exception that only two diffuse sets (2s2p2d2f) were used to give a QZ3P+2D basis. A TZ2P basis set was taken for all nonbackbone atoms of these larger molecules. The only exception to this recipe was tellurophene as no QZ3P+2D basis set was available for tellurium. In this case, a valence quadruple-ζ basis set with four polarization functions (QZ4P) was used. This basis set is optimized for use with relativistic calculations as was that applied to Br.

The geometry optimization calculations were performed with frozen cores corresponding to [He] for O, C, and N; [Ne] for S; [Ar] for Br and Se; and [Kr] for Te. No electrons were frozen in all TDDFT calculations.

The application of basis sets with diffuse functions often leads to problems because of linear dependencies. Such issues were avoided by removing linear combinations of functions corresponding to small eigenvalues of the overlap matrix ($\leq 10^{-4}$). The overall integral accuracy parameter was chosen to be 4, which is, roughly, the number of significant digits obtained when numerically integrating the ground-state electron density.[18]

MCD spectra were simulated by placing for each transition a Gaussian band shape as given by eq 5 with bandwidth W given by

$$W = 0.1605\omega_K \tag{12}$$

where ω_K is the energy of the transition in electrovolts and is also taken as the center of the band shape. The factor of 0.1605 was chosen to give calculated bandwidths similar in size to those observed experimentally. The band shape was then multiplied by the $\mathscr{B}$ term calculated for that transition to obtain the final MCD. A comparison between any simulated spectrum and experimental results in terms of absolute intensities is difficult because estimating some contributions such as the influence of the environment on the incident light and magnetic field is challenging. To get around this problem, both the experimental and simulated spectra are scaled so that the most intense band in both spectra have the same value of ΔA at their peak.

It should be noted that the current formalism is not gauge-invariant. Previous studies have found that the results should depend very weakly on the choice of gauge.[10,11] We assume that this is the case here but will return to this point in future studies.

3.1. Ethene and Propene. The first examples that we will consider are ethene and propene. Ethene is a relatively simple molecule that was included as a test case in previous ab initio calculations of MCD spectra.[9,11] Its MCD spectrum was measured by Brith-Lindner and Allen[28] a few years ago. The MCD spectrum of propene was measured more recently.[29] Note that we follow the axis convention for ethene that puts the z axis along the C−C double bond and the x axis perpendicular to the molecular plane.

As the simplest possible molecule including a C−C double bond, the absorption spectrum of ethene has been the subject of extensive investigation both experimentally and theoretically (see, for example, refs 29−37). It is generally acknowledged that the valence ($\pi \rightarrow \pi^*$, $1^1A_g \rightarrow 1^1B_{1u}$ in D_{2h} symmetry) and Rydberg ($\pi \rightarrow$ 3s, $1^1A_g \rightarrow {}^1B_{3u}$) transitions overlap in the 6−8 eV energy range. There has been some suggestion that some of the observed bands correspond to other transitions. Evidence supporting this idea was provided by the MCD spectrum of ethene,[28] where three distinct series of peaks are observed indicating the presence of at least three states. Theoretical calculations have found that the potential-energy surface of the 1^1B_{1u} state has a saddle point at the ground-state equilibrium geometry.[31,36] It would then be expected that nonadiabatic effects could play an important part in determining the absorption spectrum of ethene, and recent calculations have supported this idea.[35,37]

The calculated transition energies and oscillator strengths of the states of ethene predicted to lie 7−8 eV above the ground state are listed in Table 1 along with some previously calculated results and experimental values. The agreement between the present vertical transition energies and high-quality MRCI and equation of motion coupled−cluster calculations[36,37] is good for the most part with all but one energy being within 0.2 eV of the other quoted values. The one exception is the $1^1A_g \rightarrow 1^1B_{1g}$ transition energy, which is rather low compared to that reported by Hazra and co-workers.[37]

The $\mathscr{B}$ terms of the $1^1A_g \rightarrow 1^1B_{3u}$ and $1^1A_g \rightarrow 1^1B_{1u}$ transitions calculated by the present method are listed in Table 2. Also listed in Table 2 are the $\mathscr{B}$ terms of the same transitions calculated by other workers.

One of the features of the SOS approach to calculating properties is that, in principle, the summation runs over all possible states. It is not always clear how many states must be included in a given treatment in order to obtain suitably converged results. Listed in Table 2 are $\mathscr{B}$ terms obtained with 25, 50, 150, or 250 states included in the expansion. For the present calculations involving ethene, these numbers

Table 2. Calculated $\mathscr{B}$ Terms of Ethene[a]

	n	$1^1A_g \rightarrow 1^1B_{3u}$	$1^1A_g \rightarrow 1^1B_{1u}$
QZ3P+3D/SAOP	250	−12.77	−6.14
QZ3P+3D/SAOP	150	−12.55	−6.49
QZ3P+3D/SAOP	50	−10.93	−4.34
QZ3P+3D/SAOP	25	−11.00	−2.46
daug-cc-pVTZ/FV-CAS [b]		−13.12	−2.04
daug-cc-pVTZ/SCF [b]		−27.21	10.55
DZ/GUHF-SECI [c]			−53.45
DZ/GUHF-SECI [c]	108		34.3

[a] $\mathscr{B}$ terms in au. *n* indicates the number of states allowed to be mixed by the magnetic field in order to evaluate $\mathscr{B}$. [b] Full-valence complete active space and Hartree−Fock results from ref 9. [c] Generalized unrestricted Hartree−Fock/single-excitation configuration interaction results from ref 11.

of states correspond to all states up to 10.9, 11.7, 14.3, and 16.36 eV, respectively. Even including only 25 states gives results that are qualitatively similar to the values calculated with 250 states. The $\mathscr{B}$ terms calculated with 25 states are both negative and are similar in magnitude to those obtained with 250 states. The $\mathscr{B}$ term of the $\pi \rightarrow \pi^*$ transition is more strongly influenced by higher-lying states, but both seem to be fairly well converged once 150 states are included.

The effort required to evaluate each individual $\mathscr{B}$ term through eq 6 increases linearly with the number of states *n*. If $\mathscr{B}$ terms for all *n* states are to be evaluated, the overall cost increases quadratically. The cost of the initial TDDFT calculation also increases with *n*. It is thus to our advantage if we can keep the number of states involved as small as possible. It is worth noting that, while *n* states may be needed to give the desired accuracy, in general, far fewer than *n* $\mathscr{B}$ terms will be of interest. If this is the case, the computational cost of the MCD calculation could be alleviated significantly by evaluating only those contributions necessary for the $\mathscr{B}$ terms of interest. The more selective algorithm would scale as *Nn* (*N* is the number of $\mathscr{B}$ terms to be calculated, usually $N \ll n$) or linearly with *n* rather than n^2. This more efficient approach has not been implemented at the present time, but it is planned for the future.

The $\mathscr{B}$ terms obtained by Coriani and co-workers using quadratic response theory and states derived from a CASSCF calculation[9] agree well with the present values. Our $\pi \rightarrow$ 3s $\mathscr{B}$ term of −12.8 au is very close to the full-valence−complete active space value of −13.1 au, while the $\mathscr{B}$ term for the $\pi \rightarrow \pi^*$ transition obtained by Coriani et al. is rather less negative than ours. The discrepancy between the present work and that of Coriani et al. may be due to the lack of dynamic correlation in the latter work. The $\mathscr{B}$ terms of the $\pi \rightarrow \pi^*$ transition calculated by Honda et al. are very different from ours. This is no doubt due to the limited basis set used in that work.

In their study of the MCD spectrum of ethene, Brith-Lindner and Allen assigned the negative bands to either the $\pi \rightarrow \pi^*$ or $\pi \rightarrow$ 3s transitions and speculated that the positive bands may be due to a $\pi \rightarrow \sigma^*$ transition.[28] Snyder and co-workers offered a different assignment suggesting that the negative bands are due to $\pi \rightarrow$ 3s and $\pi \rightarrow$ 3p transitions and positive bands are due to the $\pi \rightarrow \pi^*$ transition.[29] Our results appear to be more in line with the earlier assignments of Brith-Lindner and Allen but offer no information as to what transition may be responsible for the observed positive MCD band. Given the large nonadiabatic effects found in the absorption spectrum of ethene, it seems likely that a more detailed study including such effects will be necessary to fully understand the MCD spectrum of this species.

The current approach to calculating $\mathscr{B}$ terms offers significant scope for analysis. The TDDFT calculation can be analyzed in terms of which single-electron excitations contribute to a given transition. The SOS formulation of the $\mathscr{B}$ term naturally lends itself to analysis in terms of which states provide the major contributions to eq 6.

The TDDFT calculation of ethene gives transitions that are relatively pure. All transitions are made up at least 91% by the electron transitions given in the assignments of Table 1.

Almost all of the $\mathscr{B}$ term of the $\pi \rightarrow$ 3s ($1^1A_{1g} \rightarrow 1^1B_{3u}$) transition (−12.27 out of −12.77 au) comes from mixing of the 1^1B_{1u} with the 1^1B_{3u} state. A few other states have non-negligible contributions, but these contributions almost perfectly cancel. Given the form of eq 6, this would suggest that the $\mathscr{B}$ term of the $\pi \rightarrow \pi^*$ ($1^1A_{1g} \rightarrow 1^1B_{1u}$) transition would have a significant positive contribution of 12.27 au due to mixing of the 1^1B_{3u} state with the 1^1B_{1u} state. This is indeed the case, but this positive contribution is overtaken by large negative contributions from the 3^1B_{3u} (−11.47 au) and 1^1B_{2u} (−4.41 au) states which are predicted to be found at around 9 eV above the ground state. A few other higher-lying states also provide significant negative contributions to give the final value of about −7 au.

The absorption spectrum of propene has been studied rather less extensively than that of ethene. Under the C_s symmetry of propene, all transitions become formally allowed and rather more are observed in the 6−8 eV range than was the case for ethene. The observed spectrum in this energy range consists of a broad band with an extensive fine structure.[29,38] One of the transitions found in this band is assigned as $\pi \rightarrow \pi^*$, while the others are assigned as $\pi \rightarrow$ 3*n*, where *n* is s, p, or d.[38] The MCD spectrum of propene in the range 6−8 eV is qualitatively similar to that of ethene in that it shows negative MCD at lower energies, positive MCD at intermediate energies, and negative MCD at higher energies.[29] Snyder and co-workers assign the first negative MCD band as the $\pi \rightarrow$ 3s transition, the positive MCD band as the $\pi \rightarrow \pi^*$ transition, and more tentatively, the second negative MCD band as a $\pi \rightarrow$ 3p transition.[29]

The calculated transition energies are compared with an earlier calculation and experiment in Table 3. The agreement between our results and previous work is somewhat poorer than was the case for ethene. The $\pi \rightarrow$ 3s, the first $\pi \rightarrow$ 3p, and the $\pi \rightarrow$ 3d transitions are all within 0.2 eV of experimental results, but the $\pi \rightarrow \pi^*$ transition is calculated to be 6.81 eV, while experimentally, it is measured to be 7.2 eV; one of the $\pi \rightarrow$ 3p transitions is too high in energy, while the other is too low. The two $\sigma \rightarrow \pi^*$ transitions are probably too low, as was the case for ethene.

The calculated $\mathscr{B}$ terms for the eight transitions predicted to lie below 8.2 eV as a function of the expansion size *n* are presented in Table 4. Much like the ethene case, the

Table 3. Transition Energies and Oscillator Strengths of Propene

transition	energy[a]	oscillator strength	MRCI[b]	exptl[b]
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3s)	6.58	0.008	6.89	6.55
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow \pi^*$)	6.81	0.25	7.33	7.2
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3p)	7.16	0.010	7.25	7.08
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3p)	7.33	0.019	7.81	7.76
$1^1A' \rightarrow 1A''$ ($\sigma \rightarrow \pi^*$)	7.46	0.001		
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ 3p)	8.01	0.016	7.77	7.6
$1^1A' \rightarrow 1A''$ ($\sigma \rightarrow \pi^*$)	8.05	0.002		
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3d)	8.16	0.013	8.05	8.0

[a] Energies in eV. [b] Multireference configuration interaction calculation from ref 38.

Table 4. Calculated $\mathscr{B}$ Terms of Propene[a]

	$n = 25$	$n = 50$	$n = 150$	$n = 250$
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3s)	−3.37	−2.95	−3.06	−3.12
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3p)	0.59	0.65	0.00	−0.04
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow \pi^*$)	−2.49	−3.90	−6.24	−5.84
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ 3p)	−3.99	−3.94	−4.07	−4.04
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3p)	−2.56	−2.45	−3.24	−3.28
$1^1A' \rightarrow 1A''$ ($\sigma \rightarrow \pi^*$)	−0.09	−0.06	0.05	0.08
$1^1A' \rightarrow 1A''$ ($\sigma \rightarrow \pi^*$)	1.35	1.18	1.19	1.18
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3d)	1.87	1.62	1.80	1.80

[a] $\mathscr{B}$ terms in au. n indicates the number of states allowed to be mixed by the magnetic field in order to evaluate $\mathscr{B}$.

calculated $\mathscr{B}$ term of the $\pi \rightarrow \pi^*$ transition converges rather slowly with expansion size. The $\mathscr{B}$ terms of the other transitions are described well even by a small expansion involving 25 states. With the exception of the very small $\mathscr{B}$ terms which will contribute little to the observed MCD spectrum, the other $\mathscr{B}$ terms calculated with the 25-state expansion are within 20% of the results obtained with the 250-state expansion. Four of these five "other $\mathscr{B}$ terms" are within 10% of the $n = 250$ result when calculated with the 25-state expansion.

Also similarly to ethene, we find no sign in our calculations of the transition responsible for the positive MCD in the middle of the propene spectrum. All of the transitions expected to contribute strongly to the MCD spectrum below 8 eV have negative $\mathscr{B}$ terms. Much like ethene, it appears that nonadiabatic effects will need to be considered in order to completely explain the MCD spectrum of propene. The calculations do indicate that the MCD will become positive at higher energies however.

The interpretation of the $\mathscr{B}$ terms of the $\pi \rightarrow$ 3s and $\pi \rightarrow \pi^*$ transitions in terms of contributions to eq 6 is similar to that of ethene. The $\pi \rightarrow$ 3s ($1^1A' \rightarrow 1A''$) $\mathscr{B}$ term is dominated by the contribution from mixing the $\pi \rightarrow \pi^*$ ($1^1A' \rightarrow 1A'$) state (−2.99 au out of −3.12 au). The largest contribution to the $\mathscr{B}$ term of the $\pi \rightarrow \pi^*$ transition arises from mixing of the $\pi \rightarrow$ 3s state (2.99 au), but the overall negative $\mathscr{B}$ term is caused by several smaller negative contributions. The other four calculated $\mathscr{B}$ terms that are significant in size (>1 au) are all made up of a small number of minor contributions (≈1 au). Unlike the $\pi \rightarrow \pi^*$ transition, all of these contributions come from low energy states, meaning that these $\mathscr{B}$ terms can be calculated with a relatively small expansion, as we have seen already for ethene.

Table 5. Transition Energies and Oscillator Strengths of Furan

transition	energy[a]	oscillator strengths	CC[b]	exptl[b]
$1^1A_1 \rightarrow 1^1B_2$ ($\pi \rightarrow \pi^*$)	6.00	0.17	6.32	6.06
$1^1A_1 \rightarrow 1^1A_2$ ($\pi \rightarrow$ 3s)	6.19	0.00	6.04	5.94
$1^1A_1 \rightarrow 2^1A_1$ ($\pi \rightarrow \pi^*$)	6.33	0.0002	6.57	5.8
$1^1A_1 \rightarrow 1^1B_1$ ($\pi \rightarrow 3p_y$)	6.84	0.030	6.58	6.48
$1^1A_1 \rightarrow 2^1A_2$ ($\pi \rightarrow 3p_z$)	6.97	0.00	6.73	6.61

[a] Energies in eV [b] Coupled cluster results from ref 51.

3.2. Furan, Thiophene, Selenophene, and Tellurophene. In work published in the 1970s, Thulstrup and co-workers measured the MCD spectra of a number of five-membered-ring heterocycles including furan and pyrrole, heavier heteroatom homologues of furan, and derivatives of these molecules.[39,40] These studies provide an excellent set of data that can be used to test our implementation. In this section, we will consider furan and its homologues thiophene, selenophene, and tellurophene. In the next section, pyrrole and two derivatives of pyrrole will be discussed.

The transition energies and oscillator strengths of furan, thiophene, selenophene, and tellurophene have been calculated by Hieringer and co-workers utilizing a methodology very similar to that applied here.[41] Therefore, we shall present the transition energies and oscillator strengths from the present work as they are relevant for the calculation and analysis of MCD spectra, but there will be little discussion of these results.

As a medium-sized molecule of biological importance, absorption spectra of furan and its derivatives have been studied intensively both experimentally and theoretically (see, for example, refs 40−54). The energy range of interest for the present study is 5.0−6.5 eV. In this range, the absorption spectrum of furan is dominated by a single intense band assigned as a $\pi \rightarrow \pi^*$ transition. Some of the fine structure of this band is assigned to Rydberg transitions. Increasingly more sophisticated theoretical studies have refined these assignments and have shown that a second $\pi \rightarrow \pi^*$ transition is present in this region, although it is undetectable experimentally because of its low intensity.

The MCD spectrum of furan between 5 and 6.5 eV is dominated by a single negative band with the MCD becoming slightly positive at the highest energies.[40] As we shall see, this is in contrast to its heavier homologues.

The calculated energies and oscillator strengths of all transitions of furan up to 7.0 eV are presented in Table 5 along with coupled cluster and experimental transition energies for comparison. As has been noted previously,[41] the $\pi \rightarrow \pi^*$ transition energies calculated with the SAOP functional are a little lower than those calculated at the coupled cluster level, while the Rydberg transition energies are a little higher. The $1^1A_1 \rightarrow 1^1B_2$ ($\pi \rightarrow \pi^*$) transition has a much greater oscillator strength than the other transitions and dominates the absorption spectrum in the energy range of interest.

Table 6. Calculated $\mathscr{B}$ Terms of Furan[a]

transition	$n = 25$	$n = 50$	$n = 150$	$n = 250$
$1^1A_1 \rightarrow 1^1B_2$ ($\pi \rightarrow \pi^*$)	−9.87	−12.02	−13.54	−12.59
$1^1A_1 \rightarrow 1^1A_2$ ($\pi \rightarrow$ 3s)	0.00	0.00	0.00	0.00
$1^1A_1 \rightarrow 2^1A_1$ ($\pi \rightarrow \pi^*$)	2.81	2.74	2.74	2.72
$1^1A_1 \rightarrow 1^1B_1$ ($\pi \rightarrow 3p_y$)	3.38	3.05	3.13	2.58
$1^1A_1 \rightarrow 2^1A_2$ ($\pi \rightarrow 3p_z$)	0.00	0.00	0.00	0.00

[a] $\mathscr{B}$ terms in au. n indicates the number of states allowed to be mixed by the magnetic field in order to evaluate $\mathscr{B}$.

With the exception of the $1^1A_1 \rightarrow 2^1A_1$ ($\pi \rightarrow \pi^*$) transition, all transitions are dominated (>93%) by one single-electron excitation from a π orbital to the appropriate π or Rydberg orbital. The $1^1A_1 \rightarrow 2^1A_1$ ($\pi \rightarrow \pi^*$) transition is approximately a 50/50 mix of the two $\pi \rightarrow \pi^*$ orbital excitations with A_1 symmetry.[41]

The calculated MCD $\mathscr{B}$ terms of the five lowest transitions of furan are listed in Table 6, and the simulated MCD spectrum of furan obtained from the $n = 250$ parameters is compared with the experimental spectrum in Figure 1a.

The most obvious conclusion that can be drawn from Figure 1a is that the simulated spectrum reproduces all of the important features of the observed spectrum. Both have a large negative band and slightly positive MCD at higher energies. From Table 6, it is apparent that the negative MCD is due to the significant $\mathscr{B}$ term of the $1^1A_1 \rightarrow 1^1B_2$ ($\pi \rightarrow \pi^*$) transition. The negative MCD of this transition overwhelms the lesser positive MCD at a little higher energy from the other $\pi \rightarrow \pi^*$ transition. The positive $\mathscr{B}$ term of the $1^1A_1 \rightarrow 1^1B_1$ transition is far enough away from the $1^1A_1 \rightarrow 1^1B_2$ one to be able to produce some positive MCD at the higher-energy end of the spectrum.

One difference between the calculated and simulated spectra is the energy of minimum MCD. Experimentally, it is at about 5.8 eV, while it is calculated to be just below 6 eV. This is a little surprising given the close agreement between the calculated transition energy and that observed experimentally (Table 5). The greater error in the location of the MCD band suggests that the weaker positive bands in this region, which have a non-negligible influence on the overall spectrum, are a little too low in energy in the calculation.

Table 6 lists $\mathscr{B}$ terms calculated with differing numbers of states included in the expansion given by eq 6. Much like ethene and propene, it is apparent that the MCD spectrum of furan between 5 and 7 eV can be described qualitatively by a small 25-state expansion. Of the three nonzero $\mathscr{B}$ terms, only that of the $1^1A_1 \rightarrow 1^1B_2$ transition changes by more than 1 au when going from a 25-state expansion to a 250-state expansion. A 50-state expansion reproduces the results of the 250-state expansion closely. A total of 50 states represent all states calculated to lie below 10.42 eV, while the 250 states lie below 14.87 eV.

The slower convergence of the $\mathscr{B}$ term of the $1^1A_1 \rightarrow 1^1B_2$ transition with respect to the number of states included in eq 6 indicates that a number of states make significant contributions to the perturbation expansion of this $\mathscr{B}$ term. This is indeed the case. Three states, 2^1A_1, 3^1A_1, and 10^1B_2, contribute over 1.0 au through mixing with the 1^1B_2 state with −3.3, −2.4, and −1.9 au, respectively. Although it is close in energy to the 1^1B_2 state, the 1^1B_1 state only contributes 0.4 au because the magnetic field only induces weak mixing between the two states ($\langle 1^1B_2|L_z|1^1B_1\rangle$ is −0.04 au).

Turning now to thiophene, selenophene, and tellurophene, the absorption spectra of the first two molecules are much like that of furan with a single broad band between 4.5 and 6.5 eV. The spectrum of tellurophene exhibits more structure with an intense peak at around 4.5 eV and weaker peaks near 5.0 and 6.0 eV.[40] The MCD spectra of thiophene and selenophene are similar to each other but different from that of furan. These spectra show positive MCD at lower energies and negative MCD at higher energies, giving an overall derivative shaped band characteristic of an MCD $\mathscr{A}$ term or two near-degenerate $\mathscr{B}$ terms of opposite sign.[40] These MCD spectra provided good evidence that there are at least two states that contribute to the observed absorption band. The MCD spectrum of tellurophene is more similar to that of furan as it is dominated by a single negative band. It differs from the furan spectrum, however, in that it has a weak positive band at lower energies than the negative band and at higher energies the MCD becomes more negative again.

Calculated transition energies and oscillator strengths of thiophene, selenophene, and tellurophene up to 7.0 eV are listed in Table 7. The $\pi \rightarrow \pi^*$ transitions decrease in energy from furan to tellurophene. The $1^1A_1 \rightarrow 2^1A_1$ transition decreases more rapidly so that, while for furan the $1^1A_1 \rightarrow 1^1B_2$ transition is lower in energy than the $1^1A_1 \rightarrow 2^1A_1$ transition, for thiophene the two transitions are nearly degenerate and for tellurophene the $1^1A_1 \rightarrow 2^1A_1$ transition is far enough below the $1^1A_1 \rightarrow 1^1B_2$ transition that the two bands can be resolved.[40] The $1^1A_1 \rightarrow 2^1A_1$ transition gains in intensity, and the $1^1A_1 \rightarrow 1^1B_2$ transition decreases in intensity from furan to tellurophene. The Rydberg states decrease in energy and increase in intensity as the heteroatom becomes heavier. The assignments of the $\pi \rightarrow \pi^*$ transitions excitation are clear-cut; in all cases, one MO to MO excitation dominates (>80%) the transition density. Some of the higher-lying Rydberg transition have more mixed character, but the excitation with the greatest contribution is used in the assignments.

The calculated MCD $\mathscr{B}$ terms of all transitions up to 7.0 eV can be found in Table 8, and the simulated MCD spectra of thiophene, selenophene, and tellurophene obtained from these terms are presented in Figure 1b−d. As was noted earlier, the MCD spectra of thiophene and selenophene imply the presence of overlapping oppositely signed $\mathscr{B}$ terms from two nearly degenerate states. Our calculations support this conclusion. We find that the two lowest $\pi \rightarrow \pi^*$ (2^1A_1 and 1^1B_2) states of thiophene and selenophene interact strongly in the presence of a magnetic field to give the large $\mathscr{B}$ terms listed in Table 8. The $\pi \rightarrow \pi^*$ states are close enough in energy to produce pseudo-$\mathscr{A}$ terms (Figure 1b,c). The $\mathscr{B}$ terms of the other states listed in Table 8 have little impact on the overall MCD spectrum.

Not surprisingly, the $\mathscr{B}$ terms of the 2^1A_1 and 1^1B_2 states of thiophene and selenophene are almost completely due to the interaction between these two states (±320 au for

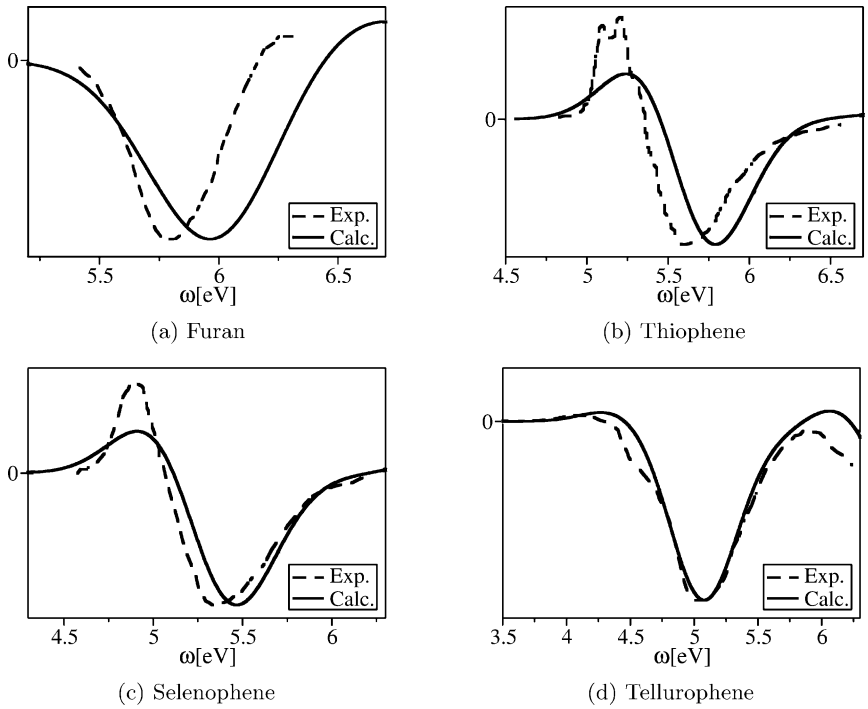


Figure 1. Experimental and simulated MCD spectra of furan, thiophene, selenophene, and tellurophene.

Table 7. Transition Energies and Oscillator Strengths of Thiophene, Selenophene, and Tellurophene

	thiophene		selenophene		tellurophene	
transition	energy[a]	oscillator strengths	energy[a]	oscillator strengths	energy[a]	oscillator strengths
$1^1A_1 \rightarrow 2^1A_1$ ($\pi \rightarrow \pi^*$)	5.56	0.062	5.14	0.088	4.42	0.100
$1^1A_1 \rightarrow 1^1B_2$ ($\pi \rightarrow \pi^*$)	5.60	0.070	5.36	0.059	5.06	0.032
$1^1A_1 \rightarrow 1^1B_1$ ($\pi \rightarrow p_y$)	5.73	0.0035	5.21	0.0022	4.41	0.015
$1^1A_1 \rightarrow 1^1A_2$ ($\pi \rightarrow p_y$)	6.03	0.00	5.28	0.00	4.06	0.00
$1^1A_1 \rightarrow 2^1A_2$ ($\pi \rightarrow s$)	6.22	0.00	6.22	0.00	5.90	0.00
$1^1A_1 \rightarrow 2^1B_1$ ($\pi \rightarrow s$)	6.53	3.69×10^{-7}	6.27	0.0033	5.50	0.0065
$1^1A_1 \rightarrow 3^1A_2$ ($\pi \rightarrow p_z$)	6.98	0.00	6.49	0.00	6.63	0.00
$1^1A_1 \rightarrow 3^1B_1$ ($\sigma \rightarrow \pi^*$)			6.59	0.0065	5.92	5.96×10^{-5}
$1^1A_1 \rightarrow 4^1B_1$ ($\pi \rightarrow p_z$)			6.68	0.0053	6.26	0.087
$1^1A_1 \rightarrow 2^1B_2$ ($\pi \rightarrow \pi^*$)			6.82	0.065	6.08	0.0069
$1^1A_1 \rightarrow 3^1A_1$ ($\pi \rightarrow \pi^*$)			6.90	0.11	6.58	0.026
$1^1A_1 \rightarrow 4^1A_2$ ($\pi \rightarrow d$)					5.99	0.00
$1^1A_1 \rightarrow 6^1A_2$ ($\pi \rightarrow p_z$)					6.24	0.00
$1^1A_1 \rightarrow 5^1B_1$ ($\pi \rightarrow d$)					5.71	0.0031
$1^1A_1 \rightarrow 5^1A_2$ ($\sigma \rightarrow \pi^*$)					6.40	0.00
$1^1A_1 \rightarrow 6^1B_1$ ($\pi \rightarrow p_y$)					6.65	0.019
$1^1A_1 \rightarrow 7^1B_1$ ($\pi \rightarrow d$)					6.71	0.026
$1^1A_1 \rightarrow 4^1A_1$ ($\pi \rightarrow p_x$)					6.77	0.040
$1^1A_1 \rightarrow 3^1B_2$ ($\sigma \rightarrow p_y$)					6.78	0.25

[a] Energies in eV.

thiophene and ±59 au for selenophene). The two $\mathscr{B}$ terms are not exactly equal in magnitude. Mixing between the higher-energy $\pi \rightarrow \pi^*$ states and the 2^1A_1 and 1^1B_2 states adds non-negligible negative contributions to the $\mathscr{B}$ terms of both the lower-energy $\pi \rightarrow \pi^*$ states. This results in $\mathscr{B}$ terms of the 1^1B_2 states that are larger in magnitude than those of the 2^1A_1 states and gives the slightly asymmetric form of the pseudo-$\mathscr{A}$ terms.

Although it is also close in energy to the 2^1A_1 and 1^1B_2 states, the 1^1B_1 state has a much smaller $\mathscr{B}$ term. The weaker MCD caused by this state is due to the lower intensity of the transition (smaller $\langle A|\mu|J\rangle$ term in eq 6) and weaker mixing between states induced by the magnetic field (smaller $\langle J|\mathbf{L}|K\rangle$ term in eq 6).

The situation is rather different in tellurophene. Here, the two lowest-energy $\pi \rightarrow \pi^*$ states are split in energy by more than 0.6 eV, but the lowest Rydberg state (1^1B_1) is very close in energy to the lowest $\pi \rightarrow \pi^*$ state (2^1A_1). Although they are almost degenerate, these two states induce only moderate $\mathscr{B}$ terms because the interaction as measured by $\langle J|\mathbf{L}|K\rangle$ is still small. The 2^1A_1 and 1^1B_2 states still interact strongly to produce the significant negative $\mathscr{B}$ term of the 1^1B_2 state.

Table 8. Calculated $\mathscr{B}$ Terms of Thiophene, Selenophene, and Tellurophene[a]

	thiophene		selenophene		tellurophene	
transition	$n = 50$	$n = 250$	$n = 50$	$n = 250$	$n = 50$	$n = 250$
$1^1A_1 \rightarrow 2^1A_1$ ($\pi \rightarrow \pi^*$)	309.80	309.08	47.25	45.24	−10.42	−11.62
$1^1A_1 \rightarrow 1^1B_2$ ($\pi \rightarrow \pi^*$)	−336.71	−338.70	−66.17	−68.15	−22.36	−23.54
$1^1A_1 \rightarrow 1^1B_1$ ($\pi \rightarrow p_y$)	6.16	6.07	−2.85	−3.09	13.08	12.94
$1^1A_1 \rightarrow 1^1A_2$ ($\pi \rightarrow p_y$)	0.00	0.00	0.00	0.00	0.00	0.00
$1^1A_1 \rightarrow 2^1A_2$ ($\pi \rightarrow s$)	0.00	0.00	0.00	0.00	0.00	0.00
$1^1A_1 \rightarrow 2^1B_1$ ($\pi \rightarrow s$)	−0.008	−0.008	−1.14	−1.04	−2.15	−1.90
$1^1A_1 \rightarrow 3^1A_2$ ($\pi \rightarrow p_z$)	0.00	0.0	0.00	0.00	0.00	0.00
$1^1A_1 \rightarrow 3^1B_1$ ($\sigma \rightarrow \pi^*$)			1.14	1.37	−0.28	−0.17
$1^1A_1 \rightarrow 4^1B_1$ ($\pi \rightarrow p_z$)			0.52	−0.13	0.24	0.18
$1^1A_1 \rightarrow 2^1B_2$ ($\pi \rightarrow \pi^*$)			20.66	18.68	−1.51	−1.40
$1^1A_1 \rightarrow 3^1A_1$ ($\pi \rightarrow \pi^*$)			−15.26	−22.10	−10.89	−13.65
$1^1A_1 \rightarrow 4^1A_2$ ($\pi \rightarrow d$)					0.00	0.00
$1^1A_1 \rightarrow 6^1A_2$ ($\pi \rightarrow p_z$)					6.40	6.79
$1^1A_1 \rightarrow 5^1B_1$ ($\pi \rightarrow d$)					0.00	0.00
$1^1A_1 \rightarrow 5^1A_2$ ($\sigma \rightarrow \pi^*$)					0.00	0.00
$1^1A_1 \rightarrow 6^1B_1$ ($\pi \rightarrow p_y$)					18.29	17.32
$1^1A_1 \rightarrow 7^1B_1$ ($\pi \rightarrow d$)					−25.98	−25.02
$1^1A_1 \rightarrow 4^1A_1$ ($\pi \rightarrow p_x$)					500.18	504.66
$1^1A_1 \rightarrow 3^1B_2$ ($\sigma \rightarrow p_y$)					−495.12	−503.34

[a] Energies in eV.

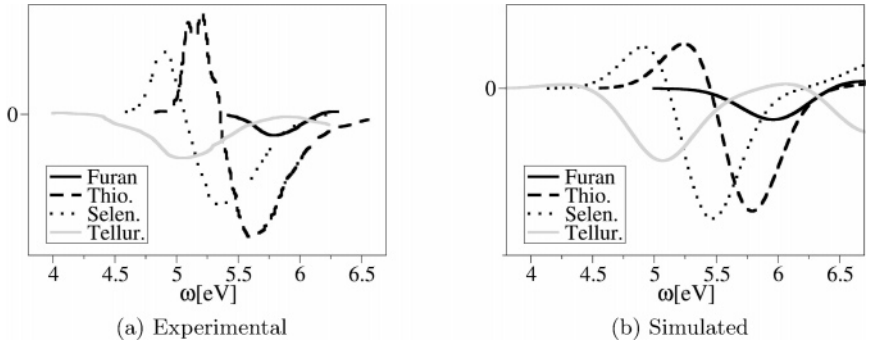


Figure 2. Experimental and simulated MCD spectra of furan, thiophene, selenophene, and tellurophene.

The higher $\pi \rightarrow \pi^*$ states again make non-negligible negative contributions to the $\mathscr{B}$ terms of the 2^1A_1 and 1^1B_2 states.

The negative MCD from the $\mathscr{B}$ terms of the 2^1A_1 and 1^1B_2 states of tellurophene nearly but not completely cancel that from the positive $\mathscr{B}$ term of the 1^1B_1 state, giving the small positive shoulder at low energies in the MCD spectrum of tellurophene (Figure 1d). At higher energies, the MCD of the two $\pi \rightarrow \pi^*$ transitions reinforce each other and a strong negative band in the MCD spectrum is observed. The trend toward negative MCD at higher energies is due to the significant negative $\mathscr{B}$ term of the second $\pi \rightarrow \pi^*$ state with A_1 symmetry that is predicted to be found just beyond the energy range of the experimental MCD spectrum.

The simulated MCD spectra of thiophene, selenophene, and tellurophene utilized $\mathscr{B}$-term calculations including 250 states in the perturbation expansions. Also listed in Table 8 are the results of calculations utilizing only the lowest 50 states. As was the case with furan, it appears that the much smaller 50-state calculations give similar results to the large 250-state calculations.

The discussion so far has focused on how well the calculations reproduce trends within a spectrum. The MCD sign, band shapes, and band maxima locations of simulated spectra have been compared with the corresponding experimental spectrum. What has been considered in much less detail is whether trends across the spectra and in particular the intensity of one spectrum as compared with another. We will discuss this issue now. The experimental MCD spectra of furan, thiophene, selenophene, and tellurophene are collected in one place in Figure 2a. The spectra are all similar in intensity in the sense that all of the major band maxima fall within 1 order of magnitude. If the greatest MCD (negative or positive) of the furan spectrum is taken as 1, the greatest MCD of the spectra follows the approximate ratio 1:5.7:4.3:2.0 for furan/thiophene/selenophene/tellurophene. The equivalent comparison of the simulated spectra can be found in Figure 2b. The data in this figure have of course not been rescaled and are derived directly from the calculated $\mathscr{B}$ terms. Figure 2a and b are remarkably similar. The only major discrepancy is that the simulated thiophene spectrum is somewhat too low in intensity as compared to the other three. The simulated spectra have a furan/thiophene/selenophene/tellurophene ratio of 1:3.9:4.1:2.3. These results suggest that not only can the features of individual MCD spectra be reproduced but so can the relationships between spectra of similar compounds.

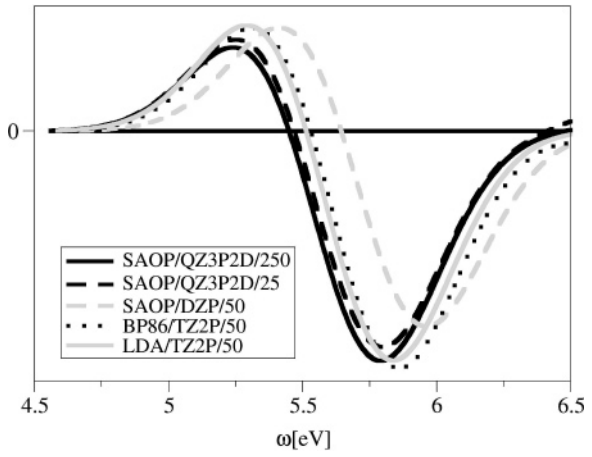


Figure 3. Simulated MCD spectra of thiophene with different functionals and basis sets. "250", "50", or "25" refers to the number of states used in the $\mathscr{B}$-term expansion.

The MCD spectra of thiophene and selenophene are examples of spectra that are ideally suited to be calculated with the approach described here. Two states have large, oppositely signed $\mathscr{B}$ terms that are produced by the mixing of a small number of low-energy states. It would therefore be expected that a TDDFT calculation involving only relatively small numbers of transitions would be required to give a reasonable simulation of the spectrum and to aid in the interpretation of the experimental spectra. In the case of thiophene and selenophene, the situation is further simplified because the two important excited states both arise because of valence excitations. Therefore, it should be possible to describe transitions to these states with TDDFT calculations involving standard generalized gradient approximations and simple basis sets rather than asymptotically corrected functionals and extensive basis sets that must include diffuse functions. To illustrate this point, we present the MCD spectrum of thiophene obtained using parameters derived from calculations with smaller numbers of states included in the expansion of eq 6, with smaller basis sets and with nonasymptotically corrected functionals (Figure 3). In this figure, BP86 is the combination of Becke's 1988 correlation functional and Perdew's 1986 exchange functional,[20−22] LDA indicates the use of the simple local density approximation, and DZP indicates a valence double-ζ basis set with a single set of polarization functions on each atom. All the spectra in Figure 3 are very similar to each other. The agreement is especially remarkable given the range of values obtained in these calculations for the $\mathscr{B}$ terms of the important two $\pi \rightarrow \pi^*$ transitions. For instance, the $\mathscr{B}$ terms obtained in the SAOP/QZ3P2D/250 calculation are on the order of $\pm$300 au, while those obtained from the SAOP/DZP/50 calculation are on the order of $\pm$100 au. The results were not rescaled, so all calculations produce a simulated MCD spectrum with very similar intensities as well as shape.

3.3. Pyrrole, *N*-Methylpyrrole, and 2,5-Dimethylpyrrole. Thulstrup and co-workers have also examined the MCD spectra of pyrrole and several of its derivatives including *N*-methylpyrrole, 1,2-dimethylpyrrole, 2,4-dimethylpyrrole, 2,5-dimethylpyrrole, and 1,2,5-trimethylpyrrole.[40] The absorption spectra of these six compounds in the range 5−6.5 eV are all very similar to each other and consist of a broad peak at around 6 eV and increasing absorption intensity at higher energies.[40]

In contrast, the MCD spectra of pyrrole and its derivatives show considerable variation. The spectrum of pyrrole between 5.5 and 6.5 eV shows only negative MCD and one peak. The MCD spectra of 1,2-dimethylpyrrole, 2,4-dimethylpyrrole, and 2,5-dimethylpyrrole have negative MCD at low energies, positive MCD at around 6 eV, and then negative MCD again at higher energies. Finally, the MCD spectra of *N*-methylpyrrole and 1,2,5-trimethylpyrrole include a positive MCD peak below 6 eV and negative MCD at higher energies.

In this section, we present calculations of the MCD spectra of pyrrole, *N*-methylpyrrole, and 2,5-dimethylpyrrole, molecules that cover the three types of MCD spectra mentioned above. The calculated transition energies and oscillator strengths of all three molecules are listed in Table 9.

As might be expected, the results for pyrrole in Table 9 share many features with the previous furan calculations. Two $\pi \rightarrow \pi^*$ transitions appear in the region of interest: the $1^1A_1 \rightarrow 1^1B_2$ transition, which is predicted to have a large oscillator strength, and the $1^1A_1 \rightarrow 2^1A_1$ transition, which is predicted to be very weak in intensity. These two states are calculated to be nearly degenerate, which will certainly have consequences for the predicted MCD spectrum of pyrrole. A number of Rydberg transitions of low intensity are also predicted to lie between 5 and 6.5 eV.

There is considerable disagreement among theoretical calculations dealing with the band in the absorption spectrum of pyrrole centered at 6 eV.[41,48,52,55−58] It is generally agreed that the lowest-energy peaks in this band are due to the $1^1A_1 \rightarrow 1^1A_2$ transition with a vertical excitation energy similar to that obtained in the present work. The most sophisticated and recent studies also agree for the most part on the location and assignment of most of the other Rydberg transitions. There is still disagreement with respect to the $\pi \rightarrow \pi^*$ transitions. In particular, there is disagreement as to the ordering of these two states, their energies, and whether they are pure $\pi \rightarrow \pi^*$ in character or whether some mixing of Rydberg character occurs. A detailed discussion of the controversy is beyond the scope of this work, but it appears that our results are most in line with the most recent CASPT2 studies of Roos and co-workers.[56] The CASPT2 transition energies from ref 56 are somewhat lower than those given here, but this study predicted that the vertical transition energies to the 1^1B_2 and 2^1A_1 states would be very similar, that the $1^1A_1 \rightarrow 1^1B_2$ transition would have a much higher oscillator strength than the other nearby transitions, and that there would be little $\pi \rightarrow \pi^*$ and Rydberg mixing. Similar conclusions could be drawn from the results of Table 9. The order of the 1^1B_2 and 2^1A_1 states given by Roos and co-workers is the reverse of that presented here, but these two states are nearly degenerate, and their ordering will be very difficult to pin down with confidence.

The transition energies and oscillator strengths calculated for *N*-methylpyrrole and 2,5-dimethylpyrrole follow a similar pattern to those of pyrrole. One of the lower-energy $\pi \rightarrow \pi^*$ transitions has a much greater intensity than all of the

Table 9. Transition Energies, Oscillator Strengths, and $\mathscr{B}$ Terms of Pyrrole, *N*-Methylpyrrole, and 2,5-Dimethylpyrrole[a]

transition	energy[a]	oscillator strengths	*B*	
			$n = 50$	$n = 250$
Pyrrole				
$1^1A_1 \rightarrow 1^1A_2$ ($\pi \rightarrow$ 3s)	5.26	0.00	0.00	0.00
$1^1A_1 \rightarrow 1^1B_1$ ($\pi \rightarrow$ 3s)	6.10	7.16×10^{-6}	0.011	0.022
$1^1A_1 \rightarrow 1^1B_2$ ($\pi \rightarrow \pi^*$)	6.20	0.19	89.49	89.01
$1^1A_1 \rightarrow 1^1A_1$ ($\pi \rightarrow \pi^*$)	6.21	1.21×10^{-4}	−91.17	−91.18
$1^1A_1 \rightarrow 2^1A_2$ ($\pi \rightarrow 3p_z$)	6.22	0.00	0.00	0.00
$1^1A_1 \rightarrow 2^1B_1$ ($\pi \rightarrow 3p_y$)	6.33	0.030	−13.37	−13.58
$1^1A_1 \rightarrow 3^1A_2$ ($\pi \rightarrow$ 3d)	6.76	0.00	0.00	0.00
$1^1A_1 \rightarrow 2^1B_1$ ($\pi \rightarrow$ 3d)	6.88	0.0018	1.77	1.70
$1^1A_1 \rightarrow 2^1B_2$ ($\pi \rightarrow 3p_x$)	6.90	0.0051	−6.61	−6.31
***N*-Methylpyrrole**				
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3s)	5.04	4.53×10^{-4}	0.009	0.000
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow \pi^*$)	5.59	0.14	10.84	10.42
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3s/3p$_z$)	5.72	0.013	1.57	1.61
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ 3s)	5.72	0.010	2.65	2.14
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow \pi^*$)	5.85	0.013	−16.49	−16.75
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow 3p_y$)	5.94	0.021	−15.87	−15.93
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ 3s/3p$_z$)	6.39	0.0056	−1.34	−1.56
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3d)	6.44	0.0048	5.07	5.04
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ 3d)	6.52	0.028	−22.36	−22.99
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3d)	6.55	0.0011	5.99	5.97
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow 3p_y$)	6.61	2.52×10^{-5}	0.02	0.04
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow 3p_x$)	6.77	0.0063	25.27	25.80
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow \pi^*$)	6.85	0.24	−57.80	−61.19
2,5-Dimethylpyrrole				
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3s)	4.60	1.17×10^{-4}	0.044	0.034
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow 3p_y$)	5.09	9.39×10^{-4}	1.72	1.23
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow \pi^*$/3s/3p$_z$)	5.46	0.13	−9.65	−9.34
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3s/3p$_z$)	5.66	0.052	−10.18	−10.72
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ 3s/π^*)	5.75	5.24×10^{-4}	−0.99	−1.00
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3s/3p$_z$)	5.76	0.020	−14.77	−15.09
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow \pi^*$/3s)	5.77	0.0069	33.15	33.17
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ 3d)	6.24	0.0056	7.93	7.75
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow 3p_y$)	6.24	8.29×10^{-4}	4.19	4.20
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow 3p_x$)	6.25	0.031	−8.24	−8.63
$1^1A' \rightarrow 1A''$ ($\pi \rightarrow$ 3d)	6.42	0.077	−12.70	−14.39
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow \pi^*$)	6.48	0.11	5.96	4.21
$1^1A' \rightarrow 1A''$	6.56	6.08×10^{-4}	−1.40	−1.37
$1^1A' \rightarrow 1A'$	6.73	0.024	5.51	5.61
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ p$_z$/s)	6.79	0.012	−9.36	−9.39
$1^1A' \rightarrow 1A'$ ($\pi \rightarrow$ p$_z$/s)	6.89	0.0077	3.64	3.45
$1^1A' \rightarrow 1A''$	6.93	0.012	−10.86	−11.71

[a] Energies in eV.

other Rydberg transitions. The transition energies decrease in magnitude going from pyrrole to *N*-methylpyrrole to 2,5-dimethylpyrrole so that further Rydberg states and an additional $\pi \rightarrow \pi^*$ valence transition are predicted to be found below 7 eV in the spectra of *N*-methylpyrrole and 2,5-dimethylpyrrole. The low-energy $\pi \rightarrow \pi^*$ transitions of these molecules are less degenerate than was the case for pyrrole. The results for 2,5-dimethylpyrrole are complicated somewhat by the greater number of states found below 7 eV and significant $\pi \rightarrow \pi^*$ and valence mixing.

The calculated MCD $\mathscr{B}$ terms of pyrrole, *N*-methylpyrrole, and 2,5-dimethylpyrrole are also listed in Table 9, and the simulated MCD spectra obtained from these parameters are compared with experimental results and are presented in Figure 4a−c.

The most striking result from the simulated spectra is that they reproduce the qualitative changes observed when going from pyrrole to *N*-methylpyrrole to 2,5-dimethylpyrrole.

For pyrrole, the presence of two nearly degenerate $\pi \rightarrow \pi^*$ transitions suggests that a pseudo-$\mathscr{A}$ term feature would be observed and the transitions to these two states have large oppositely signed $\mathscr{B}$ terms that are consistent with this observation. However, the $\pi \rightarrow 3p_y$ Rydberg transition that is also nearby in energy has a non-negligible negative $\mathscr{B}$ term that is large enough to completely cancel the positive MCD of the $\pi \rightarrow \pi^*$ ($1^1A_1 \rightarrow 1^1B_2$) transition. Although the *B* term of the Rydberg transition is rather smaller than those of the $\pi \rightarrow \pi^*$ transitions, it must be recalled that the positive and negative MCDs of the latter two transitions largely cancel, allowing the smaller MCD of the Rydberg transition to be influential.

As would be expected, the MCD of the two $\pi \rightarrow \pi^*$ transitions arises mostly from mixing among themselves. The largest contribution to the $\mathscr{B}$ term of the $\pi \rightarrow 3p_y$ transition is caused by mixing with the more intense of the two $\pi \rightarrow \pi^*$ states ($1^1A_1 \rightarrow 1^1B_2$).

The important contributions to the MCD spectrum of *N*-methylpyrrole are similar to those of pyrrole. The lowest-energy $\pi \rightarrow \pi^*$ transition provides positive MCD, while the second $\pi \rightarrow \pi^*$ and the $\pi \rightarrow 3p_y$ have significant negative $\mathscr{B}$ terms. The difference in energy between the two $\pi \rightarrow \pi^*$ transitions is greater and the $\pi \rightarrow 3p_y$ transition is further away in energy from the lowest $\pi \rightarrow \pi^*$ transition than was the case for pyrrole. The cancellation of the positive MCD of the $\pi \rightarrow \pi^*$ transition is now incomplete, leading to the positive band observed at around 5.3 eV in Figure 4b. The positive band in the simulated spectrum is somewhat weaker than is observed experimentally, suggesting that the calculated positive $\mathscr{B}$ term is too small, the negative $\mathscr{B}$ terms are too large, or the separation between the $\pi \rightarrow \pi^*$ transitions is underestimated.

The qualitative shape of the simulated MCD spectrum of 2,5-dimethylpyrrole is correct. The computed spectrum has negative MCD at the lowest energies, followed by positive MCD at higher energies and negative MCD at still higher energies, similarly to the experimental spectrum. The band minima and maximum of the simulated spectrum are shifted to higher energies, and the low-energy minimum is somewhat more intense compared with experimental results. From Table 9, it is apparent that several states contribute to the MCD observed in this region. It is therefore very difficult to discern why the simulated spectrum is in error as a number of possible reasons can be proposed, such as overestimation of the negative $\mathscr{B}$ term of the lower-energy transitions, errors in the calculated transition energies, and so on.

In the cases of pyrrole and *N*-methylpyrrole, the majority of the observed MCD was caused by the mixing of three states, the two lowest $\pi \rightarrow \pi^*$ states and the $\pi \rightarrow 3p_y$ state. The two lowest $\pi \rightarrow \pi^*$ states still make significant contributions to the MCD of 2,5-dimethylpyrrole, but the $\pi \rightarrow 3p_y$ transition only has a small $\mathscr{B}$ term. In contrast to the other two nitrogen-containing heterocycles discussed here,

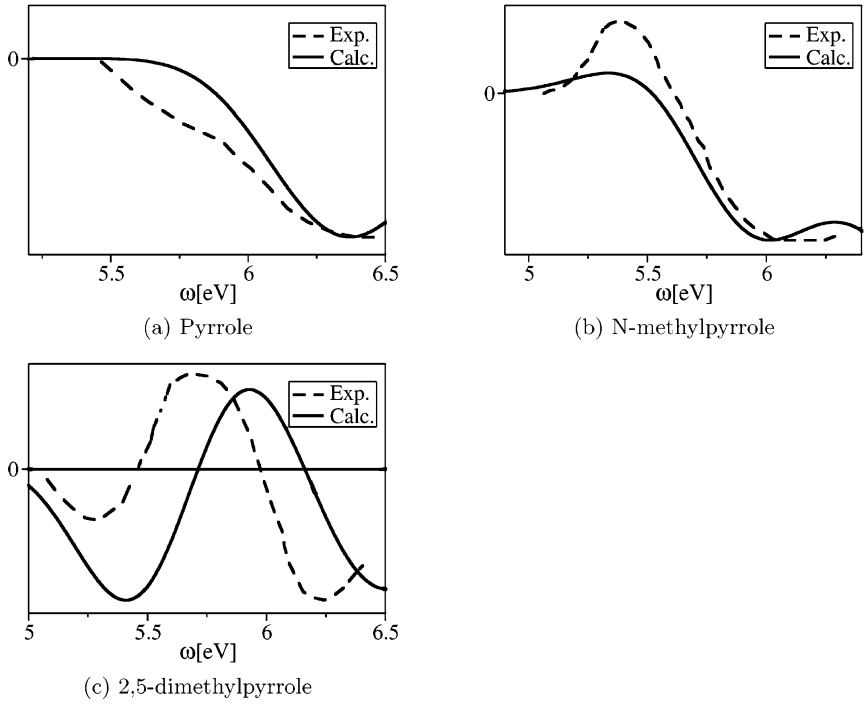


Figure 4. Experimental and simulated MCD spectra of pyrrole, *N*-methylpyrrole, and 2,5-dimethylpyrrole.

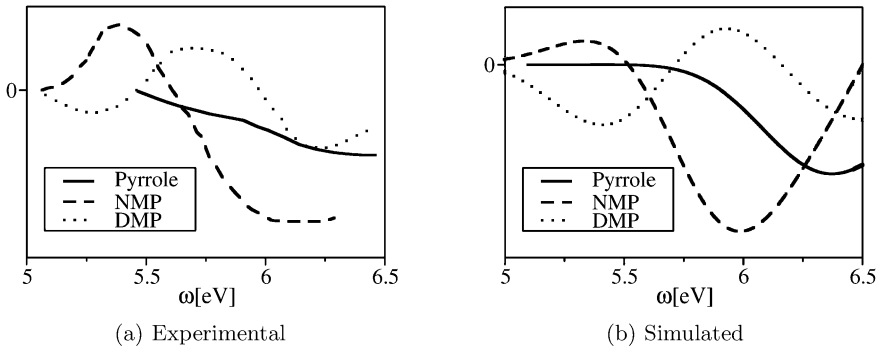


Figure 5. Experimental and simulated MCD spectra pyrrole, *N*-methylpyrrole, and 2,5-dimethylpyrrole.

the lower of the two $\pi \rightarrow \pi^*$ transitions has a negative $\mathscr{B}$ term and the higher has a positive $\mathscr{B}$ term. Several other transitions in the same region make significant contributions to the MCD, but the reversal in sign of the $\mathscr{B}$ terms of the two $\pi \rightarrow \pi^*$ transitions is chiefly responsible for the qualitative change in the spectrum when going from *N*-methylpyrrole to 2,5-dimethylpyrrole. In terms of eq 6, the changes in sign of the $\pi \rightarrow \pi^*$ $\mathscr{B}$ terms arise because the integral $\langle J|\mathbf{L}|K\rangle$ changes sign. Thus, mixing of the lower-energy $\pi \rightarrow \pi^*$ state into the higher-energy $\pi \rightarrow \pi^*$ state makes a positive contribution to the $\mathscr{B}$ term of the higher state rather than a negative contribution.

The experimental MCD spectra of pyrrole, *N*-methylpyrrole, and 2,5-dimethylpyrrole are placed alongside each other in Figure 5a, and the calculated spectra of the same three molecules are compared in Figure 5b. Much like the furan and homologues case, the calculated spectra do a reasonable job of reproducing the relative intensities of the three spectra.

4. Conclusion

In this paper, an approach for calculating the $\mathscr{B}$ terms of MCD spectroscopy was described. The method makes use of the standard formula for the $\mathscr{B}$ term[2] and takes all of the necessary integrals and energies from a TDDFT calculation. As such, it should be possible to calculate the $\mathscr{B}$ terms of any molecule for which a TDDFT calculation can be performed. Once the TDDFT calculation has been performed, the effort required to calculate the $\mathscr{B}$ term scales as the square of the number of transitions obtained and linearly with the size of the transition vectors. As such, $\mathscr{B}$-term calculations can become quite computationally expensive, especially if large numbers of states are included. Current work aimed at improving the scaling to linear in the number of transitions is in progress. The test calculations described here demonstrate that the SOS expansion of the $\mathscr{B}$ term often converges quite quickly, allowing small TDDFT calculations to be performed. MCD spectra that are dominated by pseudo-$\mathscr{A}$ terms arising because of near degeneracies, like that of thiophene, are expected to be described particularly well with this approach as they require small SOS expansions to give a good description of the spectrum.

It must be recalled that the present approach relies heavily on a reasonable starting point from a TDDFT calculation. If the TDDFT calculation is poor because of an inappropriate

choice of functional or a limited basis set, then the chances of obtaining a reasonable $\mathscr{B}$ term are not good.

The $\mathscr{B}$ terms of nine molecules, ethene, propene, furan and three of its homologues, and pyrrole and two of its derivatives, were calculated. Like other calculations, the present work was unable to find the source of the third observed MCD band in the spectra of ethene and propene. On the other hand, almost all of the features of the MCD spectra of the aromatic heterocycles could be reproduced, including the qualitative form of the spectra, many of the quantitative aspects of the spectra, and the trends in MCD intensity between spectra. These closely related molecules show quite a bit of variation in their MCD spectra, all of which could be reproduced and explained.

With the completion of this work, we have now demonstrated how to calculate all contributions to the MCD spectrum assuming that spin−orbit coupling is small and that the MCD varies linearly with the magnetic field. This covers a large number of published spectra, but many modern applications of MCD spectroscopy focus on applications where nonlinearity and spin−orbit coupling cannot be neglected. In a forthcoming publication, we shall describe our attempts to introduce spin−orbit coupling into our MCD formulation.

Acknowledgment. T.Z. would like to thank the Canadian government for a Canada research chair in theoretical inorganic chemistry. J.A. acknowledges financial support from the CAREER program of the National Science Foundation (CHE 0447321).

References

(1) Stephens, P. J. *Ann. Rev. Phys. Chem.* **1974**, *25*, 201−232.

(2) Piepho, S. B.; Schatz, P. N. *Group Theory in Spectroscopy With Applications to Magnetic Circular Dichroism*; John Wiley and Sons: New York, 1983.

(3) Mason, W. R. *Inorg. Chem.* **1989**, *28*, 2487−2489.

(4) Upton, A. H. P.; Williamson, B. E. *J. Phys. Chem.* **1994**, *98*, 71−76.

(5) Seth, M.; Ziegler, T.; Banerjee, A.; Autschbach, J.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Chem. Phys.* **2004**, *120*, 10942−10954.

(6) Seth, M.; Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2005**, *122*, 094112-1−094112-7.

(7) Seth, M.; Ziegler, T. *J. Chem. Phys.* **2006**, *124*, 144105-1−144105-12.

(8) Michl, J. *Tetrahedron* **1984**, *40*, 3845−3934.

(9) Coriani, S.; Jorgensen, P.; Rizzo, A.; Ruud, K.; Olsen, J. *Chem. Phys. Lett.* **1999**, *300*, 61−68.

(10) Honda, Y.; Hada, M.; Ehara, M.; Nakatsuji, H.; Downing, J.; Michl, J. *Chem. Phys. Lett.* **2002**, *355*, 219−225.

(11) Honda, Y.; Hada, M.; Ehara, M.; Nakatsuji, H.; Downing, J.; Michl, J. *J. Chem. Phys.* **2005**, *123*, 164113-1−164113-9.

(12) Casida, M. E. Time-Dependent Density Functional Response Theory for Molecules. In *Recent Advances in Density Functional Theory, Part 1*; Chang, D. P., Ed.; World Scientific: Singapore, 1995; pp 155−192.

(13) Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **2002**, *116*, 891−896.

(14) Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2002**, 7433−7447.

(15) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931−967.

(16) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41−51.

(17) Versluis, L.; Ziegler, T. *J. Chem. Phys.* **1988**, *88*, 322−328.

(18) te Velde, G.; Baerends, E. J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *44*, 7888−7903.

(19) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chim. Acta* **1998**, *99*, 391−403.

(20) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(21) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822−8824.

(22) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *34*, 7406.

(23) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1993**, *99*, 4597−4610.

(24) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1994**, *101*, 9783−9792.

(25) van Lenthe, E.; Ehlers, Æ.; Baerends, E. J. *J. Chem. Phys.* **1999**, *110*, 8943−8953.

(26) Gritsenko, O. V.; Schipper, P. R. T.; Baerends, E. J. *Chem. Phys. Lett.* **1999**, *302*, 199−207.

(27) Schipper, P. R. T.; Gritsenko, O. V.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Chem. Phys.* **2000**, *112*, 1344−1352.

(28) Brith-Lindner, M.; Allen, S. D. *Chem. Phys. Lett.* **1977**, *47*, 32−35.

(29) Syder, P. A.; Atanasova, S.; Hansen, R. W. C. *J. Phys. Chem. A* **2004**, *108*, 4194−4201.

(30) Merer, A. J.; Mulliken, R. S. *Chem. Rev.* **1969**, *69*, 639−656.

(31) Petrongolo, C.; Buenker, R. J.; Peyerimhoff, S. D. *J. Chem. Phys.* **1982**, *76*, 3655−3667.

(32) Sension, R. J.; Hudson, B. S. *J. Chem. Phys.* **1989**, *90*, 1377−1389.

(33) Serrano-Andrés, L.; Mechán, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. *J. Chem. Phys.* **1993**, *98*, 3151−3162.

(34) Ryu, J.; Hudson, B. S. *Chem. Phys. Lett.* **1995**, *245*, 448.

(35) Baeck, K. K.; Martinez, T. J. *Chem. Phys. Lett.* **2003**, *375*, 299−308.

(36) Barbatti, M.; Paier, J.; Lischka, H. *J. Chem. Phys.* **2004**, *121*, 11614−11624.

(37) Hazra, A.; Chang, H. H.; Nooijen, M. *J. Chem. Phys.* **2004**, *121*, 2125−2136.

(38) Walker, I. C.; Abuain, T. M.; Palmer, M. H.; Beveridge, A. J. *Chem. Phys.* **1986**, *109*, 269−275.

(39) Håkansson, R.; Nordén, B.; Thulstrup, E. W. *Chem. Phys. Lett.* **1977**, *50*, 305−308.

(40) Nordén, B.; Håkansson, R.; Pedersen, P. B.; Thulstrup, E. W. *Chem. Phys.* **1978**, *33*, 355−366.

(41) Hieringer, W.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Phys. Chem. A* **2002**, *106*, 10380−10390.

(42) Pickett, L. W. *J. Chem. Phys.* **1940**, *8*, 293−297.

(43) Pickett, N. J. H.; Liu, T. C. *J. Am. Chem. Soc.* **1951**, *73*, 4865−4869.

(44) Watanabe, K.; Nakyama, T. *J. Chem. Phys.* **1958**, *29*, 48−51.

(45) Sanche, L. *J. Chem. Phys.* **1979**, *71*, 4860−4882.

(46) Cooper, C. D.; Williamson, A. D.; Miller, J. C.; Compton, R. N. *J. Chem. Phys.* **1980**, *73*, 1527−1537.

(47) Roebber, J. L.; Gerrity, D. P.; Hemley, R.; Vaida, V. *Chem. Phys. Lett.* **1980**, *75*, 104−106.

(48) Serrano-Andrés, L.; Mechán, M.; Nebot-Gil, I.; Roos, B. O.; Fülscher, M. *J. Am. Chem. Soc.* **1993**, *115*, 6184−6197.

(49) Palmer, M. H.; Walker, I. C.; Ballard, C. C.; Guest, M. F. *Chem. Phys.* **1995**, *192*, 111.

(50) Trofimov, A. B.; Schirmer, J. *Chem. Phys.* **1997**, *224*, 175.

(51) Christiansen, O.; Jorgensen, P. *J. Am. Chem. Soc.* **1998**, *120*, 3423−3430.

(52) Burcl, R.; Amos, R. D.; Handy, N. C. *Chem. Phys. Lett.* **2002**, *355*, 8−18.

(53) Hazra, A.; Nooijen, M. *Int. J. Quantum Chem.* **2003**, *95*, 643−657.

(54) Gromov, E. V.; Trofimov, A. B.; Vitkovskaya, N. M.; Schirmer, J.; Köppel, H. *J. Chem. Phys.* **2003**, *119*, 737−751.

(55) Christiansen, O.; Gauss, J.; Stanton, J. F.; Jorgensen, P. *J. Chem. Phys.* **1999**, *111*, 525−537.

(56) Roos, B. O.; Malmqvist, P.-Å; Molina, V.; Serrano-Andrés, L.; Merchán, M. *J. Chem. Phys.* **2002**, *116*, 7526−7536.

(57) Palmer, M. H.; Wilson, P. J. *Mol. Phys.* **2003**, *101*, 2391−2408.

(58) Pastore, M.; Angeli, C.; Cimiraglia, R. *Chem. Phys. Lett.* **2006**, *422*, 522−528.

CT600283T

JCTC Journal of Chemical Theory and Computation

Bonding in Low-Coordinate Environments: Electronic Structure of Pseudotetrahedral Iron−Imido Complexes

Espen Tangen,[†] Jeanet Conradie,[†,‡] and Abhik Ghosh*[,†]

Center for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway, and Department of Chemistry, University of the Free State, 9300 Bloemfontein, Republic of South Africa

Received November 6, 2006

Abstract: A detailed density functional theory study of pseudotetrahedral $Fe^{III/IV}$−imido−phosphine complexes has yielded a host of new insights. The calculations confirm $d_{xy}^2 d_{x^2-y^2}^2 d_{z^2}^1$ (or $d_\delta^2 d_{\delta'}^2 d_\sigma^1$) electronic configurations for Fe^{III}−imido complexes of this type, as previously proposed, where the *z* direction may be identified with the Fe−N_{imido} vector. However, geometry optimization of a sterically unencumbered model complex indicated a bent (162°) imido linkage, in sharp contrast to the linear imido groups present in the sterically hindered complexes that have been studied experimentally. Under C_{3v} symmetry, the Fe^{III}−imido molecular orbital (MO) energy-level diagram indicates the existence of near-degenerate $^2A_1$ and ^2E states, and accordingly, the bending of the imido group appears to be ascribable to a pseudo-Jahn−Teller distortion. For Fe^{IV}−imido complexes, our calculations indicate a $d_{xy}^2 d_{x^2-y^2}^1 d_{z^2}^1$ (or $d_\delta^2 d_{\delta'}^1 d_\sigma^1$) electronic configuration, which is somewhat different from the $d_{xy}^1 d_{x^2-y^2}^1 d_{z^2}^2$ (or $d_\delta^1 d_{\delta'}^1 d_\sigma^2$) configuration proposed in the literature. Not surprisingly, for a sterically unencumbered Fe^{IV}−imido complex, the degenerate ^3E state (under C_{3v} symmetry) results in a mild Jahn−Teller distortion and a slightly bent (173°) imido linkage (on relaxing the symmetry constraint). The calculations also shed light on the surprising stability of the d_{z^2}-based MO, which points directly at the imido nitrogen, relative to the d_π-based MOs. The low-coordinate nature of the complexes—the absence of equatorial ligands and of a ligand trans with respect to the imido ligand—plays a key role in stabilizing the d_{z^2} orbital as well as the complexes as a whole. The electronic configurations of Fe^{IV}−imido porphyrins are radically different from that of the pseudotetrahedral complexes studied here, and we have speculated that these differences may well account for the nonobservation so far of Fe^{IV}−imido porphyrins.

The vast majority of oxo[1,2] and imido[3,4] complexes synthesized to date involve early transition metals with d^0, d^1 (d_δ^1), or d^2 (d_δ^2) electronic configurations. It has long been appreciated that the stability of these complexes owes largely to the absence of antibonding $d_\pi-p_\pi$ interactions.[5,6] By the same logic, multiply bonded metal−ligand units involving middle and late transition metals are not expected to be stable. Thus, $Fe^{IV}O$,[7,8] $Mn^{IV}O$,[9,10] and $Fe^{V}N$[11,12] porphyrins as well as $Fe^{V}O$ corroles[13,14] are all highly reactive species, while Fe^{IV}−imido porphyrins[15] remain unknown. In recent months and years, however, a handful of low-coordinate, late transition metal imido complexes have been synthesized and even crystallographically characterized.[4] By and large, high-quality density functional theory (DFT) calculations have accompanied these synthetic studies.[4] Interestingly, the calculations indicate that, although these complexes exhibit a strong tendency to avoid occupancy of

* Corresponding author fax: +47 77644065; e-mail: abhik@chem.uit.no.

[†] University of Tromsø.

[‡] University of the Free State.

10.1021/ct600318n CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/31/2007

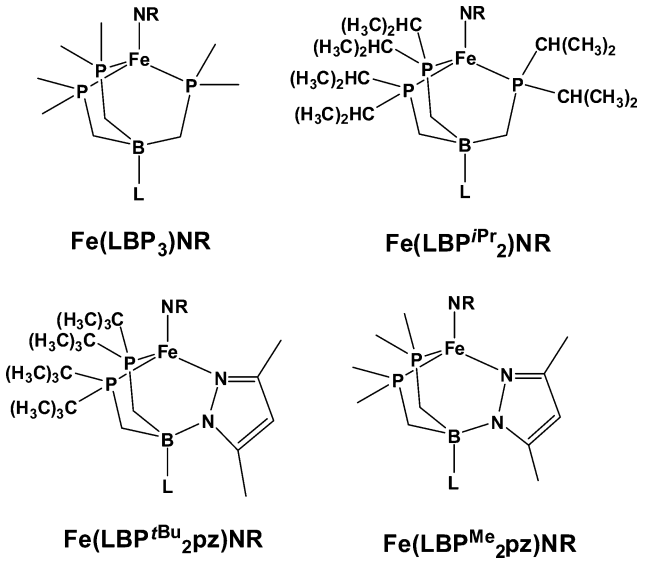


Figure 1. Different ligands considered in this study, depicted as their Fe=NR complexes. R = Me or Ad, L = Me or Ph.

the $d_\pi - p_\pi$ antibonding orbitals, a d_σ orbital, pointing directly at the imido nitrogen, is invariably occupied, either singly or doubly, in all of these complexes.[4] To understand the unexpected stability of this metal(d_σ)$-$N$_{imido}$(p_σ) antibonding orbital, we have undertaken a DFT survey of the various known classes of transition-metal$-$imido complexes. In this paper, we present the results of an aspect of our studies that has been brought to a certain conclusion, namely, an account of Fe$^{III/IV}-$imido bonding in pseudotetrahedral complexes with phosphinoborate supporting ligands.[16−19]

Figure 1 depicts the various supporting ligands used either experimentally or in our calculations. For the low-spin S = $^1/_2$ Fe$^{III}-$imido complex Fe^{III}(PhBP$^{iPr}_3$)(NAd),[18] which has a near-linear Fe$-$N$_{imido}-$C unit, DFT calculations indicated a $d_{xy}^2 d_{x^2-y^2}^2 d_{z^2}^1$ (or $d_\delta^2 d_\delta'^2 d_\sigma^1$) electronic configuration. The literature is relatively silent on why the d_{z^2} orbital, which points directly at the strongly σ-donating imido ligand, starts to fill ahead of the Fe d_π orbitals. Similarly, a $d_{xy}^1 d_{x^2-y^2}^1 d_{z^2}^2$ (or $d_\delta^1 d_\delta'^1 d_\sigma^2$) electronic configuration has been proposed for the S = 1 Fe$^{IV}-$imido complex [Fe^{IV}(PhBP$^{tBu}_2$pz)(NAd)]$^+$.[19] Although we will see that DFT calculations indicate a somewhat different $d_\delta^2 d_\delta'^1 d_\sigma^1$ configuration for this complex, the question still remains why the d_{z^2} orbital fills before the d_π orbitals. In the same vein, all known pseudotetrahedral Co$^{III}-$imido complexes exhibit S = 0 $d_{xy}^2 d_{x^2-y^2}^2 d_{z^2}^2$ (or $d_\delta^2 d_\delta'^2 d_\sigma^2$) ground states,[20,21] although in one case, an S = 1 $d_{xy}^2 d_{x^2-y^2}^2 d_{z^2}^1 d_{xz}^1$ (or $d_\delta^2 d_\delta'^2 d_\sigma^1 d_\pi^1$) excited state is thermally accessible above room temperature.[22,23] Similarly, a β-diketiminato-Co$^{III}-$imido complex also exhibits an S = 0 ground state, indicating nonoccupancy of the two d_π orbitals.[24] This, then, is the central question we sought to clarify in this study: what accounts for curious stability of the metal d_{z^2} orbital in these complexes, relative to the d_π orbitals?

Methods

All calculations were carried out using the VWN local density functional, the PW91[25] generalized gradient approximations (GGA) for both exchange and correlation, triple-ζ plus polarization Slater-type orbital basis sets, and a fine mesh for the numerical integration of matrix elements, as implemented in the ADF 2005[26] program system. As a check on the performance of the PW91 functional, the OLYP[27] GGA was also used for several calculations. In general, the PW91 GGA favors a distinctly more covalent, spin-paired description for transition-metal$-$ligand interactions, compared to OLYP.[28,29] In this study, however, both GGAs yield very similar results (geometries, spin densities, and energetics). Moreover, we did not encounter any warning signs such as unexpected bond distances or spin density profiles that might have indicated that DFT might not be a suitable approach for this study.[28,29] Accordingly, in this initial study, we have not felt the need to deploy multiconfigurational ab initio methods.[30,31]

Results

(a) Fe$^{III}-$Imido Trisphosphine Complexes. Figure 2 depicts various calculated results on the simplified, sterically unencumbered C_{3v} Fe^{III}(MeBP$_3$)(NMe) complex, while Figure 3 presents analogous results from symmetry-unconstrained optimizations of Fe^{III}(PhBP$^{iPr}_3$)(NAd), the actual molecule studied experimentally.[18] Note that the optimized geometries agree quite well with experimental results, although both PW91 and OLYP overestimate the Fe$-$N$_{imido}$ and Fe$-$P distances by about 0.02 Å. As shown in Figure 3, the optimizations confirm that the iron in Fe^{III}(PhBP$^{iPr}_3$)(NAd) exhibits close to exact C_{3v} local symmetry.[32] Accordingly, we chose to examine the simplified C_{3v} complex Fe^{III}-(MeBP$_3$)(NMe) in some detail.

Figure 2(a,b) confirms that the electronic configuration of this complex may be described as $d_{xy}^2 d_{x^2-y^2}^2 d_{z^2}^1$ (or $d_\delta^2 d_\delta'^2 d_\sigma^1$). Note that the excess spin density is almost entirely on the Fe, with a small trace of cylindrically symmetric minority spin density on the imido nitrogen. This might suggest that the singly occupied molecular orbital (SOMO) is a pure d_{z^2} orbital, but it is not: it is about 50% Fe and 12% N$_{imido}$, interacting in an antibonding manner. The reason there is no majority spin on the imido nitrogen is that it is canceled by some of the excess minority spin left there by the spatial offset between the α- and β-spin Fe(d_π)$-$N(p_π) π-bonding MOs.[33] The antibonding nature of the MO shown in Figure 2 implies an Fe$-$N$_{imido}$ bond order of 2.5, that is, 0.5 σ bonds and 2 π bonds. However, there are two interesting twists to this picture.

First, note from Figure 2a that the Fe character of the SOMO is not simply due to a d_{z^2} contribution, but also to a significant p_z contribution. The Fe p_z contribution has the effect that the "top" lobe of the d_{z^2} orbital in Figure 2a is shrunken, while the "bottom" lobe is correspondingly swollen. We believe that this specific topology of the d_{z^2} orbital goes a long way toward minimizing the Fe$-$N$_{imido}$ σ-antibonding interaction. Second, note that the top "green" lobe of the Fe d_{z^2} orbital is nearly enveloped by the equatorial "magenta" lobe (of the same orbital) as well as by the

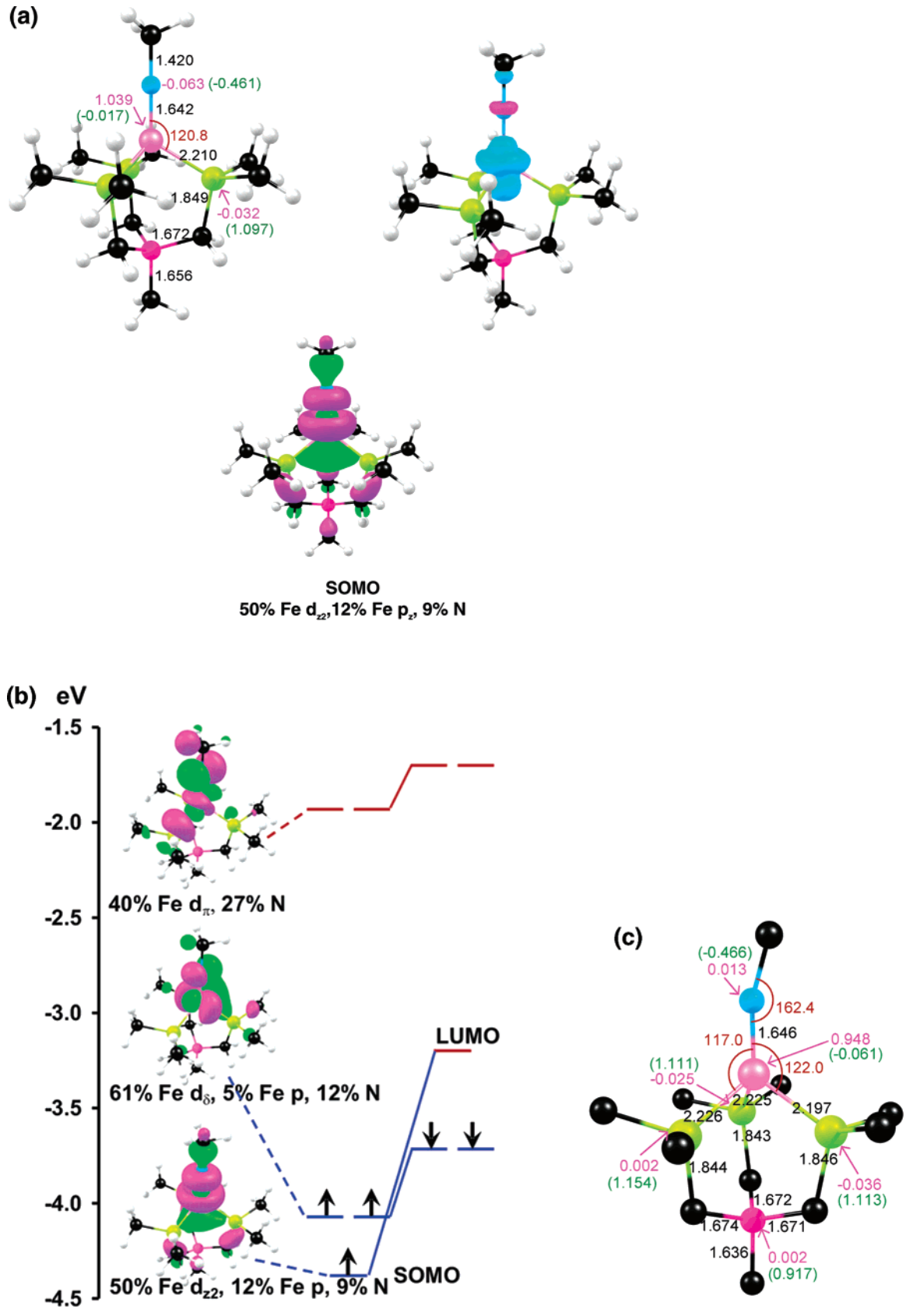


Figure 2. Selected results for $Fe^{III}(MeBP_3)(NMe)$. (a) Results from a C_{3v} symmetry-constrained optimization. Top left: Optimized distances (Å, in black), angles (deg, red), Mulliken spin populations (magenta), and charges (green, in parentheses); right: a plot of the spin density, with majority and minority spin densities indicated in cyan and magenta, respectively; bottom: a plot of the SOMO. (b) A C_{3v} valence MO energy-level diagram. (c) The symmetry-unconstrained optimized geometry, where the hydrogen atoms have been omitted for clarity.

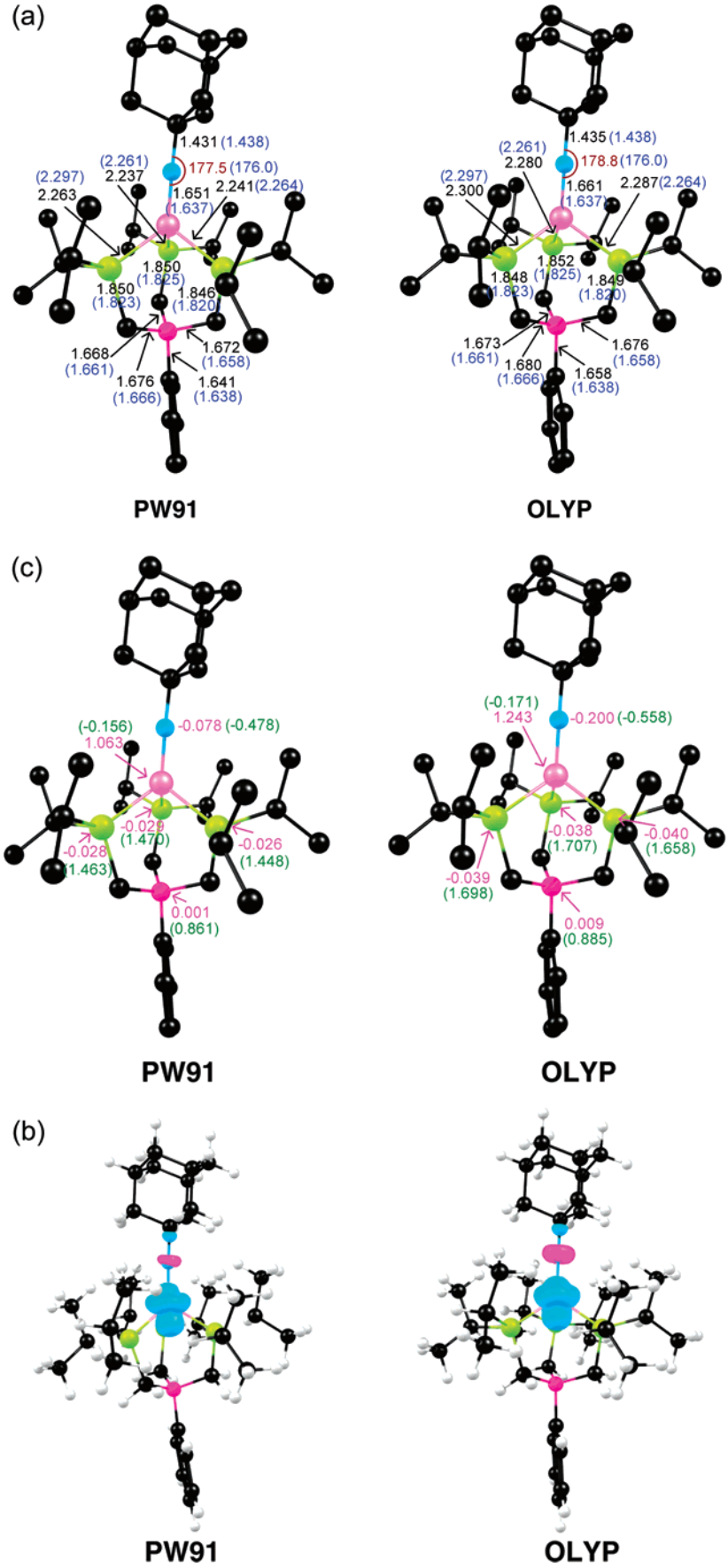


Figure 3. Selected PW91 and OLYP results from symmetry-unconstrained optimizations of $Fe^{III}(PhBP^{iPr}_3)(NAd)$, an actual molecule studied experimentally: (a) optimized distances (Å, in black), angles (deg, in red), experimental values (blue); (b) Mulliken charges (green) and spin populations (magenta); and (c) spin density profiles. For clarity, hydrogen atoms have been omitted from parts a and b.

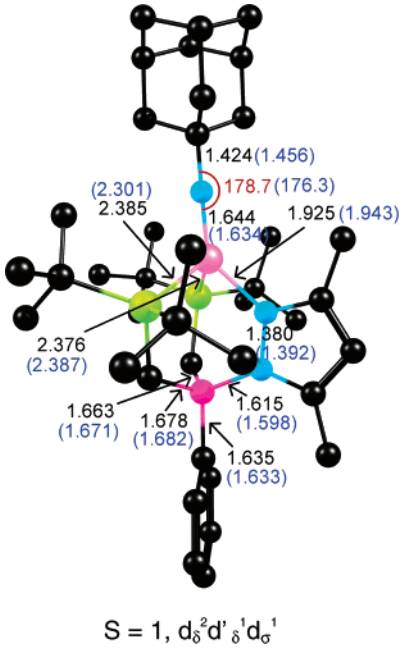


Figure 4. Selected PW91 results for [Fe^{IV}($PhBP^{tBu}_2$pz)-(NAd)]$^+$, the actual molecule studied experimentally: optimized distances (Å, in black), angles (deg, in red), and experimental values (blue). Hydrogen atoms have been omitted for clarity.

"magenta" lobe of the imido σ lone pair. This remarkable topology implies that the Fe(d_{z^2})−N(p_σ) antibonding interaction is, in reality, not particularly destabilizing, which explains the occupancy of this orbital. Accordingly, the "real" FeN bond order is somewhat above the formal value of 2.5, though it is not quite 3, as it is in an S = 0 Fe^{IV}N species (see discussion below).[34]

Besides analyzing the bonding under C_{3v} symmetry, we also carried out a PW91 symmetry-unconstrained optimization of the model complex Fe^{III}($MeBP_3$)(NMe). Key results from this calculation are shown in Figure 2c. Overall, neither the geometry nor the spin density profile change much on relaxation of the symmetry constraint, with one exception: the Fe−N_{imido}−C angle, at 162.4°, is significantly more bent in Fe^{III}($MeBP_3$)(NMe) than in the relatively sterically hindered complex Fe^{III}($PhBP^{iPr}_3$)(NAd) (see Figure 3).[18] The MO energy-level diagram in Figure 2b readily suggests a reason for this bending. Thus, note the very close spacing of the d_δ (d_{xy} and $d_{x^2-y^2}$) and d_{z2} orbitals, implying that, under C_{3v} symmetry, the $^2A_1$ and ^2E states should be very close in energy, setting the stage for a pseudo-Jahn−Teller distortion. Unfortunately, DFT being essentially a ground-state theory, we have not been able to separately converge this ^2E state. Thus, quite intriguingly, the observed near-linearity of the imido linkage[18] in Fe^{III}($PhBP^{iPr}_3$)(NAd) appears not to be an inherent feature of the Fe^{III}−imido bonding, but rather a result of a pseudo-Jahn−Teller effect suppressed by ligand steric effects.

(b) Fe^{IV}−Imido Complexes. Figure 4 compares optimized PW91 and experimental geometry parameters for [Fe^{IV}-($PhBP^{tBu}_2$pz)(NAd)]$^+$.[19] Both the metal−ligand distances and the observed linearity of the imido linkage are well-reproduced in our calculations. However, to perform a more detailed analysis of the bonding, we chose a slightly simplified C_s version of this molecule—[Fe^{IV}($PhBP^{Me}_2$pz)-(NAd)]$^+$—for which key results are shown in Figure 5. Once again, the optimized metal−ligand distances in this molecule compare well with those experimentally observed for [Fe^{IV}-($PhBP^{tBu}_2$pz)(NAd)]$^+$, and the imido linkage is essentially linear.[19] The latter strongly suggests a threefold-symmetric, Jahn−Teller-inactive electronic structure and, particularly for an S = 1 d^4 pseudotetrahedral metal ion, a $d_{xy}^1 d_{x^2-y^2}^1 d_{z^2}^2$ (or $d_\delta^1 d_\delta{}'^1 d_\sigma^2$) electronic configuration, as has indeed been proposed.[19] However, our calculations indicate a somewhat different electronic configuration.

A plot of the spin density profile of [Fe^{IV}($PhBP^{Me}_2$pz)-(NAd)]$^+$, shown in Figure 5b, provided the first clues that the electronic configuration might not be as proposed. A $d_{xy}^1 d_{x^2-y^2}^1 d_{z^2}^2$ (or $d_\delta^1 d_\delta{}'^1 d_\sigma^2$) configuration should result in a cylindrically symmetric, flattened (oblate) spheroidal blob of spin density around the iron. Instead, the blob of spin density around the iron has a distinct six-lobed shape, which appears most consistent with an unpaired d_δ electron and an unpaired d_{z^2} electron. The MO energy-level diagram shown in Figure 5c confirms this picture: the d^4 electronic configuration is accounted for by a "doubly occupied" *a′*-symmetry d_δ-based MO, a singly occupied *a″*-symmetry d_δ-based MO, and a singly occupied *a′*-symmetry d_{z^2}-based MO. The detailed views of the "open-shell MOs", given in Figure 5d, may convey a clearer sense of the orbital interaction topologies. A somewhat subtle point concerns why the *a′*-symmetry d_δ MO is "doubly occupied" (i.e., both the α- and *β*-spin MOs of this type are occupied), while the *a″*-symmetry d_δ MO is "singly occupied". The answer is that the latter MO (shown in Figure 5d) is destabilized by relatively head-on antibonding interactions with two P atoms, whereas such antibonding interactions are essentially absent in the *a′*-symmetry d_δ-based MO.

Under C_{3v} symmetry, the above electronic configuration corresponds to a Jahn−Teller-active ^3E state. Accordingly, we carried out a symmetry-unconstrained optimization of the sterically unhindered S = 1 [Fe($MeBP_3$)(NMe)]$^+$ complex, the results of which are shown in Figure 6. The FeNC angle (173°) is modestly bent, but more so than in [Fe^{IV}($PhBP^{tBu}_2$-pz)(NAd)]$^+$, the actual molecule studied experimentally.[19] As in the case of Fe^{III}−imido complexes, it appears that the sterically hindered ligands used experimentally effectively suppress the Jahn−Teller distortions that would otherwise be more pronounced. Careful examination of the optimized and crystallographic structures shows several short H···H contacts of around 2.5 Å in the NAd complexes, but none below 3.5 Å in the NMe model complexes.

(c) Fe^{IV}−Nitrido Complexes. An Fe^{IV}N trisphosphine species has also been reported and spectroscopically, but not structurally, characterized.[34] Accordingly, we modeled this species simply as Fe^{IV}($MeBP_3$)N, the calculated PW91 results being shown in Figure 7a. Not surprisingly,[34] Fe^{IV}($MeBP_3$)N exhibits a C_{3v} S = 0 ground state, corresponding to a $d_{xy}^2 d_{x^2-y^2}^2$ (or $d_\delta^2 d_\delta{}'^2$) orbital occupancy. Note the extremely short FeN distance (1.52 Å), which is even slightly shorter than that observed for an octahedral Fe^{VI}N complex (1.57),[35] reflecting the full triple-bond character of the FeN bond. Figure 7b shows that the orbital energy spacings in Fe^{IV}-

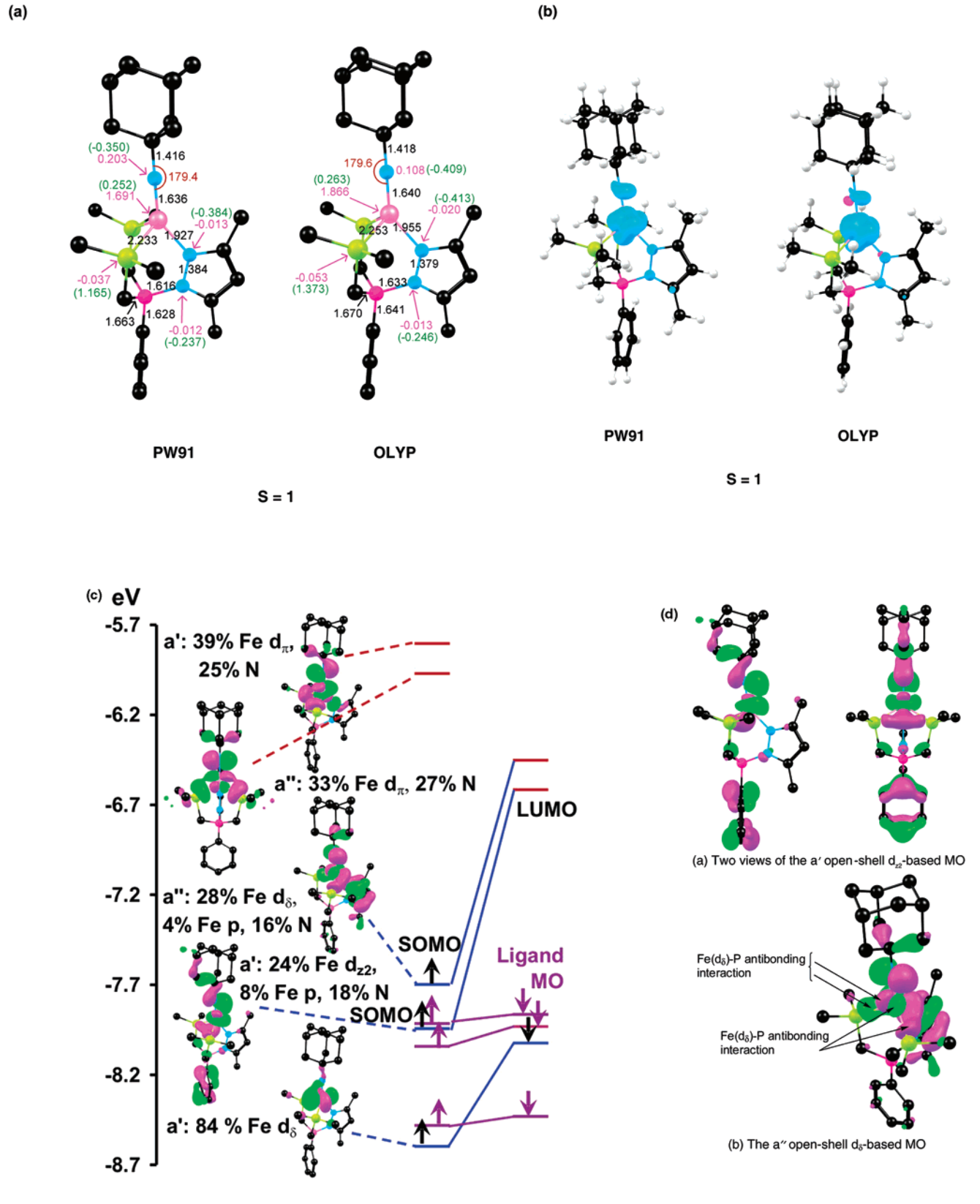


Figure 5. Selected PW91 and OLYP results for a C_s symmetry-constrained optimization of $[Fe^{IV}(PhBP^{Me}{}_2pz)(NAd)]^+$, which is a slightly simplified version of $[Fe^{IV}(PhBP^{tBu}{}_2pz)(NAd)]^+$, the actual species studied experimentally: (a) optimized distances (Å, in black), angles (deg, red), Mulliken spin populations (magenta), and charges (green, in parentheses); (b) plots of the spin density; (c) a spin-unrestricted valence MO energy-level diagram; and (d) detailed views of the two open-shell MOs. In part c, the atomic compositions of the five majority-spin d-based MOs are indicated in the same order, from top to bottom, as the corresponding MO pictures occur in the diagram. Hydrogen atoms have been omitted for clarity in parts a, c, and d.

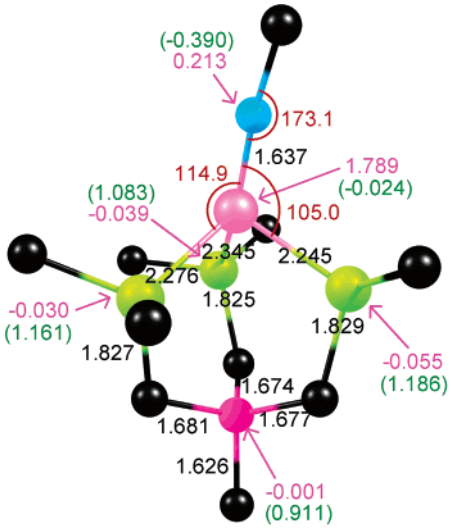


Figure 6. Selected results from a symmetry-unconstrained optimization of $Fe(MeBP_3)(NMe)]^+$: optimized distances (Å, in black), angles (deg, red), Mulliken spin populations (magenta), and charges (green, in parentheses). For clarity, hydrogen atoms have been omitted.

$(MeBP_3)N$ are very different from those in the imido complexes discussed above: specifically, the d_{z^2} orbital is very much higher in energy than the d_δ orbitals, but still lower in energy than the d_π orbitals. PW91 calculations indicate a large adiabatic singlet−triplet splitting of 0.8 eV for $Fe^{IV}(MeBP_3)N$, reflecting the high energy of the d_{z^2} orbital. These results may be contrasted with the spin-state energetics of Fe^{IV}−imido complexes. For $[Fe(MeBP_3)(NMe)]^+$, the singlet state (corresponding to a $d_{xy}^2 d_{x^2-y^2}^2$ orbital occupancy) has been found to be 0.6 eV *higher* in energy than the S = 1 ground state, according to PW91 optimizations. Likewise, for $[Fe^{IV}(PhBP^{Me}{}_2pz)(NAd)]^+$, the singlet state is *higher* in energy than the triplet ground state by 0.4 and 0.8 eV according to PW91 and OLYP optimizations, respectively.

Finally, for a high-valent transition metal complex, $Fe^{IV}(MeBP_3)N$ exhibits a very low adiabatic electron affinity of only 1.0 eV. Unlike $Fe^{III}(MeBP_3)(NMe)$, which exhibits a pronounced pseudo-Jahn−Teller distortion, the electron-added species, $[Fe^{III}(MeBP_3)N]^-$, exhibits perfect C_{3v} symmetry; the lack of a distortion reflects the fact that the $^2A_1$ state is clearly the ground state and that the lowest ^2E states are much higher in energy. As in the case of $Fe^{III}(MeBP_3)(NMe)$, the FeN bond order is still formally 2.5, but FeN π-bonding is clearly stronger with a nitride axial ligand than with an imide, so understandably, the FeN distance of 1.58 Å in $[Fe^{III}(MeBP_3)N]^-$ (see Figure 7b) is about halfway between that in $Fe^{IV}(MeBP_3)N$ (1.52 Å) and those in the imido complexes (~1.64 Å) discussed above. Another distinctive feature of the $[Fe^{III}(MeBP_3)N]^-$ anion is that the spin density is roughly evenly distributed between the Fe and the N, reflecting substantially more covalent character than in $Fe^{III}(MeBP_3)(NMe)$.

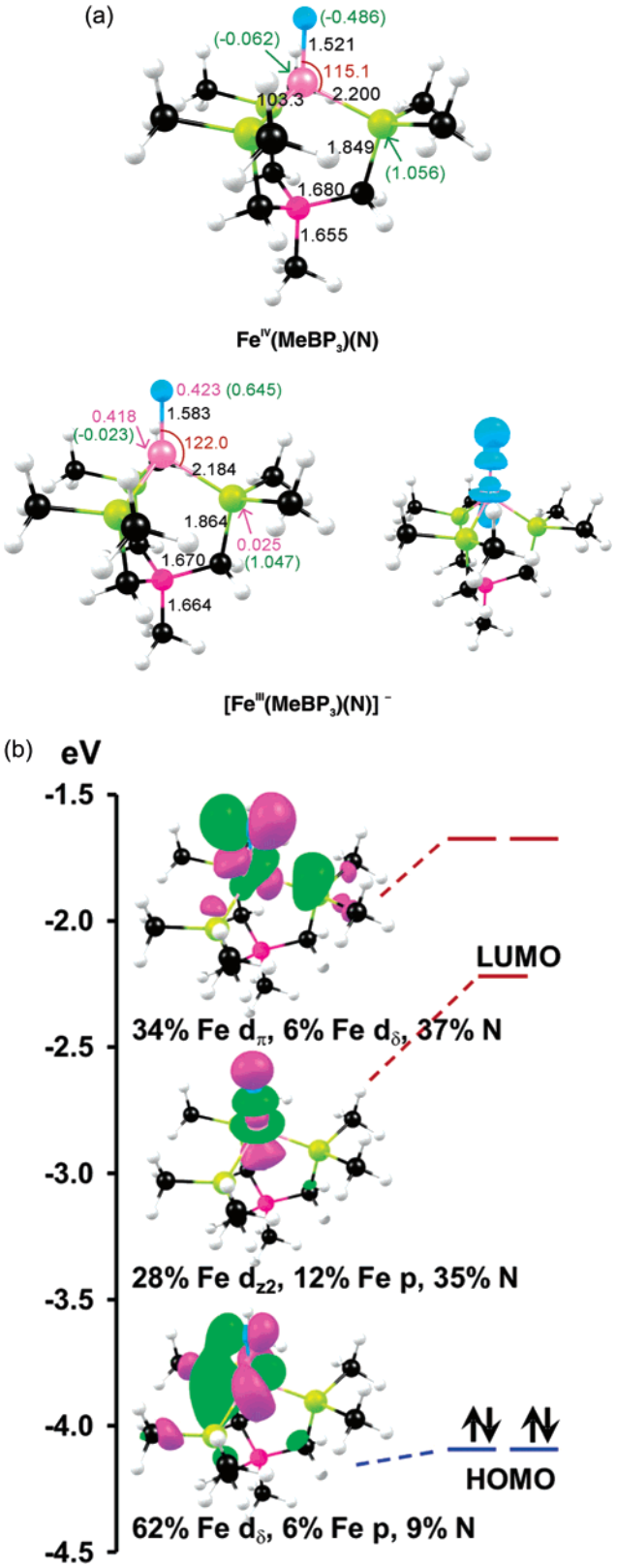


Figure 7. Selected PW91/TZP results on $Fe(MeBP_3)$−nitrido complexes: Optimized distances (Å, in black), angles (deg, red), Mulliken spin populations (magenta), and charges (green, in parentheses). (b) A valence MO energy-level diagram.

Discussion

Several aspects of the above findings need to be placed in context vis-à-vis the recent literature. An interesting point deserving of comment concerns why, unlike the Fe^{III}−imido trisphosphine complexes,[16−18] a recently reported β-diketiminato Fe^{III}−imido complex exhibits an S = $^3/_2$ ground state.

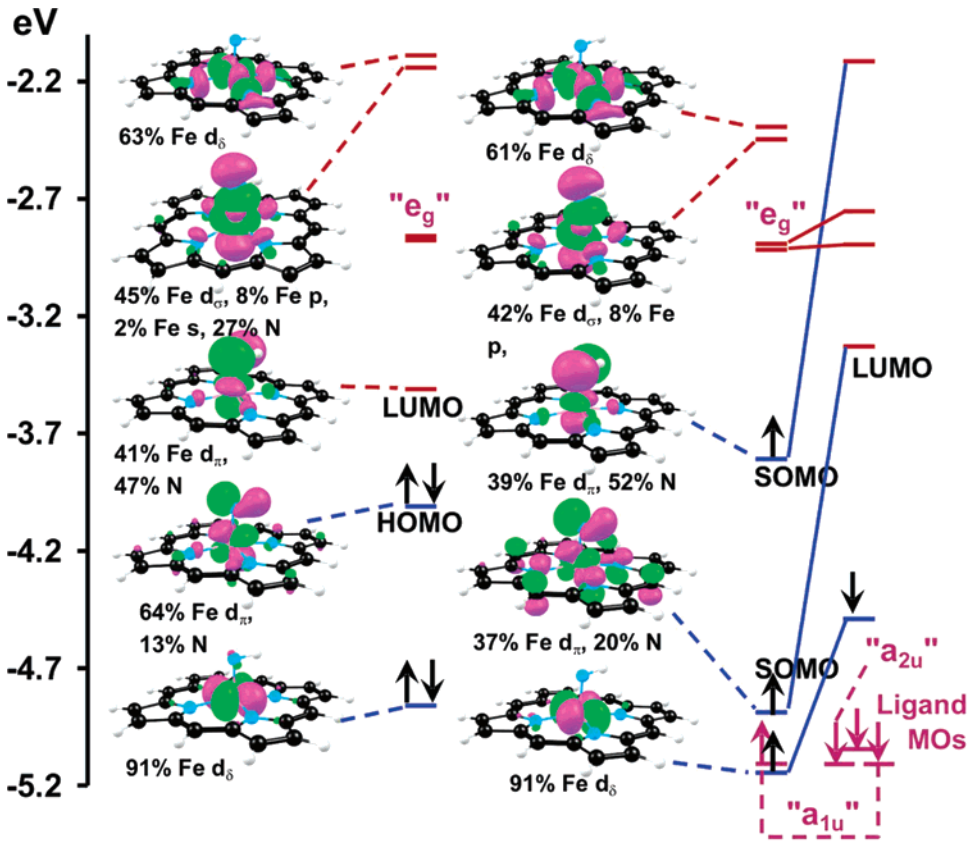


Figure 8. Frontier MO energy-level diagrams for the S = 0 (left) and 1 (right) states of $Fe^{IV}(P)(NH)$ (C_s). Selected porphyrin MOs are labeled in terms of standard D_{4h} irreps.

DFT calculations have shown that, in these trigonal-planar complexes, the "in-plane" d_π orbital (i.e., the one that is symmetric with respect to reflection across the MN_3 plane) is always of very high energy and is never occupied.[24,36,37] This orbital is destabilized not only by the metal(d_π)−N_{imido}-(p_π) antibonding interactions but also by σ-antibonding interactions involving the β-diketiminato nitrogens. However, the other d_π orbital, the one antisymmetric with respect to the MN_3 plane, does not interact much with the β-diketiminato ligand and, therefore, is not prohibitively high in energy. Nevertheless, both d_π orbitals are higher in energy than the $d_{x^2-z^2}$ orbital, which points directly at the imido nitrogen, a situation similar to that of the d_{z^2} orbital of the pseudotetrahedral complexes discussed above. These orbital spacing characteristics have the result that, while a β-diketiminato Co^{III}−imido complex avoids occupancy of *both* d_π orbitals and therefore has an S = 0 ground state,[24] an analogous Fe^{III}−imido complex only avoids the highest-energy d_π orbital and, therefore, adopts an S = $^3/_2$ ground state.[38] In contrast, both Fe d_π orbitals in the Fe^{III}−imido-phosphine complexes are high-energy unoccupied orbitals. Thus, the role of antibonding interactions involving the phosphine lone pairs in the destabilization of these d_π orbitals must not be underestimated.

Let us now return to what we view as an overarching theme of the metal−imido field, namely, that all structurally characterized middle and late transition-metal−imido complexes are low-coordinate, pseudotetrahedral, or trigonal-planar species. Iron−imido porphyrins remain unknown,[15] even after decades of intense research on metalloporphyrin-mediated catalytic processes. Indeed, to date, the sole example of an octahedral iron−imido species is a dicationic S = 1 Fe^{IV}−tosylimido intermediate, with a neutral polydentate N_5 supporting ligand. What accounts for the nonobservation so far of Fe^{IV}−imido porphyrin intermediates? Is it *simply* that the appropriate experiments have not been attempted? We believe that the answer is "no".

As alluded to above in a somewhat fragmentary fashion, low-coordinate stereochemistries facilitate metal−ligand multiple bonding involving middle and late transition metals in a number of ways. First, the lack of equatorial ligands in pseudotetrahedral and trigonal-planar metal−imido complexes results in a pair of very stable d_δ orbitals. Second, the absence of ligands both equatorial and trans to the imido group greatly stabilizes the d_{z^2} orbital. Third, as mentioned above, the d_{z^2}-based MO in imido complexes has a very special topology: the equatorial lobe of the d_{z^2} orbital seems to curve up and engage in a *bonding* interaction with the imido lone pair, thereby significantly neutralizing the head-on antibonding interaction. More formally, this special shape may be described as significant p_z character mixing into the d_{z^2} orbital. This results in an asymmetric d_{z^2} orbital with a shrunken "top" lobe and a swollen "bottom" lobe. The lack of a ligand trans to the imido group in low-coordinate complexes implies that there are no antibonding interactions to destabilize the swollen bottom lobe of the d_{z^2} orbital.

The bonding scenario in porphyrin and octahedral Fe^{IV}−imido species is radically different. The $d_{x^2-y^2}$ orbital, with its lobes pointing toward porphyrin nitrogens, is very high in energy. The equatorial "disk" of the d_{z^2} orbital too

experiences strong antibonding interactions involving the porphyrin nitrogens. Only the d_{xy} orbital is relatively stable and free of substantial antibonding interactions. Accordingly, for an iron−imido porphyrin, the d_π orbitals, of necessity, have to be (at least partially) occupied. DFT calculations with a variety of functionals on the model complex Fe^{IV}(P)(NH) predict nearly equienergetic S = 0 ($d_{xy}^2 d_{xz}^2$) and S = 1 ($d_{xy}^2 d_{xz}^1 d_{yz}$) states, with the latter very slightly lower in energy. The PW91 MO energy-level diagrams of these two states are shown in Figure 8. Of these two states, the S = 0 state in particular is expected to be highly susceptible to nucleophilic attack in view of its very low lying d_{yz} lowest unoccupied MO and a large amplitude on the imido nitrogen. Thus, even though an Fe^{IV}−imido porphyrin has an S = 1 ground state, the low-lying S = 0 state could easily provide the main channel for nucleophilic attack in processes such as aziridination, nitrene insertion, and so forth. On the other hand, triplet-state reactivity is more difficult to predict, on the basis of a MO diagram alone. However, overall, our calculations suggest that Fe^{IV}−imido porphyrins should be reactive species, considerably more so than the Fe−imidophosphine complexes discussed above, at least in part as a result of the existence of a low-lying, electrophilic S = 0 state. From this point of view, Que et al.'s recent synthesis of an S = 1 Fe^{IV}−tosylimido species[39] illustrates an ingenious approach to synthesizing an octahedral iron−imido complex: presumably, the less basic tosylimido ligand destabilizes the d_π orbitals less than an alkyl- or arylimido group would. Nevertheless, the tosylimido intermediate was found to be more reactive than the analogous $Fe^{IV}O$ species, which may (or may not) be indicative of "two-state reactivity" involving a low-lying S = 0 state. Overall, the generation and detection of an Fe^{IV}−imido or −tosylimido porphyrin intermediate remains an exciting challenge for the future.

Conclusion

A thorough DFT analysis of pseudotetrahedral $Fe^{III/IV}$−imidophosphine complexes has yielded many detailed insights, the more important of which may be summarized as follows. The complexes studied feature surprisingly low-energy, singly or doubly occupied d_{z^2} orbitals, where the *z* direction is identified with the Fe−N_{imido} axis. The low energy of this orbital appears to be due primarily to the low-coordinate nature of the complexes. The absence of equatorial ligands as well as of a ligand *trans* with respect to the imido ligand plays a key role in stabilizing the d_{z^2} orbital as well as the complexes as a whole. Moreover, certain unique topological features of the d_{z^2}-based MO ensure that the formally antibonding character of this MO is actually far less destabilizing than one might naively expect. In contrast, MO considerations suggest that iron−imido porphyrins should be substantially more reactive, at least in part because of the existence of a low-lying, electrophilic S = 0 state. Our DFT calculations also indicate a revised electronic description for the Fe^{IV}−imido complex, $[Fe^{IV}(PhBP^{tBu}{}_2pz)(NAd)]^+$.[19] Thus, instead of a $d_{xy}^1 d_{x^2-y^2}^1 d_{z^2}^2$ (or $d_\delta^1 d_\delta'^1 d_\sigma^2$) configuration, as originally proposed,[19] our calculations indicate a $d_{xy}^2 d_{x^2-y^2}^1 d_{z^2}^1$ (or $d_\delta^2 d_\delta'^1 d_\sigma^1$) configuration. Geometry optimizations with simplified trisphosphine ligands (such as $MeBP_3$) indicate that the observed linearity of the imido linkages might not be an inherent, electronically dictated phenomenon. Thus, both $Fe^{III}(MeBP_3)(NMe)$ and $[Fe^{IV}(MeBP_3)(NMe)]^+$ exhibit distinctly (albeit modestly) bent FeNC angles, 162.4° and 173.1°, respectively, which we have attributed to pseudo-Jahn−Teller and Jahn−Teller distortions, respectively. Apparently, these distortions are not seen in the experimentally studied complexes because of the sterically hindered nature of the ligands employed.[16−19] Nevertheless, the electronic near-degeneracies responsible for the bent imido units in the $MeBP_3$ complexes are important aspects of the electronic structures of pseudotetrahedral iron−imido complexes, which we should not lose sight of.

Acknowledgment. This work was supported by the Research Council of Norway, the South African National Research Foundation (grant number 2067416), and the Central Research Fund of the University of the Free State.

Supporting Information Available: Optimized Cartesian coordinates (25 pages) for the various complexes studied. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Holm, R. H. *Chem. Rev.* **1987**, *97*, 1401−1449.

(2) Shan, X. P.; Que, L., Jr. *J. Inorg. Biochem.* **2006**, *100*, 421−433

(3) Wigley, D. E. *Prog. Inorg. Chem.* **1994**, *42*, 239−482.

(4) Mehn, M. P.; Peters, J. C. *J. Inorg. Biochem.* **2006**, *100*, 634−643. See also references herein for citations to the earlier literature in this field.

(5) Mayer, J. M. *Comments Inorg. Chem.* **1988**, *8*, 125−135.

(6) Nugent, W. A.; Mayer, J. M. *Metal−Ligand Multiple Bonds*; Wiley: New York, 1998.

(7) Groves, J. T. *J. Inorg. Biochem.* **2006**, *100*, 434−447.

(8) Harris, D. L. *Curr. Opin. Chem. Biol.* **2001**, *100*, 724−735.

(9) Czernuszewicz, R. S.; Su, Y. O.; Stern, M. K.; Macor, K. A.; Kim, D.; Groves, J. T.; Spiro, T. G. *J. Am. Chem. Soc.* **1988**, *110*, 4158−4165.

(10) Ghosh, A.; Gonzalez, E. *Isr. J. Chem.* **2000**, *40*, 1−8.

(11) Nakamoto, K. *Coord. Chem. Rev.* **2002**, *226*, 153−165.

(12) Dey, A.; Ghosh, A. *J. Am. Chem Soc.* **2002**, *124*, 3206−3207.

(13) Harischandra, D. N.; Zhang, R.; Newcomb, M. *J. Am. Chem. Soc.* **2005**, *127*, 13776−13777.

(14) Wasbotten, I. H.; Ghosh, A. *Inorg. Chem.* **2006**, *45*, 4910−4913.

(15) An S = 0 Fe^{IV}−aminoimido or Fe^{II}−aminonitrene porphyrin has been reported. However, we do not view this as a simple Fe^{IV}−imido porphyrin, but rather as an Fe^{II}−diazoalkane complex, as we hope to describe in detail in a forthcoming paper. See: Mahy, J.-P.; Battioni, P.; Mansuy, D.; Fisher, J.; Weiss, R.; Mispelter, J.; Morgenstern-Badarau, I.; Gans, P. *J. Am. Chem. Soc.* **1984**, *106*, 1699−1706.

(16) Brown, S. D.; Betley, T. A.; Peters, J. C. *J. Am. Chem. Soc.* **2003**, *125*, 322−323.

(17) Brown, S. D.; Peters, J. C. *J. Am. Chem. Soc.* **2005**, *127*, 1913−1923.

(18) Betley, T. A.; Peters, J. C. *J. Am. Chem. Soc.* **2003**, *125*, 10782−10783.

(19) Thomas, C. M.; Mankad, N. P.; Peters, J. C. *J. Am. Chem. Soc.* **2006**, *128*, 4956−4957.

(20) Jenkins, D. M.; Betley, T. A.; Peters, J. C. *J. Am. Chem. Soc.* **2002**, *124*, 11238−11239.

(21) Hu, X.; Meyer, K. *J. Am. Chem. Soc.* **2004**, *126*, 16322−16323.

(22) Shay, D. T.; Yap, G. P. A.; Zakharov, L. N.; Rheingold, A. L.; Theopold, K. H. *Angew. Chem., Int. Ed.* **2005**, *44*, 1508−1510.

(23) The S = 1 ground state reported in ref 22 is in error. The ground state is S = 0, and the S = 1 state is a low-lying excited state. Corrigendum: Shay, D. T.; Yap, G. P. A.; Zakharov, L. N.; Rheingold, A. L.; Theopold, K. H. *Angew. Chem., Int. Ed.* **2006**, *45*, 7870−7870. We thank Professor Theopold for helpful discussions on this issue.

(24) Dai, X.; Kapoor, P.; Warren, T. H. *J. Am. Chem. Soc.* **2004**, *126*, 4798−4799.

(25) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Perderson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B.: Condens. Matter Mater. Phys.* **1992**, *46*, 6671−6687.

(26) Velde, G. T.; Bickelhaupt, F. M.; Baerends, E. J.; Guerra, C. F.; Van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. J. *J. Comput. Chem.* **2001**, *22*, 931−967

(27) Handy, N. C.; Cohen, A. *J. Mol. Phys.* **2001**, *99*, 403−412.

(28) Ghosh, A. *J. Biol. Inorg. Chem*. **2006**, *11*, 671−673.

(29) Ghosh, A. *J. Biol. Inorg. Chem.* **2006**, *11*, 712−724.

(30) Roos, B. O. *Acc. Chem. Res.* **1999**, *32*, 137−144.

(31) Ghosh, A.; Taylor, P. R. *Curr. Opin. Chem. Biol.* **2003**, *91*, 113−124.

(32) For a definition of the local symmetry group, see: Mislow, K.; Siegel, J. *J. Am. Chem. Soc.* **1984**, *106*, 3319−3328.

(33) Such spatially offset π-bonding interactions are well-precedented. See, e.g., Figure 4 and the related discussion on minority spin density on the NO group in S = 1/2 (d_{xy}^1) Mo(P)(NO)−(CH_3OH) (P = porphyrin) in: Tangen, E.; Ghosh, A. *J. Inorg. Biochem.* **2005**, *99*, 959−962.

(34) Betley, T. A.; Peters, J. C. *J. Am. Chem. Soc.* **2004**, *126*, 6252−6254.

(35) Berry, J. F.; Bill, E.; Bothe, E.; George, S. D.; Mienert, B.; Neese, F.; Wieghardt, K. *Science* **2006**, *312*, 1937−1941.

(36) Holland, P. L.; Cundari, T. R.; Perez, L. L.; Eckert, N. A.; Lachicotte, R. J. *J. Am. Chem. Soc.* **2002**, *124*, 14416−14424. Addition/Correction: *J. Am. Chem. Soc.* **2003**, *125*, 11772−11772.

(37) Kogut, E.; Wiencko, H. L.; Zhang, L. B.; Cordeau, D. E.; Warren, T. H. *J. Am. Chem. Soc.* **2005**, *127*, 11248−11249.

(38) Eckert, N. A.; Vaddadi, S.; Stoian, S.; Lachicotte, R. J.; Cundari, T. R.; Holland, P. L. *Angew. Chem., Int. Ed.* **2006**, *45*, 6868−6871.

(39) Klinker, E. J.; Jackson, T. A.; Jensen, M. P.; Stubna, A.; Juhasz, G.; Bominaar, E. L.; Münck, E.; Que, L., Jr. *Angew. Chem., Int. Ed.* **2006**, *45*, 7394−7397.

CT600318N

JCTC Journal of Chemical Theory and Computation

Evaluation of the SCF Combination of KS-DFT and 3D-RISM-KH; Solvation Effect on Conformational Equilibria, Tautomerization Energies, and Activation Barriers

David Casanova,[†,||] Sergey Gusarov,[‡] Andriy Kovalenko,*[,‡,§] and Tom Ziegler[||]

Departament de Química Inorgànica and Centre de Recerca en Química Teòrica, Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona, Spain, National Institute for Nanotechnology, National Research Council of Canada, 421 Saskatchewan Drive, Edmonton, Alberta, T6G 2M9, Canada, Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, T6G 2G8, Canada, and Department of Chemistry, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

Received May 20, 2006

Abstract: The effect of solvation on conformational equilibria, tautomerization energies, and activation barriers in simple S_N2 reactions is reproduced by using the self-consistent field coupling of the Kohn−Sham density functional theory (KS-DFT) for electronic structure and the three-dimensional reference interaction site model with the closure approximation of Kovalenko and Hirata (3D-RISM-KH) for molecular solvation structure. These examples are used in order to validate the implementation of the 3D-RISM-KH method in the Amsterdam Density Functional package. The computations of the free energy difference in the trans/gauche conformational equilibrium for 1,2-dichloroethane in different solvents; the relative tautomerization free energy for cytosine, isocytosine, and guanine; and the free energy activation barrier for a CH_3X-type ($X = F, Cl, Br$) S_N2 reaction exhibit agreement with the experimental data. The method is also applied to the electronic and hydration structure of carbon single-wall nanotubes of different diameters, including the effect of water located in the inner space of the nanotubes. A comparison with continuum models of solvation (including COSMO) as well as with other more precise and computationally more expensive calculations is made to demonstrate the accuracy and predictive capability of the new KS-DFT/3D-RISM-KH method.

1. Introduction

One of the challenges of contemporary quantum chemistry has been to give an accurate description of solute−solvent interactions. A number of methods have been used to account for solvent effects on molecular properties such as structure and reactivity. The methods currently in use can be divided into three categories: dielectric continuum schemes, procedures based on a combination of quantum mechanics with molecular mechanics (QM/MM) or with molecular dynamics (QM/MD and Car−Parrinello), and methods utilizing the integral equations derived in the statistical-mechanical theory of molecular liquids.

In this work, we compare the results obtained by the approaches falling into two of these categories. One is the well-known conductor-like screening model (COSMO),[6] already implemented some time ago in the Amsterdam Density Functional (ADF)[4] package. The other is the three-dimensional reference interaction site model with the closure proposed by Kovalenko and Hirata (3D-RISM-KH),[1,2] which has recently been implemented in a self-consistent field form

* Corresponding author e-mail: andriy.kovalenko@nrc-cnrc.gc.ca.
† Universitat de Barcelona.
‡ National Research Council of Canada.
§ University of Alberta.
|| University of Calgary.

10.1021/ct6001785 CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/22/2007

in ADF.[6] Our objective in this paper is to assess how well the 3D-RISM-KH method, more complex computationally but significantly more advanced theoretically, works in practical calculations in comparison with the empirical COSMO approach.

2. Theory

Here, we present a brief background theory for the two solvation methods that are used in the present study.

2.1. Dielectric Continuum Model of Solvation. The conductor-like screening model (COSMO) by Klamt and Schüürmann[5] belongs to the category of dielectric continuum methods. The solute molecule is considered as being situated in a cavity inside a structureless medium/continuum, with the cavity shapes and medium parameters empirically fitted to reproduce solvation thermochemistry data for a reference set of pure solvents. In the QM/COSMO approach, the original Hamiltonian is modified by adding a solvent potential which is constructed on the basis of the electrostatic potential of the charge distribution $\rho_s(\mathbf{r}_S)$ induced on the cavity surface S by the nuclear charges Z_A and the electronic density distribution $n(\mathbf{r})$ of the solute molecule.[6] The energy E^{es} due to the electrostatic interaction between the solute and cavity surface charges as well the self-interaction of the surface charges constitute the total electrostatic solvation energy given by

$$E^{es} = \int_S \sum_A \frac{Z_A\, \rho_s(\mathbf{r}_S)}{|\mathbf{R}_A - \mathbf{r}_S|}\, d\mathbf{r}_S + \int_S \int_S \frac{\rho_s(\mathbf{r}_S)\, \rho_s(\mathbf{r}_S)\, (\mathbf{r}'_S)}{|\mathbf{r}_S - \mathbf{r}'_S|}\, d\mathbf{r}_S\, d\mathbf{r}'_S + \int_V \int_S \frac{n(\mathbf{r})\, \rho_s(\mathbf{r}_S)}{|\mathbf{r} - \mathbf{r}_S|}\, d\mathbf{r}\, d\mathbf{r}_S \quad (1)$$

In addition to the electrostatic contribution, there is also a nonelectrostatic term $E^{non-elst}$ arising from cavitation, dispersion, and repulsion. These contributions are usually modeled as functions of the area of the cavity surface.[6]

The COSMO approach readily yields the solvation chemical potential of a given solute in the pure solvents that have been parametrized at certain (typically ambient) conditions, provided the COSMO molecular surface can be constructed for the given solute molecule. (The latter is meaningless, for instance, inside a nanotube channel.[7]) However, COSMO does not provide the complete physical picture of the solvation structure. Furthermore, the whole approach is not transferable to a given solvent at given thermodynamic conditions, not present in the database. Nor does it work for solvent mixtures or electrolyte solutions.[8]

The above limitations are partially obviated in the COSMO for real solvents (COSMO-RS) method developed by Klamt which gives quite accurate thermochemical information on liquids and mixtures.[8,9] It is based on the statistics of solvation shells coarse-grained into walls (or double layers) between COSMO cavities encapsulating molecules in the liquid. The solvation chemical potential is obtained as the statistical average of the interaction of wall charges over the distribution in the space of screening charge density values (σ profile). The interaction energy functional is approximated by the self-energy of the total charge density on the two adjacent sides of a wall segment (electrostatic misfit energy) plus the hydrogen-bonding energy, whereas the interaction between different segments is neglected. The ensemble of interacting molecules is thus reduced to an ensemble of independently interacting surface segments, with each molecule represented by a histogram of cavity surface area with respect to screening charge density. This enormously simplifies the problem and yields the solvation chemical potential in a simple calculation. The adjustment parameters for each type of chemical element constituting molecules are determined by using full QM/COSMO calculations for about 800 small molecules to fit experimental physicochemical data for their pure compounds and for the partition of solvent mixtures (typically, water with octanol, benzene, hexane, CCl_4, ether, and ethylacetate) at temperatures around $T = 260-340$ K and ambient pressure.

As seen from the above, COSMO-RS is a post-COSMO method interpolating the thermochemistry of a molecular liquid or a mixture between the QM/COSMO fits for a number of simple, pure compounds at thermodynamic conditions around ambient values. The interpolation is based, in fact, on very simplified and yet rather reasonable approximations: (i) coarse-graining three-dimensional solvation shells into two-dimensional molecular surfaces and (ii) representing the solvent-mediated effective potential between two molecules by local interaction dependent just on the screening charge densities at the contact of their molecular surfaces, with the partition coefficient resolved in the space of screening charge density values. That is, the two "labeled" molecules of the liquid effectively interact through a local contact of their molecular surfaces. This essentially contracts the orientational and, in part, translational degrees of freedom of a "third" molecule in the overlapping solvation shells mediating the effective interaction (a field vertex in the diagrammatic language). It is hence clear where this approach works and where it does not. COSMO-RS gives good results for liquid−liquid, vapor−liquid, and solid−liquid equilibrium properties and the solubility of many compounds and mixtures. Currently, the parametrization is available for molecules that contain only hydrogen, carbon, nitrogen, oxygen, fluorine, phosphorus, sulfur, chlorine, bromine, and iodine.[8,10] However, COSMO and COSMO-RS produce poor results for the phase equilibria of a number of substances, in particular, for polymers and highly polar compounds of complex shape like amines.[10] This is apparently caused by a more ordered solvation structure of such compounds, which is mistreated by the COSMO-RS coarse-graining. Furthermore, COSMO-related approaches are not able to deal with volumetric properties, as the COSMO representation of the solvation structure by a molecular surface totally discards the information about the structure in the dimension across the solvation shells. It reduces the degrees of freedom across the shells to average values on the molecular surface. We emphasize that, although such a mean field can always be introduced in the equations, it cannot be derived in a transferable form within COSMO for a number of cases when the 3D solvation structure is important. It has been shown that the second solvation shell contributes equally to

or even more than the first one into the partial molar volume and compressibility of the solute, which are related to excluded volume and electrostriction.[11] The physics of these effects is beyond COSMO-RS. Moreover, COSMO-related approaches are not applicable at all to systems with confined geometry for which the notion of a molecular surface just doesn't make sense, such as a solvent inside and in pockets of an organic rosette nanotube[12] and inside a carbon nanotube (see section 4.3.3 below).

Of crucial importance in computational chemistry is the description of chemical reactions in solution. Concerning this task, COSMO-RS has significant difficulty with handling geometry optimization, as it requires rerunning the COSMO calculation for the solute molecule at each step of geometry modification.[13] Meanwhile, a single calculation of a screening charge distribution for large molecules with hundreds of atoms becomes so time-consuming[3] that even a self-consistent field QM/COSMO calculation becomes unfeasible and is usually replaced with one COSMO iteration, thus treating the solvent effect just as a perturbation. However, a principal limitation of COSMO-related approaches in treating chemical reactions lies in their inability to accurately yield transition states. The analytical gradients of the solvation free energy are available in COSMO only for the electrostatic term and not for the nonelectrostatic contributions arising from cavitation, dispersion, and repulsion, which are empirically constructed.[6] COSMO also requires additional approximations to avoid discontinuities and singularities in the derivatives arising from the polyhedral representation of the molecular surface. This is done in some practical realizations like the smooth solvation model,[14] which can reproduce the potential energy surface for most of the cases, with some exceptions.[14] However, transition states with the elements having the electron density strongly different from that in their stable states fall out of the COSMO-RS parametrization database. They are practically not amenable to the COSMO-RS parametrization procedure because of the necessity to possess reference solvation thermodynamics data from simulations (or perhaps experiments) for all of the COSMO-RS large training set compounds but with their elements in transition states with a range of different partial charges and sizes.

To address all of the above principal drawbacks of COSMO-related empirical approaches to solvation, one needs to resort to the integral equation theory of molecular liquids, or molecular theory of solvation, which stands on the first principles of statistical mechanics and thus reproduces the essential physics of solvation.

2.2. Interaction Site Model. The reference interaction site model (RISM) is one of the commonly used methods of the integral equation theory of liquids.[15] The RISM integral equation, or site−site Ornstein−Zernike integral equation, was first introduced in 1972 by Chandler and Andersen[16] to study the classical solvation structure of molecular liquids. It is an extension of the Ornstein−Zernike equation for atomic liquids[15] to molecular liquid with atoms strongly bound by intramolecular correlations representing chemical bonds. The RISM integral equation yields the site−site radial distribution functions between interaction sites of molecules of liquid, completely averaged over orientations of the molecules.

As distinct, the 3D-RISM[1,2,17,18] method treats a solute molecule fully in a three-dimensional manner, whereas orientational averaging around interaction sites is applied to solvent molecules. This yields the three-dimensional classical density distributions of interaction sites of solvent molecules around a solute particle of arbitrary shape. The 3D-RISM integral equation can be derived from the molecular Ornstein−Zernike equation[15] for the six-dimensional solute−solvent correlation functions by reducing the orientations of solvent molecules.[1,2] Alternatively, the 3D-RISM integral equation can be obtained from the 3D-RISM in the hypernetted chain approximation (3D-RISM-HNC theory)[17] constructed on the basis of the density functional theory of nonuniform polyatomic liquids.[19] The 3D-RISM integral equation gives the 3D density correlations representing the response of a molecular solvent to the external field of the solute:

$$h_\gamma(\mathbf{r}) = \sum_\alpha \int d\mathbf{r}'\, c_\alpha(\mathbf{r} - \mathbf{r}')\, \chi_{\alpha\gamma}(r') \qquad (2)$$

where $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ are respectively the 3D total and direct correlation functions of solvent site γ around the solute and $\chi_{\alpha\gamma}(r)$ is the site−site susceptibility of the pure solvent. The 3D total correlation functions $h_\gamma(\mathbf{r})$ are related to the 3D distribution function $g_\gamma(\mathbf{r}) = h_\gamma(\mathbf{r}) + 1$, which gives the normalized probability of finding solvent site γ at position $\mathbf{r}$ with respect to the solute molecule. Outside the molecular core, the 3D direct correlation function $c_\gamma(\mathbf{r})$ has the asymptotics of the interaction potential $u_\gamma(\mathbf{r})$ between solvent site γ and the solute, $c_\gamma(\mathbf{r}) \sim -u_\gamma(\mathbf{r})/(k_B T)$, where $k_B T$ is the Boltzmann constant times the solvent temperature. The solvent susceptibility breaks up into the intra- and intermolecular parts

$$\chi_{\alpha\gamma}(\mathbf{r}) = \omega_{\alpha\gamma}(r) + \rho_\alpha\, h_{\alpha\gamma}(r) \qquad (3)$$

where $\omega_{\alpha\gamma}(r)$ is the intramolecular pair correlation matrix function defining the geometry of the solvent molecule, $h_{\alpha\gamma}(r)$ is the radial total correlation function between sites α and γ in the pure solvent, and ρ_α is the mean number density of solvent site α. The intramolecular matrix of a rigid molecule has the form $\omega_{\alpha\gamma}(r) = \delta(r - z_{\alpha\gamma})/(4\pi r_{\alpha\gamma}^2)$, or $\omega_{\alpha\gamma}(k) = \sin(kz_{\alpha\gamma})/(kz_{\alpha\gamma})$ in the reciprocal space, where $z_{\alpha\gamma}$ are the Z matrix elements. Equations 2 and 3 split the effective interaction between two molecules of liquid up into the direct intermolecular interaction and the medium induced correlation propagating through other molecules of liquid.

In liquid state theory, the integral equation for the total and direct correlation functions must be complemented with an additional relation between them, called "closure". The exact closure is nonlocal and extremely cumbersome; therefore, it is usually replaced with one of the known approximations, for instance, the hypernetted chain (HNC) closure and the mean spherical approximation (MSA).[15] Effects predicted by an integral equation theory of liquid stem from singularities and asymptotics inherent in a closure approximation involved. Kovalenko and Hirata[1,2] proposed a closure (KH approximation) which nontrivially combines HNC and MSA, appropriately applying them to regions of

density depletion and enrichment. This properly accounts for density distribution peaks and long-range tails brought about by chemical specificities of molecules, as well as for density depletion within their repulsive cores.[2,20,21] In this study, we complement the 3D-RISM integral eq 2 for the 3D total and direct correlations functions with the 3D-KH closure which reads[1,2]

$$g_\gamma(\mathbf{r}) = \begin{cases} \exp(d_\gamma(\mathbf{r})) & \text{for } d_\gamma(\mathbf{r}) \le 0 \\ 1 + d_\gamma(\mathbf{r}) & \text{for } d_\gamma(\mathbf{r}) > 0 \end{cases} \quad (4)$$

where

$$d_\gamma(\mathbf{r}) = -\frac{u_\gamma(\mathbf{r})}{k_B T} + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})$$

To properly treat electrostatic forces in electrolyte solution with a polar solvent when numerically solving the 3D-RISM-KH eqs 2 and 4, the Coulomb singularities of all the correlation functions $h_\gamma(\mathbf{r})$, $c_\gamma(\mathbf{r})$, and $\chi_{\alpha\gamma}(\mathbf{r})$ are separated out in both the direct and reciprocal spaces and accounted for analytically.[2,22] This includes correction for the supercell periodicity artifacts arising from the 3D fast Fourier transform employed to calculate the convolution in eq 2.

To obtain the radial total correlation functions of solvent $h_{\alpha\gamma}(\mathbf{r})$ constituting the intermolecular part of the solvent susceptibility in eq 3, we use the dielectrically consistent RISM theory (DRISM) developed by Perkyns and Pettitt,[23,24] which consistently describes the dielectric properties of an electrolyte solution comprising ions and a polar molecular solvent. We solve the DRISM integral equation complemented with the KH closure, with the long-range electrostatic asymptotics of the total and direct correlation functions being separated out and treated analytically.[2,22] The computational cost of solving the DRISM integral equations is orders of magnitude smaller than that for 3D-RISM, and the solution can be stored for further use.

Similarly to the RISM-HNC[25] and 3D-RISM-HNC[1,2] theories, the 3D-RISM-KH eqs 2 and 4 lead to the closed analytical expression for the excess chemical potential of solvation[1,2]

$$\Delta\mu^{\mathrm{KH}} = k_B T \sum_\gamma \rho_\gamma \int d\mathbf{r} \left\{ \frac{1}{2} (h_\gamma(\mathbf{r}))^2 \Theta(-h_\gamma(\mathbf{r})) - c_\gamma(\mathbf{r}) - \frac{1}{2} h_\gamma(\mathbf{r})\, c_\gamma(\mathbf{r}) \right\} \quad (5)$$

In what follows, we denote the changes due to solvation by Δ for all quantities.

Unlike the COSMO approach constructing the electrostatic potential of the solvent cavity,[6] our method combined with Kohn−Sham density functional theory (KS-DFT/3D-RISM-KH) yields the self-consistent field of solvent from the first principles by introducing the Helmholtz free energy functional defined as

$$A = E^{\mathrm{el}} + \Delta\mu[n(\mathbf{r}),\rho_\alpha(\mathbf{r})] \quad (6)$$

where E^{el} is the internal energy of the solute including translational, rotational, and vibrational degrees of freedom and $\Delta\mu_{\mathrm{solv}}$ is the part of the solvation free energy (excess chemical potential of solvation) due to the solute−solvent interaction and solvent reorganization. The solvent potential term $v(\mathbf{r})$ in the KS equation as well as the classical potential $u_\alpha(\mathbf{r})$ of the solute molecule acting on site α of the solvent molecules are obtained analytically by variational differentiation of the functional (eq 6) with respect to the 3D density of solute valence electrons $n(\mathbf{r})$ and the 3D site density distributions of solvent molecules $\rho_\alpha(\mathbf{r}) = \rho_\alpha g_\alpha(\mathbf{r})$:

$$v(\mathbf{r}) = \frac{\delta\Delta\mu}{\delta n(\mathbf{r})} \quad (7)$$

$$u_\alpha(\mathbf{r}) = \frac{\delta\Delta\mu}{\delta\rho_\alpha(\mathbf{r})} \quad (8)$$

With the term $\Delta\mu_{\mathrm{solv}}$ given by eq 5, the electronic and classical potentials (eqs 7 and 8) coupling the solute electronic and solvent classical subsystems are obtained in a closed analytical form in terms of pseudopotentials $v_\alpha^{(\mathrm{ps})}(\mathbf{r})$ representing the total electronic density of interaction sites of solvent molecules:[1,2]

$$v(\mathbf{r}) = \sum_\alpha \int d\mathbf{r}'\, \rho_\alpha(\mathbf{r} - \mathbf{r}')\, v_\alpha^{(\mathrm{ps})}(\mathbf{r}') \quad (7')$$

$$u_\alpha(\mathbf{r}) = \int d\mathbf{r}'\, n(\mathbf{r} - \mathbf{r}')\, v_\alpha^{(\mathrm{ps})}(\mathbf{r}') \quad (8')$$

In the simplest way, the solvent site pseudopotentials $v_\alpha^{(\mathrm{ps})}(\mathbf{r})$ can be represented as a sum of the core repulsion, dispersion, and electrostatic terms.[1,2] This results in the subdivision of the classical interaction potentials $u_\alpha(\mathbf{r})$ between the whole solute and solvent interaction sites into the Lennard-Jones and Coulomb parts.

Notice that, unlike COSMO, the KS-DFT/3D-RISM-KH theory is based on the proper physical concept following from the first principles and statistical physics. The self-consistency is ensured for both the solute electronic structure and the solvent distribution determining the shape of the excluded volume and solvation shell. We emphasize that the latter replaces the concepts of a solvent cavity and molecular surface in continuum solvation models.

Another important advantage is that the 3D-RISM-KH theory allows one to analytically calculate the derivatives of the free energy functional.[3] This makes the method readily applicable to molecular geometry optimization and the study of chemical reactions in solutions. The analytical gradients from the RISM theory in combination with KS-DFT for molecules in solution were first derived by Sato et al.[26] It should be noted that our analytical gradients following from the 3D-RISM-KH theory are consistent since they are based on the solvation potential $v(\mathbf{r})$ obtained directly from the definition (eq 7). The latter is not possible for 1D-RISM, which gives $\Delta\mu_{\mathrm{solv}}$ in terms of the radial correlations dependent on partial site charges. It thus requires contraction of the electron density by using the population operator, which, in addition, can be defined in different ways. Another approach to the analytical gradients within the 3D-RISM method is based on separation of the space around the solute into long- and short-range regions.[27] The resulting estimate

expressions for the analytical gradients are more complicated and depend on a definition of the regions.

2.3. Thermodynamic Calculations. For comparison with experimental results, we compute the Gibbs free energies. In the gas phase, the Gibbs free energy of the solute is related to its enthalpy by

$$G_{gas} = H_{gas}(T) - TS_{gas}^{trans} - TS_{gas}^{rot} - TS_{gas}^{vib} \quad (9)$$

The part $H_{gas}(T)$ contains the internal potential energy of the ideal gas and the standard contributions due to the zero-point temperature corrections and the finite temperature corrections to the enthalpy. The entropic terms for translational, rotational, and vibrational motion are obtained from the standard expressions.[28] The free energy of a particle in solution G_{sol} can be split up into that in the gas phase and the solvation free energy:

$$G_{sol} = G_{gas} + \Delta G \quad (10)$$

In COSMO, the solvation free energy is broken up as

$$\Delta G = \Delta G^{solute} + E^{es} + E^{non-es} \quad (11)$$

where the electrostatic energy E^{es} of the solute−solvent interaction is given by eq 1 and the nonelectrostatic contribution E^{non-es} is modeled empirically as a function of the cavity surface area. Notice that eq 11 does not include the change in the translational entropy of solvent ΔS induced by the solute. The term ΔG^{solute} represents the changes in the solute and is composed in COSMO as follows:

$$\Delta G^{solute} = \Delta E^{el} + \Delta H^{rot}(T) + \Delta H^{vib}(T) - T\Delta S^{rot} - T\Delta S^{vib} \quad (12)$$

where ΔE^{el} is the energy of solute electronic reorganization upon transfer from the gas phase to solution, including the response to the change in the solute geometry. The other terms $\Delta H^{rot,vib}(T)$ and $T\Delta S^{rot,vib}$ are the changes induced by the solvent in the enthalpy and entropy of rotational and vibrational degrees of freedom of the solute. The energy and enthalpy terms in eq 12 are replaced altogether with $\Delta H(T)$, obtained from the same functional form $H_{gas}(T)$ but with the inertia moments and vibrational frequencies modified by solvation. Furthermore, one has to construct the rotational and vibrational entropic terms $T\Delta S^{rot,vib}$, as well as the translational term $T\Delta S$ missing in eq 11. They are usually estimated on the basis of the entropies in the gas phase by using the procedure of Wertz.[29] For a water solvent, this procedure gives the following empirical expression:[30]

$$S_{sol} = 0.258 + 0.54\, S_g^{\circ} \quad (13)$$

In sections 4.1 and 4.2, where we deal with the relative free energies between molecules with similar modes, ΔE^{el} is the major contribution and the other terms in eq 12 are neglected.[31]

With the 3D-RISM methodology, the solvation free energy ΔG is split up into the solute electronic reorganization energy ΔE^{el}, the change in the rotational and vibrational free energies of the solute upon transfer from the gas phase to solution ΔG^{rot} and ΔG^{vib}, and the excess chemical potential of solvation $\Delta\mu$ due to the solute−solvent interaction and solvent reorganization:

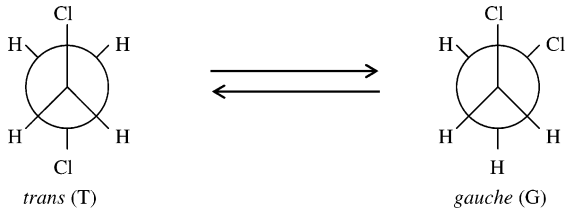


Figure 1. Trans/gauche equilibrium for 1,2-dichloroethane.

$$\Delta G = \Delta E^{el} + \Delta G^{rot} + \Delta G^{vib} + \Delta\mu \quad (14)$$

The excess chemical potential of solvation $\Delta\mu$ obtained by expression 5 contains the solute−solvent (uv) interaction energy E^{uv}, the energy of solvent reorganization (vv) around the solute ΔE^{vv}, and the solvent translational entropy change ΔS induced by the solute:

$$\Delta\mu = E^{uv} + \Delta E^{vv} - T\Delta S \quad (15)$$

The change in the vibrational terms is small ($\Delta G^{vib} < 1$ kcal/mol) and can be neglected, especially in the calculation of relative free energies as in sections 4.1 and 4.2. We can readily obtain the solvent effect on the vibrational free energy ΔG^{vib} by the standard frequency calculation. However, it is small and is not included in the results presented below.

3. Computational Details

All calculations were carried out using the ADF program package.[4] We used both the local-density approximation (LDA)[32] and the BP86 scheme in which LDA is augmented with gradient corrections for exchange[33] and correlation.[34] The majority of the results presented in this work were computed using the BP86 scheme. The calculations carried out with LDA are indicated explicitly. The basis set was of triple-ζ quality with one polarization function added (TZP). In the KS-DFT/3D-RISM-KH calculations, the 3D-FFT grid size was $64 \times 64 \times 64$ in the supercell of size $32 \times 32 \times 32$ Å^3. Further refinement of the grid did not result in any considerable changes. In one particular case of the reaction $Br^- + CH_3Cl \rightarrow CH_3Br + Cl^-$, the geometry of the species and the shape of the solvation potential required a doubling of the grid resolution.

The COSMO implementation is that introduced into ADF by Pye and Ziegler,[6] whereas the 3D-RISM-KH scheme has been recently implemented in ADF by Gusarov et al.[3]

4. Results and Discussion

We shall now evaluate the self-consistent field combination of KS-DFT with the 3D-RISM-KH method in a number of applications studying the effect of solvation on several molecular properties and compare the results with experimental data as well as other solvation approaches, including COSMO.

4.1. Solvent Effect on Conformational Equilibria. The free energy difference between the two conformers (trans and gauche) of the 1,2-dichloroethane molecule (Figure 1)

$$\Delta G^{T\rightarrow G} = G^G - G^T \quad (16)$$

Table 1. Free Energy Difference $\Delta G^{T\to G}$ (kcal/mol) between the trans and gauche Conformers of 1,2-Dichloroethane in the Gas Phase (Upper Part) and in Solvents (Lower Part)[a]

	this work	Capelli [b]	experiment [c]		
gas phase	1.83	1.5	1.35		
solvent	dielectric constant[c]	COSMO	PCM[d]	3D-RISM-KH	experiment[c]
n-hexane	2.02	1.34	0.95	1.57	1.15
1,4-dioxane	2.20	1.28	0.82	0.04	0.54
carbon tetrachloride	2.23	1.27	0.81	0.32	1.03
tetrachloroethylene	2.50	1.18	0.79	1.20	0.95
diethyl ether	4.34	0.84	0.61	0.54	0.77
ethyl acetate	6.00	0.70	0.48	0.63	0.50
THF	7.52	0.62	0.38	−0.11	0.40
acetone	20.70	0.4	0.13	0.16	0.18

[a] Predictions of the KS-DFT coupled with COSMO, PCM, and 3D-RISM-KH methods, versus experimental results. [b] B3LYP and 6-31G**, ref 36. [c] From refs 35 and 36. [d] From ref 36.

Table 2. Solvent Effect on the 1,2-Dichloroethane trans/gauche Conformational Free Energy Difference, eq 17, $\Delta\Delta G^{T\to G}_{gas\to sol}$ (kcal/mol)

solvent	dielectric constant[a]	COSMO	PCM[b]	3D-RISM-KH[c]	experiment[a]
n-hexane	2.02	0.49	0.55	0.13	0.20
1,4-dioxane	2.20	0.55	0.68	1.66	0.81
carbon tetrachloride	2.23	0.56	0.69	1.38	0.32
tetrachloroethylene	2.50	0.65	0.71	0.50	0.40
diethyl ether	4.34	0.99	0.89	1.16	0.58
ethyl acetate	6.00	1.13	1.02	1.07	0.85
THF	7.52	1.21	1.12	1.81	0.95
acetone	20.70	1.43	1.37	1.54	1.17

[a] From refs 35 and 36. [b] From ref 36. [c] Without the vibrational correction.

has been computed in different solvents using the COSMO and 3D-RISM-KH methods, as well as in the gas phase. The values are compared with the results of previous calculations[35−37] and experimental data[35,36] for several solvents in Table 1. This is a good test for theoretical predictions; the conformational equilibrium of the 1,2-dichloroethane molecule in solution is quite sensitive to the solute reorganization energy and the solvation shell structure, as was shown for instance in the case of 1,2-dichloroethane in water.[38]

The experimental value for $\Delta G^{T\to G}_{gas}$ in the gas phase is positive. Thus, the trans conformer is the most stable in the gas phase, as one would expect on the basis of steric grounds. The value of $\Delta G^{T\to G}_{gas}$ calculated in this work is 1.83 kcal/mol, which is about 0.5 kcal/mol larger than the experimental estimate[35,36] (Table 1). Fortunately, this error cancels out when we turn to the point of interest for our study, namely, the solvent effect on the trans/gauche equilibrium. We expect that the errors in the electronic structure calculation for $\Delta G^{T\to G}_{gas}$ in the gas phase and $\Delta G^{T\to G}_{sol}$ in solution are similar and almost cancel out in their difference, giving the solvent effect on the conformational stability.

In solution, the free energy difference between the two conformers $\Delta G^{T\to G}_{sol}$ becomes smaller (Table 1). This is understood, as only the gauche conformer has a permanent dipole, and is likely to be stabilized more than the trans conformer by solvation. Compared to the experiment, all the utilized theoretical methods seem to reproduce the overall trend in $\Delta G^{T\to G}_{sol}$ with respect to different solvents.

In order to highlight the solvent effects, we also computed the change in the conformational energy difference $\Delta G^{T\to G}$ upon transfer from the gas phase to solution:

$$\Delta\Delta G^{T\to G}_{gas\to sol} = \Delta G^{T\to G}_{gas} - \Delta G^{T\to G}_{sol} \qquad (17)$$

Table 2 makes a comparison of the theoretical and experimental results for the same solvents as in Table 1. The values of $\Delta\Delta G^{T\to G}_{gas-sol}$ obtained using the polarizable continuum model (PCM) by Capelli et al.,[36] COSMO, and 3D-RISM-KH are close to the experimental ones. In most cases, they all slightly overestimate $\Delta\Delta G^{T\to G}_{gas\to sol}$. The results of COSMO and PCM are close to each other. The COSMO values are within a 0.4 kcal/mol range from the experimental data. However, the COSMO and PCM curves run monotonically with the solvent dielectric constant ϵ and do not reproduce some features. In Figure 2, we visualize the data from Tables 1 and 2, making the comparison between the theoretical models and experimental results for $\Delta\Delta G^{T\to G}_{gas\to sol}$ as well as $\Delta G^{T\to G}_{sol}$ as a function of the dielectric constant ϵ of the solvent. The curves for the continuum models show the expected increase of $\Delta\Delta G^{T\to G}_{gas\to sol}$ with the value of ϵ : $f(\epsilon) = (\epsilon - 1)/(\epsilon + 0.5)$. However, solvation effects in molecular solvents with a low ϵ can significantly deviate from the monotonic behavior predicted by continuum models. Indeed, the experimental data for both the conformation energy difference $\Delta G^{T\to G}$ and its change $\Delta\Delta G^{T\to G}_{gas\to sol}$ in a 1,4-dioxane solvent clearly stand out of the smooth dependence

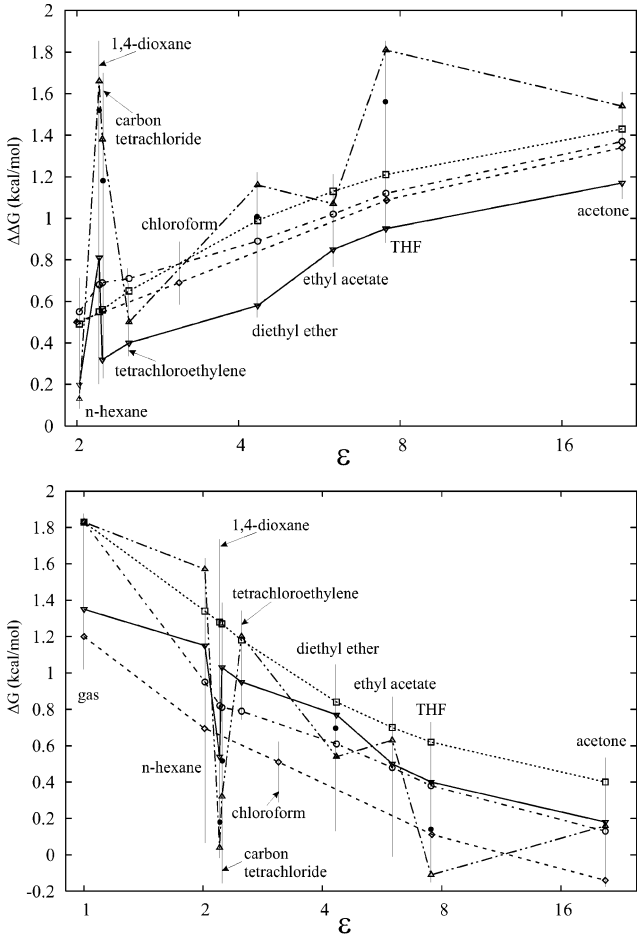


Figure 2. Conformational free energy difference between the trans and gauche conformers of 1,2-dichloroethane in solution $\Delta G^{T\rightarrow G}$ (lower part) as well as its change $\Delta\Delta G^{T\rightarrow G}$ upon transfer from the gas phase to solution (upper part) given by eq 17. Predictions of the KS-DFT coupled with continuum models (COSMO = squares and dotted lines, PCM = circles and dash-dotted lines; the COSMO results are without the translational, vibrational, and rotational thermodynamic corrections to the free energy, given by eq 12) and with the 3D-RISM-KH theory[b] (united-atom solvent models = triangles and dash-double-dotted lines, all-atom models = filled circles; the 3D-RISM values are without the vibrational free energy corrections). Experimental data (references from Capelli[35,36] = diamonds, from Colominas[54] = upside down triangles, and solid lines). The connecting lines are added for the eye.

for the other solvents with similar low ϵ values. This effect cannot be explained simply as dielectric screening by a cavity in a structureless solvent and, obviously, involves short-range interactions of the solute and solvent molecules in the solvation shell, which requires resolution of the three-dimensional solvation structure and a proper account of both energetic and entropic contributions into the solvation free energy. The quality of the electronic basis set and the three-dimensional detail of the solvation model are particularly important for the conformational equilibrium of solvated 1,2-dichlorethane.[38] The 3D-RISM-KH method adequately represents the 3D structure of a molecular solvation shell and, therefore, readily reproduces, in agreement with experimental results, the off-trend values of $\Delta G^{T\rightarrow G}$ and $\Delta\Delta G^{T\rightarrow G}_{gas\rightarrow sol}$ in the 1,4-dioxane solvent. The qualitative agreement with experimental results improves when the united-atom model utilized for the solvent throughout this work is replaced with the more precise all-atom one (filled pentagons in Figure 2). The remaining discrepancy is largely due to the quality of the electronic part, and we were able to further refine the agreement by optimizing the basis set and functional used (data not presented here). We stress that, for this solvent with a low ϵ, the 3D-RISM-KH method readily reproduces the deviation from the monotonic behavior peculiar to dielectric screening, which is missed by continuum solvation models but is indeed observed in experimental results.[35,36]

4.2. Solvent Effect on Tautomerization Energies and Structures. Since the work by Watson and Crick,[39] the problem of tautomeric equilibria has been studied widely. Also, tautomerism has been used in several models to explain the spontaneous mutation of DNA.[40,41] For this reason, a great amount of experimental[42−51] and theoretical[52−57] work based on the study of the tautomerism of nucleic bases has been published. In spite of these efforts, there are still unresolved details. The main theoretical problem has been obtaining accurate results from quantum mechanical calculations in the aqueous phase.

In the analysis of the solvent effect on tautomerization energies and structures of cytosine and isocytosine, we compare our method with the available experimental data and other theoretical studies. As there are no experimental data related to the stability of guanine, protonated cytosine, and protonated guanine tautomers, we compare our calculations just with the theoretical studies available in the literature.

4.2.1. Cytosine. We begin with studying the relative stability of the seven possible tautomers of the cytosine molecule (Figure 3). Experimental IR spectra[42−51,58] and theoretical[52,53,55−57] studies in the gas phase have shown that cytosine exists in three tautomeric forms. The two most stable forms are the aminooxo (canonic) and iminohydroxo forms (CYT2a and CYT2b in Figure 3), while the third most stable form is the aminooxo species, which is observed in small amounts. The aminohydroxo form has no biological significance due to the fact that in DNA the proton at the N1 position is substituted by a sugar moiety.

The energies obtained in this work (Table 3) predict relative stabilities that agree quite well with previous experimental[41−45] and theoretical studies:[52−57]

$$\text{CYT1} \geq \text{CYT4a} \geq \text{CYT2a} \geq \text{CYT4b} \approx \text{CYT2b} \gg \text{CYT3} \gg \text{CYT5}$$

Our calculations predict that in the gas phase the cytosine molecule exists at equilibrium conditions as a mixture of the following tautomers: aminooxo (CYT1), iminooxo (CYT4a and CYT4b), and aminohydroxo (CYT2a and CYT2b). The energy difference between the iminooxo and the aminooxo forms found in the present work is quite small (1.6 kcal/mol), which is in agreement with the previous results (Table 3). Also, we found, in line with other studies,[54] that the enol-imino form (CYT5) is less stable than other tautomers by 15−20 kcal/mol. In a polar solvent like water, there is a change in the relative stability of the cytosine

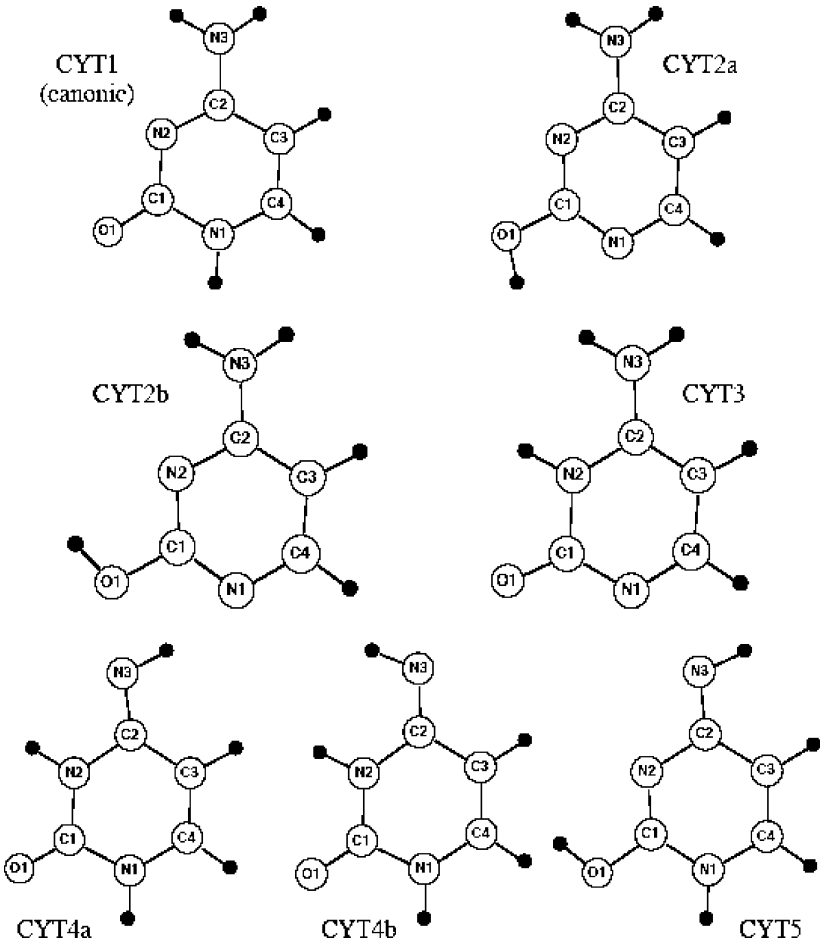


Figure 3. Cytosine tautomers.

Table 3. Free Energy (kcal/mol) of the Cytosine Tautomers in the Gas Phase with Respect to the CYT1 Tautomer

tautomer	this work	Gorb[a]	Colominas[b]	Kobayashi[c]	experiment[d]
CYT 1	0.00	0.0	0.0/0.0	0.0/0.0	
CYT 2a	2.62	0.5	−0.8/1.3	0.330/1.545	0.965
CYT 2b	3.30		0.0/2.0	0.660/0.727	
CYT 3	6.41	7.3	7.0/7.1		
CYT 4a	1.80	1.4	3.0/3.7	−1.45/0.335	
CYT 4b	3.10		1.6/2.2	3.274/2.919	
CYT 5	21.68	20.7			

[a] From ref 56. [b] MP4/6-311++G(d,p)//MP2/6-31G(d) and B3LYP(6-311++G(d,p))//MP2/6-31G(d), respectively.[54] [c] Electronic energies calculated at the CCSD(T)//MP2 and DFT//DFT (cc-pvtz(-f) basis) levels of theory, respectively.[52] [d] From ref 43.

tautomers; only the canonic forms have been observed experimentally.[46−51]

Here, we assess the accuracy of the 3D-RISM-KH method by comparing its predictions for the relative stabilities with experimental data and other theories,[54,56] including the COSMO scheme. Our calculations for the solvent effects by the 3D-RISM-KH method and continuum model (COSMO) predict that the aminooxo (CYT1) is the only tautomer observable in solution (Table 4), in agreement with experimental results for IR spectra in a water solution.[46−51] The consonance of the 3D-RISM-KH and COSMO results with those of previous theoretical works[54,56] is quite remarkable. All the methods find that the canonical form CYT1 is much more stable in solution than the other tautomers. Most of the calculations seem to identify the tautomer CYT3 as the second most stable in solution. The most important stabilization in solution takes place in the CYT3 tautomer, which has the largest dipole moment (Table 5).

Table 4. Free Energy (kcal/mol) of the Cytosine Tautomers in Aqueous Solution with Respect to the CYT1 Tautomer

tautomer	COSMO	3D-RISM-KH	Gorb[a]	Gorb[b]	Colominas[c]
CYT 1	0.00	0.00	0.0	0.0	0.0
CYT 2a	10.86	8.17	5.6	5.0/3.2	6.8
CYT 2b	10.34	7.31			7.1
CYT 3	2.63	3.75	2.9	7.8	5.6
CYT 4a	6.63	7.59	5.4	4.5/7.0	6.1
CYT 4b	7.11	8.81			6.1
CYT 5	23.35	22.24	22.7	20.5	

[a] SCRF model.[56] [b] Tautomer plus one water molecule.[56] [c] SCRF model.[54]

The free energy of solvation for the canonic form is much larger than that for the other stable tautomers in the gas phase (CYT2a, CYT2b, CYT4a, and CYT4b). This explains why the iminooxo and aminohydroxo forms are not detected in solution. In general, as we can expect, there is a correlation between the solvation free energy and the dipole moment of the tautomers. The computed dipole moments in solution are always by several kilocalories per mole larger than those in the gas phase because of the polarity of the water solvent. Notice that the dipole moment values obtained by using the

Table 5. Solvent Effect on the Difference of the Free Energy of the Cytosine Tautomers with Respect to the CYT1 Tautomer, $\Delta\Delta G_{gas\rightarrow sol}^{CYT1\rightarrow t}$, and Dipole Moments μ of the Cytosine Tautomers in the Gas Phase and in Aqueous Solution

	$\Delta\Delta G_{gas\rightarrow sol}^{CYT1\rightarrow t}$ (kcal/mol)			μ (Debye)		
tautomer *t*	COSMO	3D-RISM-KH	Colominas[a]	gas phase	COSMO	3D-RISM-KH
CYT 1	0.00	0.00	0.0	6.50	10.96	10.93
CYT 2a	8.23	5.55	7.6	3.50	5.69	5.57
CYT 2b	7.04	4.02	7.1	4.86	7.74	7.61
CYT 3	−3.78	−2.67	−1.4	8.22	14.01	13.83
CYT 4a	4.83	5.79	4.5	4.69	7.42	7.38
CYT 4b	4.01	5.71	4.5	2.41	4.28	4.15
CYT 5	1.67	0.56		5.52	9.79	9.87

[a] From ref 54.

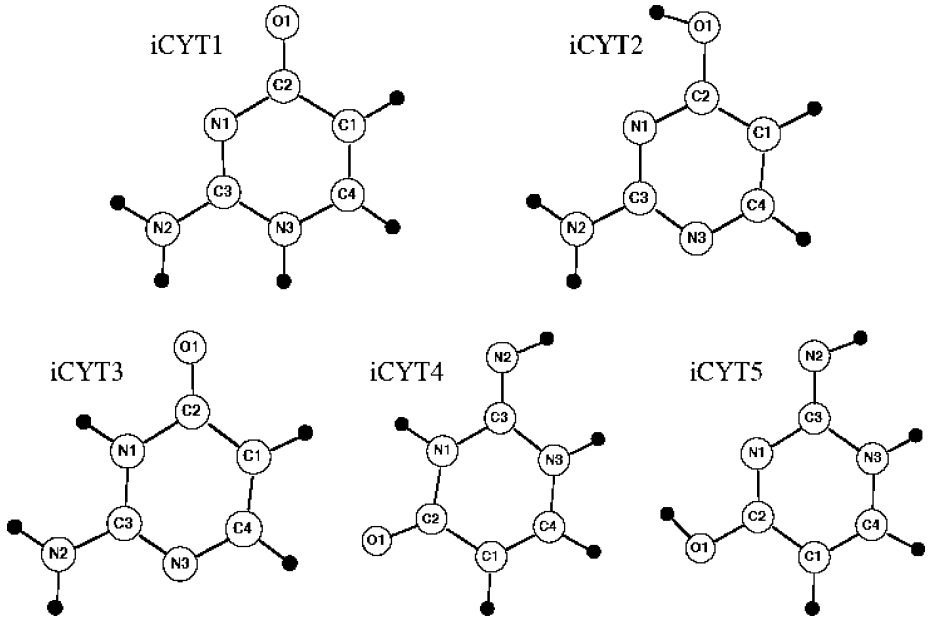


Figure 4. Isocytosine tautomers.

3D-RISM-KH method and the COSMO scheme are very similar and show the same trend (Table 5).

4.2.2. Isocytosine. Isocytosine, in contrast to cytosine, has not been detected in natural DNA or RNA. On the other hand, the C nucleoside of isocytosine is known as an antileukemia drug.[59] Moreover, isocytosine has been incorporated enzymatically into both DNA and RNA.[60−67]

We have studied the different stabilities of the five tautomers of the isocytosine molecule (Figure 4) in the gas phase and in solution. The experimental data in the gas phase[68−72] show that the isocytosine exists in a mixture of two forms, iCYT2 and iCYT3, where the aminohydroxo (iCYT2) seems to be the most stable one. Our calculations in the gas phase (Table 6) reproduce this trend and predict iCYT3 and iCYT2 as the most stable tautomers, although the former is 0.5 kcal/mol more stable than the latter. We determined all the other tautomers to be considerably less stable, in agreement with the fact that iCYT2 and iCYT3 are the only detectable tautomers in the gas phase. In general, our calculations of the relative free energy follow the same trend as that in previous work by Gorb et al.[56] (Table 6). Much as for the tautomers of cytosine, the enol-imino form (iCYT5) is less stable by 10 kcal/mol.

In aqueous solution, the relative stabilities are quite different. Experimentally, only the aminooxo forms (iCYT1 and iCYT3) have been found.[41,46−51,58,73,74] Previous self-consistent reaction field (SCRF)[75] calculations on these tautomers did not reproduce the relative stability order observed experimentally. The KS-DFT/3D-RISM-KH method predicts the existence of three tautomers in solution: iCYT1, iCYT2, and iCYT3. The results of our COSMO calculations give a substantial difference between the relative stabilities of iCYT1, iCYT2, and iCYT3. These values are not quite in agreement with the experimental observation of the three species. Meanwhile, 3D-RISM-KH predicts the aminooxo forms as the most stable in solution. The difference in stability between the two aminooxo forms iCYT1 and iCYT3 is quite small (1.3 kcal/mol), and their free energy in solution is lower than iCYT2 by about 4 and 5 kcal/mol, respectively

Table 6. Free Energy (kcal/mol) of the Isocytosine Tautomers in Gas Phase with Respect to the iCYT2 Tautomer

tautomer	this work	Gorb[a]
iCYT 1	8.98	9.7
iCYT 2	0.00	0.0
iCYT 3	−0.50	1.2
iCYT 4	5.55	7.0
iCYT 5	18.39	22.4

[a] From ref 56.

Table 7. Free Energy (kcal/mol) of the Isocytosine Tautomers in Aqueous Solution with Respect to the iCYT3 Tautomer

tautomer	COSMO	3D-RISM-KH	Gorb[a]	Gorb[b]
iCYT 1	−4.70	1.26	1.9	8.4
iCYT 2	8.93	5.37	1.3	0.9
iCYT 3	0.00	0.00	0.0	0.0
iCYT 4	5.60	6.44	5.7	8.2/9.4
iCYT 5	12.38	16.50	23.7	20.7

[a] SCRF model.[56] [b] Tautomer plus one water molecule.[56]

(Table 7). Thus, the 3D-RISM-KH method seems to give a more consistent account of the experimental fact that only the two aminooxo forms are detectable.

The calculations by Gorb et al.[56] used only a single water molecule to represent the solvent. They predict the iCYT1 and iCYT2 tautomers to be the most stable species, in disagreement with experimental data. Although they differ by about 5−10 kcal/mol from the KS-DFT calculations with the COSMO and 3D-RISM-KH treatments of solvation, all of these calculations agree that iCYT5 is the least stable tautomer in solution.

The relative free energy of solvation for the five tautomers of isocytosine (Table 8) calculated with 3D-RISM-KH and COSMO shows that the iCYT1 tautomer is the most stable in aqueous solution. The hydration thus changes the relative order of stability among the isocytosine tautomers in solution, compared to that found in the gas phase.

An important stabilizing factor for the tautomer iCYT1 in water is its dipole moment interacting with the solvent. On the other hand, the iCYT2 tautomer has the smallest solvation free energy, in agreement with its modest dipole moment. Much as for the cytosine tautomers, the calculated dipole moments increase in aqueous solution because of polarization by water.

4.2.3. Guanine. We shall next turn to a study of the relative stability for some of the tautomers of the guanine molecule. We have studied five tautomers of guanine (Figure 5): two keto-amino (G19 and G17) and three enol-amino (G96c, G96t, and G76c) species. The enol-imino forms are not included in this work, as it is well-known that they are very unstable with respect to the keto-amino and enol-amino tautomers.

The relative free energies computed in the gas phase show that the two keto-amino forms G19 and G17 are the most stable (Table 9). In our calculations, G17 appears to be the most stable and G19 slightly differs in energy by ∼0.5 kcal/mol. These values are very close to the DFT calculations combined with self-consistent reaction field and Monte Carlo simulations carried out by Colominas et al.[54]

The enol-amino forms G96c and G96t are ∼2.5 and ∼3.0 kcal/mol less stable than the G17 tautomer. Although these values differ by about 1 kcal/mol from the estimates of Colominas et al., the stability order predicted is the same. Finally, the tautomer G76c is the least stable in the gas phase; it has a free energy more than 5 kcal/mol higher than that of the most stable tautomer G17.

In aqueous solution, the order of relative stability remains nearly the same, but the difference in stability between some tautomers increases (Table 10). The results obtained with 3D-RISM-KH and COSMO agree with the order of tautomers stability computed previously by Colominas et al.[54] using optimized versions of the continuum model developed by Miertus et al.[76] Much as in the gas phase, the G17 and G19 tautomers are the most stable in aqueous solution. Their relative stability differs by just about 1 kcal/mol, with G17 being the most stable in the gas phase and G19 favored in solution. The most stable tautomer G19 has the largest dipole moment (Table 11).

The lower stability of the enol-amino species G96c, G96t, and G76c relative to the most stable keto-amino forms G17 and G19 is emphasized in aqueous solution, the difference considerably increasing by 7 kcal/mol (Table 11). The G76c tautomer is the least stable in water.

4.2.4. Protonated Cytosine. We have studied just two protonated cytosine molecules (Figure 6) derived from the most stable neutral cytosine tautomers (keto-amino and enol-amino forms). A previous theoretical study of the other protonated enol forms[54] revealed that they are more than 9 kcal/mol less stable. Further, the protonated imino species were shown to be very unstable by some 29 kcal/mol compared to the keto-amino derivatives.

The energy difference between the two studied protonated species pCYT13 and pCYT12c is very small in the gas phase (Table 12). The order of stability we obtained agrees with the previous DFT study by Colominas et al.;[54] however, the difference is larger by 1.5 kcal/mol.

The difference between the solvation free energy of pCYT13 and pCYT12c (Table 13) increases the stability gap between the two tautomers. Both 3D-RISM-KH and COSMO computations show that in aqueous solution the keto-amino form pCYT13 is much more stable than the enol-amino form pCYT12c. The difference between the solvation free energy of these protonated forms is likely related to the negative charge on the keto-oxygen (Table 14).

4.2.5. Protonated Guanine. Five protonated tautomers of guanine (Figure 7) have been studied in the gas phase (Table 15) and in aqueous solution. Colominas et al.[54] determined that all tautomers are within 6 kcal/mol. We found a slightly larger difference (about 6.6 kcal/mol) between the most and least stable tautomers pG179 and pG376c, respectively. The second most stable tautomer is pG376c, while pG137 and pG796c have quite similar stabilities.

In aqueous solution, the difference in the tautomers' stabilities increases (Table 16). Much as in the gas phase, pG179 and pG196t are the most and least stable tautomers in water. The pG137 tautomer has the larger hydration energy (Table 17) and, therefore, becomes the second most stable tautomer (very close to pG179). Again, the stabilization of the tautomers correlates quite well with the negative charge on the oxygen atom (Table 17).

4.3. Solvent Effect on Activation of S_N2 Reactions with CH_3X (X = F, Cl, Br). In organic chemistry, the bimolecular nucleophilic S_N2 substitution reaction

$$Y^- + RX \rightarrow YR + X^- \qquad (18)$$

has been studied extensively by both theoretical[77−88] and experimental[88−94] means. Experimental studies have shown that the rate of S_N2 reactions involving CH_3X species with

Table 8. Solvent Effect on the Difference of the Free Energy of the Isocytosine Tautomers with Respect to the iCYT1 Tautomer, $\Delta\Delta G_{gas\rightarrow sol}^{iCYT1\rightarrow t}$, and Dipole Moments μ of the Isocytosine Tautomers in the Gas Phase and in Aqueous Solution

	$\Delta\Delta G_{gas\rightarrow sol}^{iCYT1\rightarrow t}$ (kcal/mol)		μ (Debye)		
tautomer *t*	COSMO	3D-RISM-KH	gas phase	COSMO	3D-RISM-KH
iCYT 1	0.00	0.00	8.56	14.90	14.68
iCYT 2	17.02	13.10	1.41	1.98	1.87
iCYT 3	8.58	8.23	4.50	7.42	7.20
iCYT 4	8.14	8.62	5.69	9.15	8.97
iCYT 5	7.67	5.83	4.85	8.19	8.27

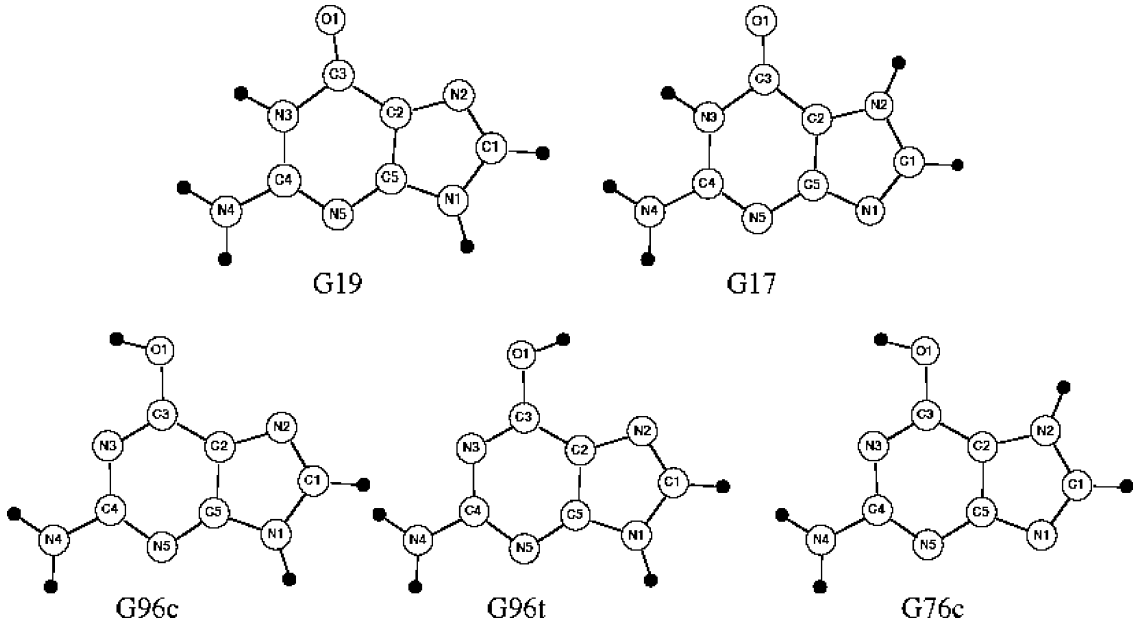


Figure 5. Guanine tautomers.

Table 9. Free Energy (kcal/mol) of the Guanine Tautomers in the Gas Phase with Respect to the G19 Tautomer

tautomer	this work	Colominas[a]	Colominas[b]
G19	0.00	0.0	0.0
G17	−0.57	0.2	−0.4
G96c	2.09	1.1	1.1
G96t	2.53	1.8	1.8
G76c	4.49	4.4	3.7

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54] [b] B3LYP(6-311++G(d,p))//MP2/6-31G(d).[54]

Table 10. Free Energy (kcal/mol) of the Guanine Tautomers in Aqueous Solution with Respect to the G19 Tautomer

tautomer	COSMO	3D-RISM-KH	Colominas[a]
G19	0.00	0.00	0.00
G17	0.40	1.10	1.00
G96c	10.11	7.33	7.20
G96t	10.57	7.18	8.00
G76c	10.94	9.32	8.80

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54]

halogen X = F, Cl, and Br are quite different in the gas phase and in solution. Our aim here is to demonstrate how well the 3D-RISM-KH method reproduces this difference, compared to the COSMO solvation model.

4.3.1. Identity S_N2 Reactions. The S_N2 reaction both in the gas phase and in solution is characterized by a typical double-well potential energy profile, although the depth of the potential well in solution is marginal (Figure 8). We shall first study the case in which the leaving group X is identical to the incoming nucleophile Y in eq 18. It is convenient to introduce the difference RC = R(X−C) − R(Y−C) as a reaction coordinate of the S_N2 reaction, where X and Y are the leaving and incoming halogens, respectively. Figure 8 displays the potential energy profile against the reaction coordinate RC.

The two minima of the double-well potential for the gas-phase identity S_N2 reaction correspond to the ion−dipole complexes. The minima are separated by a central energy barrier representing the transition state of the whole reaction. In the transition state at RC = 0, the carbon atom is in the center of a bipyramid, with the two halogens in the axial positions equally distant from the carbon (Figure 9). Typical free energy parameters that define an S_N2 reaction in the gas phase are the complexation free energy ΔG^C, the overall barrier ΔG^B relative to the reactants, and the intrinsic barrier ΔG^{IB}, as shown in Figure 8.

In solution, the potential energy profile is quite different. In this case, there are almost no ion−dipole minima and the reaction can be defined in terms of the overall barrier between the reactant and transition states. Solvation makes

Table 11. Solvent Effect on the Difference of the Free Energy of the Guanine Tautomers with Respect to the G19 Tautomer, $\Delta\Delta G_{gas\rightarrow sol}^{G19\rightarrow t}$, and Dipole Moments μ of the Guanine Tautomers in the Gas Phase and in Aqueous Solution

	$\Delta\Delta G_{gas\rightarrow sol}^{G19\rightarrow t}$ (kcal/mol)			μ (Debye)		
tautomer *t*	COSMO	3D-RISM-KH	Colominas[a]	gas phase	COSMO	3D-RISM-KH
G19	0.00	0.00	0.0	6.73	11.15	10.84
G17	0.97	1.67	0.8	2.17	3.72	3.85
G96c	8.02	5.25	6.1	3.30	4.96	4.92
G96t	8.04	4.66	6.2	3.50	5.56	5.16
G76c	6.46	4.84	4.4	3.62	6.39	6.55

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54]

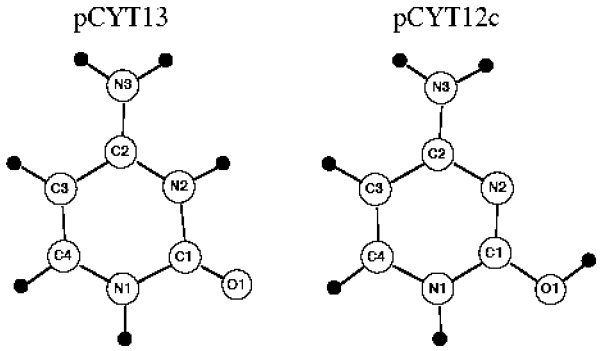


Figure 6. Protonated cytosine tautomers.

Table 12. Free Energy (kcal/mol) of the Protonated Forms of Cytosine in the Gas Phase with Respect to the pCYT13 Tautomer

tautomer	this work	Colominas
pCYT 13	0.00	0.0
pCYT 12c	1.96	−0.4[a]/0.3[b]

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54] [b] B3LYP(6-311++G(d,p))//MP2/6-31G(d).[54]

Table 13. Free Energy (kcal/mol) of the Protonated Forms of Cytosine in Aqueous Solution with Respect to the pCYT13 Tautomer

tautomers	COSMO	3D-RISM-KH	Colominas[a]
pCYT 13	0.00	0.00	0.0
pCYT 12c	11.97	8.43	10.7

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54]

a substantial contribution to the activation energy and can even change the relative order of stability between reactants and the transition state.

In this study, we consider three different cases of the methyl-transfer S_N2 identity reaction with X = Y = F, Cl, and Br. The angle Y−C−X assumes the values in the range from 160 to 180°, as one might expect from an S_N2-type reaction with inversion of the configuration. Moreover, the optimized geometries for the different ion−dipole complexes and the transition state are similar to those obtained in previous works.[85−87,95] The free energies in the gas phase and in aqueous solution were calculated by introducing the thermodynamic terms as detailed above. We computed the difference between the free energy barrier in aqueous solution and in the gas phase, which is shown in Figure 8:

$$\Delta\Delta G_{gas\rightarrow sol}^{B} = \Delta G_{sol}^{B} - \Delta G_{gas}^{B} \qquad (19)$$

For COSMO, the deviation in $\Delta\Delta G_{gas\rightarrow sol}^{B}$ from experimental results amounts to about 1 kcal/mol, whereas for 3D-RISM-KH, it is 3 kcal/mol (see Table 18). As is seen, the 3D-RISM-KH results considerably depend on the accuracy of the Lennard-Jones parameters used. Also presented in Table 18 are the results of the COSMO calculation without the translational, rotational, and vibrational thermodynamic corrections. In this case, the discrepancy with the experiment becomes much larger, which shows the importance of these terms.

4.3.2. Nonidentity S_N2 Reactions. We have also studied some methyl-transfer nonidentity S_N2 reactions. The transition structure is no longer a bond-symmetric arrangement at the midpoint of the reaction coordinate connecting reactants and products (RC ≠ 0). One of the most important distinctions between identity and nonidentity S_N2 reactions is the presence of a "thermodynamic driving force" in the latter due to a free energy difference between the reactants and products.

We obtained a typical potential energy profile and relevant structures both in the gas phase and in solution by using COSMO and 3D-RISM-KH (see Table 19). Notice that the COSMO calculation for $\Delta\Delta G_{gas\rightarrow sol}^{B}$ without the translational, rotational, and vibrational thermodynamic corrections by eq 12 is in poor agreement with experimental results. This shows that these corrections are important and should be accounted for in most of the cases by standard methods. The optimized structures do not differ appreciably from those obtained in previous calculations.[84,86] The prediction for the overall free energy difference $\Delta\Delta G_{gas\rightarrow sol}^{B}$ given by eq 19 for X = F and Y = Cl obtained with COSMO is very close to the experimental value (the deviation is about 0.5 kcal/mol), whereas it is 3.5 kcal/mol for 3D-RISM-KH. For the two other cases of X = F/Y = Br and X = Cl/Y = Br, the values of $\Delta\Delta G_{gas\rightarrow sol}^{B}$ obtained with COSMO are overestimated as 6 and 3 kcal/mol, respectively. The 3D-RISM-KH calculation for X = F and Y = Cl achieves an even better precision of 0.24 kcal/mol. For X = Cl and Y = Br, the value obtained with 3D-RISM-KH is at least 5−7 kcal/mol lower than the experimental values of$\Delta\Delta G_{gas\rightarrow sol}^{B}$. Unlike the case of X = F, 3D-RISM-KH considerably overemphasizes the experimentally observed decrease of $\Delta\Delta G_{gas\rightarrow sol}^{B}$ for X = Cl. This discrepancy can be eliminated upon further optimization of the Lennard-Jones parameters of the solute, as they substantially change in the reaction process. To illustrate this, we also show in parentheses the considerably improved values of $\Delta\Delta G_{gas\rightarrow sol}^{B}$ we obtained by switching in the transition state from the Lennard-Jones parameters of

Table 14. Water Solvent Effect on the Difference of the Free Energy of the Protonated Cytosine Tautomers with Respect to the pCYT13 Tautomer, $\Delta\Delta G_{gas\rightarrow sol}^{pCYT13\rightarrow t}$, and Partial Charge on the Oxygen Atom of the Protonated Cytosine

	$\Delta\Delta G_{gas\rightarrow sol}^{pCYT13\rightarrow t}$ (kcal/mol)			oxygen charge (atomic units)	
tautomer *t*	COSMO	3D-RISM-KH	Colominas[a]	COSMO[b]	3D-RISM-KH[b]
pCYT 13	0.00	0.00	0.0	−0.72	−0.65
pCYT 12c	10.01	6.70	11.1	−0.56	−0.53

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54] [b] Obtained from the Mulliken charge analysis.

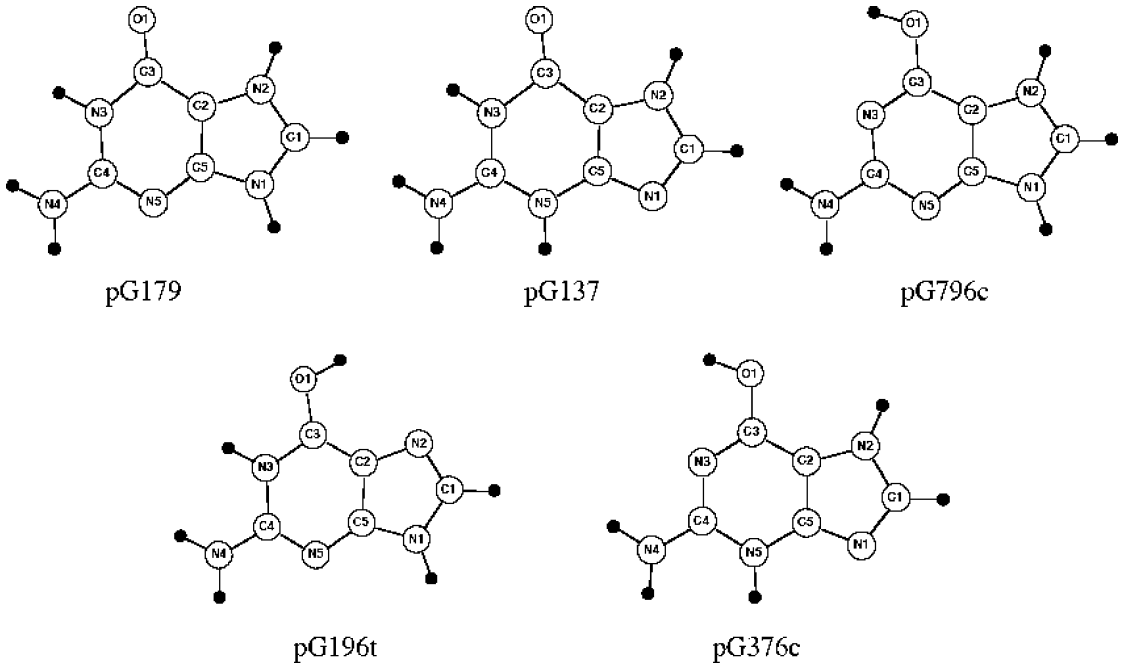


Figure 7. Protonated guanine tautomers.

Table 15. Free Energy (kcal/mol) of the Protonated Forms of Guanine in the Gas Phase with Respect to the pG179 Tautomer

tautomer	this work	Colominas
pG179	0.00	0.0
pG137	4.34	3.5[a]/4.8[b]
pG796c	4.84	3.6[a]/3.6[b]
pG196t	6.59	5.1[a]/5.9[b]
pG376c	3.35	1.8[a]/1.9[b]

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54] [b] B3LYP(6-311++G(d,p))//MP2/6-31G(d).[54]

Table 16. Free Energy (kcal/mol) of the Protonated Forms of Guanine in Aqueous Solution with Respect to the pG179 Tautomer

tautomer	COSMO	3D-RISM-KH	Colominas[a]
pG179	0.00	0.00	0.0
pG137	0.18	2.49	1.1
pG796c	11.95	8.98	10.7
pG196t	12.16	11.94	12.7
pG376c	10.02	9.25	9.8

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54]

the Br^- ion to those of the neutral Br atom. This outlines a possible way to further improve the description by interpolating the Lennard-Jones parameters in an intermediate state by projecting onto the initial and final states.

4.3.3. Water Distribution inside Carbon Nanotubes. Important effects can be revealed by exploring the solvent confined in inner spaces of nanomaterials.[96] In particular, introducing guest species into the inner phase of a carbon nanotube may give rise to altered composite system properties through spontaneous innerphase charge transfer and electrostatic interactions.[97] The energetics, mechanism, and dynamics of chemical reactions may be significantly altered inside carbon nanotubes because of their large electronic polarizabilities and because of the severely decreased reaction volume.[98] Aqueous solution confined in such nanosystems inside a molecular size volume shows a number of unique properties and processes which can be qualitatively different from the regular bulk environment. The prediction of these properties from the first principles is of significant interest since confined water can be found in many important systems, for example, in pockets of proteins, in channels of various biological and synthetic organic membranes, and inside various types of nanotubes.

Theoretical investigation of a solution confined in nanospaces is complicated by the fact that it cannot be treated by continuum solvation models because the concept of surface charge distribution in general loses meaning in this case.[7] So far, molecular dynamics and Monte Carlo simulations have been widely used for studying water in a nanospace environment.[99−101] These methods provide a realistic and accurate picture but are very expensive com-

Table 17. Water Solvent Effect on the Difference of the Free Energy of the Tautomers of Protonated Guanine with Respect to the G179 Tautomer, $\Delta\Delta G_{gas\rightarrow sol}^{pG179\rightarrow t}$, and Partial Charge on the Oxygen Atom of the Protonated Guanine

	$\Delta\Delta G_{gas\rightarrow sol}^{pG179\rightarrow t}$ (kcal/mol)			oxygen charge (atomic units)	
tautomer	COSMO	3D-RISM-KH	Colominas[a]	COSMO[b]	3D-RISM-KH[b]
pG179	0.00	0.00	0.0	−0.75	−0.68
pG137	−4.16	−2.39	−2.5	−0.70	−0.65
pG796c	7.11	4.56	7.5	−0.58	−0.54
pG196t	5.58	4.42	8.0	−0.55	−0.51
pG376c	6.66	5.93	7.1	−0.56	−0.52

[a] MP4/6-311++G(d,p)//MP2/6-31G(d).[54] [b] Obtained from the Mulliken charge analysis.

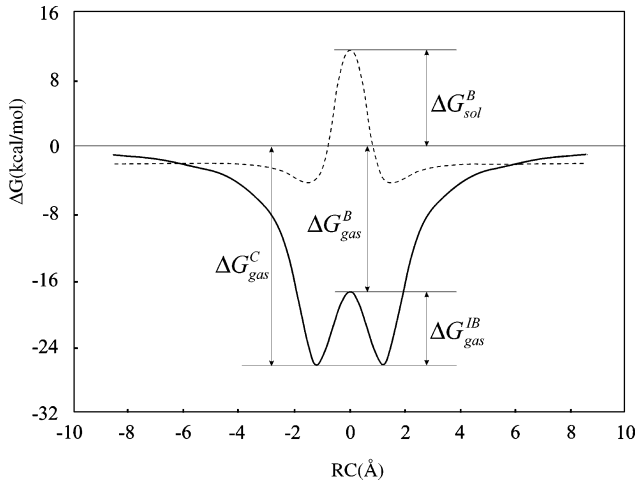


Figure 8. Free energy profiles (FEPs) for the identity S_N2 reaction $F^- + CH_3F$ in the gas phase (solid line) and in solution (dashed line). The energy parameters defining the FEP in the gas phase and in aqueous solution: the complexation free energy ΔG^C, the overall free energy barrier ΔG^B, and the intrinsic free energy barrier ΔG^{IB}.

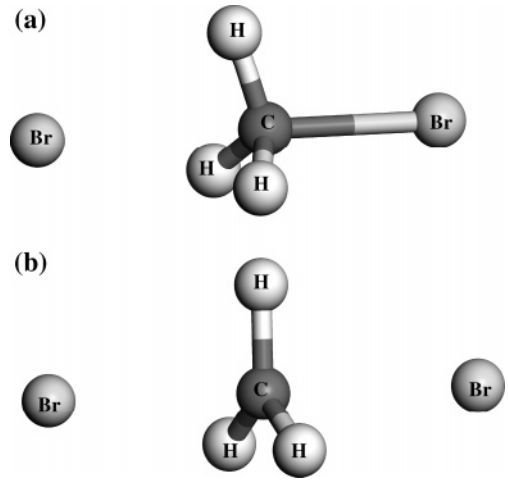


Figure 9. Ion-complex (a) and transition state (b) geometries for the $Br^- + BrCH_3 \rightarrow BrCH_3 + Br^-$ S_N2 reaction.

putationally. As an advantage, the KS-DFT/3D-RISM-KH theory readily yields a self-consistent field description with the relevant part of the nanosystem treated at the quantum chemical level and the properties of the confined solvent predicted from the first principles of statistical mechanics by the molecular theory of solvation.

Table 18. Difference $\Delta\Delta G_{gas\rightarrow sol}^{B}$ (kcal/mol) between the Free Energy of the Transition State of the Identity S_N2 Reactions in Aqueous Solution and in the Gas Phase, eq 19[a]

X = Y	COSMO[b]	COSMO[c]	3D-RISM-KH	experiment[c]
F	21.13	26.46	18.74	22.11
Cl	12.80	18.69	10.67	13.74
Br	13.14	16.61	16.18	14.97

[a] Theoretical predictions (calculated with the GGA functional) versus experimental results. The experimental values in the gas phase of the intrinsic barrier correspond to enthalpies. These are corrected with the calculated entropies to obtain free energies.[87−91] [b] With the translational, rotational, and vibrational thermodynamic corrections given by eq 12. [c] Without the translational, rotational, and vibrational thermodynamic corrections.

Table 19. Difference $\Delta\Delta G_{gas\rightarrow sol}^{B}$ (kcal/mol) Between the Free Energy of the Transition State of the Nonidentity S_N2 Reactions in Aqueous Solution and in the Gas Phase[a]

X, Y	COSMO[b]	COSMO[c]	3D-RISM-KH	experiment
F, Cl	31.24	36.54	35.27	31.71
F, Br	38.29	43.77	32.60	32.36
Cl, Br	26.86	32.66	7,6 (10.2)	15.95/23.80

[a] Theoretical predictions (calculated with the LDA functional) versus experimental results. The experimental values of the intrinsic barrier in the gas phase are derived from the measured enthalpy corrected by the calculated entropy to obtain the free energy.[87−91] [b] With the translational, rotational, and vibrational thermodynamic corrections given by eq 12. [c] Without the translational, rotational, and vibrational thermodynamic corrections.

To illustrate the capabilities of the KS-DFT/3D-RISM-KH theory for a solvent confined in inner spaces, we applied it to the simplest example of such a system: (6,6), (8,8), (12,12), and (20,20) carbon single-wall nanotubes (SWNTs) of different lengths immersed in water. So far, such a description level has been achieved only in the ONIOM[102] and QM/MD[102] calculations, which are significantly more expensive computationally. The whole nanotube is treated by KS-DFT with the LDA functional and the DZP basis set, which allows for the nanotube polarization. The computational load can be further decreased by applying the quantum mechanical treatment just to the most important part of the system (for instance, reactants inside the nanotube) and modeling the electronic response of the nanotube by the classical polarizability of its atoms.

The systems we studied included the SWNTs of five different diameters: (6,6), (8,8), (10,10), (12,12), and (20,-20). The nanotube length of ∼14 Å was chosen to be large enough to eliminate the end effects on the density distribu-

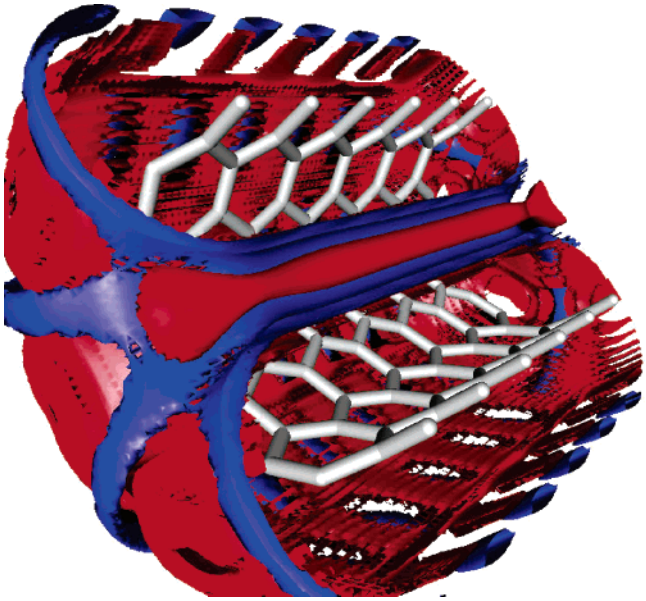


Figure 10. First hydration shell of the (6,6) single-wall carbon nanotube. Isosurfaces of the 3D distribution functions $g(\mathbf{r}) > 1.1$ of water oxygen (red) and hydrogen (blue) in the inner part as well as around the nanotube. Hidden is part of the nanotube and hydration shell.

tions in the central part the nanotube. Figure 10 presents the predictions of the KS-DFT/3D-RISM-KH theory for the 3D density distributions $g_\gamma(\mathbf{r})$ of water inside the (6,6) SWNT. The theory yields the hydration structure in 3D detail including the orientations of water molecules, both around the nanotube and in its inner part. Water molecules in contact with the outer surface of the nanotube are located mostly at the on-top and bridge-site positions of the surface lattice. Their hydrogens are oriented outward from the nanotube, with the HOH molecular plane at small angles to the nanotube axis. This alignment of water dipoles outward from the surface is caused by the strong polarization of the nanotube.

In Figure 11, we make a comparison between the predictions of the KS-DFT/3D-RISM-KH theory for the axial profiles of water sites (O = solid line, H = dashed line) inside the (6,6) SWNT and the MD simulations[99] (O = dash-dotted line, H = dotted line) available for this system. The 3D-RISM-KH results are in rather good agreement with the simulations and predict a number of features of the hydration structure. In the narrow channel of the (6,6) nanotube close to the water molecule size, water oxygens are almost centered at the nanotube axis, whereas water hydrogens are oriented mostly outward from the axis. As the OH intramolecular bond length is 1 Å, the hydrogen maximum at a distance of 0.6 Å from the channel axis suggests that water molecules are oriented with the dipole moment around normal to the axis and the HOH molecular plane at small angles to the axis. The relative position of the oxygen and hydrogen maxima is clearly seen also from the 1D axial distribution functions $g_\gamma(r)$ shown in Figure 11. In agreement with the MD simulations,[99] the theory predicts a low peak of the oxygen distribution at $r = 1.6$ Å for water molecules against the inner surface of the nanotube and a small peak of the hydrogen distribution at $r = 0$. The latter gives some

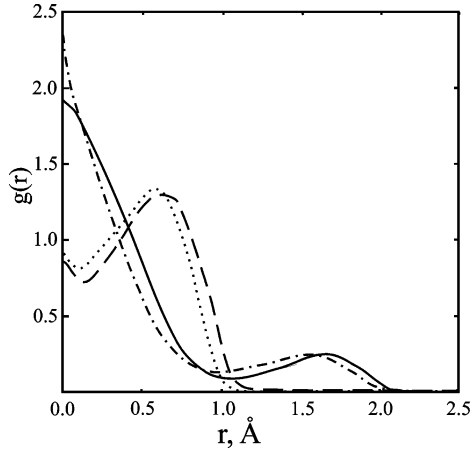


Figure 11. Water distribution profiles inside (6,6) SWNT as a function of the distance from the nanotube axis, predicted by the KS-DFT/3D-RISM-KH theory (O = solid line, H = dashed line) versus MD simulation[99] (O = dash-dotted line, H = dotted line).

probability of finding hydrogens around the channel axis, with a possibility to get in contact with oxygen cores of the adjacent water molecules and thus form hydrogen bonds. However, the integration $2\pi\rho_H \int_0^{r_1} r\, dr\, g_H(r)$ up to the distance $r_1 \approx 0.3$ Å estimating the region where hydrogens can be in contact with adjacent oxygens gives approximately just one hydrogen bond per 20 water molecules in the channel. We stress again that this level of description is comparable to molecular simulations and cannot be achieved by any continuum solvation model.

Further, the most probable orientations of a "labeled" water molecule for any given space region outside or inside the nanotube can be obtained explicitly in the form of the distribution function $g(\mathbf{r},\Omega)$ dependent on its position $\mathbf{r}$ and orientation Ω with respect to the nanotube by using the 3D-RISM approach to orientationally dependent potentials of mean force between solution molecules.[103]

Figure 12 presents the density profiles of water inside the nanotubes as a function of the distance from the nanotube center, obtained by azimuthal averaging of the 3D distribution functions $g_\gamma(\mathbf{r})$. In the (20,20) SWNT with the wide channel, the profile consists of four hydration shells. The height of the peaks is largest against the nanotube wall and decreases toward the nanotube center. All the peaks of water hydrogen are lower but wider than those of water oxygen, and the positions of the oxygen and hydrogen peaks are very close in the first and second hydration shells. This is typical for hydrophobic hydration with hydrogens aligned mainly along the hydrophobic surface but some assuming orientations toward and outward from it. This picture gets distorted for the narrower (12,12) SWNT, which holds about three hydration shells, the third one of the increasing height in the nanotube center. The even narrower (10,10) SWNT fits two hydration shells, the second one in the nanotube center elevated to the same height as that against the nanotube wall. Finally, the narrowest (8,8) SWNT accommodates just one hydration shell, with the complete void of water distribution in the nanotube center. The position of the hydrogen peak at $r = 1.7$ Å is shifted with respect to the oxygen one at r

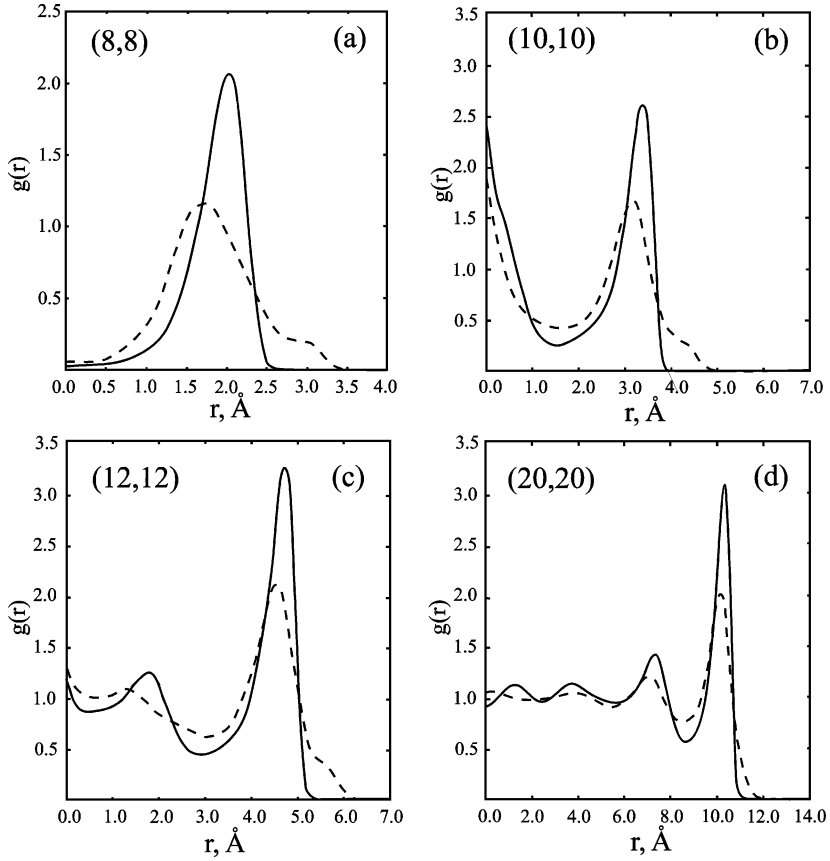


Figure 12. Water oxygen (solid line) and hydrogen (dashed line) distribution profiles inside SWNT as a function of the distance from the nanotube symmetry axis. Parts a−d: (8,8), (10,10), (12,12), and (20,20) SWNT, respectively.

= 2.0 Å, revealing the tilt of the OH intramolecular bonds by about 15−20° off the orientation along the nanotube wall toward the nanotube center. The considerable shoulder of the hydrogen profile at $r = 3.0$ Å corresponds apparently to the orientation of second hydrogens of such water molecules in the hydration shell in normal toward the nanotube wall.

Many theoretical investigations of water distributions inside carbon nanotubes are available,[99−102,104] and we shall not further discuss the electronic and classical effects in these systems presented in detail in the literature, such as the polarization of solvated nanotubes of different diameters. Appropriate for the goals of this article is to demonstrate that the KS-DFT/3D-RISM-KH theory reproduces the results of molecular simulations with a high level of accuracy, which is evident from Figure 11, making a comparison with MD simulation[99] for the hydration structure of (6,6) SWNT. This renders the 3D-RISM method self-consistently coupled with DFT or multireference methods as a valuable tool for the prediction of the properties of inner phases in nanomaterials, including chemical reactions in solution confined in inner spaces.

Conclusion

To validate the newly implemented KS-DFT/3D-RISM-KH theory by comparison with experimental results and other solvation approaches, we applied it to predict the effect of various solvents on the conformational equilibrium of 1,2-dichloroethane; the relative stability of some tautomers of cytosine, isocytosine, and guanine molecules in aqueous solution; and the solvation effect on the activation free energy of some S_N2 reactions. The results are quite similar to those obtained by high-level ab initio methods and continuum solvation approaches such as COSMO, when such calculations are available. Some differences have been detected in a few cases. We found that in general the KS-DFT/3D-RISM-KH method well-reproduces the free energy and predicts the solvation effects in agreement with experimental results.

We currently use the Lennard-Jones parametrization to represent the exchange and dispersive part of the classical potential between the atomic sites of solute and solvent molecules. Because these parameters considerably affect the results, it is crucial to take them from an appropriate force field optimized for liquids of the corresponding molecules rather than just for individual molecules in aqueous solvent. We took the Lennard-Jones parameters of both solute and solvent interaction sites, as well as the site charges of solvent molecules from the OPLS force field. Alternatively, these parameters can be determined from a post Hartree−Fock procedure[105] which is relatively inexpensive computationally and can be included in a self-consistent calculation. For chemical reactions, the Lennard-Jones parameters in an intermediate state can be interpolated between the initial and final states.

We found that the KS-DFT/3D-RISM-KH theory well-reproduces the conformational preferences of 1,2-dichloroethane in different solvents observed in experiments. It also predicts well the relative stabilities of several tautomers of

cytosine, isocytosine, and guanine molecules, determined experimentally. Moreover, KS-DFT/3D-RISM-KH has been the only method able to reproduce the experimental findings for the relative stability of the isocytosine tautomers in aqueous solution.

Most of the results obtained in the calculation of the solvent effects for the identity and nonidentity S_N2 reactions are in agreement with the latest experimental results. We obtained only one considerable discrepancy, which might be eliminated by parametrizing the Lennard-Jones parameters in intermediate states, as described above.

In general, the computational cost of the 3D-RISM-KH method is comparable with COSMO. While slower for small molecules, 3D-RISM-KH becomes more efficient in the case of macromolecules with hundreds of atoms.[3] The most computationally expensive part of 3D-RISM is the 3D fast Fourier transform, which scales as $N \log N$ with the number of 3D grid nodes N. The latter is roughly proportional to the number of atoms in the solute molecule and stays within reasonable size even for such large macromolecules as nanotubes. On the other hand, the scaling of COSMO is determined by the inversion of the matrix of size $M \times M$, where the number of nodes M covering the molecular surface with screening charge is proportional to the number of atoms in the solute. In the general case, the latter is the operation scaling as M^3. Therefore, for large macromolecules, COSMO gets increasingly inefficient and even unusable in converging the SCF loop.[106] Further linear-scaling modifications of COSMO are based on linear approximations in the calculation of the gradients.[106] However, a profound drawback of such methods is that the linear approximations shift the minimum of the potential energy surface in an uncontrolled way. This disadvantage of linear approximations added to conjugate gradient methods is well-known in numerical methods.[107] There is no guarantee in a general case that such linear-scaling modifications to COSMO reach the global minimum, and they should be used with considerable caution. As distinct, the 3D-RISM is precise in that respect and can be applied without reservation to *any* large system.

The 3D-RISM-KH theory presents principal advantages, compared to empirical continuum models such as COSMO, and is free from the principal limitations of post-COSMO thermodynamic methods like COSMO-RS. The 3D-RISM-KH theory has a predictive capability of a first-principle statistical-mechanical method. It can treat a given molecular solvent or mixture in a given thermodynamic state and is transferable. The theory gives the full physical view of molecular solvation: the 3D structure of the successive solvation shells, including association structures such as hydrogen bonds; the solvation free energy and all its thermodynamic derivatives such as entropy, partial molar volume, and so forth; the potentials of mean force between all species in solution, including the medium-induced effective interaction between a ligand and a solute supra- or macromolecule. The theory provides from the first principles the analytical gradients including all the solvation contributions, both electrostatic and nonelectrostatic (such as cavitation, dispersion, and repulsion). This allows one to accurately determine the potential energy surface and optimize molecular geometry in solution. The theory yields the solvent effect on the transition states, which enables prediction of the chemical reactions in solution. We emphasize that a crucial advantage of the KS-DFT/3D-RISM-KH method is a first-principle, physical view on electronic structure in solution. With reasonable computational efforts, it allows one to predict the electronic structure and geometry of solvated macromolecules, their thermochemistry, and chemical reactions in solution, including the effect of solvent in inner spaces of nanosystems such as channels of nanotubes. This renders it an attractive alternative to continuum solvation methods in theoretical chemistry.

Acknowledgment. This work was supported by the National Research Council (NRC) of Canada and by the Natural Science and Engineering Research Council (NSERC) of Canada. D.C. thanks the Generalitat de Catalunya for a scholarship. T.Z. thanks the Canadian Government for a Canada Research Chair.

References

(1) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095.

(2) Kovalenko, A. Three-Dimensional RISM Theory for Molecular Liquids and Solid−Liquid Interfaces. In *Molecular Theory of Solvation*; Hirata, F., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003; pp 169−275.

(3) Gusarov, S.; Ziegler, T.; Kovalenko, A. *J. Phys. Chem. A* **2006**, *110*, 6083.

(4) Te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca, C. G.; Van Gisbergen, S. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931.

(5) Klamt, A.; Shüürmann, G. *J. Chem. Soc. Trans.* **1993**, *2*, 799.

(6) Pye, C. C.; Ziegler, T. *Theor. Chem. Acc.* **1999**, *101*, 396.

(7) Britz, D.; Khlobystov, A. *Chem. Soc. Rev.* **2006**, *35*, 637.

(8) Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: New York, 2005.

(9) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224.

(10) Mullins, E.; Oldland, R.; Liu, Y.; Wang, S.; Sandler, S.; Chen, C.; Zwolak, M.; Seavey, K. *Ind. Eng. Chem. Res.* **2006**, *45*, 4389.

(11) Imai, T.; Nomura, H.; Kinoshita, M.; Hirata, F. *J. Phys. Chem.* **2002**, *106*, 7308.

(12) Moralez, J. G.; Raez, J.; Yamazaki, T.; Motkuri, R. K.; Kovalenko, A.; Fenniri, H. *J. Am. Chem. Soc.* **2005**, *127*, 8307.

(13) Eckert, F.; Klamt, A. *J. Comput. Chem.* **2006**, *27*, 11.

(14) Gregersen, B.; Khandogin, J.; Thiel, W.; York, D. *J. Phys. Chem. B* **2005**, *109*, 9810.

(15) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*; Academic: London, 1986.

(16) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 193.

(17) Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 7821.

(18) Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **1998**, *290*, 237.

(19) Chandler, D.; McCoy, J.; Singer, J. *J. Chem. Phys.* **1996**, *85*, 5971.

(20) Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **2001**, *349*, 496.

(21) Kovalenko, A.; Hirata, F. *J. Theor. Comput. Chem.* **2002**, *1*, 381.

(22) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 10391.

(23) Perkyns, J. S.; Pettitt, B. M. *J. Chem. Phys.* **1992**, *97*, 7656.

(24) Perkyns, J. S.; Pettitt, B. M. *Chem. Phys. Lett.* **1992**, *190*, 626.

(25) Singer, S. J.; Chandler, D. *Mol. Phys.* **1985**, *55*, 621.

(26) Sato, H.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1996**, *105*, 1546.

(27) Yoshida, N.; Hirata, F. *J. Comput. Chem.* **2006**, *27*, 453.

(28) McQuarrie, D. A. *Statistical Thermodynamics*; University Science Books: New York, 1973.

(29) Wertz, D. H. *J. Am. Chem. Soc.* **1980**, *102*, 5316.

(30) Cooper, J.; Ziegler, T. *Inorg. Chem.* **2002**, *41*, 6614.

(31) Yu, B.; Privalov, P.; Hodges, R. *Biophys. J.* **2001**, *81*, 1632.

(32) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(33) Perdew, J. P.; Wang, Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *46*, 6671.

(34) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.

(35) Wiberg, K. B.; Keith, T. A.; Frisch, M. J.; Murcko, M. *J. Phys. Chem.* **1995**, *99*, 9072.

(36) Capelli, C.; Corni, S.; Tomasi, J. *J. Phys. Chem. A* **2001**, *105*, 10807−10815.

(37) Christiansen, O.; Mikkelsen, K. V. *J. Chem. Phys.* **1999**, *110*, 1365.

(38) Lee, J.; Yoshida, N.; Hirata, F. *J. Phys. Chem. B* **2006**, *110*, 16018.

(39) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 964.

(40) Topal, M.; Fresco, J. R. *Nature* **1976**, *260*, 285.

(41) Kwiatkowski, J. S.; Pullman, B. *Adv. Heterocycl. Chem.* **1975**, *18*, 199.

(42) Gould, I. R.; Vincent, M. A.; Hiller, I. H.; Lapinski, L.; Nowak, M. J. *Spectrochim. Acta, Part A* **1992**, *48*, 811.

(43) Jaworski, A.; Szczepaniak, M.; Kubulat, K.; Person, W. B. *J. Mol. Struct.* **1990**, *63*, 223.

(44) Nowak, M. J.; Lapinski, L.; Fulara, J. *Spectrochim. Acta, Part A* **1989**, *45*, 229.

(45) Radchenko, E. D.; Sheina, G. G.; Smorygo, N.; Blagoi, Y. P. *J. Mol. Struct.* **1984**, *116*, 387.

(46) Brown, D. J.; Teitei, T. *Aust. J. Chem.* **1965**, *18*, 559.

(47) Helene, C.; Dozou, P. *CR Acad. Sci. (Paris)* **1964**, *259*, 4387.

(48) Kaito, A.; Hatano, M.; Ueda, T.; Shibuya, S. *Bull. Chem. Soc. Jpn.* **1980**, *53*, 3073.

(49) Lowe, P. R.; Schwalbe, C. H.; Williams, G. J. B. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1987**, *43*, 330.

(50) Morita, H.; Nagakura, S. *Theor. Chim. Acta* **1968**, *11*, 279.

(51) Sharma, B. D.; McConnel, J. F. *Acta Crystallogr.* **1965**, *19*, 797.

(52) Kobayashi, R. *J. Phys. Chem. A* **1998**, *102*, 10813.

(53) Les, A.; Adamowicz, L.; Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 4001.

(54) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811−6821.

(55) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1987**, *111*, 2308.

(56) Gorb, L.; Podolyan, Y.; Leszczynski, J. *THEOCHEM* **1999**, *487*, 47.

(57) Podolyan, Y.; Gorb, L.; Leszczunski, J. *Int. J. Mol. Sci.* **2003**, *4*, 410−421.

(58) Saenger, W. *Principles of Nucleic Acid Structure;* Springer-Verlag: New York, 1983.

(59) Burchenal, J. H.; Ciovacco, K.; Kalaher, K.; O'Toole, T.; Dowling, M. H.; Chu, C. K.; Watanabe, K. A.; Fox, J. J. *Cancer Res.* **1982**, *36*, 1520.

(60) Roberts, C.; Bandaru, R.; Switzer, C. Y. *Tetrahedron Lett.* **1995**, *36*, 3601.

(61) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322.

(62) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489.

(63) Bain, J. D.; Switzer, C. Y.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537.

(64) Strobel, S. A.; Cech, T. R.; Usman, N.; Beigelman, L. *Biochemistry* **1994**, *33*, 13824.

(65) Horn, T.; Chang, C. A.; Collins, M. L. *Tetrahedron Lett.* **1995**, *36*, 2033.

(66) Tor, Y.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4461.

(67) Musier-Forsyth, K.; Shi, J.-P.; Henderson, B.; Bald, R.; Furste, J. P.; Erdmann, V. A.; Schimmel, P. *J. Am. Chem. Soc.* **1995**, *117*, 7253.

(68) Sheina, G. G.; Stepanian, S. G.; Radchenko, E. D.; Blagoi, Yu. P. *J. Mol. Struct.* **1987**, *158*, 275.

(69) Vranken, H.; Smets, J.; Maes, G.; Lapinski, L.; Nowak, M. J.; Adamowicz, L. *Spectrochim. Acta, Part A* **1994**, *50*, 875.

(70) Szczesniak, M. Ph.D. Thesis, University of Warsaw: Warsaw, Poland, 1985.

(71) Szczesniak, M.; Nowak, M. J.; Szczesniak, K. *J. Mol. Struct.* **1984**, *115*, 221.

(72) Stepanian, S. G.; Radchenko, E. D.; Sheina, G. G.; Blagoi, Yu. P. *J. Mol. Struct.* **1990**, *216*, 77.

(73) Kwiatkowski, J. S.; Leszczunski, J. *J. Phys. Chem.* **1996**, *100*, 941.

(74) Beak, P. *Acc. Chem. Res.* **1997**, *10*, 186.

(75) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.

(76) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.

(77) Shi, Z.; Boyd, R. J. *J. Am. Chem. Soc.* **1990**, *112*, 6789.

(78) Tucker, S. C.; Truhlar, D. G. *J. Phys. Chem.* **1989**, *93*, 8138.

(79) Vetter, R.; Zülicke, L. *J. Am. Chem. Soc.* **1990**, *112*, 5136.

(80) Cernusák, I.; Urban, M. *Collect. Czech. Chem. Commun.* **1988**, *53*, 2239.

(81) Dedieu, A.; Veillard, A. *J. Am. Chem. Soc.* **1972**, *94*, 6730.

(82) Dedieu, A.; Veillard, A. *Chem. Phys. Lett.* **1970**, *5*, 328.

(83) Sato, H.; Sakaki, S. *J. Phys. Chem. A* **2004**, *108*, 1629−1634.

(84) Yang, S.-Y.; Fleurat-Lessart, P.; Hristov, I.; Ziegler, T. *J. Phys. Chem. A* **2004**, *108*, 9461.

(85) Alemán, C.; Maseras, F.; Lledós, A.; Duran, M.; Bertrán, J. *J. Phys. Chem.* **1989**, *2*, 611.

(86) Gonzales, J. M.; Pak, C.; Cox, R. S.; Allen, W. D.; Shaefer, H. F., III; Császár, A. G.; Tarczay, G. *Chem.—Eur. J.* **2003**, *9*, 2173.

(87) Parthiban, S.; de Oliveira, G.; Martin, J. M. L. *J. Phys. Chem. A* **2001**, *105*, 895−904.

(88) Shaik, S. S.; Schlegel, H. B.; Wolfe, S. *Theoretical Aspects of Physical Organic Chemistry: The SN2 Mechanism*; John Wiley & Sons: New York, 1992.

(89) Bathgate, R. H.; Moelwyn-Huches, E. A. *J. Chem. Soc.* **1959**, 2642.

(90) McLennan, D. J. *Aust. J. Chem.* **1978**, *31*, 1897.

(91) Wladkowski, B. D.; Brauman, J. I. *J. Phys. Chem.* **1993**, *97*, 13158.

(92) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154.

(93) Pellerite, M. J.; Brauman, J. I. *J. Am. Chem. Soc.* **1983**, *105*, 2672.

(94) Wilbur, J. L.; Wladkowski, B. D.; Brauman, J. I. *J. Am. Chem. Soc.* **1993**, *115*, 10823.

(95) Glukhovtsev, M. N.; Pross, A.; Schlegel, H. B.; Bach, R. D.; Random, L. *J. Am. Chem. Soc.* **1996**, *118*, 11258.

(96) Lu, G. Q.; Zhao, X. S. *Nanoporous Materials: Science and Engineering*; Imperial College Press: London, 2005.

(97) Knox, J. E.; Halls, M. D.; Schlegel, H. B. *J. Comput. Nano. Sci.* **2006**, *3*, 398.

(98) Halls, M. D.; Schlegel, H. B. *J. Phys. Chem. B* **2002**, *106*, 1921.

(99) Gordillo, M. C.; Marti, J. *Chem. Phys. Lett.* **2000**, *329*, 341.

(100) Hanasaki, I.; Nakatani, A. *J. Chem. Phys.* **2006**, *124*, 174714.

(101) Moulin, F.; Devel, M.; Picaud, S. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 165401.

(102) Xu, S.; Irle, S.; Musaev, D. G.; Lin, M. C. *J. Phys. Chem. A* **2005**, *109*, 9563.

(103) Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **1999**, *103*, 7942.

(104) Grujicic, M. *Appl. Surf. Sci.* **2003**, *206*, 167.

(105) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 154104.

(106) York, D. M.; Lee, T.-S.; Yang, W. *Chem. Phys. Lett.* **1996**, *263*, 297.

(107) Press, W. H.; Teukolsky, A. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes*; Cambridge University Press: New York, 1992.

CT6001785

JCTC Journal of Chemical Theory and Computation

The Effect of the Basis-Set Superposition Error on the Calculation of Dispersion Interactions: A Test Study on the Neon Dimer

Antonio Monari,*,† Gian Luigi Bendazzoli,† Stefano Evangelisti,*,‡ Celestino Angeli,§ Nadia Ben Amor,‡ Stefano Borini,§ Daniel Maynau,‡ and Elda Rossi||

Dipartimento di Chimica Fisica e Inorganica, Universita di Bologna, Viale Risorgimento 4, I-40136 Bologna, Italy, Laboratoire de Physique Quantique, UMR 5626, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex, France, Dipartimento di Chimica, Universita di Ferrara, via Borsari 46, I-44100 Ferrara, Italy, and CINECA, Consorzio Interuniversitario, via Magnanelli 6/3, I-46100 Casalecchio di Reno (BO), Italy

Received October 6, 2006

Abstract: The dispersion interactions of the Ne_2 dimer were studied using both the long-range perturbative and supramolecular approaches: for the long-range approach, full CI or string-truncated CI methods were used, while for the supramolecular treatments, the energy curves were computed by using configuration interaction with single and double excitation (CISD), coupled cluster with single and double excitation, and coupled-cluster with single and double (and perturbative) triple excitations. From the interatomic potential-energy curves obtained by the supramolecular approach, the C_6 and C_8 dispersion coefficients were computed via an interpolation scheme, and they were compared with the corresponding values obtained within the long-range perturbative treatment. We found that the lack of size consistency of the CISD approach makes this method completely useless to compute dispersion coefficients even when the effect of the basis-set superposition error on the dimer curves is considered. The largest full-CI space we were able to use contains more than 1 billion symmetry-adapted Slater determinants, and it is, to our knowledge, the largest calculation of second-order properties ever done at the full-CI level so far. Finally, a new data format and libraries (Q5Cost) have been used in order to interface different codes used in the present study.

1. Introduction

On both the theoretical and the experimental sides, the interest of chemists and physicists in clusters involving rare-gas (Rg) atoms increases.[1−3] In order to be able to perform simulations on medium-size clusters, very accurate two-body potentials are needed. These potentials can be conveniently computed by using high-level quantum-chemistry algorithms on Rg dimers, Rg_2.

For two identical interacting atoms (system with spherical symmetry), the long-range tail of the interaction potentials, as far as exchange can be neglected, is conveniently expanded in a series of inverse powers of the internuclear distance

$$E(R) = E(\infty) + \sum_n \frac{C_{2n}}{R^{2n}} \tag{1}$$

where the dispersion coefficients C_{2n} depend only on the electric properties of the atoms involved. Such a formulation,

* Corresponding authors. E-mail: stefano@irsamc1.ups-tlse.fr (S.E.), amonari@fci.unibo.it (A.M.).
† Universita di Bologna.
‡ Université Paul Sabatier.
§ Universita di Ferrara.
|| CINECA, Consorzio Interuniversitario.

10.1021/ct6003005 CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/09/2007

like for instance the well-known Lennard-Jones potential, allows a very simple and efficient parametrization of the two-body potential. Because of the smallness of the interaction, a very accurate value of the asymptotic limit of the potential-energy curve is extremely important. Since Rg_2 dimers are closed-shell systems that dissociate into two closed-shell atoms, a single determinant gives a qualitatively correct description of the dimer for any value of the internuclear distance. For this reason, single-reference methods can be successfully used even in the dissociation region, contrary to what happens for most chemical systems. In this context, one must be very cautious toward truncated configuration-interaction (CI) methods, as they suffer from the well-known size-consistency (SC) problem. On the other hand, coupled-cluster (CC) methods are size-consistent, and therefore they are not affected by this kind of problem.

In order to obtain accurate values of the potential-energy curves, it is absolutely necessary to take into account the basis-set superposition error (BSSE), and to accordingly modify the energy values. BSSE arises because the wave function (WF) of the dimer at a finite internuclear distance is better described than the WF of the separated atoms, since the orbitals *of the two atoms* are simultaneously used. For this reason, BSSE is particularly important for small- or medium-size basis sets, while it goes to zero in the limit of complete basis sets.

It should be stressed that a rigorous solution to this problem is at the moment not available. Indeed, one should use exactly the same basis set for all the calculations at the different geometries. However, in the linear combination of atomic orbitals (LCAO) formalism, the atomic orbitals are generally centered on nuclei, and therefore the basis set depends on the geometry of the system. A possibility one could think of would be the use of plane-wave orbitals, as it is done for periodic and extended systems. Unfortunately, plane waves are unable to correctly describe the behavior of the electronic WF close to the nuclei (the "nuclear cusp"), particularly in the case of heavy atoms.

The most diffuse procedure to overcome this problem is the use of the "counterpoise correction", proposed by Boys and Bernardi.[4] In this procedure, a series of atomic energies is computed, by using a basis set (b_t) composed of the atomic orbitals of all the atoms of the system, and the differences between these energies and the atomic values are subsequently used to correct the energy surface of the system. In other words, the energy $E(b_t)$ of the complex, computed by using the basis set b_t is corrected by adding a geometry-dependent energy shift Δ, which is given by

$$\Delta = \sum_I \Delta_I \tag{2}$$

where I labels the atoms in the complex

$$\Delta_I = E_I(b_I) - E_I(b_t) \tag{3}$$

Here, b_I represents the atomic basis set of the atom I, while b_t is the total LCAO basis, given by the union of all the atomic basis sets. This procedure, although not an exact one, gives a satisfactory approximation and leads to a satisfactory approximation to the BSSE. In the present article, the effect of the BSSE on the calculation of dispersion coefficients at the CI and CC levels is investigated. As already well-known, the BSSE must be corrected in order to obtain reliable potential-energy curves of van der Waals (VdW) systems. However, although the BSSE-corrected curves obtained by the different methods are at first sight qualitatively rather similar, CI and CC methods show very different behaviors as far as the long-range dispersion coefficients are concerned. Indeed, we found that, while CC approaches are well-adapted for these types of calculations, the corresponding truncated-CI values are completely useless. The reason for this striking difference can be traced back to the lack of size consistency of the truncated-CI methods.

Our dimer calculations were compared with full-CI (FCI) results obtained on a single atom by means of a perturbative scheme. In our largest FCI calculation, the CI space contains more than 1 billion partly symmetry-adapted and spin-adapted Slater determinants. This represents, to the best of our knowledge, by far, the largest calculation of second-order properties ever done at the full-CI level.

2. Computational Details

In the present section, the basis sets and computational methods, which have been used in the present study, are described. The use of the interface Q5COST between different computational codes is also illustrated and discussed.

2.1. Basis Sets. The computation of molecular dispersion interaction is very sensitive to the quality of the computed polarizabilities of the constituent atoms. These properties are critically dependent on the quality of the basis sets and, in particular, on the presence of diffuse atomic orbitals. For this reason, diffuse orbitals are usually added to the standard atomic basis sets in these circumstances. Unfortunately, this fact can have the consequence of an even larger effect on the BSSE. Some authors[5] report the use of the so-called midbond functions, instead of diffuse ones, for the computation of neon dimer potential-energy curves; we decided to discard this possibility mainly in order to use the same basis set in supramolecular and long-range perturbative approaches. The calculations are performed with the correlation consistent basis sets, optimized by Dunning and co-workers.[6−8] In particular, the following two basis sets, retrieved from the Pacific Northwest Laboratory basis set library EMSL,[9] have been used: triply augmented valence double-ζ (taug-vDZ) and quadruply augmented valence triple-ζ (qaug-vTZ). Since these are valence basis sets, it does not make sense to correlate core electrons, and in all the correlated computations, the 1s orbitals of the two neon atoms have been kept frozen at the Hartree−Fock level. This fact also presents the advantage of a considerable savings in computation time.

2.2. Computational Methods. The following methods have been used in the present study:

1. Long-Range Perturbative Approach (LRPT). LRPT is an approach where the atom−atom interaction[10] is treated by perturbation theory starting from the product of isolated fragments' wave functions.

The adopted computation strategy was the following: when the full or string-truncated CI formalism is used, it is

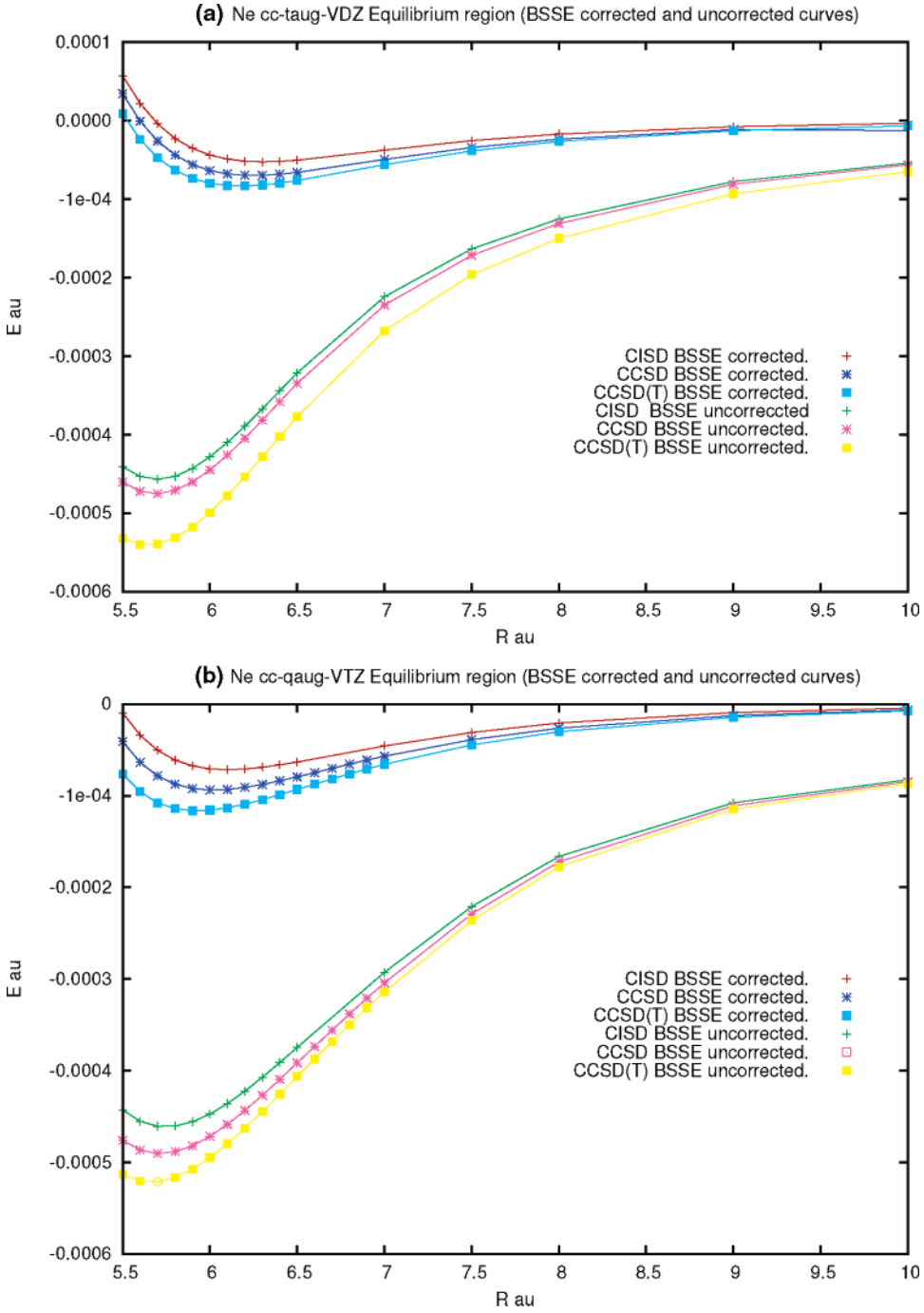


Figure 1. CISD, CCSD, and CCSD(T) potential-energy curves as a function of the internuclear distance. (a) taug-vDZ. (b) qaug-vTZ. Units: distances in bohr and energies in hartree.

possible to immediately get the values of the dispersion coefficients via the use of an innovative perturbation-variational formalism.[11] Note that this technique involves only the use of isolated atoms' wave functions, so values obtained can be considered as BSSE-free and size-consistent by construction. Moreover, values of neon atom dipole and quadrupole polarizabilities are obtained as byproducts. The formalism involved implies solving the perturbative equations for the dispersion interaction by expanding the solution as a linear combination of tensor products of suitable FCI vectors. In the present computation, the latter were chosen to be the so-called Cauchy vectors,[11] strictly related to the FCI computation of Cauchy moments. An expansion set of 10 Cauchy vectors provided satisfactory convergence. Both FCI and string-truncated CI calculations were obtained with the use of the program VEGA.[12] Molecular orbitals and their integrals were computed with the MOLPRO2000 code.[13]

In the string-truncated CI formalism,[14] the determinants formed by strings having up to a given level of excitation are retained in the CI space: single excitations (CIS); single and double excitations (CISD); and single, double, and triple excitations (CISDT). Notice that, if up to quadruply excited strings are considered, in the case of the neon atom, one gets frozen-core FCI.

2. Supramolecular Approach. In this approach, potential-energy curves for the Ne_2 dimer are computed using (a)

Table 1. CISD, CCSD, CCSD(T) BSSE-Counterpoise-Uncorrected and -Corrected Minimum and Vibrational Frequencies[a]

	R_{min}^{uncorr}	E_{min}^{uncorr}	R_{min}^{corr}	E_{min}^{corr}	N_{bs}	ΔE_0	ω
			taug-VDZ				
CISD	5.7025	−456.71	6.3037	−52.675	1	7.2243	
CCSD	5.6862	−475.50	6.2454	−70.022	2	8.2618	6.6990
CCSD(T)	5.6527	−540.83	6.1741	−83.488	2	9.1751	8.3160
			qaug-VTZ				
CISD	5.7514	−461.30	6.0995	−71.563	2	8.5749	6.9332
CCSD	5.7188	−490.46	6.0157	−93.766	2	9.9941	9.4897
CCSD(T)	5.6518	−521.52	5.9269	−116.44	2	11.430	11.960
			Experiment				
24			5.84	−134			
25[b]			5.85	−134	2	12.56	13.76

[a] R_{min}^{uncorr} interpolated value of the BSSE-uncorrected energy curve minimum (a_0 bohr); E_{min}^{uncorr} BSSE-uncorrected potential-energy well depth (μE_h); R_{min}^{corr} interpolated value of the BSSE-corrected energy curve minimum (a_0 bohr); E_{min}^{corr} BSSE-corrected potential-energy well depth (μE_h); N_{bs} number of bound states for BSSE-corrected curves; ΔE_0 zero-point energy calculated from the BSSE-corrected well depth (cm^{-1}); ω anharmonic vibrational frequency from BSSE-corrected curves (cm^{-1}). [b] R_{min} and E_{min} are obtained fitting a model potential, see text for details.

Table 2. Ne Atom, Taug-VDZ, and Qaug-VTZ Basis Set: Full and String-Truncated CI Properties and Dispersion Coefficients[a]

	N_{CI}	E	α_{dip}	α_{quad}	C_6	C_8
			taug-vDZ			
CIS	2.929×10^3	−128.663 720	2.436 792	3.097 065	−5.9899	−7.2528
CISD	1.926×10^6	−128.708 024	2.649 742	3.605931	−6.3270	−19.4611
CISDT	1.319×10^8	−128.709 878	2.680 308	3.666 439	−6.3996	−19.7892
FCI	1.044×10^9	−128.709 923	2.680 788	3.667 532	−6.4008	−19.7955
			qaug-vTZ			
CI-sd	7.100×10^7	−128.810 697	2.649	7.005	−6.354	−35.550
			Interpolated taug-vDZ			
CCSD					−5.8849	−21.9760
CCSD(T)					−6.5433	−28.4863
			Interpolated qaug-vTZ			
CCSD					−6.1717	−37.4064
CCSD(T)					−7.1054	−37.8797
experiment[29,30]			2.669	7.52	−6.383	

[a] Dispersion coefficient interpolated from BSSE-corrected potential-energy curves. N_{CI} is the number of CI determinants in the D_{2h} symmetry point group; E is the total energy of the atom (E_h hartree); α_{dip} is the dipole polarizability (atomic units a_0^3 where a_0 bohr); α_{quad} is the quadrupole polarizability (atomic units a_0^5); C_6 and C_8 are the R^{-6} and R^{-8} dispersion coefficients, respectively ($E_h a_0^6$ and $E_h a_0^8$). When available, the experimental or previous computed best values are also reported.

single-and-double truncated CI, CISD, program CASDI;[15] (b) single-and-double truncated CC, CCSD, DALTON package;[16] (c) single-and-double truncated CC with noniterative triple correction, CCSD(T), DALTON package.[16]

At CISD, CCSD, and CCSD(T), the energy curve has been obtained performing energy computation at various values of interatomic separation; the curve has been subsequently counterpoise (CP)-corrected and linearized as described in the next section in order to get the values of the dispersion coefficients. From the energy curves, we also derived values of minimum energy distance, well-depth energy, zero-point energy, and anharmonic vibrational frequency. As concerns the vibrational levels, the computations were performed by the Numerov method[17] in matrix form as formulated by Lindberg[18] implemented in a code described in ref 19.

2.3. The Use of Q5Cost Wrappers. The CASDI program used for the computation of CISD energy curves was originally interfaced with the Molcas[20] program suite via the wrapper MOLCOST. In order to perform such a computation in the same environment as the CC ones, we decided to interface CASDI with DALTON, using the newly developed Q5Cost data format[21] and library.[22] Q5Cost is a new data format and set of Fortran libraries developed by us that allows the easy exchange of the so-called "quantum-chemistry binary data" (mainly molecular integrals) among different codes. In particular, atomic basis integrals produced by Dalton[16] after self-consistent field calculations were processed with a four-indices transformation to get them in molecular orbitals basis and written in Q5Cost format. Subsequently, a wrapper (Q5MOLCOST) was designed and written in order to write molecular orbitals in a MOLCOST format directly accessible by the CASDI program.

3. Results

In Figure 1 (a and b), the potential-energy curves are reported for the different correlated methods and obtained by using the two basis sets. The CISD, CCSD, and CCSD(T) curves are rather similar. It can be seen that the position of the

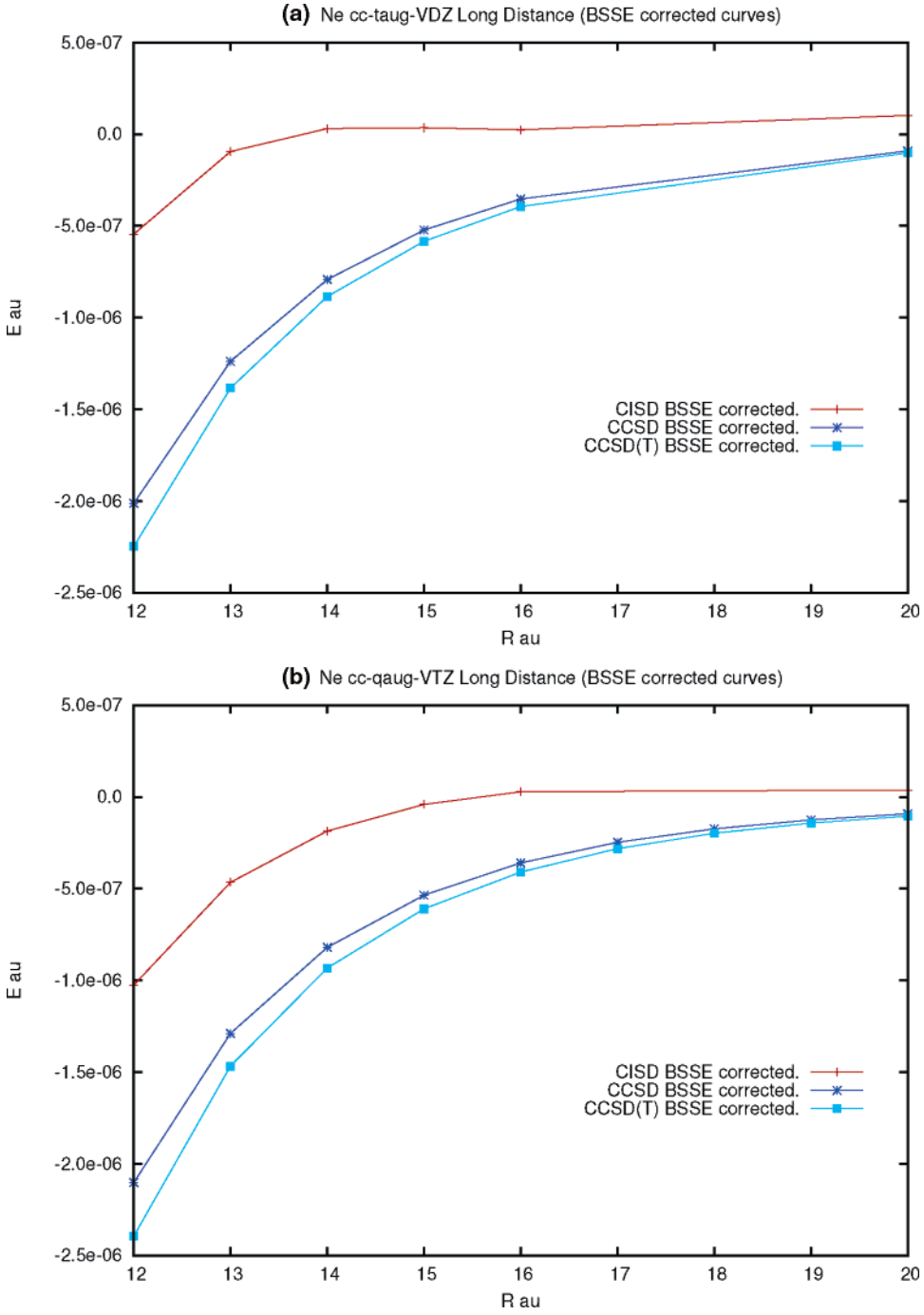


Figure 2. CISD, CCSD, and CCSD(T) BSSE-corrected potential-energy curves in the asymptotic region. (a) taug-vDZ. (b) qaug-vTZ. Units: distances in bohr and energies in hartree.

minimum is not strongly affected by the BSSE correction, while the energy-well shape and depth are completely changed by the BSSE. The curves obtained by using vDZ (Figure 1a) and vTZ (Figure 1b) are extremely similar, a fact that indicates that the BSSE converges very slowly to zero as a function of the basis-set size, probably because of the presence of diffuse functions[23] (we remind the reader that BSSE vanishes for a complete basis set). In Table 1, BSSE-uncorrected and -corrected equilibrium distances and energy-well depths are reported together with the zero-point energy, the number of vibrational bound states, and the anharmonic vibrational frequency, determined after counterpoise correction. Again, these parameters show the same behavior as the ones previously described.

In Figure 2 (a and b), the same curves are displayed, relative to the asymptotic region (from 12.0 to 20.0 bohr). Again, the curves obtained by using vDZ (Figure 2a) and vTZ (Figure 2b) are very similar, but it appears now that the behavior of CI is extremely different than that of CC.

As discussed in the Introduction, the leading terms of the asymptotic energy are given by the equation

$$E(R) = E_\infty + C_6R^{-6} + C_8R^{-8} \qquad (4)$$

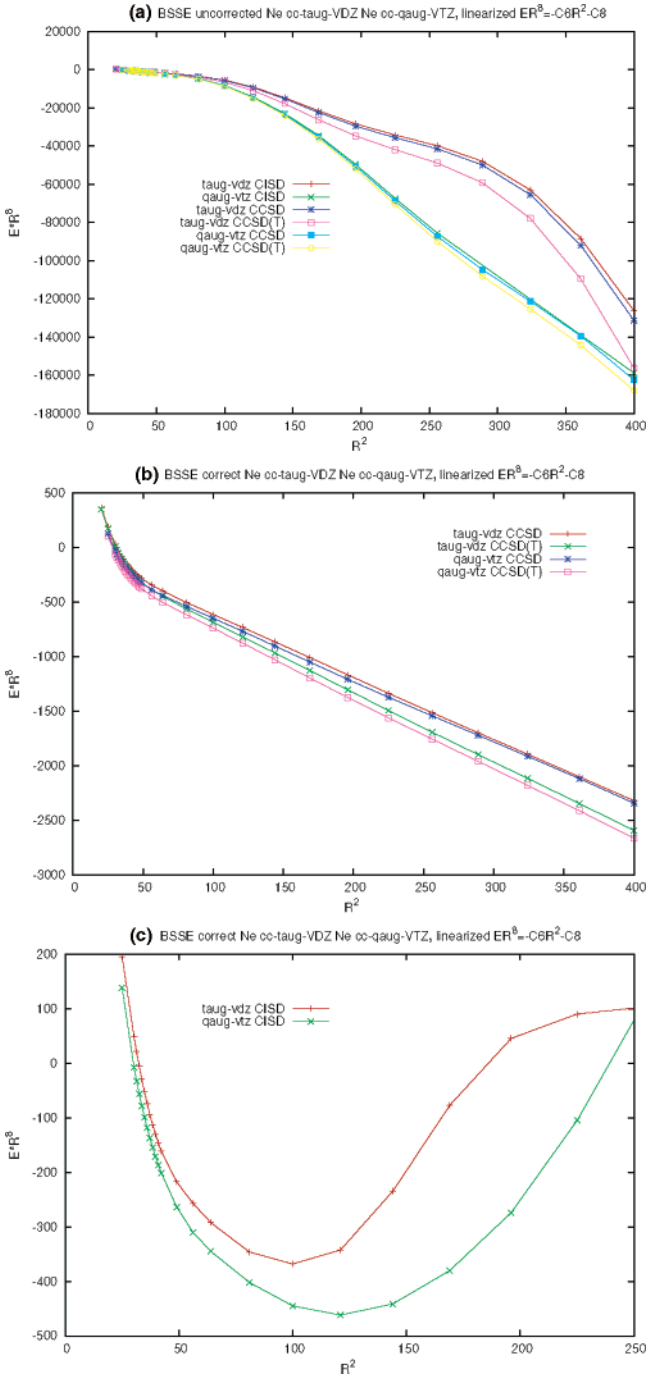


Figure 3. *ER*^{8} as a function of R^2 (see text), in the asymptotic region. (a) taug-vDZ and qaug-VTZ CISD, CCSD, and CCSD-(T) BSSE-uncorrected. (b) taug-vDZ and qaug-VTZ CCSD and CCSD(T) BSSE-corrected. (c) taug-vDZ and qaug-VTZ CISD BSSE-corrected. Units: bohr^2 versus hartree·bohr^8.

By multiplying this expression by R^8 and rearranging the different terms, one gets

$$[E(R) - E_\infty]R^8 = C_6R^2 + C_8 \qquad (5)$$

This means that, if one plots the quantity $[E(R) - E_\infty]R^8$ as a function of R^2, the result should be a straight line for large values of R. The results of these plots are shown in Figure 3, for four different cases: uncorrected CI and CC (3a) and BSSE-corrected CC (3b) and CI (3c). It is clear that, in the case of uncorrected energies, either CI or CC, the long-range curves are far from being straight lines. This implies that the BSSE completely masks the correct long-range behavior of the potential energy for this VdW species. On the other hand, once the BSSE has been corrected, the CC results [either CCSD or CCSD(T)] have a correct linear behavior. Rather surprisingly, however, this is not true for the CISD results, as it could have been guessed from the long-range tail of the potential, Figure 2. When a linear least-square regression was used, it was possible to obtain values of the dispersion coefficients from the coupled cluster, BSSE-corrected, potential curves: results are collected in Table 2. These values can be compared with the results obtained from LRPT treatment and are collected and, with experimental ones, reported again in Table 2. In Table 2, computed or experimental values of polarizabilities are presented too (see also Figure 4).

4. Discussion

Two main aspects can be underlined from the analysis of the data: the long-range behavior of the potential-energy curves with the determination of dispersion coefficients, in particular, for the failure of CISD, and the determination of spectroscopic properties from the analysis of the equilibrium region of the curves.

4.1. Dispersion Coefficients: The Failure of CISD. The remarkable difference in the long-distance part of potentials that are overall substantially similar is rather unexpected. The reason can be traced back to a subtle interplay between two different sources of error that affect CI calculations: basis-set superposition error and size-consistency error (SCE). SCE originates from the fact that, in truncated-CI calculation, determinants that are present in the product of the monomer WF are absent in the dimer WF. For this reason, the CISD energy of two fragments separated by such a large distance, that they are physically noninteracting, is different from the sum of the CISD energies of the isolated fragments. The SCE is far from being negligible: in fact, the CISD energy of two noninteracting neon atoms is about 0.15 hartree higher than the sum of the corresponding energies of isolated atoms. However, the SCE depends only weakly on the geometry: once the BSSE has been taken into account via the counterpoise correction, the CISD values for the equilibrium distance and dissociation energy are in reasonable accord with the corresponding CCSD values (which are SCE-free) and also in a reasonable accord with the experimental[24,25] and previously computed values.[26,27] For this reason, CISD can be used to compute the spectroscopic quantities of a VdW dimer as Ne_2, although the obtained results are certainly less accurate than those obtained from CCSD, and much less accurate than those from CCSD(T). As expected, the use of diffuse functions appears to be of great importance to improve the computation of dispersion interactions; as an example, we can consider the Ne_2 CISD BSSE-corrected energy-well depth computed with the vDZ basis set during a preliminary study: in that case, a value of about 30 μhartree was obtained to be compared with 83.5 μhartree for taug-vDZ (experimental value 134 μhartree).

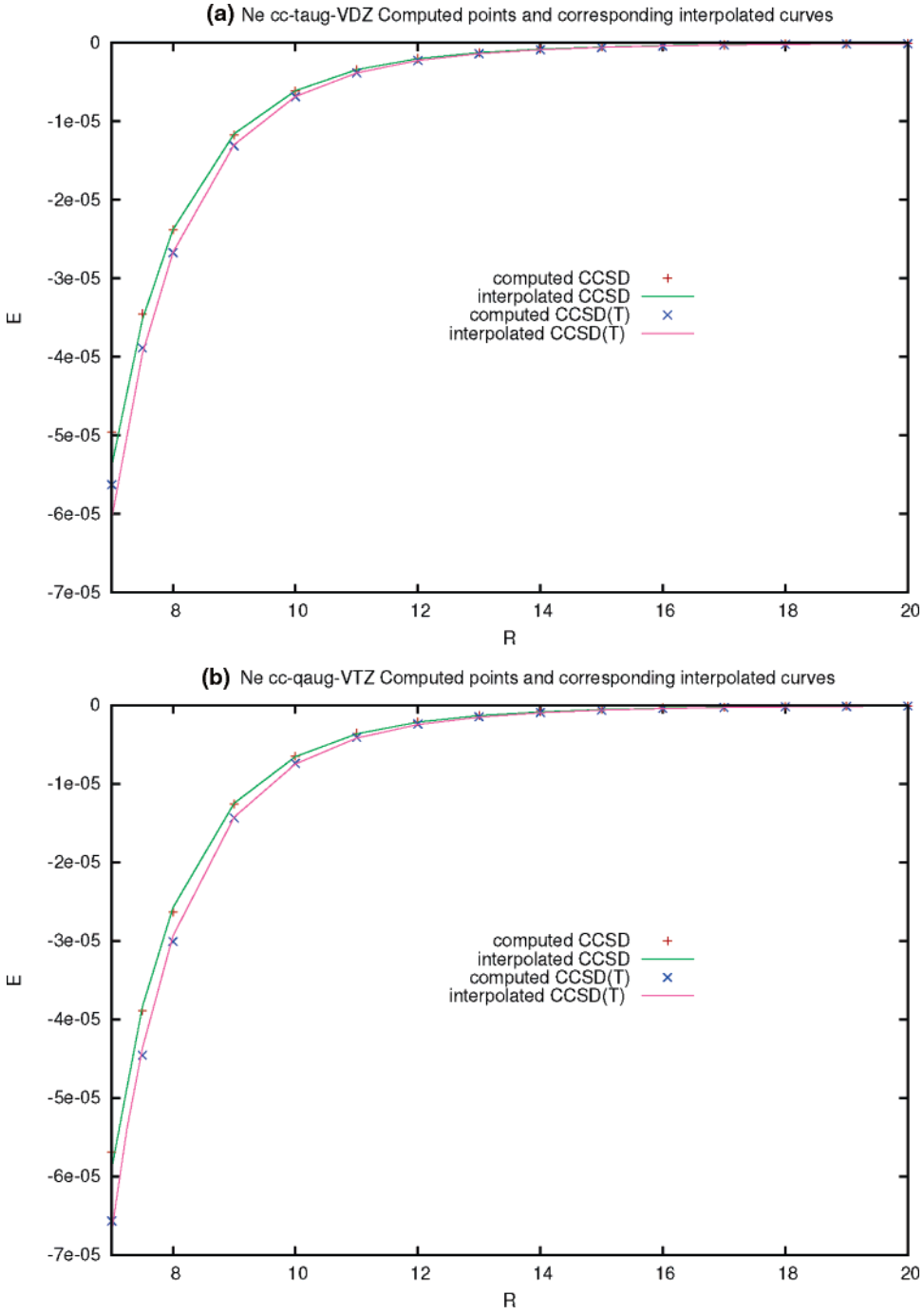


Figure 4. Computed points and the corresponding interpolated curves, in the asymptotic region. (a) taug-vDZ. (b) qaug-vTZ. Units: distances in bohr and energies in hartree.

Let us consider now the dispersion coefficients. Before being corrected to take into account the BSSE, the long-range tail of the potential-energy curves gives absolutely unreliable results. Once the BSSE has been taken into account via the CP correction, the CC curves fit very well into the long-range expression, and the values of the dispersion coefficients are in good agreement with both the FCI and experimental values. The situation is completely different for the CISD calculations, which cannot be fitted with the theoretical expression at large distances. In this case, the CP correction overcorrects the energy values, which become even higher than the corresponding asymptotic values. This is because the CP correction is extracted from **atomic** calculation, while it is used to correct **molecular** energies. The (relatively small) error due to the lack of size consistency of the CI results has a dramatic effect on the long-range tail of the potential-energy curves. In fact, the sum of the atomic energies is larger than the energy of noninteracting atoms, giving therefore a too-large correction. For this reason, the CP correction **overestimates** the effect of BSSE, thus giving a long-range tail of the potential that is completely artifactual.

4.2. Spectroscopic Properties. As the spectroscopic properties are concerned, as already stated, our values can be compared with a recent experimental work by Wüest and Merkt.[25] In that paper, the authors determine the position of

rovibrational energy levels of the Ne_2 dimer using vacuum ultraviolet laser spectroscopy. The potential curve for the ground electronic state was subsequently determined by means of a nonlinear fitting of a model interaction potential to the measured position of the rovibrational levels. It is quite interesting to see how the zero-point energy level lies very high in energy; in fact, it accounts for about 40% of the well depth, leading to a very low binding of the complex; this fact is anyway confirmed by experimental results. Moreover, Wüest and Merkt[25] observe only two vibrational levels, in agreement with our results, but from the analysis of the potential, they predict the existence of a third vibrational level with a very low binding energy. The existence of this level is, anyways, still uncertain and depends strongly on the energy-well depth and from the form of the long-range tail of the potential due to the high diffuse nature of the vibrational WF. The computed spectroscopic properties can be improved toward the basis set limit using a two-point basis set extrapolation formula.[28] Applying this formula to the CCSD(T) BSSE-corrected results, we obtain $E_{min} = -130$ μhartree, $R_{min} = 5.82$ bohr, $\Delta E_0 = 12.38$ μhartree, and $\omega = 13.5$ cm^{-1}.

5. Conclusions

It has been shown that the BSSE plays a key role in the numerical calculation of the dispersion coefficients of VdW species. No reasonable value of dispersion constants or equilibrium properties can be obtained for the Ne_2 dimer if one does not take into account the BSSE correction. The use of counterpoise correction allows one to obtain satisfactory results provided one uses size-consistent methods for the computation of the potential-energy curves of the dimer. Values obtained in such a way with CCSD or CCSD(T) agree quite well with experimental values and with the BSSE-free LRPT values. On the other hand, the application of the counterpoise correction to curves obtained with non-size-consistent methods gives quite good values for the equilibrium properties but totally wrong dispersion coefficients. This fact is due to a subtle interplay between basis-set-superposition and size-consistency errors. By using the potential-energy curves obtained at CI and CC levels, we computed the zero-point energy and the anharmonic vibrational frequency for the fundamental electronic state of Ne_2, showing the existence of two bound vibrational states; our results agree quite well with spectroscopic experiments. Finally, we show that a new data format and set of libraries (Q5Cost) were used in our calculations in order to interface the different codes used in the present study.

Acknowledgment. This work was partly supported by the European Community through the COST actions D23 and D26. One of us (S.E.) would like to thank the CINECA consortium for financial support and kind hospitality within the HPC−Europa program.

References

(1) Haberland, H. *Clusters of Atoms and Molecules*; Springer: Berlin, 1994.

(2) Sebastianelli, F.; Baccarelli Di Paola, C.; Gianturco, F. A. *J. Chem. Phys.* **2003**, *119*, 5570−5582.

(3) Filippone, F.; Gianturco, F. *Europhys. Lett.* **1998**, *44*, 585−591.

(4) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.

(5) Spelsberg, D.; Lorentz, T.; Meyer, W. *J. Chem. Phys.* **1993**, *99*, 7845−7858.

(6) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(7) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem Phys.* **1992**, *96*, 6796−6806.

(8) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 2975−2988.

(9) Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, version 02/25/04, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory, which is part of the Pacific Northwest Laboratory, P. O. Box 999, Richland, Washington 99352, and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multiprogram laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76.RLO 1830. Contact Karen Schuchardt for further information. http://www.emsl.pnl.gov/forms/basisform.html (accesed Dec 6, 2004).

(10) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford, U. K., 1996.

(11) Bendazzoli, G. L. *Int. J. Quantum Chem.* **2005**, *104*, 38−51.

(12) Bendazzoli, G. L.; Evangelisti, S. *J. Chem. Phys.* **1993**, *98*, 3141−3150.

(13) MOLPRO is a package of ab initio programs: Werner, H.-J.; Knowles, P. J.; Almlöf, J.; Amos, R. D.; Berning, A.; Deegan, M. J. O.; Eckert, F.; Elbert, S. T.; Hampel, C.; Lindh, R.; Meyer, W.; Nicklass, A.; Peterson, K.; Pitzer, R.; Stone, A. J.; Taylor, P. R.; Mura, M. E.; Pulay, P.; Schütz, M.; Stoll, H.; Thorsteinsson, T.; Cooper, D. L. *MOLPRO*; University College Cardiff Consultants Limited: Cardiff, Wales, U. K., 2006.

(14) Bendazzoli, G. L.; Evangelisti, S. *Gazz. Chim. Ital.* **1996**, *126*, 619−623.

(15) *CASDI*; Toulouse University: Toulouse, France. See also: Ben Amor, N.; Maynau, D. *Chem. Phys. Lett.* **1998**, *286*, 211.

(16) DALTON, a molecular electronic structure program, release 2.0 (2005). http://www.kjemi.uio.no/software/dalton/dalton.html (accesed Dec 6, 2006).

(17) Numerov, B. *Publ. Obs. Central Astrophys. Russ.* **1923**, *2*, 188−288.

(18) Lindberg, B. *J. Chem. Phys.* **1988**, *88*, 3805−3810.

(19) Passarini, F. *Tesi di Laurea*; Bologna University: Bologna, Italy.

(20) Andersson, K.; Blomberg, M. R. A.; Fülscher, M. P.; Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Neogrády, P.; Olsen, J.; Roos, B. O.; Sadlej, M.; Schütz, A. J.; Seijo, L.; Serrano-Andrés, L.; Siegbahn, P. E. M.; Widmark, P.-O. *Molcas 4*; Lund University: Lund, Sweden, 1997.

(21) Angeli, C.; Bendazzoli, G. L.; Borini, S.; Cimiraglia, R.; Emerson, A.; Evangelisti, S.; Maynau, D.; Monari, A.; Rossi, E.; Sanchez-Marin, J.; Szalay, P. G.; Tajti, A. Submitted.

(22) Borini, S.; Monari, A.; Rossi, E.; Tajti, A.; Angeli, C.; Bendazzoli, G. L.; Cimiraglia, R.; Emerson, A.; Evangelisti, S.; Maynau, D.; Sanchez-Marin, J.; Szalay, P. G. Submitted.

(23) Hanspeter, H.; Kirchner, B.; Solca, J.; Steinebrunner, G. *Chem. Phys. Lett.* **1997**, *266*, 388−390.

(24) Aziz, R. A.; Slamam, M. J. *Chem. Phys.* **1989**, *130*, 187−194.

(25) Wüest, A.; Merkt, F. *J. Chem. Phys.* **2003**, *118*, 8807−8812.

(26) Woon, D. E. *J. Chem. Phys.* **1994**, *100*, 2838−2850.

(27) Gdanitz, R. J. *Chem. Phys. Lett.* **2001**, *348*, 67−74.

(28) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, E.; Olsen, J.; Wlison, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243−252..

(29) Kumar, A.; Meath, W. J. *Can. J. Chem.* **1985**, *63*, 1616−1630.

(30) Nicklass, A.; Dolg, M.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1995**, *102*, 8942−8951.

CT6003005

JCTC Journal of Chemical Theory and Computation

Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms

Kwangho Nam,[†,‡] Qiang Cui,[§] Jiali Gao,[†,‡] and Darrin M. York*[,†]

Department of Chemistry and Supercomputing Institute and the Digital Technology Center, University of Minnesota, Minneapolis, Minnesota 55455−0431, and Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, 1101 University Avenue, Madison, Wisconsin 53706

Received July 30, 2006

Abstract: A semiempirical AM1/d Hamiltonian is developed to model phosphoryl transfer reactions catalyzed by enzymes and ribozymes for use in linear-scaling calculations and combined quantum mechanical/molecular mechanical simulations. The model, designated AM1/d-PhoT, is parametrized for H, O, and P atoms to reproduce high-level density-functional results from a recently constructed database of quantum calculations for RNA catalysis (http://theory.chem.umn.edu/Database/QCRNA), including geometries and relative energies of minima, transition states and reactive intermediates, dipole moments, proton affinities, and other relevant properties. The model is tested in the gas phase and in solution using a QM/MM potential. The results indicate that the method provides significantly higher accuracy than MNDO/d, AM1, and PM3 methods and, for the transphosphorylation reactions, is in close agreement with the density-functional calculations at the B3LYP/6-311++G(3df,2p) level with a reduction in computational cost of 3−4 orders of magnitude. The model is expected to have considerable impact on the application of semiempirical QM/MM methods to transphosphorylation reactions in solution, enzymes, and ribozymes and to ultimately facilitate the design of improved next-generation multiscale quantum models.

1. Introduction

Phosphorylation and dephosphorylation reactions play a central role in biochemical systems.[1] The formation and hydrolysis of ATP constitute a central mechanism for the storage and utilization of chemical energies in metabolic pathways.[2,3] Furthermore, transphosphorylations provide a key regulatory mechanism in eukaryotic cellular signaling.[4] These processes are catalyzed by enzymes such as the proton gradient-driven ATPases and protein kinases and phosphatases. Of particular interest is the study of RNA enzymes or ribozymes.[5] The understanding of the molecular mechanisms of ribozyme catalysis has been greatly facilitated by the study of small prototype self-cleaving RNA[6] such as the hammerhead,[7,8] hairpin,[9,10] and hepatitis delta virus[11−13] ribozymes.

The importance of transphosphorylation reactions has stimulated extensive theoretical and experimental investigations aimed at the identification and characterization of the underlying catalytic mechanisms.[3,14−17] Experimentalists have widely used small model compounds, such as phosphate mono- and diesters with variety of leaving groups, to carry out kinetic experiments. Many different types of experiments led to the consensus that, in solution, hydrolysis of phosphate monoesters follows a predominantly unimolecular dissociative mechanism, whereas phosphate diesters and triesters are

* To whom correspondence should be addressed. E-mail: york@chem.umn.edu.

† Department of Chemistry and Supercomputing Institute, University of Minnesota.

‡ Digital Technology Center, University of Minnesota.

§ Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin.

10.1021/ct6002466 CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/09/2007

hydrolyzed by a bimolecular associative mechanism via pentacovalent intermediates, transition states, or both.[3,14,16−18] Although these general conclusions have been drawn for phosphoryl-transfer reactions in solution, there is no consensus about the mechanism in enzymes and ribozymes. Unfortunately, kinetic models are not always able to discern between multiple mechanistic pathways that fit experimental data equally well,[19−21] which underscores the need for theoretical studies to help interpret experimental data.

Ab initio electronic structure methods and density functional theory (DFT) have been widely used to study phosphoryl-transfer mechanisms and to help interpret kinetic data.[19−42] However, most of these calculations have been performed on small model reactions in the gas phase or with the aid of implicit solvation models. Although these approaches provide insight into the nature of the reactions, it is of great interest to study the reactions using explicit models in aqueous solution and in the active sites of enzymes or ribozymes, where specific electrostatic interactions, hydrogen bonds, and solvent dynamics are particularly important.[43−50] Phosphoryl-transfer reactions typically involve highly charged species that undergo considerable changes in hybridization along the reaction path. This necessitates the use of a large, explicit solvation and counterion environment coupled with an accurate quantum mechanical treatment of biological phosphates that includes a d-orbital description of phosphorus. The computational cost of the calculations, however, currently precludes simulations using high-level density-functional methods if sufficient sampling is carried out.

Semiempirical quantum methods such as the modified neglect of diatomic overlap (MNDO),[51,52] Austin model 1 (AM1),[53] and parametric method 3 (PM3)[54] methods are typically 3−4 orders of magnitude faster than DFT methods but currently are not sufficiently accurate to model phosphoryl-transfer reactions. This is, in part, because the parametrization of these models does not include transition state data in the training set and because of the lack of d orbitals for phosphorus. Nonetheless, the remarkable computation efficiency of these methods allows very large scale problems to be addressed using linear-scaling electronic structure methods[55−58] and explicit solvent dynamics through molecular dynamics (MD) simulations using combined quantum mechanical/molecular mechanical (QM/MM) potentials.[48,59−62] The d orbital extension of the MNDO method (MNDO/d)[63,64] has recently been applied to QM/MM simulations of phosphoryl transfer reactions in solution.[45,46,65] Other recently developed semiempirical methods that show promise include the OMx models,[66,67] the PDDG/PM3 model,[68−70] the PM3-MAIS and PM3-PIF models,[71,72] the SCC-DFTB method,[73−76] and the very recent NO-MNDO model.[77] One problem that is prevalent in almost all semiempirical models that utilize a minimal valence basis methods is that they typically underpredict molecular polarizabilities as a function of charge. Recently a method has been proposed that greatly improves the modeling of charge-dependent response properties without increasing the atomic orbital basis and with minimal computational overhead.[78] Although there were efforts to design semiempirical models for specific reactions in phosphatases,[79,80] currently, none of these models yet provides sufficient accuracy for the phosphoryl-transfer reactions of interest in the present work. An important step toward the development of new-generation quantum models that afford significantly greater accuracy and transferability is to first identify and quantify the accuracy limits of existing models for important biological reactions. It is the hope that in this way complementary quantum models for QM/MM calculations such as semiempirical and SCC-DFTB methods can be directly compared and their advantages and disadvantages characterized and understood. These efforts will ultimately help lay the groundwork for the design of more robust and efficient quantum models for biocatalysis.

The aim of the present work is to develop a new semiempirical Hamiltonian model, designated AM1/d-PhoT, that will allow accurate calculation of a wide range of phosphoryl-transfer reactions by determination of a set of specific reaction parameters (SRP)[81] for phosphoryl transfer within the AM1 formulation with d-orbital extension. The new model has been designed to accurately reproduce high-level DFT results such as geometries, dipole moments, proton affinities, and relative energies for a large set of molecules, complexes, and chemical reactions relevant to biological phosphoryl transfer. These data were taken from a recently constructed database of quantum calculations for RNA catalysis (QCRNA).[82,83] The resulting AM1/d-PhoT Hamiltonian is tested in the gas phase and in solution using a combined QM/MM potential and is demonstrated to provide a very good agreement with high-level DFT results for transphosphorylation reactions and offers a considerable improvement over the AM1, PM3, and MNDO/d methods.

2. Theory

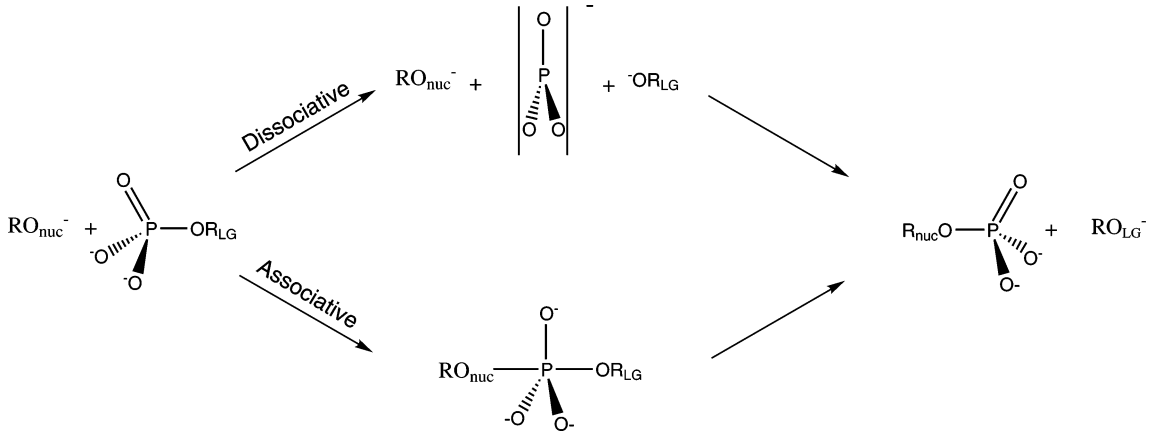
Semiempirical MNDO and AM1 types of Hamiltonians share many common features and have been discussed extensively elsewhere.[84−87] In this section, the major differences of these models are outlined, and a description of the slightly modified AM1/d-PhoT model is provided.

2.1. Semiempirical Core−Core Repulsion Term. The MNDO, AM1, PM3, and MNDO/d semiempirical models discussed here are all based on the neglect of the diatomic differential overlap (NDDO) approximation, as is the new AM1/d-PhoT model. The MNDO and MNDO/d Hamiltonians differ from the AM1/PM3 Hamiltonians in the way that core−core repulsion interactions are treated. Additionally, the MNDO/d method includes a set of d orbitals on second-row elements. In the MNDO and MNDO/d methods, the repulsion between two nuclear cores (A and B) is determined by

$$E_{\mathrm{N}}^{\mathrm{MNDO}}(\mathrm{A},\mathrm{B}) = Z_{\mathrm{A}}Z_{\mathrm{B}}\langle s_{\mathrm{A}}s_{\mathrm{B}}|s_{\mathrm{A}}s_{\mathrm{B}}\rangle(1 + e^{-\alpha_{\mathrm{A}}R_{\mathrm{AB}}} + e^{-\alpha_{\mathrm{B}}R_{\mathrm{AB}}}) \quad (1)$$

where Z_{A} and Z_{B} are the effective core charges, $\langle s_{\mathrm{A}}s_{\mathrm{B}}|s_{\mathrm{A}}s_{\mathrm{B}}\rangle$ is a Coulomb repulsion integral between two s orbitals centered on atoms A and B, and α_{A} and α_{B} are parameters that decrease screening of the nuclear charge by the electrons at small interatomic distances. For the O−H and N−H bonds, a slightly different screening form is used, a detailed

Scheme 1. Reaction Schemes of Dissociative Mechanism and Associative Mechanism for Transesterification Reaction



description for which is provided in the original paper[51] and elsewhere.[88]

In the AM1 and PM3 Hamiltonians, the core−core term includes an additional set of Gaussian terms that depend parametrically on the nature of atoms and takes the form

$$E_{\mathrm{N}}(\mathrm{A,B}) = E_{\mathrm{N}}^{\mathrm{MNDO}}(\mathrm{A,B}) + \frac{Z_{\mathrm{A}}Z_{\mathrm{B}}}{R_{\mathrm{AB}}}\left(\sum_k a_{\mathrm{A}k}\mathrm{e}^{-b_{\mathrm{A}k}(R_{\mathrm{AB}}-c_{\mathrm{A}k})^2} + \sum_k a_{\mathrm{B}k}e^{-b_{\mathrm{B}k}(R_{\mathrm{AB}}-c_{\mathrm{B}k})^2}\right) \quad (2)$$

The addition of Gaussian core−core terms leads to significant improvements in the performance of semiempirical models in energies, especially for interactions at the hydrogen-bonding distances.[54] The MNDO and MNDO/d models are known to be problematic in the description of noncovalent intermolecular interactions because of the excessive repulsion just outside bonding distances. The Gaussian core−core terms in eq 2 compensate for these undesirable repulsions. The Gaussian core−core terms are empirical adjustments to the potential devoid of any rigorous physical meaning. Previous studies showed that serious artifacts can be introduced if one only changes the empirical Gaussian core−core repulsion functions.[72,89] Nevertheless, several new core−core interaction potentials or parameters have been proposed, including the PDDG/PM3 model,[68] the PM3-MAIS and PM3-PIF models,[71,72] and the $\mathrm{PM3_{BP}}$ model.[88] These models have been designed to offer improvements in the description of proton-transfer reactions and hydrogen bonds.

2.2. AM1/d-PhoT Model Employing the Modified AM1 Formalism. It has been well-established that the study of hypervalent phosphates, such as the transition states and reactive intermediates formed in biological transphosphorylation, require an explicit *d* orbital representation.[63,64,90] In particular, transphosphorylation reactions typically involve a change of valency on phosphorus along the reaction path: it changes from tetracovalent to tricovalent in the dissociative mechanism or changes from tetracovalent to pentacovalent in the associative and concerted mechanisms (Scheme 1). The MNDO/d model[63,64] has been demonstrated to perform reliably in the study of transphosphorylation under basic conditions,[45,46] where there is no hydrogen bonding or proton transfer that occurs within the quantum region. However, the MNDO/d model was found to be of limited reliability for transphosphorylation reactions in the neutral to acidic pH range where hydrogen bonding and proton transfer play major roles.

The derivation of a new model for phosphoryl transfer reactions under more general conditions requires components of both the AM1/PM3 and MNDO/d models. The AM1 and PM3 models offer a significant improvement for hydrogen bonding relative to MNDO and MNDO/d models but have the problem that they over-stabilize hypervalent structures because of the artificially attractive core−core interactions.[64,70,90] Therefore, in the design of an AM1/d model for phosphoryl transfer reaction, it is desirable to keep the core−core interactions for hydrogen bonding but to turn these interactions off for phosphorus bonding where the d orbitals allow proper hybridization and accurate representation of hypervalent species. In this way, a balanced model may be achieved that accurately models reactive intermediates in transphosphorylation with, at the same time, improved treatment of hydrogen bonding. Toward this end, a scale factor was introduced into the Gaussian core−core terms in the present AM1/d-PhoT model as

$$E_{\mathrm{N}}(\mathrm{A,B}) = E_{\mathrm{N}}^{\mathrm{MNDO}}(\mathrm{A,B}) + \frac{Z_{\mathrm{A}}Z_{\mathrm{B}}}{R_{\mathrm{AB}}} G_{\mathrm{scale}}^{\mathrm{A}} G_{\mathrm{scale}}^{\mathrm{B}}\left(\sum_k a_{\mathrm{A}k}\mathrm{e}^{-b_{\mathrm{A}k}(R_{\mathrm{AB}}-c_{\mathrm{A}k})^2} + \sum_k a_{\mathrm{B}k}\mathrm{e}^{-b_{\mathrm{B}k}(R_{\mathrm{AB}}-c_{\mathrm{B}k})^2}\right) \quad (3)$$

where $G_{\mathrm{scale}}^{\mathrm{A}}$ and $G_{\mathrm{scale}}^{\mathrm{B}}$ are scaling parameters for atoms A and B, and in the present work, they vary from zero to one (values of 0 recover the conventional MNDO core−core model, whereas values of 1 recover the AM1 core−core model). Alternatively, the product $G_{\mathrm{scale}}^{\mathrm{A}}\,G_{\mathrm{scale}}^{\mathrm{B}}$ can be made into pairwise terms for specific atom pairs. It is worthwhile to mention here that setting Gaussian core−core parameters $a_{\mathrm{A}k}$ to be zero for a certain atom, such as P, *would not* eliminate the Gaussian core−core interactions involving P atoms since all terms from the other atoms would remain. The G_{scale} scaling parameters hence provide the flexibility to attenuate (or even shut off) Gaussian core−core interactions between certain atoms and offers a simple mechanism for interconverting between AM1-like models and MNDO-

Table 1. Optimized AM1/d-PhoT Parameters for Hydrogen, Oxygen, and Phosphorus Atoms, along with Original AM1 Parameters for Comparison[a]

	H		O		P	
parameters	AM1/d	AM1	AM1/d	AM1	AM1/d	AM1
U_{ss} (eV)	−10.934610	−11.396427	−96.760676	−97.830000	−46.250810	−42.029863
U_{pp} (eV)			−78.776203	−78.262380	−40.712918	−34.030709
ζ_s (au)	1.143846	1.188078	3.057965	3.108032	1.909168	1.981280
ζ_p (au)			2.515332	2.524039	2.008466	1.875150
β_s (eV)	−5.911108	−6.173787	−29.472306	−29.272773	−11.194791	−6.353764
β_p (eV)			−28.515785	−29.272773	−11.985621	−6.590709
α (1/Å)	2.884915	2.882324	4.404417	4.455371	1.883237	2.455322
G_{ss} (eV)	13.737453	12.848000	14.234714	15.420000	14.645747	11.560005
G_{pp} (eV)			14.454530	14.520000	11.694918	7.877589
G_{sp} (eV)			14.539451	14.480000	5.689654	5.237449
G_{p2} (eV)			12.942259	12.980000	10.328696	7.307648
H_{sp} (eV)			4.339705	3.940000	1.175115	0.779238
U_{dd} (eV)					−24.504161	
ζ_d (au)					0.840667	
β_d (eV)					−2.360095	
$\tilde{\zeta}_s$ (au)					2.085120	
$\tilde{\zeta}_p$ (au)					1.535336	
$\tilde{\zeta}_d$ (au)					1.236266	
ρ_{core} (au)					1.185130	
G_{scale}[b] (unitless)	1.000000		1.000000		0.353722	
a_1 (unitless)	0.106238	0.122796	0.288526	0.280962	−0.344529	−0.031827
b_1 (1/Å^2)	5.735290	5.000000	4.883265	5.000000	3.034933	6.000000
c_1 (Å)	1.261430	1.200000	0.850910	0.847918	1.134275	1.474323
a_2 (unitless)	0.004043	0.005090	0.061586	0.081430	−0.021847	0.018470
b_2 (1/Å^2)	7.080122	5.000000	4.435791	7.000000	1.684515	7.000000
c_2 (Å)	2.084095	1.800000	1.353681	1.445071	2.716684	1.779354
a_3 (unitless)	−0.002800	−0.018336			−0.036003	0.033290
b_3 (1/Å^2)	0.739913	2.000000			5.243357	9.000000
c_3 (Å)	3.649474	2.100000			1.924175	3.006576

[a] Standard notation for parameters taken from refs 84 and 63. [b] Scale factor of Gaussian core−core repulsion interactions.

like models. As will be shown below, adjustment of the G_{scale} parameter allows significantly improved behavior of AM1/d-PhoT model for phosphoryl transfer reactions along with the specific reaction parametrizations.

3. Parametrization Procedure

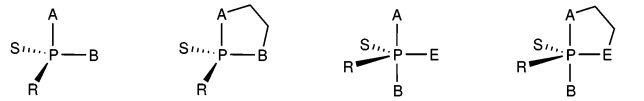
This section describes the methods and procedures employed to develop the specific reaction parameters for the AM1/d-PhoT model. The first subsection describes the DFT dataset used as the reference data. The second subsection describes the details of the parameter optimization procedure. The final parameters for the AM1/d-PhoT model are listed in Table 1.

3.1. Density Functional Calculations. All of the DFT datasets in this work, including the full structural and thermodynamic quantities, are available in the recently constructed QCRNA database[82] which is available on-line.[83] All DFT calculations were carried out using Kohn−Sham density functional theory with the three-parameter hybrid exchange functional of Becke[91,92] and the Lee, Yang, and Parr correlation functional[93] (B3LYP). Energy minimum and transition state geometry optimizations were carried out with default convergence criteria, while the stability conditions of the restricted closed-shell Kohn−Sham determinant for each final structure were verified.[94,95] Frequency calculations were performed to establish the nature of all stationary points and to evaluate thermodynamic quantities. The 6-31++G-(d,p) basis set was used in the calculations.

Electronic energies and dipole moments were refined by single-point energy calculations using the 6-311++G-(3df,2p) basis set and the B3LYP hybrid density functional at the optimized geometries. All single-point calculations were run with the convergence criteria on the SCF wave function tightened to 10^{-8} au to ensure high precision for properties sensitive to the use of diffuse basis functions.[96] The protocol applied to obtain the (refined energy)// (geometry and frequencies) is designated by the abbreviated notation B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G-(d,p). All DFT calculations were performed with the GAUSSIAN03 suite of programs.[97] This density-functional protocol has been extensively tested and applied to model biological phosphorus compounds.[23−25,47,98,99] The gas-phase proton affinities at 298.15 K for the DFT dataset were calculated by the procedure described by Range et al. without empirical correction.[100]

3.2. AM1/d-PhoT Parameter Optimization. This section presents details for the AM1/d-PhoT parametrization procedure for biological phosphorus compounds. The first section describes the construction of a set of compounds and their properties that are used as the target in the parameter

Scheme 2. Nomenclature Convention for Ligand Designations in Acyclic and Cyclic Phosphate and Phosphorane Compounds of Biological Interest[a]



P(R)(S)(A)(B) P(R)(S)(AB) P(R)(S)(E)(A)(B) P(R)(S)(EA)(B)

[a] This nomenclature is consistent with the naming convention used for similar compounds in previous work.[25,98]

optimization. The second subsection describes the construction of χ^2 merit function that was used in the parametrization procedure. The third subsection briefly outlines the parametrization strategy that was applied to arrive at the final optimized semiempirical parameters (Table 1).

3.2.1. Construction of Training Dataset. The primary focus of the current work is to develop specific reaction parameters that accurately reproduce the structures, energetics, and other properties for the transphosphorylation reactions. Taking into account of the catalytic mechanisms of several ribozymes, protein kinases, and phosphatases, which involve general acid and base catalysis, we produced a molecular set which was used to construct the training set for parametrizations and which includes the following list of properties.

Molecules. Eleven non-phosphorus-containing molecules and polyatomic ions and 16 phosphorus-containing molecules and polyatomic ions that included metaphosphates, phosphates, and phosphoranes in various protonation states, and one diphosphate ion were included in the training set. (See Scheme 2 for the nomenclature convention for phosphorus-containing compounds.)

Complexes. Three hydrogen-bonded complexes involving water with water, methanol, and acetate were included in the dataset.

Reaction Mechanisms. Three dissociative reactions and five associative reactions of various phosphates and all stationary points of these reactions were included.

Potential Energy Surfaces. One hundred fifty intermediate configurations along both one and two-dimensional potential energy surfaces (PES) for three reactions were included in the final stages of parameter refinement, one of which was chosen from the five associative reactions mentioned above, and two additional reactions were added.

All compounds and structures selected in the parameter training sets, energies, and relevant other properties are available from the on-line QCRNA database via the Internet (see ref 83). The properties considered in the parametrizations included the following.

Heat of Formation. Fourteen experimental values of heats of formation were used, including three phosphorus containing compounds. They are listed in Table 2.

Proton Affinity. Seven experimental gas-phase proton affinity (PA) values were included. In addition, 14 relative PA values for phosphate and phosphorane compounds calculated with DFT were used. These values are shown in Table 3.

Dipole Moment. The dipole moments for 11 neutral molecules and four hydrogen-bonded complexes determined from DFT calculations were used and are shown in Table 4.

Structure. Geometrical parameters including bond lengths, angles, and dihedrals were considered for all molecules in the training set. If the molecules or complexes form hydrogen bonds, the hydrogen bond distances and angles were also included in the parametrizations. The results are summarized in Table 5.

Intermolecular Interaction Energy. The intermolecular interaction energies of three hydrogen-bonded complexes were included and are listed in Table 6.

Relative Reaction Energy. The relative energies of intermediates and transition states for all reactions in the training

Table 2. Experimental (exptl) and Computational Heats of Formation for Compounds Used in Parametrizations (kcal/mol)

	exptl	AM1/d		AM1		PM3		MNDO/d	
molecule[a]	ΔH_f	ΔH_f[b]	error[c]	ΔH_f[b]	error[c]	ΔH_f[b]	error[c]	ΔH_f[b]	error[c]
H_3O^+	138.0[d]	140.1	1.2	143.5	4.6	159.1	20.2	134.2	−4.7
H_2O	−57.8[d]	−56.8	1.0	−59.2	−1.4	−53.4	4.4	−60.9	−3.1
HO^-	−33.2[e]	−26.8	6.4	−14.1	19.1	−17.5	15.7	−5.8	27.4
CH_3OH	−48.1[f]	−53.7	−5.6	−57.0	−8.9	−51.9	−3.8	−57.4	−9.3
CH_3O^-	−36.0[e]	−35.9	0.1	−38.5	−2.5	−37.9	−1.9	−39.7	−3.7
C_2H_5OH	−56.2[f]	−53.5	2.7	−62.7	−6.5	−56.9	−0.7	−63.0	−6.8
$C_2H_5O^-$	−47.5[e]	−38.1	9.4	−45.5	−2.0	−44.8	2.7	−45.3	2.2
C_6H_5OH	−23.0[f]	−19.2	3.8	−22.2	0.8	−21.7	1.3	−26.7	−3.7
$C_6H_5O^-$	−40.5[e]	−38.2	2.3	−41.0	−0.5	−44.1	−3.6	−42.2	−1.7
CH_3CO_2H	−103.3[f]	−102.5	0.8	−103.0	0.3	−102.0	1.3	−101.1	2.2
$CH_3CO_2^-$	−122.5[e]	−123.7	−1.2	−115.4	7.1	−119.7	2.8	−110.0	12.5
$P(CH_3)_3$	−24.2[g]	−16.7	7.5	−22.0	2.2	−29.8	−5.6	−28.1	−3.9
$(CH_3)_3PO$	−103.8[h]	−105.8	−2.0	−101.5	2.3	−82.7	21.1	−105.5	−1.7
PO_3^-	−225.4[d]	−206.1	19.3	−203.0	22.4	−197.2	28.2	−257.3	−31.9
MSE			3.3		2.9		5.9		−1.8
MUE			4.5		5.8		8.1		8.1

[a] Bold molecules are included in the parametrization as test sets. [b] The heat of formation is determined by $\Delta H_f = E_{el}^{mol} + E_{core}^{mol} - \sum_A E_{el}^A + \sum_A \Delta H_f^A$, where E_{el}^{mol} and E_{core}^{mol} are the molecular electronic and core−core repulsion energy, respectively, and E_{el}^A and ΔH_f^A are the electronic energy and the experimental heat of formation for isolated atom A, respectively, where the summations run over all the atoms A in the molecule. [c] Error values are computed as $\Delta H_f^{calcd} - \Delta H_f^{Exptl}$. [d] Ref 131. [e] Ref 132. [f] Ref 133. [g] Ref 134. [h] Ref 135.

Table 3. Gas-Phase Proton Affinity (PA) of Selected Molecules (kcal/mol)[a]

	ref		error[e]				
molecule (AH)[b]	exptl[c]	CBS[d]	AM1/d	AM1	PM3	MNDO/d	DFT[f]
H_3O^+	165.0	164.1	3.8	−2.0	−11.8	5.6	−1.1
H_2O	390.3	392.1	5.4	20.5	11.3	30.6	0.1
CH_3OH	381.5	382.8	2.0	2.7	−1.9	1.8	−2.2
C_2H_5OH	378.2	379.0	2.9	4.7	−0.4	5.2	−2.2
C_6H_5OH	350.1	349.5	−3.4	−3.1	−6.9	0.0	−2.4
CH_3CO_2H	347.2	347.5	−2.7	6.1	0.9	9.6	−0.8
P(O)(OH)(OH)(OH)	330.5	328.1	−3.4	7.6	15.0	−12.2	−2.4
MSE (vs exptl)[g]		0.0	0.7	5.2	0.9	5.8	−1.6
MUE (vs exptl)[g]		1.2	3.4	6.7	6.9	9.3	1.6
HPO_3		310.6	1.5	20.6	35.1	−3.6	−0.1
P(O)(O)(OH)(OH)$^-$		458.9	−1.9	16.8	24.7	−2.8	−1.1
P(O)(O)(O)(OH)2−		581.1	10.4	33.7	36.4	16.3	−1.7
P(O)(OH)(OH)(OCH_3)		329.3	0.3	7.2	14.9	−12.0	0.4
P(O)(O)(OH)(OCH_3)$^-$		454.9	0.7	16.5	22.8	−7.6	−1.4
P(O)(OH)(OCH_3)(OCH_3)		329.4	1.8	7.3	12.3	−14.1	0.7
P(O)(OH)(OCH_2CH_2O)		329.5	−0.2	7.6	11.8	−17.1	−0.1
P(OH)(OH)(OH)(OH*)(OH)		351.0	3.0	9.0	8.3	−1.3	−0.4
P(OH*)(OH)(OH)(OH)(OH)		341.0	1.8	13.6	9.0	−8.7	−1.8
P(OH)(OH)(OCH_2CH_2O)(OH*)		351.9	1.2	5.9	1.7	−11.8	−0.9
P(OH*)(OH)(OCH_2CH_2O)(OH)		343.2	−2.5	8.0	−0.5	−17.4	−1.1
P(OH*)(OCH_3)(OCH_2CH_2O)(OH)		345.2	−3.5	3.6	−2.3	−20.2	−0.7
P(OH)(OCH_3)(OCH_2CH_2O)(OH*)		352.0	2.3	5.4	−0.4	−27.0	−0.8
P(OH*)(OH)(OCH_2CH_2O)(OCH_3)		343.5	−0.7	6.2	−0.9	−19.5	−1.1
MSE (vs CBS-QB3)			1.0	11.5	12.4	−10.5	−0.7
MUE (vs CBS-QB3)			2.3	11.5	12.9	12.8	0.9

[a] PA is defined as the negative of the enthalpy change (ΔH) of the process of $A^-_{(g)} + H^+_{(g)} \rightarrow AH_{(g)}$. [b] Bold molecules and the deprotonated form (A^-) are included in parametrization as training set molecules. [c] The experimental (exptl) and calculated (CBS) reference data were taken from refs 136 and 100, respectively. In computing semiempirical PA values, the experimental heats of formation of proton (365.7 kcal/mol) is used.[137] [d] CBS is a value computed from CBS-QB3 method reported in reference.[100] [e] Error values are computed as $PA_{calcd} - PA_{ref}$, where ref is the exptl values whenever experimental values are available. Otherwise, CBS-QB3 values are used as reference values. [f] DFT proton affinity is computed as a sum of adiabatic energy contribution at B3LYP/6-311++G(3df,2p) level and the enthalpic correction computed from B3LYP/6-31++G(d,p) level. The geometries are optimized at the level of B3LYP/6-31++G(d,p) as described in section 3. [g] The MSE and MUE values are computed for molecules with known experimental values indicated with vs exptl.

set were considered in the parameter optimization. The reference data were relative "adiabatic" DFT energies. The strategy here is to have the semiempirical model reproduce adiabatic potential energy surfaces from which quantum mechanical zero-point corrections, tunneling effects, and thermodynamic corrections can be accounted for explicitly in QM/MM simulations.[101−104] The relative reaction energies are listed in Tables 8−13.

It should be noted that, in the final parameter refinement procedure, all structures (including transition states) in the training set were explicitly optimized along all unconstrained degrees of freedom for every trial semiempirical parameter set. For structures that were minima, all degrees of freedom were fully optimized; transition state structures were fully optimized (transition state search using gradient norm minimization) to a first-order saddle point (with the exception of the dianionic dissociative reactions that, for stability, had the cleaved P−O bond constrained to the DFT values), and for PESs all degrees of freedom excluding the PES variables were fully optimized (i.e., the PESs were adiabatic).

3.2.2. Construction of χ^2 Merit Function. The optimized AM1/d-PhoT parameters are determined by nonlinear optimization of the χ^2 merit function that measures the quality of properties predicted by the trial parameter set, λ, against the target data summarized above. The $\chi^2(\lambda)$ merit function is defined as

$$\chi^2(\lambda) = \sum_{i}^{\text{mol}} \sum_{\alpha}^{\text{prop}(i)} w_{i\alpha}(Y_{i\alpha}^{\text{AM1/d}}(\lambda) - Y_{i\alpha}^{\text{Ref}})^2 \qquad (4)$$

where the first summation with index i runs over molecules, complexes, or reactions (mol), and the second summation with index α runs over properties associated with the ith "molecule" (prop(i)). The term $Y_{i\alpha}^{\text{AM1/d}}(\lambda)$ is the value of the property α for molecule i calculated with a set of trial parameters, λ, $Y_{i\alpha}^{\text{Ref}}$ is the corresponding target value from experiment or from density-functional calculations, and $w_{i\alpha}$ is the associated weight in the regression. The weights $w_{i\alpha}$ are defined as the inverse square of the $\sigma_{i\alpha}$ values in eq 5

$$w_{i\alpha} = (\sigma_{i\alpha}{}^2)^{-1} \qquad (5)$$

in which the $\sigma_{i\alpha}$ values have the same units as the molecular property $Y_{i\alpha}$, and specifies the significance or importance of this particular property.

3.2.3. Nonlinear Parameter Optimizations. In the optimization of the χ^2 merit function, a suite of integrated

Table 4. Gas-Phase Dipole Moments (debye) of Selected Molecules and Hydrogen-Bonded Complexes

molecules[a]	DFT[b]	AM1/d	AM1	PM3	MNDO/d
		isolated molecules			
H_2O	1.91	2.39	1.86	1.74	1.78
CH_3OH	1.69	2.33	1.62	1.49	1.48
C_2H_5OH	1.61	2.25	1.55	1.45	1.40
C_6H_5OH	1.29	1.90	1.23	1.14	1.17
CH_3CO_2H	1.82	2.69	1.89	1.84	1.68
$P(CH_3)_3$	1.24	1.98	1.52	1.08	1.20
$(CH_3)_3PO$	4.52	5.76	5.01	3.92	3.57
P(O)(O)(OH)	3.21	3.35	2.37	1.80	3.05
P(O)(OH)(OH)(OH)	0.51	0.13	3.74	3.16	1.58
P(O)(OH)(OH)(OCH_3)	0.98	1.18	1.06	3.15	1.86
P(O)(OH)(OCH_3)(OCH_3)	1.24	1.37	1.10	2.18	1.97
P(O)(OCH_3)(OCH_3)(OCH_3)	1.09	0.47	0.84	0.43	2.23
P(O)(OH)(OCH_2CH_2O)	4.10	5.37	4.02	2.30	3.52
P(O)(OCH_3)(OCH_2CH_2O)	3.88	5.07	3.75	2.30	3.70
MSE		0.51	0.18	−0.08	0.08
MUE		0.65	0.42	0.91	0.47
		hydrogen bond complexes			
H_2O:H_2O[c]	2.89	3.05	2.53	2.48	2.71
CH_3OH:H_2O[c]	2.87	2.74	2.23	2.23	2.90
H_2O:P(O)(OH)(OH)(OCH_3)[c]	1.90	1.87	1.70	0.99	1.84
H_2O:P(O)(OH)(OCH_3)(OCH_3)[c]	3.32	2.66	1.10	2.18	1.97
H_2O:P(O)(OCH_3)(OCH_3)(OCH_3)[c]	3.57	2.98	3.54	3.33	4.75
CH_3OH:P(O)(OH)(OH)(OCH_3)[c]	1.93	1.82	1.60	1.22	1.90
CH_3OH:P(O)(OH)(OCH_3)(OCH_3)[c]	3.43	2.87	1.92	2.98	2.94
CH_3OH:P(O)(OCH_3)(OCH_3)(OCH_3)[c]	3.78	3.67	1.93	3.39	2.63
H_2O:P(O)(OH)(OCH_2CH_2O)[c]	2.85	4.32	4.42	2.34	2.17
MSE		−0.06	−0.62	−0.60	−0.30
MUE		0.42	0.97	0.60	0.57

[a] Bold molecules/complexes are included in parametrization as a training set molecules/complexes. [b] Reference DFT dipole moments are computed from B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) calculations. [c] Dipole moments of hydrogen-bonded complex.

nonlinear optimization methods were used[90] that included (1) a genetic algorithm and (2) a direction set of minimization methods. Genetic algorithms[105,106] have been demonstrated elsewhere to be useful in semiempirical parameter optimization.[81,107−109] The implementation of the genetic algorithm was based on the description by Goldberg[105] and used tournament selection and multidimensional phenotypic (parameter set) niching. The quadratically convergent direction set optimization utilized the method of Powell and has been described in detail elsewhere.[110] In the present work, we follow a stepwise approach of (1) initial coarse-grained parameter optimization, followed by (2) final parameter refinement and testing. The final optimized parameters are listed in Table 1. All properties were computed using a modified version of the MNDO97 program.[111]

3.2.3.1. Initial Coarse-Grained Parameter Optimization. On the basis of the original AM1 parameters, the hydrogen and oxygen atomic parameters were adjusted first. The optimization was carried out using a quadratically convergent direction set optimization method[110] with narrow parameter bands to avoid large changes in atomic parameters from their starting AM1 values (approximately within 10%). In the present specific reaction parameter optimization, although we restrict the computed absolute heats of formation to be close to the experimental values, it is important to obtain a good estimate of the absolute proton affinities and to balance the quality of relative energies for acid−base reactions to model general acid and base catalysis. Since the H and O atoms are directly involved in the protonation reaction in the PA calculations, we decided not to change the AM1 parameters for carbon. It was also assumed that the Gaussian core−core parameters on H and O atoms were sufficiently well balanced in the original AM1 model. Thus, the scaling parameters G_{scale}^{H} and G_{scale}^{O} were held fixed at 1.0.

After obtaining an initial optimized set of parameters for hydrogen and oxygen, we turned into the optimization of the phosphorus atomic parameters, which includes a set of functions for d orbitals. In view of the fact that the MNDO/d model was successful in transphosphorylation reactions in the absence of H-bonding and proton transfer, the initial phosphorus parameters were taken from the MNDO/d model except two one-center two-electron integral parameters (G_{sp} and H_{sp}) and the Gaussian core−core parameters, which were taken from AM1 for the consistency with the other atoms. The starting G_{scale}^{P} value was set to 1.0. Starting from this set of initial parameters, we performed several steps of direction set optimization, followed by optimization using a genetic algorithm to optimize the Gaussian core−core parameters and the G_{scale} value of P. The details of the implementation of genetic algorithms for nonlinear parameter optimizations will be described elsewhere.[112] In the present

Table 5. Comparison of Gas-Phase Geometries for the Reactions in the Training and Test Sets[a]

	AM1/d		AM1		PM3	
	MSE	MUE	MSE	MUE	MSE	MUE
			reactants and products			
bond (P−O)	0.000	0.009	−0.020	0.026	0.042	0.051
bond ($P-O_{Nuc/LG}$)[b]	−0.166	0.179	0.224	0.255	0.083	0.153
bond ($H-O_{Ac}$)[c]	0.098	0.114	0.410	0.447	0.039	0.163
angle (O−P−O)	−0.11	1.44	−0.18	2.34	−0.16	1.81
angle ($O_{Nuc}-P-O_{LG}$)[b]	0.70	13.77	−5.85	7.47	−8.10	9.63
angle ($O_{Do}-H-O_{Ac}$)[c]	−21.83	21.83	−22.35	24.24	−0.60	7.28
			phosphorane intermediates			
bond ($P-O_{ax}$)[d]	−0.026	0.026	−0.086	0.086	−0.002	0.022
bond ($P-O_{eq}$)[d]	−0.002	0.014	−0.029	0.029	0.038	0.045
angle ($O_{ax}-P-O_{ax}$)[d]	−7.01	7.75	−5.60	6.32	−19.76	19.76
angle ($O_{ax}-P-O_{eq}$)[d]	0.55	2.51	0.43	2.31	1.25	7.32
angle ($O_{eq}-P-O_{eq}$)[d]	−0.01	5.28	0.00	5.10	−0.33	7.68
			transition states			
bond ($P-O_{Nuc/LG}$)[b]	0.052	0.121	0.190	0.445	0.158	0.397
bond ($P-O_{ax}$)[d]	−0.006	0.017	0.039	0.056	0.040	0.050
bond ($P-O_{eq}$)[d]	−0.010	0.012	0.102	0.102	0.031	0.043
bond ($H-O_{Ac/Do}$)[e]	0.089	0.105	0.042	0.139	−0.003	0.110
angle ($O_{ax}-P-O_{Nuc/LG}$)[b,d]	−2.14	3.82	−4.62	6.79	−9.52	10.44
angle ($O_{eq}-P-O_{Nuc/LG}$)[b,d]	−0.82	3.25	−1.23	6.31	−0.06	7.19
angle ($O_{eq}-P-O_{eq}$)[d]	−0.53	2.05	−1.37	3.69	−0.64	2.79
angle ($O_{Do}-H-O_{Ac}$)[e]	−16.39	16.39	−9.11	15.88	−9.23	16.63

[a] All error values are computed as $Prop_{calcd} - Prop_{DFT}$, where reference geometries are from B3LYP/6-31++G(d,p) optimized geometries. the unit for the bond length is angstroms, and for the angle, it is degrees. [b] Nuc and LG are the oxygen atom in the nucleophilic and leaving group, respectively, and Nuc/LG stands for either nucleophile or leaving group. [c] Ac and Do are the hydrogen bond acceptor and donor group, respectively. [d] ax and eq are the oxygen atoms at the axial and equitorial position of phosphorane geometry, respectively. [e] Do and Ac are the donor and acceptor at the transition state for the proton transfer.

Table 6. Intermolecular Interaction Energies (kcal/mol) for Hydrogen-Bonded Complexes

complex[a]	DFT[b]	AM1/d	AM1	PM3	MNDO/d	ref
H_2O:H_2O	−4.8	−4.2	−5.4[c]	−3.5	−0.7	−5.4[d]
						−5.0[e]
H_2O:CH_3OH	−4.6	−3.8	−5.1	−3.2	−0.6	
H_2O:$CH_3CO_2^-$	−15.7	−16.5	−13.8	−13.5	−6.7	
H_2O:P(O)(O)$(O)^-$	−14.2	−16.4	−13.1	−9.1	−4.5	
H_2O:P(O)(O)(OH)$(OH)^-$	−16.0	−19.1	−14.1	−12.3	−6.2	
H_2O:P(O)(O)(O)$(OH)^{2-}$	−30.4	−31.2	−20.6	−24.7	−9.7	
H_2O:P(O)(O)$(OCH_2CH_2O)^-$	−15.3	−17.2	−13.2	−10.9	−4.9	
MSE[f]		−1.1	2.1	3.4	9.7	
MUE[f]		1.5	2.6	3.4	9.7	

[a] Bold complexes are included in parametrization as a training set hydrogen bond complexes. [b] B3LYP/6-311++G(3df,2p) evaluated adiabatic hydrogen bond energy. [c] AM1 water dimerization energy is for the bifurcated geometry. In the case of nonplanar C_s structure, the interaction energy is −3.3 kcal/mol. [d] Experimental hydrogen bond energy of water dimer.[138] [e] Benchmark ab initio hydrogen bond energy of water dimer.[139] [f] Errors are computed as $E_{calcd} - E_{DFT}$, where E is hydrogen bond energy.

work, genetic algorithm runs were performed using population sets of 128 members that were allowed to evolve over 100 generations. The parameter sets with the highest fitness were then refined with the quadratically convergent direction set method for 100 iterations. During the optimizations, parameters were typically allowed to vary by only around 5−10% of their initial values. This combined genetic algorithm−direction set optimization procedure was repeated, sometimes adjusting the parameter value bounds, until satisfactory results were obtained. It should be mentioned that in the process of nonlinear optimizations, multiple minima in the parameter space were identified and systematically eliminated on the basis of further testing and evaluation using the extended (testing) dataset.

3.2.3.2. Final Parameter Refinement. Up to this point, the training set only included stationary and transition state geometries. After the coarse-grained parameter optimization passed, additional geometries taken from the relaxed potential energy surfaces were included in the training set to map the potential energy surface along the reaction pathway explicitly. Because of the extreme sensitivity of the reaction energy profile and the existence of phosphorane intermediates on the protonation and alkyl substitution level, the parametrizations only with stationary point geometries do not guarantee the correct curvature and shape of potential energy surface for the reactions, except in the most simple reactions.[81] Therefore, the inclusion of geometries from potential energy surfaces for the reactions is critical to improve the accuracy of AM1/d-PhoT model with correct energy curvature near the transition states. All H, O, and P atomic parameters were refined simultaneously under narrow parameter bounds (typically 5% or less) using the direction set optimization scheme to arrive at the final optimized parameter set.

4. Results and Discussion

This section presents an analysis and discussion of the results from the semiempirical AM1/d-PhoT parametrizations for biological transphosphorylation described above. The optimized AM1/d-PhoT parameters for the H, O, and P atoms are shown in Table 1, along with the original AM1 parameters for comparison. The results are compared with experimental data, high-level DFT calculations from the QCRNA database,[82] and results from other semiempirical models including AM1, PM3, and MNDO/d. In some cases, convergence problems in transition state optimizations were encountered with the MNDO/d model, particularly for reaction steps that involved proton transfers, which disallowed comparison with the present results. In all tables and figures, the AM1/d-PhoT model is simply referred to as AM1/d.

4.1. Heats of Formation. The main focus of this paper is to develop a semiempirical model that is highly accurate in describing phosphoryl transfer reactions, including the structures and relative energies of the transition state and reactive intermediates. Nonetheless, to create a more robust model, we included a selected set of experimental heat of formation data. Table 2 shows the heats of formation (experimental, AM1/d-PhoT, AM1, PM3, and MNDO/d) obtained for the training set compounds. The overall performance of the AM1/d-PhoT model in computing heats of formation of molecules is similar with other semiempirical models. The mean signed errors (MSE) in the computed heat of formation for the AM1, PM3, MNDO/d, and AM1/d-PhoT models are 2.9, 5.9, −1.8, and 3.3 kcal/mol, respectively, whereas the corresponding values in the mean unsigned error (MUE) are 5.8, 8.1, 8.1, and 4.5 kcal/mol. In particular, the AM1, PM3, and MNDO/d have been known to overestimate the heats of formation of the hydroxide ion by 15−27 kcal/mol, but the AM1/d-PhoT model reduces the error to 6.4 kcal/mol. This offers rather significant improvement but still underscores the need to consider functional forms that allow a more balanced treatment of the stability and response properties of small anions.[78]

4.2. Gas-Phase Proton Affinity. The protonation states, pK_a values, and pH-dependent rate profiles are major issues in the identification of the underlying catalytic mechanisms of biological transphosphorylation both in enzymatic and in nonenzymatic systems. Hence, the inclusion of the absolute and relative proton affinities of molecules into the semiempirical parametrization is important in the study of biological phosphoryl transfer.[100] The calculated proton affinities (PAs) are compared with experimental and CBS-QB3[113,114] calculated values in Table 3, and Figure 1 shows the distribution of computed PA values of neutral and monoanionic species against reference CBS-QB3 calculated PAs. Previous studies have shown that the CBS-QB3 calculations have high chemical accuracy in computing absolute proton affinity.[100] The computed results from the standard AM1, PM3, and MNDO/d models, as well as DFT calculations, are also shown for comparison. The AM1/d-PhoT results exhibit considerable improvement over other semiempirical models. For the molecules with available experimental PA values, the performance of the AM1/d-PhoT, PM3, and DFT is similar in term of MSE values, which are 0.7, 0.9, and −1.6 kcal/mol, respectively, whereas the AM1 and MNDO/d models have MSE values greater than 5 kcal/mol. However, the comparisons of MUE values show that the AM1/d-PhoT model (3.4 kcal/mol) considerably outperforms the other semiempirical models (6.7 kcal/mol or greater). For molecules with phosphorus, which lack experimental PA values, the calculated proton affinities are compared against the CBS-QB3 results. The MSE and MUE values from AM1/d-PhoT calculations are only slightly worse (1.0 and 2.3 kcal/mol, respectively) than the DFT results (−0.7 and 0.9 kcal/mol, respectively), whereas other semiempirical methods shows much larger errors (over 10 kcal/mol in magnitude). When the potential inaccuracy in the evaluation of the energies of highly negatively charged molecules from semiempirical models employing minimal basis is considered, the error analysis with exclusion of the PO_4^{3-}/HPO_4^{2-} pair, which shows a maximum error in PA, further decreases the MSE and MUE values to 0.3 and 1.6 kcal/mol, respectively. The AM1/d-PhoT model is even more encouraging in the comparison of the relative PA values. For example, the proton transfer from acetic acid to the phenolate ion, which is found in several enzymatic reactions including protein tyrosine phosphatase,[79] is exothermic from the experiment, CBS-QB3, DFT, and AM1/d-PhoT calculations, while the AM1, PM3, and MNDO/d models yield endothermic reaction energies.

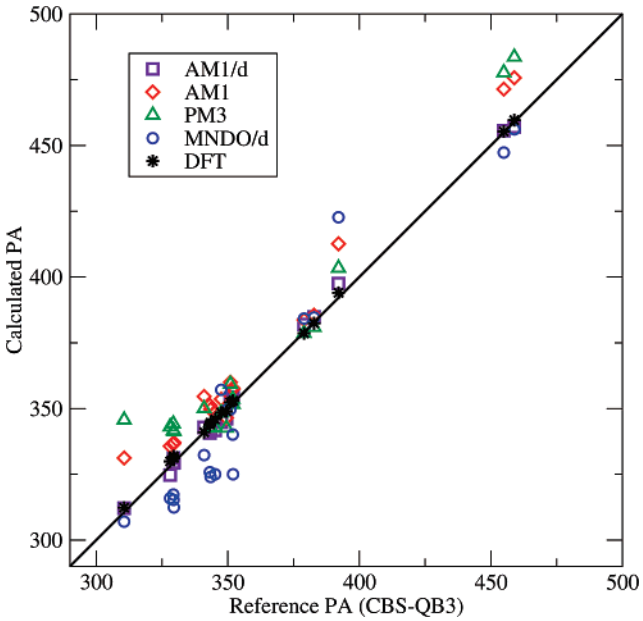


Figure 1. Gas-phase proton affinities of neutral and monoanionic species. The reference proton affinities are from CBS-QB3 calculations. Methods used in the calculations are indicated by various labels in the figure, and units are kilocalories per mole.

Table 3 also includes phosphorane compounds and the PA results at the axial and equitorial protonation sites. The phosphorane compounds are formed as intermediates in the associative reaction paths. Overall, the PAs are larger in phosphoranes than in similar phosphate compounds. The relative trends of absolute PAs between phosphorane and phosphate pairs are correctly reproduced from the AM1/

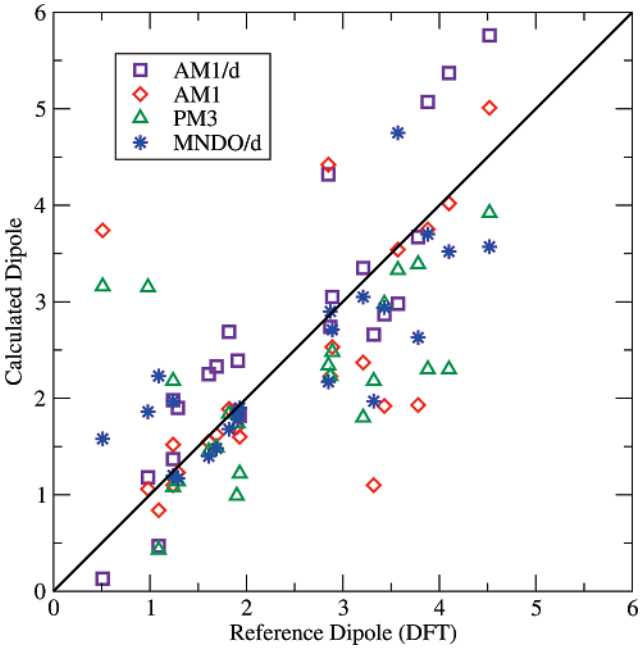


Figure 2. Scatter plot of gas-phase dipole moment. The reference dipole moments are from calculations at the B3LYP/6-311++G(3df,2p) level. The unit of dipole moments is debye.

d-PhoT model, whereas the other semiempirical methods show larger variations. The AM1 and PM3 models are reasonably reliable for the relative PAs between similar phosphorane/phosphate pairs. However, for proton transfers between dissimilar functional groups, as in enzymes, ribozymes, and solution reactions, the predictive capability of the methods is considerably reduced. It is paramount to have a quantum model that is able to predict correct absolute PA values to ensure transferability between diverse sets of proton donor−acceptor pairs. In addition, the AM1 and PM3 models show larger error with increasing negative charge on the phosphate compounds, but the AM1/d-PhoT model has smaller, more balanced errors.

4.3 Dipole Moments. The calculated dipole moments for 14 molecules and 9 hydrogen bond complexes are shown in Table 4, and Figure 2 shows the scatter plot of predicted dipole moments against reference dipole moments from the DFT calculations. In the parametrizations, the DFT dipole moments were used in the reference dataset. The error analysis shows that the MSE and MUE values for the dipole moments of isolated molecules from the AM1/d-PhoT model are 0.51 and 0.65 debye, respectively. The results are comparable to values from other semiempirical methods. The MSE and MUE values are 0.18 and 0.42 debye from AM1, −0.08 and 0.91 debye from PM3, and 0.08 and 0.47 debye from MNDO/d, respectively. The MNDO/d method is best compared to DFT dataset, while the AM1/d-PhoT model slightly overestimates the dipole moments of isolated molecules. It is sometimes the case that the deviation of the dipole moments with respect to the DFT dataset values arises from the differences in the optimized geometries at the semiempirical level. It is known that the semiempirical methods systematically underestimate the molecular polarizability because of the limitations inherent in the use of a minimal valence basis set.[78] Thus, in the absence of a quantum model that predicts accurate polarization response, it might be desirable to have the dipole moment slightly overestimated in the gas-phase such that the semiempirical models reproduce correct interaction energy in the condensed phase. The comparisons of hydrogen bond complexes support this. The computed dipole moments from the AM1/d-PhoT calculations reproduce the DFT dataset results, while other semiempirical models underestimate the interaction energy by 0.30−0.62 debye. The results suggest that the AM1/d-PhoT model can be a good method for interactions in aqueous solution. Alternately, next-generation models that allow for more accurate charge-dependent response properties[78] would be a preferable alternative. It should be noted here that proper optimizations of the QM/MM van der Waals parameters are also critical to reproduce correct relative solvation free energies from condensed-phase simulations using combined QM/MM potentials.[115−117] This aspect will be addressed in more detail in future work in the application of the AM1/d-PhoT model to study biological transphosphorylation reactions.

4.4. Geometries. In Table 5, the analysis on the structures of phosphorus compounds are presented for the reactions in training and test sets. (For details of the reactions, refer to section 4.6 and Tables 8−11.) The AM1/d-PhoT, AM1, and PM3 optimized geometries are compared with DFT optimized geometries. Since the MNDO/d model fails at the optimization of transition states for most reactions, they are not compared in this section. The geometries are compared separately at the reactant and product states, phosphorane intermediates, and transition states.

For phosphate geometries of reactants and products, the AM1/d-PhoT model performs better than the AM1 and PM3 models, but the geometries are well reproduced from all three methods tested (Table 5). The situation for the hydrogen bond geometries is somewhat different among various models at reactant and product complexes. In particular, the angle of $O_{Nuc}-P-O_{LG}$, in which the O_{Nuc} and O_{LG} are the nucleophilic and leaving oxygens in the reaction, has large deviation from the DFT values for all three models tested (Table 5). The hydrogen bond lengths in the complexes are reasonably close to the DFT geometries for the AM1/d-PhoT and PM3 models, but the errors in bond angle are large. For hydrogen bond angles, the PM3 model performs better than the AM1/d-PhoT and AM1 models. For phosphoranes, the three methods reproduce DFT geometry accurately. The AM1 model yields the smallest MSE and MUE, which is followed by the AM1/d-PhoT model, whereas the PM3 model underestimates the bond angle by as much as 20° in some cases (Table 5).

For the transition state, the most important geometrical parameters are the bond lengths of the P atom from the oxygen atoms of the nucleophile or the leaving group for nucleophilic substitution and addition−elimination reactions, and the distance from the transferring proton to donor and acceptor atoms ($H-O_{Ac/Do}$). The AM1/d-PhoT model has the smallest MSE and MUE values of 0.052 and 0.121 for the $P-O_{Nuc}$ and $P-O_{LG}$ distances and 0.089 and 0.105 for $H-O_{Ac}$ and $H-O_{Do}$ lengths, respectively. Although the PM3 model predicts quite accurate geometries for proton-transfer

Table 7. Comparison of Geometries for Hydrogen-Bonded Complexes Given in Table 6[a]

	AM1/d		AM1		PM3		MNDO/d	
	H···O[b]	O−H−O[c]	H···O[b]	O−H−O[c]	H···O[b]	O−H−O[c]	H···O[b]	O−H−O[c]
MSE	0.010	−9.40	0.838	−31.97	0.450	−12.93	1.586	−19.27
MUE	0.073	9.62	0.838	37.83	0.751	16.19	1.586	20.14

[a] All error values are computed as $Prop_{calc} - Prop_{DFT}$, where reference geometries are from B3LYP/6-31++G(d,p) optimized geometries. The unit for the bond length is angstroms, and that for the angle is degrees. [b] Hydrogen to O_{Ac} distance. [c] The angle formed by O_{Do}, hydrogen, and O_{Ac} atoms.

reactions, the errors in $P-O_{Nuc/LG}$ bond are too large. For example, the maximum error of PM3 for $P-O_{Nuc/LG}$ bond length is 1.255 Å, while the maximum error for the $P-O_{Nuc/LG}$ bond from AM1/d-PhoT model is 0.347 Å.

4.5. Hydrogen-Bonding Energy and Geometry. For bimolecular complexes, the performance of the AM1/d-PhoT model is compared with other semiempirical models. Although the semiempirical interaction energies are considered to be the enthalpies of association,[51,53,54,68] we compare the results with adiabatic interaction energies without including temperature and zero-point vibrational energies. Our purpose in the development of the present specific reaction parameters is to use a semiempirical model to accurately reproduce the potential energy surfaces for transphosphorylation reactions in condensed phase simulations. The results are presented in Tables 6 and 7. The AM1/d-PhoT model shows an improvement in the hydrogen bond geometries and interaction energies compared to the standard AM1, PM3, and MNDO/d methods. The MUE for the intermolecular interaction energy with the AM1/d-PhoT model is 1.5 kcal/mol, whereas the AM1, PM3 and MNDO/d models have corresponding MUE values of 2.6, 3.4, and 9.7 kcal/mol, respectively. Importantly, the AM1/d-PhoT model overcomes the well-documented problem in the AM1 dimer structure for water, which has bifurcated geometry. The interaction energy for a water dimer is −4.2 kcal/mol from the AM1/d-PhoT model, and the hydrogen bond distance between hydrogen bond donor and acceptor is 2.97 Å, which is comparable to the 2.89 Å value from DFT dataset. Table 6 presents six other hydrogen bond complexes including four phosphorus-containing complexes. In particular, the hydrogen bond energy of water with an ion increases with the increase of charge of ions, and the AM1/d-PhoT model performs better than other semiempirical models, in which existing semiempirical models considerably underestimate the interaction energy between H_2O and $P(O)(O)(O)(OH)^{2-}$.

The results from an error analysis of the hydrogen bond geometries are presented in Table 7. The AM1/d-PhoT model results in more accurate hydrogen bond geometries. The MSE and MUE values from the AM1/d-PhoT model are 0.010 and 0.073 Å for the hydrogen bond distance and −9.40 and 9.62° for the angle formed by hydrogen bond donor, H atom, and acceptor. The AM1 and MNDO/d models predict quite different hydrogen bond geometries from the reference geometries. However, compared with Table 5, the PM3 model performs quite reliably, although it is problematic in balancing the geometries and the interaction energies. One example is the hydrogen bond between $P(O)(O)(O)(OCH_2-CH_2O)^-$ and the water molecule. The DFT and AM1/d-PhoT model predict a doubly hydrogen-bonded complex, while the PM3 model predicts a singly hydrogen-bonded complex and causes a large error in angle.

4.6 Reaction Energetics. The reaction energies from the AM1/d-PhoT model for a series of reactions with total charge ranging from 0 to −2 are listed in Tables 8−11. The reaction dataset includes reaction energies, barrier heights, and relative energies of intermediates for over 19 associative reactions and 3 dissociative reactions. Note that although these reactions cover a fairly broad range of mechanisms in the gas phase, they do not consider more complex reaction mechanisms that may occur in the aqueous phase, such as water-assisted pathways[37,118,119] or complex water relays.[120] These mechanisms often require fairly sophisticated transition path sampling techniques[121,122] and sometimes intricate bridging water chains that are currently not amenable to efficient automated parameter optimization. The computed values from the AM1 and PM3 models are also compared with the calculated DFT values. Although the MNDO/d model was successful in the transphosphorylation reaction in the absence of hydrogen bond and proton transfer, this model fails in optimizing the geometries at the transition state involving proton transfers, which makes it difficult to compare the performance with other semiempirical methods and is not included in the comparisons. The results from error analysis are summarized in Tables 12 and 13, and Figure 3 shows a scatter plot of computed energies against the DFT reference values. The reactions are classified on the basis of the total charge and mechanism. Thus, associative (neutral, monoanionic, and dianionic) and dissociative (monoanionic) mechanisms are considered (Scheme 1). It is assumed that all associative reactions proceed via in-line reaction pathways, whenever the reaction involves a proton transfer from a nucleophile or to a leaving group. Figure 4 shows the reaction energy profiles of representative reactions for each reaction type. In this figure, the calculations using the MNDO/d model are carried out from relaxed potential energy surface scans constrained at different values of the $P-O_{Nuc/LG}$ and $H-O_{Ac/Do}$ lengths to obtain an approximate transition state, while the direct transition state optimizations are carried out for other semiempirical models. The dianionic dissociative reactions are not listed because of the failure of semiempirical methods in locating the products after the formation of the P−O bond.

The AM1/d-PhoT model performs noticeably better than the AM1 and PM3 models in the associative reactions (see Table 12). The PM3 method severely underestimates the barrier heights. In contrast, the AM1 model performs well for certain reactions and poorly for others. The computed MSE values for neutral and monoanionic reactions are 3.28 and 0.24 kcal/mol for the AM1 model and −18.76 and

Table 8. In-line Associative Neutral Reactions and Energies (kcal/mol) Used in Parametrizations and Their Tests[a]

reaction[b]	type	DFT	error ($E_{calcd} - E_{DFT}$) AM1/d	AM1	PM3	label
H_2O + P(O)(OH)(OH)(OCH_3) ⇌ CH_3OH + P(O)(OH)(OH)(OH)	R	11.39	4.54	−2.24	−9.19	$\mathbf{A}_1^0$
	$TS_{2'}$	35.01	−1.30	−1.16	−14.35	
	I	16.64	−1.59	−49.56	−31.57	
	$TS_{5'}$	36.05	−4.21	1.30	−14.79	
	PS	0.87	−1.34	−7.34	−7.30	
	P	12.94	3.83	−13.51	−18.66	
H_2O + P(O)(OH)(OCH_3)(OCH_3) ⇌ CH_3OH + P(O)(OH)(OH)(OCH_3)	R	5.65	5.87	5.05	−1.31	$\mathbf{A}_2^0$
	$TS_{2'}$	34.69	2.73	4.83	−17.28	
	I	15.19	−0.96	−43.17	−27.01	
	$TS_{5'}$	37.59	4.25	5.39	−24.17	
	PS	0.60	−3.32	−4.94	−7.88	
	P	6.92	5.22	−2.96	−10.71	
H_2O + P(O)(OCH_3)(OCH_3)(OCH_3) ⇌ CH_3OH + P(O)(OH)(OCH_3)(OCH_3)	R	6.40	3.91	−15.86	−1.44	$\mathbf{A}_3^0$
	$TS_{2'}$	40.70	−0.98	2.67	−17.93	
	I	20.36	−3.51	−45.09	−30.54	
	$TS_{5'}$	40.81	−5.41	3.75	−22.67	
	PS	1.32	−1.92	−6.04	−9.65	
	P	7.43	3.31	−23.78	−12.06	
CH_3OH + P(O)(OH)(OH)(OCH_3) ⇌	R	11.74	5.18	−3.40	−11.13	$\mathbf{A}_4^{0,c}$
	$TS_{2'}$	35.26	−2.07	9.24	−20.38	
	I	19.84	−2.17	−42.20	−26.02	
CH_3OH + P(O)(OH)(OCH_3)(OCH_3) ⇌	R	5.99	4.36	1.44	−2.24	$\mathbf{A}_5^{0,c}$
	$TS_{2'}$	35.45	−0.72	12.72	−6.47	
	I	17.66	−2.41	−38.02	−20.80	
CH_3OH + P(O)(OCH_3)(OCH_3)(OCH_3) ⇌	R	6.32	3.98	−0.21	−2.08	$\mathbf{A}_6^{0,c}$
	$TS_{2'}$	40.61	−3.55	10.54	−8.23	
	I	22.61	−3.05	−39.54	−21.81	
P(O)(OH)(OH)(OCH_2CH_2OH) ⇌ H_2O + P(O)(OH)(OCH_2CH_2O)	$TS_{2'}$	29.23	5.68	9.76	−19.66	$\mathbf{A}_7^0$
	I	10.76	2.59	−40.49	−26.04	
	$TS_{5'}$	35.99	9.17	5.44	−26.72	
	PS	−2.59	4.85	5.05	2.74	
	P	9.48	6.81	0.19	−6.91	
P(O)(OCH_3)(OCH_3)(OCH_2CH_2OH) ⇌ CH_3OH + P(O)(OCH_3)(OCH_2CH_2O)	$TS_{2'}$	31.19	−0.34	−4.10	−28.27	$\mathbf{A}_8^0$
	I	15.33	2.60	−40.21	−29.08	
	$TS_{5'}$	38.88	6.58	−15.10	−22.92	
	P	10.95	5.51	−9.57	−17.52	

[a] All the energies are relative to the reactant hydrogen-bonded complex. The reference DFT energies are computed as adiabatic energies from B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level. [b] Bold reactions are included in the parametrization as a training set reaction. [c] Symmetric reactions.

−13.01 kcal/mol from the PM3 model, respectively. In this regard, the MSE values are 0.76 and −1.12 kcal/mol from the AM1/d-PhoT model for both reaction types. The same conclusion is drawn from comparison of the MUE values. The relative energies of intermediates with respect to the reactant complexes are fairly satisfactory from the AM1/d-PhoT calculations, but they are severely underestimated using AM1 and PM3 models. These two methods produce a huge stabilization in the energy of phosphorane compounds, resulting from the use of a minimal valence basis and the unbalanced use of Gaussian core−core interactions for phosphate and phosphorane. The artificial stabilizations of the pentacoordinate phosphorus compounds are transferred into the energy at the transition state. This causes systematic underestimation of barrier heights for most reactions using the PM3 model and several reactions from the AM1 model (Tables 8, 9, and 12). This is further exacerbated in the relative energies between hydrogen-bonded reactant complexes and reaction intermediates. On the other hands, the AM1/d-PhoT model accurately reproduces the energetics from DFT calculations within several kilocalories per mole in the reaction energies, barrier heights, and relative energies of reaction intermediates. This shows the importance of the d orbitals in the handling of hypervalent phosphorus compounds and the effectiveness of scaled Gaussian core−core interactions.

For dianionic associative reactions, similar artificial stabilizations of hypervalent phosphorus compounds from the AM1 and PM3 calculations are observed (Table 10). The barrier heights from the AM1 and PM3 models are still underestimated considerably. Again, the AM1/d-PhoT model performs well in the prediction of reaction energies and barrier heights.

It is interesting to note that there are large deviations of reaction energies for the associative reactions computed from the semiempirical AM1 and PM3 models and that many

Table 9. In-line Associative Monoanionic Reactions and Energies (kcal/mol) in Parametrizations and Their Test[a]

reaction[b]	type	DFT	error ($E_{calcd} - E_{DFT}$) AM1/d	AM1	PM3	label
H_2O + P(O)(O)(OH)(OCH_3)$^-$ ⇌ CH_3OH + P(O)(O)(OH)(OH)$^-$	R	16.25	2.55	−1.85	−4.70	$\mathbf{A}_1^-$
	$TS_{2'}$	44.59	1.99	−13.36	−5.42	
	I	37.23	−7.55	−48.16	−41.91	
	$TS_{5'}$	41.83	−1.43	−10.07	−24.25	
	PS	1.31	−0.45	−1.31	−6.89	
	P	15.97	1.13	−9.73	−10.98	
H_2O + P(O)(O)(OCH_3)(OCH_3)$^-$ ⇌ CH_3OH + P(O)(O)(OH)(OCH_3)$^-$	R	17.31	1.00	−10.60	−10.93	$\mathbf{A}_2^-$
	$TS_{2'}$	39.20	1.62	1.61	−7.32	
	I	36.69	−5.02	−54.38	−44.78	
	$TS_{5'}$	40.94	−6.68	−4.38	−46.03	
	PS	−0.06	−1.88	−9.72	−11.72	
	P	15.68	0.58	−14.82	−16.03	
CH_3OH + P(O)(O)(OH)(OCH_3)$^-$ ⇌	R	15.76	1.79	−5.18	−4.65	$\mathbf{A}_3^-$[d]
	$TS_{2'}$	43.08	−0.42	−30.73	−14.30	
	I	36.76	−6.89	−43.63	−35.20	
CH_3OH + P(O)(O)(OCH_3)(OCH_3)$^-$ ⇌	R	14.56	2.55	−9.26	−2.76	$\mathbf{A}_4^-$[d]
	$TS_{2'}$	41.40	−4.51	−9.91	−18.12	
	I	39.62	−4.91	−20.24	−32.00	
CH_3O^- + P(O)(OCH_3)(OCH_3)(OCH_3) ⇌	R	20.04	4.89	−2.38	1.40	$\mathbf{A}_5^-$[c,d]
	$TS_{2'}$	10.30	−5.95	28.82	12.23	
	I	1.69	−5.61	−45.48	−18.25	
P(O)(OH)(OCH_3)(OCH_2CH_2O)$^-$ ⇌ HO^- + P(O)(OCH_3)(OCH_2CH_2O)	R	5.43	7.38	4.53	−2.19	$\mathbf{A}_6^-$[c]
	$TS_{2'}$	2.81	2.16	3.74	−5.33	
	I	−2.27	−2.19	−38.56	−30.01	
	$TS_{5'}$	23.84	2.59	12.83	6.17	
	P	37.32	12.97	19.15	5.53	
P(O)(O)(OCH_3)(OCH_2CH_2OH)$^-$ ⇌ CH_3OH + P(O)(O)(OCH_2CH_2O)$^-$	$TS_{2'}$	32.21	−1.03	17.54	−14.18	$\mathbf{A}_7^-$
	I	32.16	−1.71	−43.76	−37.36	
	$TS_{5'}$	40.80	−0.71	6.59	−26.57	
	PS	3.77	−2.34	−6.11	−21.32	
	P	18.43	−0.40	−12.43	−16.08	

[a] All the energies are relative to the reactant hydrogen-bonded complex. The reference DFT energies are computed as adiabatic energies from the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level. [b] Bold reactions are included in the parametrization as a training set reaction. [c] No proton transfer is involved in the reaction. [d] Symmetric reactions.

Table 10. In-line Associative Dianionic Reactions and Energies (kcal/mol) in Parametrizations and Their Test[a]

reaction[b]	type	DFT	error ($E_{calcd} - E_{DFT}$) AM1/d	AM1	PM3	label
H_2O + P(O)(O)(O)(OCH_3)$^{2-}$ ⇌ CH_3OH + P(O)(O)(O)(OH)$^{2-}$	R	29.42	0.89	−8.85	−5.37	$\mathbf{A}_1^{2-}$[c,e]
	$TS_{5'}$	50.43	−3.47	−48.56	−36.74	
	PS	1.99	−0.68	−5.85	−3.41	
	P	32.84	−2.82	−16.78	−10.61	
CH_3OH + P(O)(O)(O)(OCH_3)$^{2-}$ ⇌	R	29.98	−1.18	−10.14	−6.44	$\mathbf{A}_2^{2-}$[c,d,e]
	$TS_{5'}$	50.56	−5.95	−33.90	−33.73	
CH_3O^- + P(O)(O)(OCH_3)(OCH_3)$^-$ ⇌	$TS_{5'}$	94.38	−4.66	−10.85	−24.12	$\mathbf{A}_3^{2-}$[d]
P(O)(O)(OCH_3)(OCH_2CH_2O)$^{2-}$ ⇌ CH_3O^- + P(O)(O)(OCH_2CH_2O)$^-$	$TS_{5'}$	43.07	−0.43	−7.14	−38.92	$\mathbf{A}_4^{2-}$[e]
	P	−44.10	0.84	−12.93	−21.53	

[a] The reference energy for the $\mathbf{A}_1^{2-}$ and $\mathbf{A}_2^{2-}$ reactions and the $\mathbf{A}_3^{2-}$ and $\mathbf{A}_4^{2-}$ reactions is the energy of the reactant hydrogen-bonded complex and the separated reactant state, respectively. The reference DFT energies are computed as adiabatic energies from the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level. [b] Bold reactions are included in the parametrizations as a training set reaction. [c] Proton transfer is involved in the reaction. [d] Symmetric reactions. [e] Several points on the 1-D/2-D potential energy surfaces are included in the parametrizations to ensure the curvature of the potential energy surface near the transition states.

reactions in the reaction set consist of a water and methanol pair as nucleophile and leaving group. The errors of these two molecules in heat of formation coincide with the errors in the computed reaction energies from the AM1 and PM3 models. Even though the errors are smaller than other semiempirical methods tested, there exist similar errors from the AM1/d-PhoT model. Other defects of the AM1 and PM3 models are also evident. For example, there is a systematic increase of error in the calculated heats of formations as the charge on phosphorus containing compounds increase (Table

Table 11. Dissociative Monoanionic Reactions and Energies (kcal/mol) in Parametrizations and Their Test[a]

reaction[b]	type	DFT	error ($E_{calcd} - E_{DFT}$) AM1/d	AM1	PM3	label
H_2O + P(O)(O)(O)$^-$ ⇌ P(O)(O)(OH)(OH)$^-$	R	14.16	2.20	−1.03	−5.01	**D_1^-**,[c]
	TS	24.15	10.93	5.69	−16.50	
	P	10.01	7.78	−29.59	−20.60	
CH_3OH + P(O)(O)(O)$^-$ ⇌ P(O)(O)(OH)(OCH_3)$^-$	R	13.17	0.74	−3.88	−2.19	**D_2^-**
	TS	22.52	6.92	12.85	−2.50	
	P	−10.71	7.73	−24.57	−11.49	
C_6H_5OH + P(O)(O)(O)$^-$ ⇌ P(O)(O)(OH)(OC_6H_5)$^-$	R	20.09	−2.29	−6.92	−4.17	**D_3^-**,[c]
	TS	23.80	−7.17	−7.57	−23.52	
	P	−6.32	0.88	−27.39	−16.33	

[a] All the energies are relative to the reactant hydrogen-bonded complex. The reference DFT energies are computed as adiabatic energies from the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level. [b] Bold reactions are included in the parametrizations as a training set reaction.

Table 12. Summary of Gas-Phase Reaction Energetics of Associative Reactions in the Training and Test Sets[a]

	neutral rxn			monoanionic rxn			dianionic rxn		
	AM1/d	AM1	PM3	AM1/d	AM1	PM3	AM1/d	AM1	PM3
				reaction energy[b]					
no. rxn		5			4			2	
MSE	2.07	−7.32	−10.78	0.84	−2.48	−4.94	−1.44	−9.00	−2.96
MUE	2.86	7.39	10.78	1.96	9.79	8.80	2.28	9.00	5.65
				activation energy					
no. TS		13			11			4	
MSE	0.76	3.28	−18.76	−1.12	0.24	−13.01	−3.63	−25.11	−33.38
MUE	3.61	6.62	18.76	2.64	12.69	16.36	3.63	25.11	33.38
				relative energy of intermediates					
No. int		8			7				
MSE	−1.06	−42.29	−26.61	−4.84	−42.03	−34.22			
MUE	2.36	42.29	26.61	4.84	42.03	34.22			

[a] All error values are computed as $E_{calcd} - E_{DFT}$, where the DFT values are B3LYP/6-311++G(3df,2p) single-point energies at a given B3LYP/6-31++G(d,p) optimized geometries. All energies are given in kilocalories per mole. [b] Symmetric reactions are ignored in computing reaction energies.

Table 13. Summary of Gas-Phase Reaction Energies (kcal/mol) of Three Dissociative Reactions in Trial Set and in the Training Set[a]

error of reaction	reaction energy			activation energy		
	AM1/d	AM1	PM3	AM1/d	AM1	PM3
MSE	5.25	−23.24	−12.35	3.56	3.66	−14.17
MUE	5.25	23.24	12.35	8.34	8.70	14.17

[a] All error values are computed as $E_{calcd} - E_{DFT}$, where DFT values are B3LYP/6-311++G(3df,2p) single-point energies at a given B3LYP/6-31++G(d,p) optimized geometries.

3). The errors in predicted PA values increase with the increase of charge on phosphorus compounds. This explains the relatively large error in the last reaction presented in Table 10.

The performance of the AM1/d-PhoT model on dissociative reactions is less impressive than for the associative reactions, which might be related to the slight unbalance of the heats of formation of metaphosphate and phosphate compounds. All the semiempirical methods, however, perform considerably poorly. Nonetheless, the AM1/d-PhoT model is more accurate than AM1 and PM3 by a factor of 2−3. The MSE values are 5.25 and 3.56 kcal/mol from the AM1/d-PhoT, −23.24 and 3.66 kcal/mol from the AM1, and −12.35 and −14.17 kcal/mol from the PM3 calculations for reaction energy and reaction barrier, respectively. If the dissociation reaction is the reverse reaction presented in Table 11, the reaction barriers from AM1/d-PhoT model are comparable to DFT dataset.

4.7. Free-Energy Simulations in Water. Although the AM1/d-PhoT model has been developed and tested for gas-phase processes, our goal is to study transphosphorylation reactions in aqueous solution and in enzymes or ribozymes. Thus, it is necessary to demonstrate an application of the AM1/d-PhoT model in condensed phase QM/MM simulations. To this end, molecular dynamics simulations employing umbrella sampling[123] have been carried out to compute potentials of mean force (PMF) for a model transphosphorylation reaction in water.

The transphosphorylation reaction model tested is the A_7^{-1} reaction presented in Table 9. The solute molecule, which has a total charge of −1 e, is embedded in a 40 × 40 × 40 Å^3 cubic box consisting of 2039 TIP3P water molecules.[124] In addition, we include one Na^+ ion to neutralize the charge of the system. The solutes were represented quantum-mechanically by the AM1/d-PhoT model, and the rest of the system constitutes the classical region. The geometry of water was held rigid by using the

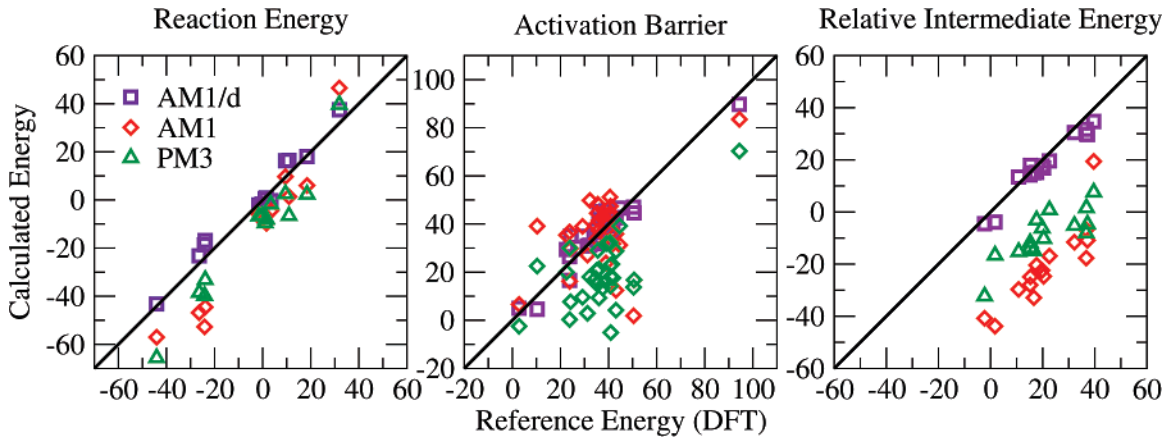


Figure 3. Scatter plots of gas-phase reaction energy (left), activation barrier (middle), and relative energy of intermediate (right). The reference reaction energies are from calculations at B3LYP/6-311++G(3df,2p) level, and symmetric reactions are not presented for reaction energy. The unit of energy is kilocalories per mole.

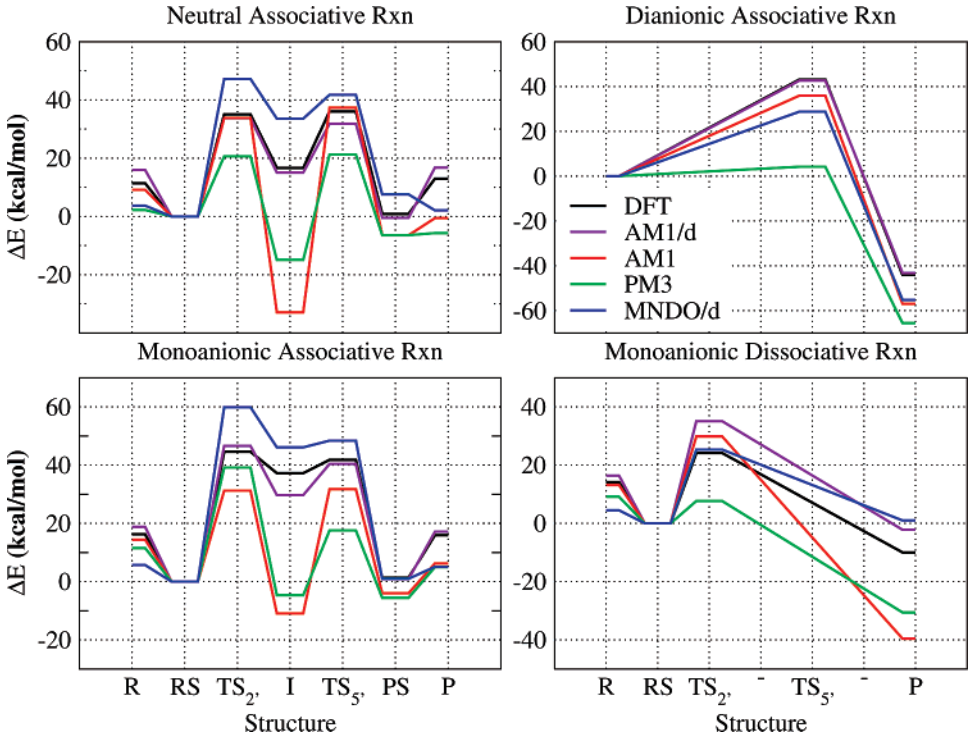


Figure 4. Comparison of performance of semiempirical methods on representative reactions. The plots are for the neutral associative reaction of A_1^0, the monoanionic associative reaction of A_1^-, the dianionic associative reaction of A_4^{2-}, and the dissociative reaction of D_1^-. The labels are shown in Table 8−11. The values for the MNDO/d model are obtained from relaxed potential energy surface as approximate energies, while the direct transition state optimizations are carried out for other semiempirical and DFT results.

SHAKE algorithm in all simulations.[125] To handle long-range electrostatic interactions, the QM/MM−Ewald scheme[65] was used along with the particle mesh Ewald (PME) method[126,127] for the electrostatic interactions between pure MM atom pairs and a 10.0 Å cutoff with potential shift in computing van der Waals interactions (refer ref 65 for further details on the simulations).

The reactions are divided into two steps: intramolecular nucleophilic attack and exocyclic cleavage. Each step involves a proton transfer, the first from the nucleophile to the nonbridging oxygen and the second from the nonbridging oxygen to the leaving group. The reaction is assumed to follow an in-line reaction pathway. Thus, each step requires a 2-dimensional umbrella sampling run to cover the entire free-energy surface. The computed results are shown in Figure 5. The contours are computed from 338 separate umbrella sampling windows, each consisting of over 30 ps averaging with at least 5 ps equilibration. Thus, the free-energy contour was constructed from a total of 12 ns of MD simulation. The computed barrier for the first step of the reaction ($TS_{2'}$) is about 32.0 kcal/mol and for the second step ($TS_{5'}$) is approximately 37.0 kcal/mol from combined QM/MM simulations (Figure 5). A phosphorane intermediate (I) has a PMF value of 29.0 kcal/mol, with barriers to collapse to reactants and products of 3.0 and 8.0 kcal/mol, respectively. The lifetime of such intermediates are im-

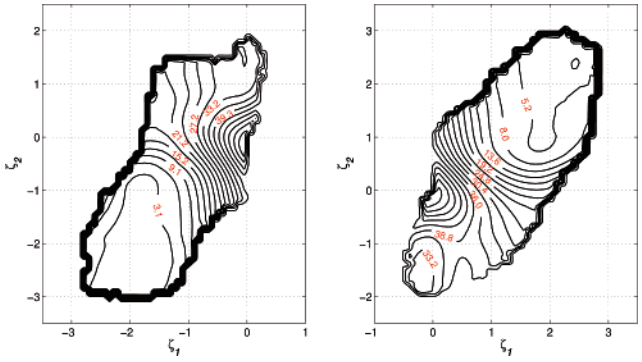


Figure 5. Computed potential of mean force (PMF) on the transphosphorylation reaction of A_7^- shown in Table 9. The reactions have two separate steps: intramolecular nucleophilic attack leading to formation of an endocyclic bond (left side) and exocyclic cleavage of the phosphorane intermediate leading to departure of the methanol leaving group (right side). The ζ_1 is defined as a distance difference in $R(O_B-P)$ and $R(P-O_A)$, where O_A and O_B are oxygens on the nucleophile and leaving group, respectively, presented in Scheme 2. The ζ_2 is the proton-transfer reaction coordinate also defined as distance differences between $R(O_A-H)$ and $R(H-O_R)$ (left plot) and between $R(O_R-H)$ and $R(H-O_B)$ (right plot), where H is the transferred proton and O_R is one of nonbridging oxygen on cyclic phosphate in Scheme 2. The unit of energy is kcal/mol, and that of distance is Å.

portant in modeling phosphoryl transfer reactions that might undergo alternate reactions after pseudorotation of the phosphorane (e.g., RNA migration).[25,47] The computed results are in reasonably good accord with the results from the B3LYP/6-311++G(3df,2p) calculations with implicit solvation modeled with the polarizable continuum model (PCM)[128,129] and a conductor-like screening model (COSMO).[130] These calculations yielded the barriers with respect to the unimolecular reactant for $TS_{2'}$ and $TS_{5'}$ of 28.2 and 38.0 kcal/mol, respectively, with PCM and 26.9 and 37.8 kcal/mol, respectively, with COSMO.[24] The DFT results identify a stable intermediate with barrier to collapse to reactants and products of 2.0 and 11.8 kcal/mol, respectively, with PCM and 2.1 and 13.0 kcal/mol, respectively, with COSMO. These values are in reasonable agreement with the QM/MM results, the main difference being that the $TS_{2'}$ transition state and I intermediate in the QM/MM calculations are shifted relative to the DFT values by approximately 3−5 kcal/mol.

5. Conclusion

A specific reaction parameter AM1/d Hamiltonian, AM1/d-PhoT, has been developed to model transphosphorylation reactions. The model is parametrized for H, O, and P atoms to reproduce high-level density-functional results from a recently constructed database of quantum calculations for RNA catalysis, including geometries and relative energies of minima, transition states and reactive intermediates, dipole moments, proton affinities, and other relevant properties. The model has been tested in the gas phase and in solution using a QM/MM potential, and demonstrated to provide overall very good accuracy with respect to the DFT results for a reduction of over 3−4 orders of magnitude in computational cost. The model offers a significant improvement over the MNDO/d, AM1, and PM3 methods for the transphosphorylation reactions. The ultimate goal of this work is to make strides toward the development of highly accurate semiempirical quantum models for biological reactions that can be used in conjunction with linear-scaling electronic structure and combined QM/MM methods to address complex problems of biocatalysis that simultaneously span large spatial domains and long time scales. The current work makes considerable progress in the development of fast, accurate quantum models for phosphoryl transfer reactions in solution and catalyzed by enzymes and ribozymes. Future work will involve critical comparison of AM1/d-PhoT with new SCC-DFTB models to assess the advantages and disadvantages of each method and facilitate the design of more robust quantum models for QM/MM calculations. It is the hope that through continued development and testing of semiempirical and SCC-DFTB models over a broad range of biological applications the next-generation of improved quantum methods for multiscale modeling of biocatalytic processes may emerge.

Acknowledgment. This work was partially supported by grants from the National Institutes of Health (GM62248 and GM46736), from the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement DAAD19-01-2-0014, and from the Office of Naval Research (ONR) under Grant N00012-05-01-0538. Computational resources were provided by the Minnesota Supercomputing Institute and the Army High Performance Computing Research Center.

References

(1) Westheimer, F. H. *Science* **1987**, *235*, 1173−1178.

(2) Boyer, P. D.; Cross, R. L.; Momsen, W. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 2837−2839.

(3) Admiraal, S.; Herschlag, D. *Chem. Biol.* **1995**, *2*, 729−739.

(4) Ahn, N. *Chem. Rev.* **2001**, *101*, 2207−2208.

(5) Scott, W. G. *Curr. Opin. Struct. Biol.* **1998**, *8*, 720−726.

(6) Takagi, Y.; Ikeda, Y.; Taira, K. *Top. Curr. Chem.* **2004**, *232*, 213−251.

(7) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065−2069.

(8) Scott, W. G. *Q. Rev. Biophys.* **1999**, *32*, 241−294.

(9) Walter, N. G.; Burke, J. M. *Curr. Opin. Chem. Biol.* **1998**, *2*, 24−30.

(10) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferré-D'Amaré, A. R. *Science* **2002**, *298*, 1421−1424.

(11) Shih, I.-h.; Been, M. D. *Biochemistry* **2000**, *39*, 9055−9066.

(12) Shih, I.-h.; Been, M. D. *EMBO J.* **2001**, *20*, 4884−4891.

(13) Shih, I.-h.; Been, M. D. *Annu. Rev. Biochem.* **2002**, *71*, 887−917.

(14) Perreault, D. M.; Anslyn, E. V. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 432−450.

(15) Zhou, D.-M.; Taira, K. *Chem. Rev.* **1998**, *98*, 991−1026.

(16) Oivanen, M.; Kuusela, S.; Lönnberg, H. *Chem. Rev.* **1998**, *98*, 961−990.

(17) Hengge, A. C. *Acc. Chem. Res.* **2002**, *35*, 105−112.

(18) Hengge, A. C.; Cleland, W. W. *J. Am. Chem. Soc.* **1990**, *112*, 7421−7422.

(19) Florián, J.; Åqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 11524−11525.

(20) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719−734.

(21) Åqvist, J.; Kolmodin, K.; Florian, J.; Warshel, A. *Chem. Biol.* **1999**, *6*, R71−R80.

(22) Liu, Y.; Gregersen, B. A.; Hengge, A.; York, D. M. *Biochemistry* **2006**, *45*, 10043−10053.

(23) Liu, Y.; Lopez, X.; York, D. M. *Chem. Commun.* **2005**, *31*, 3909−3911.

(24) Liu, Y.; Gregersen, B. A.; Lopez, X.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 19987−20003.

(25) López, C. S.; Faza, O. N.; de Lera, A. R.; York, D. M. *Chem.—Eur. J.* **2005**, *11*, 2081−2093.

(26) Lim, C.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 5872−5873.

(27) Uchimaru, T.; Tanabe, K.; Nishikawa, S.; Taira, K. *J. Am. Chem. Soc.* **1991**, *113*, 4351−4353.

(28) Lim, C.; Tole, P. *J. Phys. Chem.* **1992**, *96*, 5217−5219.

(29) Mercero, J. M.; Barrett, P.; Lam, C. W.; Fowler, J. E.; Ugalde, J. M.; Pedersen, L. G. *J. Comput. Chem.* **2000**, *21*, 43−51.

(30) Arantes, G. M.; Chaimovich, B. *J. Phys. Chem. A* **2005**, *109*, 5625−5635.

(31) Dejaegere, A.; Lim, C.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 4353−4355.

(32) Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **1993**, *115*, 5316−5317.

(33) Tole, P.; Lim, C. *J. Phys. Chem.* **1993**, *97*, 6212−6219.

(34) Tole, P.; Lim, C. *J. Am. Chem. Soc.* **1994**, *116*, 3922−3931.

(35) Chang, N.; Lim, C. *J. Phys. Chem. A* **1997**, *101*, 8706−8713.

(36) Florián, J.; Warshel, A. *J. Am. Chem. Soc.* **1997**, *119*, 5473−5474.

(37) Hu, C.-H.; Brinck, T. *J. Phys. Chem. A* **1999**, *103*, 5379−5386.

(38) Lopez, X.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2001**, *123*, 11755−11763.

(39) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 5010−5018.

(40) Lopez, X.; Dejaegere, A.; Leclerc, F.; York, D. M.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 11525−11539.

(41) Chen, X.; Zhan, C.-G. *J. Phys. Chem. A* **2004**, *108*, 6407−6413.

(42) Xu, D.; Guo, H.; Liu, Y.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 13827−13834.

(43) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11−27.

(44) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.

(45) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2003**, *125*, 7178−7179.

(46) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 7504−7513.

(47) López, C. S.; Faza, O. N.; Gregersen, B. A.; Lopez, X.; de Lera, A. R.; York, D. M. *Chem. Phys. Chem.* **2004**, *5*, 1045−1049.

(48) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.

(49) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *J. Mol. Biol.* **2003**, *327*, 549−560.

(50) Garcia-Viloca, M.; Poulsen, T. D.; Truhlar, D. G.; Gao, J. *Protein Sci.* **2004**, *13*, 2341−2354.

(51) Dewar, M. J.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.

(52) Dewar, M. J.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907−4917.

(53) Dewar, M. J. S.; Zoebisch, E.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(54) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.

(55) Khandogin, J.; Hu, A.; York, D. M. *J. Comput. Chem.* **2000**, *21*, 1562−1571.

(56) Khandogin, J.; York, D. M. *J. Phys. Chem. B* **2002**, *106*, 7693−7703.

(57) Khandogin, J.; Musier-Forsyth, K.; York, D. M. *J. Mol. Biol.* **2003**, *330*, 993−1004.

(58) Khandogin, J.; York, D. M. *Proteins* **2004**, *56*, 724−737.

(59) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(60) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(61) Gao, J. *Rev. Comput. Chem.* **1995**, *7*, 119−185.

(62) Gao, J. *Curr. Opin. Struc. Biol.* **2003**, *13*, 184−192.

(63) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391−404.

(64) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616−626.

(65) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2−13.

(66) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775−789.

(67) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495−506.

(68) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.

(69) Tubert-Brohman, I.; Guimaraes, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2003**, *25*, 138−150.

(70) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2005**, *1*, 817−823.

(71) Bernal-Uruchurtu, M.; Ruiz-López, M. *Chem. Phys. Lett.* **2000**, *330*, 118−124.

(72) Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C.; C. Millot, M. F. R.-L. *J. Comput. Chem.* **2000**, *21*, 572−581.

(73) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260−7268.

(74) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. Status Solidi B* **2000**, *217*, 357−376.

(75) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569−585.

(76) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458−6469.

(77) Sattelmeyer, K. W.; Tubert-Brohman, I.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2006**, *2*, 413−419.

(78) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2005**, *123*, 164108.

(79) Alhambra, C.; Wu, L.; Zhang, Z.-Y.; Gao, J. *J. Am. Chem. Soc.* **1998**, *120*, 3858−3866.

(80) Arantes, G. M.; Loos, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 347−353.

(81) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231−236.

(82) Giese, T. J.; Gregersen, B. A.; Liu, Y.; Nam, K.; Mayaan, E.; Moser, A.; Range, K.; Nieto Faza, O.; Silva Lopez, C.; de Lera, A. R.; Schaftenaar, G.; Lopez, X.; Lee, T.; Karypis, G.; York, D. M. *J. Mol. Graphics Modell.* **2006**, *25*, 423−433.

(83) QCRNA, http://theory.chem.umn.edu/Database/QCRNA.

(84) Stewart, J. J. P. *Rev. Comput. Chem.* **1990**, *1*, 45−81.

(85) Thiel, W. In *New Methods in Computational Quantum Mechanics*; Advances in Chemical Physics 93; Prigogine, I., Rice, S. A., Eds.; John Wiley and Sons: New York, 1996; pp 703−757.

(86) Winget, P.; Selçuki, C.; Horn, A.; Martin, B.; Clark, T. *Theor. Chem. Acc.* **2003**, *110*, 254−266.

(87) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons: Chichester, U.K., 2004.

(88) Giese, T. J.; Sherer, E. C.; Cramer, C. J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 1275−1285.

(89) Csonka, G. I. *J. Comput. Chem.* **1993**, *14*, 895−898.

(90) Lopez, X.; York, D. M. *Theor. Chem. Acc.* **2003**, *109*, 149−159.

(91) Becke, A. D. *Phys. Rev. A.* **1988**, *38*, 3098−3100.

(92) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(93) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785−789.

(94) Bauernschmitt, R.; Ahlrichs, R. *J. Chem. Phys.* **1996**, *104*, 9047−9052.

(95) Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1977**, *66*, 3045−3050.

(96) Frisch, Æ.; Frisch, M. J. *Gaussian 98 User's Reference*, 2nd ed.; Gaussian, Inc.: Pittsburgh, PA, 1999.

(97) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(98) Range, K.; McGrath, M. J.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 1654−1665.

(99) Mayaan, E.; Range, K.; York, D. M. *J. Biol. Inorg. Chem.* **2004**, *9*, 807−817.

(100) Range, K.; Riccardi, D.; Cui, Q.; Elstner, M.; York, D. M. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3070−3079.

(101) Alhambra, C.; Carochado, J.; Sánchez, M. L.; Garcia-Viloca, M.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. B* **2001**, *105*, 11326−11340.

(102) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2001**, *114*, 9953−9958.

(103) Truhlar, D. G.; Gao, J.; Garcia-Viloca, M.; Alhambra, C.; Corchado, J.; Sanchez, M. L.; Poulsen, T. D. *Int. J. Quantum Chem.* **2004**, *100*, 1136−1152.

(104) Major, D.; York, D. M.; Gao, J. *J. Am. Chem. Soc.* **2005**, *127*, 16374−16375.

(105) Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.

(106) Coley, D. A. *An Introduction to Genetic Algorithms for Scientists and Engineers*; World Scientific: River Edge, NJ, 1999.

(107) Cundari, T. R.; Deng, J.; Fu, W. *Int. J. Quantum Chem.* **2000**, *77*, 421−432.

(108) Voityuk, A. A.; Rösch, N. *J. Phys. Chem. A* **2000**, *104*, 4089−4094.

(109) Hutter, M. C.; Reimers, J. R.; Hush, N. S. *J. Phys. Chem. B* **1998**, *102*, 8080−8090.

(110) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, W. P. *Numerical Recipes in Fortran*, 2nd ed.; Cambridge University Press, Cambridge, U.K., 1992.

(111) Thiel, W. *MNDO97*, version 5.0; University of Zurich: Zurich, Switzerland, 1998.

(112) Giese, T. J.; York, D. M. In preparation.

(113) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822−2827.

(114) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532−6542.

(115) Gao, J.; Xia, X. *Science* **1992**, *258*, 631−635.

(116) Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386−395.

(117) Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467−6478.

(118) Gerratana, B.; Sowa, G. A.; Cleland, W. W. *J. Am. Chem. Soc.* **2000**, *122*, 12615−12621.

(119) Zhan, C.-G.; Landry, D. W. *J. Phys. Chem. A* **2001**, *105*, 1296−1301.

(120) König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. A* **2006**, *110*, 548−563.

(121) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.

(122) Li, G.; Cui, Q. *J. Mol. Graphics Modell.* **2005**, *24*, 82−93.

(123) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.

(124) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(125) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(126) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(127) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Hsing, L.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(128) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(129) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43−54.

(130) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.

(131) Hunter, E.; Lias, S. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413−457.

(132) Bartmess, J. E.; McIver, R. T. *Gas Phase Ion Chemistry*; Academic Press: New York, 1979; Vol. 2.

(133) Cox, J. O.; Pilcher, G. *Thermochemistry of Organic and Organometallic Compounds*; Academic Press: New York, 1970.

(134) Pilcher, G. In *The Chemistry of Organophosphorus Compounds*; Hartley, F. R., Ed.; Wiley: New York, 1990; Vol. 1, Chapter 5, pp 127−136.

(135) Pilcher, G.; Skinner, H. A. In *The Chemistry of the Metal−Carbon Bond*; Hartley, F. R., Patai, S., Eds.; Wiley: New York, 1982; Vol. 1, Chapter 2, pp 43−90.

(136) Linstrom, P., Mallard, W., Eds. *NIST Chemistry WebBook*; NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 2003; 20899; http://webbook.nist.gov (March 2003).

(137) Lide, D. R., Ed. *CRC Handbook of Chemistry and Physics*, 83rd ed.; CRC Press LLC: Boca Raton, FL, 2003.

(138) Dyke, T. R.; Mack, K. M.; Muenter, J. S. *J. Chem. Phys.* **1977**, *66*, 498−510.

(139) Feyereisen, M. W.; Feller, D.; Dixon, D. A. *J. Phys. Chem.* **1996**, *100*, 2993−2997.

CT6002466

JCTC Journal of Chemical Theory and Computation

Reaction Mechanism of 1,3,5-Trinitro-*s*-triazine (RDX) Deciphered by Density Functional Theory

Matthew J. Swadley and Tonglei Li*

Pharmaceutical Sciences, University of Kentucky, Lexington, Kentucky 40536

Received June 15, 2006

Abstract: 1,3,5-Trinitro-*s*-triazine, or cyclotrimethylene trinitramine, or RDX, is a sensitive, secondary explosive, which has been the subject of a number of studies regarding the sensitivity and mechanism of decomposition in energetic materials. Several initial mechanistic steps have been proposed for RDX decomposition, with no conclusive agreement upon any one as the definitive pathway. Our research utilizes density functional theory (DFT)-based calculations and concepts, particularly the nuclear Fukui function, to analyze the effects of additive/depletive electronic perturbation upon vapor conformers and crystal RDX structures. Since the nuclear Fukui function is a measure of the physical stress that a nucleus encounters upon a change in the electron population, it may provide useful information regarding the role of each atom in unimolecular decomposition. The results illustrate that both homolytic cleavage of N−N bonds and elimination of HONO from RDX exhibit favorability as initial steps in the decomposition of RDX in either phase. The nuclear Fukui function proved a valuable tool for gaining insight into the initial steps of unimolecular reactions.

Introduction

1,3,5-Trinitro-*s*-triazine (RDX) is a highly symmetric, energetic compound often used as a secondary explosive.[1] Its unimolecular decomposition has recently been the subject of many mechanistic and sensitivity studies, as it has served as a model system for energetic molecules.[2] Despite the growing body of work regarding the compound, however, there has been no conclusive agreement to what the initial mechanistic step is for its decomposition.[3]

Past experimental evidence of RDX decomposition has been interpreted into three distinct initial steps, namely concerted fission, N−N homolysis, and HONO elimination (Figure 1). Concerted fission proceeds through the simultaneous breakage of three C−N bonds, separating the RDX molecule into three molecules of CH_2NNO_2.[4,5] N−N homolysis is self-explanatory, proceeding through the even breakage of a N−N bond into nitrogen dioxide radical and remaining RDX amine radical products.[4,6−9] HONO elimination proceeds by simultaneous breakage of N−N and neighboring C−H bonds, along with the formation of an O−H bond, resulting in the formed HONO.[4,6,7,9] One particular research group has performed a variety of experiments regarding the thermal breakdown of crystal RDX and found that the initial stages of the decomposition may actually occur in the gas and liquid states.[6,7,9] Moreover, decomposition mechanisms in the gas and liquid phase were found to differ from that in the crystal. Depending on if the temperature is above or below RDX's melting point, either melted or vaporized molecules are decomposed first, with their byproducts initiating the breakdown of the crystal proper. These reactions appear to occur through multiple pathways, including those initiated by both HONO elimination and N−N homolysis. Many experiments involving the shock or UV photolysis detonation of crystal RDX point to the initial reaction step of N−N homolysis.[8,10−13] One in particular involved the pyrolysis of a thin film of RDX with a pulsed CO_2 laser, yielding N_2O_4, the product of two NO_2 radicals formed from RDX N−N homolysis.[8] Another study dealt with the shock-induced decomposition of crystal RDX and found a decrease in the ratio of nitro-group nitrogen atoms to ring-bound amine nitrogen atoms through X-ray

* Corresponding author phone: (859)257-1472; fax: (859)257-7585; e-mail: tonglei@uky.edu. Corresponding author address: 514 College of Pharmacy, University of Kentucky, 725 Rose Street, Lexington, KY 40536-0082.

10.1021/ct600202e CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/06/2007

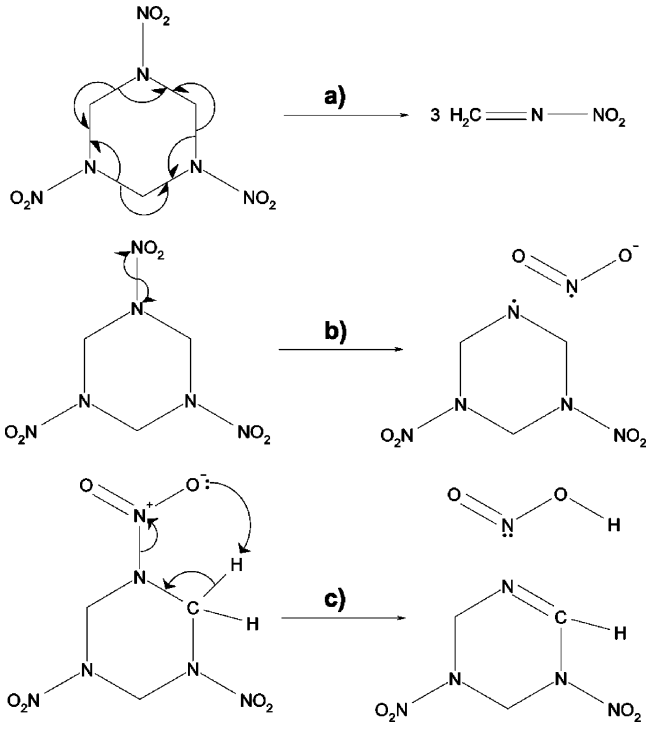


Figure 1. Summary of proposed mechanistic steps for RDX decomposition, concerted fission (a), N−N homolysis (b), and HONO elimination (c).

photoelectron spectroscopy (XPS) after the shock of a sample crystal.[12] This, combined with electron paramagnetic resonance (EPR) evidence of nitro group radical products from the reaction, indicated to the authors that N−N homolysis may indeed be the initial step in mechanically impacted RDX crystals.[12] A study by Chakraborty et al. in which the transition state energies of each decomposition pathway (concerted fission, N−N homolysis, and HONO elimination) were calculated with quantum mechanical methods yielded insight into the mechanistic feasibility of these paths in the gas phase.[4] In this study, it was found that the HONO elimination mechanism was most favorable, with N−N cleavage also having energetic favorability. The concerted fission mechanism, proposed from earlier experimental evidence,[5] was ruled out as an initial step to detonation due to its transition state existing at a much higher energy than the other pathways.

There have been many recent investigations that study the crystal structure of RDX as a means to identify trends in its reactivity. Many of these studies have focused upon the RDX lattice and its role in the ease of electronic excitation (perturbation). Electronic excitation (resulting from shock, heat, or UV energy) is often discussed in terms of narrowing and disappearance of the gap between valence and conduction bands in crystalline solids, which can initiate chemical reactions such as detonation through conformational changes and bond breakage.[14,15] Experimental evidence showing increases in the conductivity of RDX crystals just before detonation support this assertion, as closing of the band gap would allow for free electron movement throughout the compound.[16−19] RDX crystals containing defects have been shown to be more detonation shock-sensitive than perfect crystals.[20] Localized regions of focused energy from shock within a crystal known as "hot spots" have recently been studied.[21−24] These studies have theoretically calculated the narrowing of the valence-conduction band gap of the RDX crystal in the presence of several types of defects (including edge dislocations, vacancies and cracks) at various unit cell volume compressions. It was found that unit cell compression, a product of mechanical shock, and edge dislocations significantly reduce the band gap in RDX crystals.[22−24] Another study analyzed the feasibility of charge-transfer (CT) pair formation in RDX crystals and their relation to detonation sensitivity in perfect and defect-containing crystals.[1] CT pairs are a potential mechanism of electronic perturbation in crystals where an electron from one molecule is transferred to another, resulting in neighboring cationic and anionic molecules.[1] It was found that the energy released from a relaxing CT pair in a crystal containing vacancy defects is enough to break an $N-NO_2$ bond in RDX, potentially inducing detonation. Recently, studies have also been performed to evaluate the mechanical strength of the RDX lattice through its elasticity constants.[25−29] It has been proposed that stiffer lattices imply less sensitive shock detonation in energetic materials, as stiffness rankings in RDX homologues are the reverse of detonation sensitivity.[26]

In a previous study by Luty et al., the nucleophilic and electrophilic nuclear Fukui functions, chemical reactivity index, and nuclear stiffness were calculated to analyze the effects of electronic perturbation upon RDX nuclei.[2] The results yielded novel insight into how the responses of individual nuclei to perturbation can be interpreted with mechanistic implications. It was found that while the geometric structure of an RDX molecule had a C_{3v} symmetric structure, the electronic structure was reduced to C_s symmetry upon electronic perturbation, evidenced by asymmetric chemical reactivity index and nuclear stiffness magnitudes on symmetrically similar nitrogen atoms. These asymmetrical responses focused strain resulting from electronic perturbation on a particular set of bonded nitrogen atoms, which were concluded to displace from one another more easily than the others within the molecule. This focused strain was also related to the decomposition sensitivity exhibited by RDX as well as evidence for the N−N homolysis mechanism.

The subject of the previous research by Luty et al. was limited to the triaxial (AAA) conformer of RDX and did not yet account for effects of the crystal electronic structure. Vapor-phase RDX has been shown to exist as a superposition of several conformers, denoted by the nitro group positions (as either axial or equatorial) and ring conformations (as chair, boat, or twist).[4,30−32] The molecular energies of such conformers exist at comparable energies, indicating that a high degree of interconversion is possible.[4,30−32] Previous literature has also shown that the most stable conformers appear to be the chair triaxial (AAA) form and the chair diaxial form (AAE), with energy rankings differing in separate accounts.[4,30−32] Infrared spectra of experimental RDX vapor correlates well with simulated spectra for the triaxial (AAA) conformer, indicating that it may be the best representative of the structure of gaseous RDX.[32] Single-crystal neutron-diffraction, however, has shown crystal RDX has a diaxial (AAE) conformation.[33] Structural differences

between interconverting molecular conformers as well as intermolecular influences from the RDX lattice may imply that more calculations are necessary to better understand gaseous and solid-state perturbation responses in RDX.

Our research seeks to utilize conceptual density functional theory (DFT)-based concepts, most notably the nuclear Fukui function, to shed light on how electronic perturbation can reveal the initial decomposition mechanism of RDX vapor and crystals. Previous investigations from our research group have studied reactivity in crystal systems using conceptual DFT methods, with an emphasis upon reactivity differences in polymorphs of the same compound.[34,35] The nuclear Fukui function has proven to be a key tool for identifying the role of a compound's electronic structure in its reaction mechanism. In one particular study,[35] we compared nuclear Fukui function values of α- and γ-indomethacin, which were experimentally shown to have different reaction rates with ammonia gas, α-form being faster. Upon analysis of the nucleophilic nuclear Fukui function, it was found that a particular set of symmetrically similar indomethacin molecules in the α-form had significantly larger responses upon the reactive carboxylic hydrogen atom than any in the γ-form. This was concluded to be evidence that the electronic structure of the α-form allows easier dissociation of the acidic hydrogen, hence accounting for its greater reactivity. We believe that the insight gained through analysis of nuclear Fukui functions in indomethacin crystals can be applied to better understand the initial decomposition mechanism of RDX.

Methods

The nuclear Fukui function is derived from DFT, which states that electron density is the fundamental quantity for describing atomic and molecular ground states and that energy is a functional of electron density.[36−38] A full derivation of the nuclear Fukui function can be seen in our previous reports.[34,35] In brief, the nuclear Fukui function has its roots in the Hellmann-Feynmann force, denoted as the force upon nucleus α within a molecular system[39,40]

$$\mathbf{F}_\alpha = Z_\alpha \left[\int \rho(\mathbf{r}) \frac{\mathbf{r}_\alpha}{r_\alpha^3} d\mathbf{r} - \sum_{\beta \neq \alpha} Z_\beta \frac{\mathbf{R}_{\alpha\beta}}{R_{\alpha\beta}^3} \right] \quad (1)$$

where $\rho(\mathbf{r})$ is the electron density at point $\mathbf{r}$, $\mathbf{r}_\alpha$ is equal to the displacement vector between the position of nucleus α and point $\mathbf{r}$, and $\mathbf{R}_{\alpha\beta}$ is the displacement vector between nucleus α and nucleus β with their charges as Z_α and Z_β, respectively. It is evident that from this equation $\mathbf{F}_\alpha$ is dependent in part upon the molecular system's electronic structure. The concept of the nuclear Fukui function (Φ_α) is developed from the response of $\mathbf{F}_\alpha$ to a perturbation of the electronic structure, defined as[41]

$$\Phi_\alpha = \left(\frac{\partial \mathbf{F}_\alpha}{\partial N} \right)_\nu \quad (2)$$

where N is the number of electrons in the system and $\nu(\mathbf{r})$ is the external potential defined by nuclear charges, $\{Z_\alpha\}$, and nuclear position vectors, $\{\mathbf{R}_\alpha\}$. Therefore, the nuclear Fukui function characterizes the force that nucleus α experiences upon a change in the electronic number or density due to perturbation or exchange with neighboring systems. Since it is expected that large nuclear Fukui function force responses can identify atomic displacement, they can be used as a means to investigate chemical reactivity. Because N has to be integers, in practice, the smallest change in N is one. As such, the nuclear Fukui function concept can be further developed into nucleophilic and electrophilic components, numerically calculated by the finite differences[41]

$$\Phi^+{}_\alpha = \mathbf{F}^+{}_\alpha - \mathbf{F}^0{}_\alpha$$

$$\Phi^-{}_\alpha = \mathbf{F}^0{}_\alpha - \mathbf{F}^-{}_\alpha \quad (3)$$

where $\Phi^+{}_\alpha$ is the nuclear Fukui function for nucleophilic attack on nucleus α, $\Phi^-{}_\alpha$ is the nuclear Fukui function for electrophilic attack, and $\mathbf{F}^+{}_\alpha$, $\mathbf{F}^-{}_\alpha$, and $\mathbf{F}^0{}_\alpha$ are the Hellman-Feynmann forces on the anionic, cationic, and neutral compound species, respectively. It should be noted that for molecules at equilibrium (in any phase), $\mathbf{F}^0{}_\alpha$ is close to zero, so $\mathbf{F}^+{}_\alpha$ and $\mathbf{F}^-{}_\alpha$ may independently characterize nucleophilic or electrophilic attack, respectively.[42] In the case of unimolecular decomposition, electronic perturbation may occur through the momentary addition or depletion of electron density to/from the molecular structure, so both the nucleophilic and electrophilic nuclear Fukui functions may be useful as a means of assessing reactivity and identifying the initial bond-breaking site(s) in RDX. It should be noted that nuclear Fukui functions of a molecular system are calculated with nucleus positions remaining fixed, while the electronic structure is perturbed. The applicability of these functions for understanding a chemical reaction where old bonds are broken and new ones are formed follows the assumption that changes in the electron density lead to the reorganization of the nuclear geometry.[43]

Single RDX molecules of AAA, AAE, AEE, and EEE conformers were optimized using the program package Gaussian 03 (Gaussian, Inc., Wallingford, CT) at a B3LYP/6-311G(2d,p)++ level of accuracy. In addition the α-form RDX crystal (*Pbca*, $a = 13.182$, $b = 11.574$, $c = 10.709$Å, $\alpha = \beta = \gamma = 90°$, $Z = 8$)[33] was optimized with constant unit cell parameters at a B3YLP/6-21G(d,p) level of accuracy using the Crystal 03 code package.[44] Nuclear Fukui function calculations were made from the optimized structures at the same levels of accuracy. All calculations were performed on a 28-CPU Linux cluster. The magnitudes of crystal RDX nuclear Fukui function values have been normalized by a factor of 8 ($Z = 8$ in the α-RDX crystal unit cell) as a means to enhance comparability of trends within each system. This normalization should not be misconstrued to be useful in the comparison of raw values between single molecules and the crystal structure, as the calculations for each have been performed at different levels of accuracy and have been perturbed in different manners (adding/depleting one electron per molecule vs adding/depleting one electron per volume of the unit cell).

Results and Discussion

Vapor-Phase RDX. As stated earlier, it is believed that vapor-phase RDX exists as a superposition of several

Table 1. Nucleophilic (Φ^+_α) and Electrophilic (Φ^-_α) Nuclear Fukui Function Magnitudes of Vapor-Phase RDX in the Triaxial (AAA), Diaxial (AAE), Diequatorial (AEE), and Triequatorial (EEE) Conformations[a]

	$\lvert\Phi^+_\alpha\rvert$				$\lvert\Phi^-_\alpha\rvert$			
	AAA	AAE	AEE	EEE	AAA	AAE	AEE	EEE
C1	0.6827	0.5939	0.5965	0.6537	1.7682	1.1046	2.7983	2.1930
C2	0.6834	0.6715	0.6066	0.6550	0.7988	2.5481	1.0588	2.1910
C3	0.6831	0.5939	0.6066	0.6540	1.8626	1.1010	1.0608	2.1921
H1	0.1767	0.1827	0.3661	0.4176	0.1543	0.1514	0.2178	0.1242
H2	0.1721	0.1106	0.1048	0.1136	0.1571	0.1649	0.7965	0.4520
H3	0.1740	0.1854	0.1135	0.1138	0.1463	0.1415	0.1937	0.4525
H4	0.1762	0.1088	0.2304	0.4181	0.2462	0.0581	0.2659	0.1246
H5	0.1767	0.1827	0.2303	0.4173	0.1471	0.1518	0.2652	0.1242
H6	0.1723	0.1105	0.1135	0.1137	0.1551	0.1649	0.1925	0.4521
N1	0.8574	0.5112	0.6342	0.7282	2.8836	0.3836	2.1790	2.3179
N2	0.8701	1.0421	0.6340	0.7305	0.8757	2.6350	2.1843	2.3237
N3	0.8654	1.0418	1.2290	0.7341	0.9792	2.6346	1.6068	2.3165
N4	2.2475	2.4715	2.6007	2.6643	4.3974	0.4358	3.5831	4.0688
N5	2.2710	2.5791	2.6002	2.6635	1.3913	4.1012	3.5897	4.0674
N6	2.2664	2.5782	2.8767	2.6709	1.5595	4.1007	3.0082	4.0631
O1	3.0257	2.8081	2.9601	2.9905	1.1426	0.6962	1.2616	1.1550
O2	3.0265	2.8080	3.0166	2.9910	1.1449	0.6972	0.9794	1.1543
O3	3.0560	3.0195	2.9594	2.9799	0.7536	1.0048	1.2624	1.1550
O4	3.0610	3.2731	3.0157	2.9802	0.6190	1.1928	0.9812	1.1540
O5	3.0563	3.0188	3.0623	2.9866	0.7751	1.0048	0.8524	1.1535
O6	3.0606	3.2725	3.0623	2.9864	0.6425	1.1927	0.8526	1.1536

[a] Unit: nN.

conformers.[4,30−32] In order to assess the initial decomposition step in the single RDX molecule, nuclear Fukui functions of each axial−equatorial chair conformation were calculated (Table 1). Energetically, the AAE conformation was found to be most stable, with the AAA and AEE conformations existing at slightly higher energies (about 3.05 and 3.53 kJ/mol higher, respectively). The EEE conformer is less stable, with a total energy 18.87 kJ/mol higher than the AAE conformation. Vapor-phase RDX, therefore, may have a minimal probability of existing in the EEE conformer at any instant in time. These values support the idea that interconversion may exist between at least the AAA, AAE, and AEE conformers in the gas phase.

Several trends can be interpreted from the nuclear Fukui function data (vector representations shown in Figures 2 and 3) as they pertain to the initial mechanistic steps proposed in the literature. First, the data suggest that concerted fission (Figure 1a) is not a likely initial step in RDX vapor. The concerted fission mechanism proceeds through the simultaneous breakage of C−N bonds within the ring of RDX. If this mechanism were likely, the largest nuclear Fukui function values would occur at carbon and ring-bound nitrogen atoms. The nucleophilic nuclear Fukui function magnitudes ($\lvert\Phi^+_\alpha\rvert$) provide no such evidence in any conformation, as the largest values occur in nitro group nitrogen and oxygen atoms. Carbon atoms values (in the range of 0.59−0.69 nN) and ring-bound amine nitrogen atom values (in the range of 0.51−1.23 nN) in general have far smaller response values than these nitro group atoms (nitrogen atoms between 2.24 and 2.88 nN, oxygen atoms between 2.80 and 3.28 nN). This indicates that reactivity

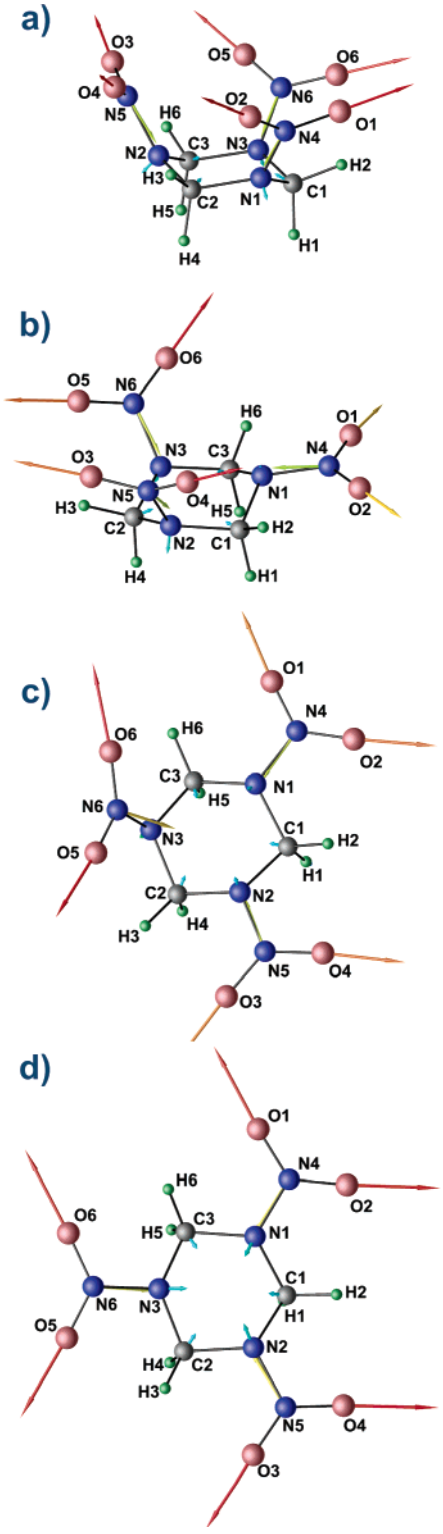


Figure 2. Vector representations of nucleophilic nuclear Fukui functions (Φ^+_α) for single RDX molecules of the AAA (a), AAE (b), AEE (c), and EEE (d) conformers. Vector lengths (small to large) and colors (blue to red) represent relative magnitudes of the nuclear Fukui functions within each particular molecule.

due to the addition of an electron to a single molecule of RDX will tend to involve the nitro groups more than ring-bound carbon and nitrogen atoms. The electrophilic nuclear Fukui function magnitudes ($\lvert\Phi^-_\alpha\rvert$) provide similar evidence, as concerted fission does not appear to be a likely pathway

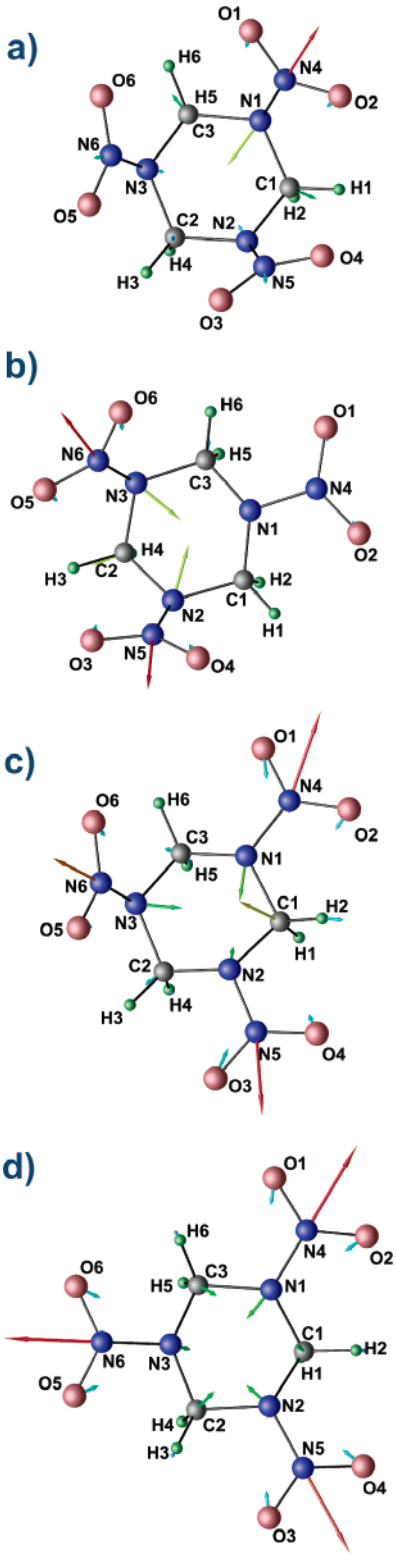


Figure 3. Vector representations of electrophilic nuclear Fukui functions (Φ^-_α) for single RDX molecules of the AAA (a), AAE (b), AEE (c), and EEE (d) conformers. Vector lengths (small to large) and colors (blue to red) represent relative magnitudes of the nuclear Fukui functions within each particular molecule.

for electron depleted systems. The electrophilic carbon and ring-bound nitrogen atom values are larger than their nucleophilic counterparts but are still outweighed in general by values within the nitro groups. In the AAA conformer, the largest four electrophilic nuclear Fukui function forces occur at N4 (4.40 nN), N1 (2.88 nN), C3 (1.86 nN), and C1 (1.77 nN). Even though there are large force responses between the bonded atoms N1−C1 and N1−C3, the most likely mechanistic step would indicate the breakage of the N1−N4 bond in this conformer. The other conformers show similar trends, as even with some carbon atoms having large electrophilic responses, the largest force responses occur at bonded nitrogen regions (N2−N5, N3−N6 in AAE, N1−N4, N2−N5 in AEE, all N−N bonds in EEE). From the Φ^+_α and Φ^-_α magnitudes, it is evident that the concerted fission mechanism is not the most likely initial step in the decomposition of RDX gas.

Both $|\Phi^+_\alpha|$ and $|\Phi^-_\alpha|$ provide evidence that N−N breakage may occur in RDX decomposition. This is evidenced by large nucleophilic responses for nitro group nuclei and strong electrophilic responses upon bonded nitrogen atoms. This breakage is necessary for both the N−N homolysis and HONO elimination mechanisms (Figure 1b,c).[4] HONO elimination requires the additional breakage of a C−H bond accompanied by formation of an O−H bond. Distinguishing these mechanisms from one another may therefore depend upon nuclear Fukui function magnitudes on carbon and hydrogen atoms as well as O−H proximity within each conformer.

In the AAA and AAE conformers, noted from the literature as the most stable conformers,[4,30−32] there is no evidence of a significantly large hydrogen atom electrophilic or nucleophilic force response, as Φ^+_α and Φ^-_α magnitudes are less than 0.25 nN. However, $|\Phi^-_\alpha|$ in the AEE conformer do provide some evidence of HONO elimination. H2 has a Φ^-_α magnitude of 0.7965 nN (about 4 times larger than in the AAA or AAE conformer and at least 3 times larger than any other hydrogen atom in the same conformer) and is bonded to C1, whose magnitude is relatively large at 2.7983 nN. While the H2 force is not as large as those upon other atoms within the molecule, its effect upon displacement of the hydrogen atom may be significant due to hydrogen's small mass. These significantly large C1 and H2 values indicate that bond breakage may occur at this site. Moreover, H2 is in close proximity of O2 and O4, members of nitro groups with significantly large nitrogen Φ^-_α values (N4 and N5). From these data, it certainly seems plausible that HONO elimination of H2−O2−N4−O1 or H2−O4−N5−O3 could occur. In the EEE conformer, HONO elimination seems a possible mechanism as well, with H2, H3 and H6 having larger Φ^-_α values (all about 0.45 nN) paired with carbon atom Φ^-_α values of about 2.2 nN. However, the feasibility of mechanisms in this conformer may be hampered by the large relative energy as compared to conformers with axial nitro groups. In the vapor phase, it seems that the AAA and AAE conformers are more likely to decompose through the N−N homolysis route, while the AEE and EEE conformers seem to exhibit evidence of HONO elimination initiated decomposition.

Symmetry also seems to play a role in the reactivity of compounds. One of the most intriguing findings of previous RDX research was the asymmetric responses of seemingly similar bonded nitrogen atoms found by Luty et al., which would focus perturbation response strain in the AAA

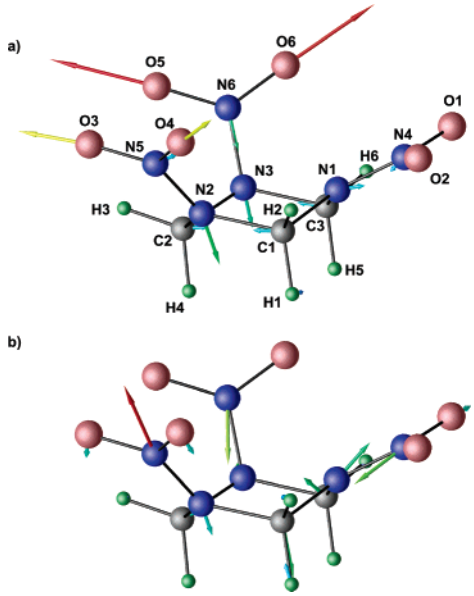


Figure 4. Vector representations of the nucleophilic (a) and electrophilic (b) nuclear Fukui functions for a molecule in crystal RDX. Vector lengths (small to large) and colors (blue to red) represent relative magnitude within the molecule.

conformer onto one bond.[2] Each RDX conformer exhibits structural symmetry, with the AAA and EEE conformers exhibiting C_{3v} symmetry and the AAE and AEE conformers exhibiting C_s symmetry. Their findings for the AAA conformer are echoed in this report, as the electrophilic nuclear Fukui function has far larger magnitudes upon the bonded N1 and N4 atoms than other nitrogen atoms within the same molecule. This indicates that the perturbed electronic structure of AAA RDX does not have the same symmetry as the geometric structure. The shape of the highest occupied molecular orbital (HOMO) of AAA RDX (Figure 5a) further emphasizes this point, as while the molecular structure has C_{3v} symmetry, the orbital is only C_s symmetric. When an electron is removed from a molecule, it is removed from the HOMO, which helps to explain the asymmetric Φ^-_α values.

The electronic symmetries of the other RDX conformers, however, seem to match the symmetries in their skeletal structures. In the EEE conformer, which exhibits structural C_{3v} symmetry, each symmetrically similar atom has similar nuclear Fukui function values (i.e., each carbon, hydrogen, and nitro group experiences similar force responses to additive or depletive electronic perturbation, Figures 2 and 3). In Figure 5g,h, the shapes of the HOMO and LUMO of the EEE conformer exhibit C_{3v} symmetry, illustrating the similar responses for symmetrically similar atoms upon electron addition/depletion. The AAE and AEE skeletal structures have C_s symmetry, with reflection planes defined by C2, N1, and N4 in AAE and C1, N3, and N6 in AEE. Nuclear Fukui function directions and magnitudes show that the symmetries of the electronic structures match those of the skeletal structures in these conformers. In the AAE conformer, similar atoms within the axial nitro groups have similar Φ^+_α and Φ^-_α values, as do C1/C3 and O1/O2. In the AEE conformer, similar atoms within the equatorial nitro groups have similar Φ^+_α and Φ^-_α values, as do C2/C3 and

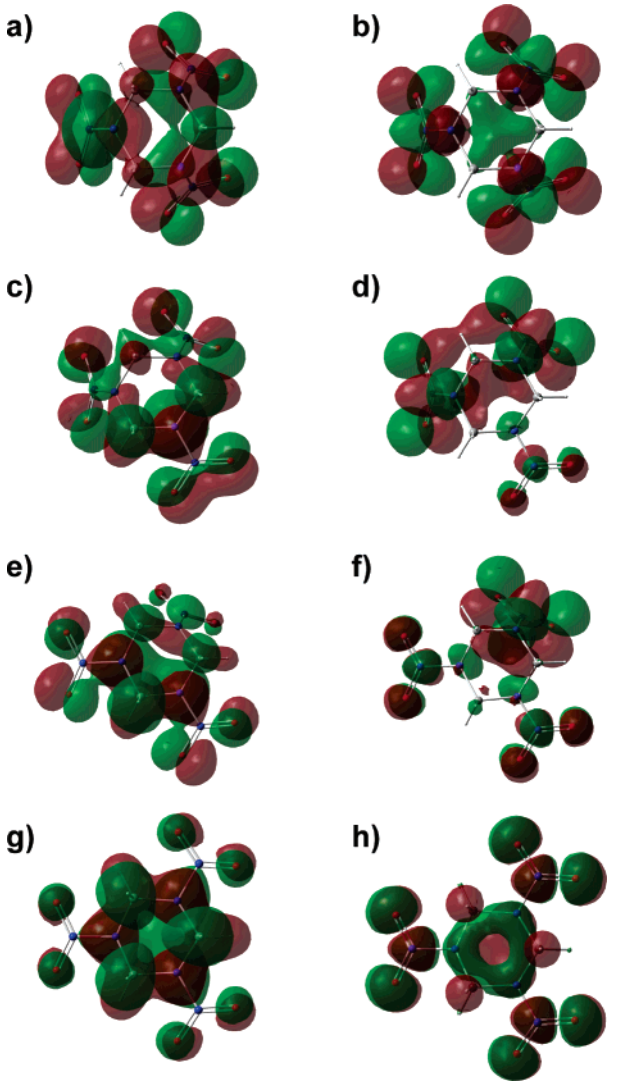


Figure 5. Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) visualizations for AAA (a, b), AAE (c, d), AEE (e, f), and EEE (g, h) conformers of vapor-phase RDX.

O5/O6. These symmetry trends are also seen in the shapes of the HOMOs and LUMOs (Figure 5c−f).

RDX Crystal. First, the values of the nuclear Fukui functions (Table 2, vector representations in Figure 4) in the α-form RDX crystal do not seem to suggest that concerted fission is a likely initial mechanistic step for decomposition, as initial, simultaneous breakage of C−N bonds seems unlikely. Upon analysis of the nucleophilic nuclear Fukui function (Φ^+_α) magnitudes, it can be seen that the largest values occur at oxygen atoms in the axial positions (O5 > O6 > O3 > O4), followed by N2, N6, and N3 in order of strength. The three carbon atoms follow after these values, with magnitudes of about 1.14−1.18 nN. These data suggest that reactivity as a result of adding an electron to the system would more likely affect axial nitro groups, rather than C−N bonding, not to mention that multiple breakages of C−N bonds are required to generate the CH_2NNO_2 products. The electrophilic nuclear Fukui function (Φ^-_α) magnitudes also preclude concerted fission as a mechanism, as the strongest magnitudes reside on bonded nitrogen pairs, rather than C−N pairs. The largest force magnitudes are upon atoms N5, N6,

Table 2. Nucleophilic (Φ^+_α) and Electrophilic (Φ^-_α) Nuclear Fukui Function Magnitudes for the Vapor-Phase AAE Conformer and α-Crystal[a]

	$\lvert\Phi^+_\alpha\rvert$		$\lvert\Phi^-_\alpha\rvert$	
	vapor	crystal	vapor	crystal
C1	0.5939	1.1849	1.1046	1.2585
C2	0.6715	1.1449	2.5481	0.9331
C3	0.5939	1.1878	1.1010	0.9693
H1	0.1827	0.4386	0.1514	0.4581
H2	0.1106	0.2812	0.1649	0.3895
H3	0.1854	0.2795	0.1415	0.1579
H4	0.1088	0.3329	0.0581	0.1319
H5	0.1827	0.2766	0.1518	0.1795
H6	0.1105	0.2986	0.1649	0.1818
N1	0.5112	1.1229	0.3836	1.0337
N2	1.0421	2.1837	2.6350	0.7651
N3	1.0418	1.6533	2.6346	1.3260
N4	2.4715	0.7777	0.4358	1.4564
N5	2.5791	0.8980	4.1012	2.2703
N6	2.5782	1.8595	4.1007	1.4622
O1	2.8081	0.2897	0.6962	0.7452
O2	2.8080	0.2040	0.6972	0.2171
O3	3.0195	3.1171	1.0048	0.8051
O4	3.2731	3.0391	1.1928	0.7158
O5	3.0188	4.1742	1.0048	0.2396
O6	3.2725	4.1195	1.1927	0.3215

[a] Atom naming corresponds with Figure 4. Unit: nN.

N3, N4, and C1 in order of decreasing magnitude. The bonded C−N pair with the strongest $\lvert\Phi^-_\alpha\rvert$ values, C1−N1, (combined magnitudes of about 2.26 nN) seems less likely to break than does N1−N4 (combined about 2.5 nN), N3−N6 (about 2.79 nN), or N2−N5 (about 3 nN). Also, no other carbon atom has a value large enough to indicate a second C−N breakage that would form the proper CH_2NNO_2 products resulting from the concerted fission mechanism.

Nuclear Fukui function data do seem to indicate that N−N breakage is likely to occur as a result of electronic perturbation in RDX crystals. Upon the addition of electron density to the crystal structure, the largest force responses occur within the axial nitro groups (N6−[O5,O6] and N5−[O3,-O4]) and their connecting ring-bound nitrogen atoms (N2 and N3). Upon depletion of electron density, the largest force responses occur between connected nitrogen atoms (N1−N4, N2−N5, and N3−N6), with axial nitro groups having slightly larger responses. Breakage of the N−N bonds is critical to both the N−N homolysis and HONO elimination mechanisms. As in the RDX vapor conformers, the presence of large nuclear Fukui function responses upon bonded carbon-hydrogen pairs combined with proximity to reactive nitro groups may be used to discriminate HONO elimination over N−N homolysis in the RDX crystal.

The nucleophilic nuclear Fukui function provides only slight evidence of such an interaction. H1 has an asymmetrically large magnitude (0.4386 nN), about 50% larger than the average of the other hydrogen atoms within the crystal (about 0.29 nN). However, H1 does not appear to be in a position to form a bond with an oxygen atom of a reactive axial nitro group, the closest being O4. As seen in Figure 4, H1 is on the opposite side of the ring as O4, making the possibility of bond formation unlikely. This seems to indicate that N−N homolysis is the most likely mechanistic route of decomposition due to additive electron perturbation. The electrophilic nuclear Fukui function shows that both H1 and H2 have significantly larger responses upon removal of an electron than other hydrogens within the crystal. $\lvert\Phi^-_\alpha\rvert$ on H1 (0.4581 nN) is nearly 3 times larger than those on hydrogen atoms not bonded to C1, and H2 (0.3895 nN) is about 2.3 times larger. As seen in Figure 4, H2 is much closer to O4 than H1, indicating that it may be able to form a bond with O4 in the event of perturbation. This seems to indicate that HONO elimination, in addition to N−N homolysis, may be a reaction path in crystal RDX.

The crystal lattice of α-RDX has a profound effect upon the symmetry of the electronic structure of its individual molecules. As shown before, the electronic structure of the HOMO and LUMO in the vapor-phase AAE had the same symmetry as its molecular skeleton, and this fact was reinforced by the magnitudes and directions of the nuclear Fukui functions. Despite their AAE molecular conformation, the electronic structures of RDX crystal molecules seem to be skewed from such a C_s symmetry. Evidence of this is seen throughout the RDX crystal molecule in its nuclear Fukui functions, as symmetrically similar atoms do not have similar magnitudes and directions (Figure 4). The axial nitro groups (N2−N5−[O3, O4] and N3−N6−[O5, O6]) do not have similar values for either the nucleophilic or electrophilic nuclear Fukui function, and the methyl groups that connect each axial nitro group to the equatorial one (C1−[H1, H2] and C3−[H5, H6]) also experience asymmetric responses to depletive electronic perturbation. The asymmetry upon axial bonded nitrogen atoms provides interesting force responses upon perturbation, especially at the N2−N5 bonded region. Upon addition of electron density, N2 has the largest response upon a non-oxygen nucleus within the molecule (2.1837 nN), but its bonded counterpart, N5, has a relatively small value (0.8980 nN). Upon the removal of electron density, however, N5 has the largest response of any nucleus within the molecule (2.2703 nN), while N2's is surprisingly small (0.7651 nN). The other axial bonded nitrogen atoms, N3 and N6, have similar responses upon either type of perturbation. Despite the sums of the responses for each of these bonded pairs being comparable for both the nucleophilic and electrophilic nuclear Fukui functions, it certainly appears that perturbation affects each in a quite different manner.

In summary, we have calculated the optimized structures, energies, and nuclear Fukui functions of RDX AAA, AAE, AEE, and EEE conformations as well as of α-RDX single crystals in an effort to understand the initial step of its unimolecular decomposition. In the vapor phase, RDX showed favorability for both the N−N homolysis and HONO elimination mechanisms. The calculated nuclear Fukui functions for AAA and AAE conformers provided the best support for N−N homolysis, as the largest responses occurred at bonded N−N sites, and no large responses were found on hydrogen atoms within the molecular structure. The AEE and EEE conformers, however, showed that HONO elimination may indeed be a viable reaction path, as strong N−N

force responses were paired with significantly large C−H magnitudes and proper H orientation for O−H formation. Data for the EEE conformer, however, may not be as applicable to the mechanism of vapor-phase RDX breakdown, as its higher energy may indicate that the conformer contributes less to the overall interconverting structure. Moreover, solid α-RDX showed favorability for the N−N homolysis mechanism. This is due to large nucleophilic nuclear Fukui function magnitudes upon atoms in axial $N-NO_2$ groups and strong electrophilic N−N responses. There was also some evidence that HONO elimination could occur, as the C1 and H2 nuclei had asymmetrically large electrophilic responses paired with H2 proximity to the O4 atom of a reactive axial nitro group. The nuclear Fukui function provided useful insight into the mechanistic effects of electronic perturbation in both the vapor phase and solid-state RDX. The concept of force derivative with respect to the electronic structure (eq 2) may be explored to use higher-order derivatives (e.g., with respect to both the electronic structure and nuclear position) for further understanding the solid-state reactions.

Acknowledgment. The research was supported by NSF (DMR-0449633). The authors would like to thank Dr. Shaoxin Feng for his technical support in this project.

References

(1) Tsiaousis, D.; Munn, R. W. Energy of charged states in the RDX crystal: Trapping of charge-transfer pairs as a possible mechanism for initiating detonation. *J. Chem. Phys.* **2005**, *122*, 184708.

(2) Luty, T.; Ordon, P.; Eckhardt, C. J. A model for mechanochemical transformations: Applications to molecular hardness, instabilities, and shock initiation of reaction. *J. Chem. Phys.* **2002**, *117*, 1775−1785.

(3) Ryzhkov, L. R.; Toscano, J. P. Crystal Lattice Effects on the Orientation and Orbital Degeneracy of Nitric Oxide Trapped in Nitramine Single Crystals. *Cryst. Growth Des.* **2005**, *5*, 2066−2072.

(4) Chakraborty, D.; Muller, R. P.; Dasgupta, S.; Goddard, W. A. The mechanism for unimolecular decomposition of RDX (1,3,5-trinitro-1,3,5-triazine), an ab initio study. *J. Phys. Chem. A* **2000**, *104*, 2261−2272.

(5) Zhao, X. S.; Hintsa, E. J.; Lee, Y. T. Infrared Multiphoton Dissociation of RDX in a Molecular-Beam. *J. Chem. Phys.* **1988**, *88*, 801−810.

(6) Behrens, R.; Bulusu, S. Thermal-Decomposition of Energetic Materials .3. Temporal Behaviors of the Rates of Formation of the Gaseous Pyrolysis Products from Condensed-Phase Decomposition of 1,3,5-Trinitrohexahydro-S-Triazine. *J. Phys. Chem.* **1992**, *96*, 8877−8891.

(7) Behrens, R.; Bulusu, S. Thermal-Decomposition of Energetic Materials .4. Deuterium-Isotope Effects and Isotopic Scrambling (H/D, C-13/O-18, N-14/N-15) in Condensed-Phase Decomposition of 1,3,5-Trinitrohexahydro-S-Triazine. *J. Phys. Chem.* **1992**, *96*, 8891−8897.

(8) Botcher, T. R.; Wight, C. A. Explosive Thermal-Decomposition Mechanism of RDX. *J. Phys. Chem.* **1994**, *98*, 5441−5444.

(9) Maharrey, S.; Behrens, R. Thermal decomposition of energetic materials. 5. Reaction processes of 1,3,5-trinitrohexahydro-s-triazine below its melting point. *J. Phys. Chem.* A **2005**, *109*, 11236−11249.

(10) Choi, M.; Kim, H.; Chung, C. FT-IR Spectra of Photochemical-Reaction Products of Crystalline RDX. *J. Phys. Chem.* **1995**, *99*, 15785−15789.

(11) Dremin, A. N. Shock Discontinuity Zone Effect - the Main Factor in the Explosive Decomposition Detonation Process. *Philos. Trans. R. Soc. London, Ser. A* **1992**, *339*, 355−364.

(12) Owens, F. J.; Sharma, J. X-Ray Photoelectron-Spectroscopy and Paramagnetic-Resonance Evidence for Shock-Induced Intramolecular Bond Breaking In Some Energetic Solids. *J. Appl. Phys.* **1979**, *51*, 1494.

(13) Pace, M. D. EPR-Spectra of Photochemical NO_2 Formation in Monocyclic Nitramines and Hexanitrohexaazaisowurtzitane. *J. Phys. Chem.* **1991**, *95*, 5858−5864.

(14) Gilman, J. J. *Chem. Propul. Inf. Agency* **1992**, *589*, 379.

(15) Gilman, J. J. Mechanochemistry. *Science* **1996**, *274*, 65.

(16) Chambers, G. P.; Lee, R. J.; Oxby, T. J.; Perger, W. F. *Shock Compression of Condensed Matter*; AIP Conf. Proc. No. 620, 2002, Melville.

(17) Ershov, A. P. Ionization During Detonation of Solid Explosives. *Combust. Expl. Shock Waves* **1975**, *11*, 798−803.

(18) Ershov, A. P.; Zubkov, P. I.; Lukyanchikov, L. A. Measurements of Electrical-Conductivity Profile in Detonation Front of Solid Explosives. *Combust. Expl. Shock Waves* **1974**, *10*, 776−782.

(19) Zubkov, P. I.; Lukyanch, La; Novoselo, Bs. Electrical-Conductivity in Detonation Zone of Condensed Explosives. *Combust. Expl. Shock Waves* **1971**, *7*, 253−256.

(20) Caulder, S. M.; Buess, M. L.; Garroway, A. N.; Miller, P. J. NQR Line Broadening Due to Crystal Lattice Imperfections and Its Relationship to Shock Sensitivity; Shock Compression of Condensed Matter - 2003: Proceedings of the Conference of the American Physical Society Topical Group on Shock Compression of Condensed Matter, 2004, Portland, Oregon, U.S.A.

(21) Kuklja, M. M. On the initiation of chemical reactions by electronic excitations in molecular solids. *Appl. Phys. A* **2003**, *76*, 359−366.

(22) Kuklja, M. M.; Aduev, B. P.; Aluker, E. D.; Krasheninin, V. I.; Krechetov, A. G.; Mitrofanov, A. Y. Role of electronic excitations in explosive decomposition of solids. *J. Appl. Phys.* **2001**, *89*, 4156−4166.

(23) Kuklja, M. M.; Kunz, A. B. Compression-induced effect on the electronic structure of cyclotrimethylene trinitramine containing an edge dislocation. *J. Appl. Phys.* **2000**, *87*, 2215−2218.

(24) Kuklja, M. M.; Stefanovich, E. V.; Kunz, A. B. An excitonic mechanism of detonation initiation in explosives. *J. Chem. Phys.* **2000**, *112*, 3417−3423.

(25) Haussuhl, S. Elastic and thermoelastic properties of selected organic crystals: acenaphthene, trans-azobenzene, benzophenone, tolane, trans-stilbene, dibenzyl, diphenyl sulfone, 2,2′-biphenol, urea, melamine, hexogen, succinimide, pentaerythritol, urotropine, malonic acid, dimethyl malonic acid, maleic acid, hippuric acid, aluminium acetylacetonate, iron acetylacetonate, and tetraphenyl silicon. *Z. Kristallogr.* **2001**, *216*, 339−353.

(26) Haycraft, J. J.; Stevens, L. L.; Eckhardt, C. J. The elastic constants and related properties of the energetic material cyclotrimethylene trinitramine (RDX) determined by Brillouin scattering. *J. Chem. Phys.* **2006**, *124*, 024712.

(27) Schwarz, R. B.; Hooks, D. E.; Dick, J. J.; Archuleta, J. I.; Martinez, A. R. Resonant ultrasound spectroscopy measurement of the elastic constants of cyclotrimethylene trinitramine. *J. Appl. Phys.* **2005**, *98*, 056106.

(28) Sewell, T. D. Monte Carlo calculations of the hydrostatic compression of hexahydro-1,3,5-trinitro-1,3,5-triazine and beta-octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine. *J. Appl. Phys.* **1998**, *83*, 4142−4145.

(29) Ye, S.; Tonokura, K.; Koshi, M. A Raman study of energy transfer processes between phonon and vibron in RDX and β-HMX at low temperatures. *Kayaku Gakkaishi* **2002**, *63*, 104.

(30) Boyd, S.; Gravelle, M.; Politzer, P. Nonreactive molecular dynamics force field for crystalline hexahydro-1,3,5-trinitro-1,3,5 triazine. *J. Chem. Phys.* **2006**, *124*, 104508.

(31) Harris, N. J.; Lammertsma, K. Ab initio density functional computations of conformations and bond dissociation energies for hexahydro-1,3,5-trinitro-1,3,5-triazine. *J. Am. Chem. Soc.* **1997**, *119*, 6583−6589.

(32) Rice, B. M.; Chabalowski, C. F. Ab initio and nonlocal density functional study of 1,3,5-trinitro-s-triazine (RDX) conformers. *J. Phys. Chem. A* **1997**, *101*, 8720−8726.

(33) Choi, C. S.; Prince, E. Crystal-Structure of Cyclotrimethylene-Trinitramine. *Acta Crystallogr. B* **1972**, *B28*, 2857−2862.

(34) Feng, S.; Li, T. Understanding solid-state reactions of organic crystals with density functional theory-based concepts. *J. Phys. Chem. A* **2005**, *109*, 7258−7263.

(35) Li, T.; Feng, S. Study of crystal packing on the solid-state reactivity of indomethacin with density functional theory. *Pharm. Res.* **2005**, *22*, 1964−1969.

(36) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *B136*, B864−B871.

(37) Kohn, W.; Becke, A. D.; Parr, R. G. Density functional theory of electronic structure. *J. Phys. Chem.* **1996**, *100*, 12974−12980.

(38) Parr, R. G.; Yang, W. T. Density-functional theory of the electronic structure of molecules. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701−728.

(39) Feynman, R. P. Forces in Molecules. *Phys. Rev.* **1939**, *56*, 340.

(40) Hellmann, H. *Einführung in die Quantenchemie*; Deuticke: Vienna, 1937.

(41) Cohen, M. H.; Gandugliapirovano, M. V.; Kudrnovsky, J. Electronic and Nuclear-Chemical Reactivity. *J. Chem. Phys.* **1994**, *101*, 8988−8997.

(42) DeProft, F.; Liu, S. B.; Geerlings, P. Calculation of the Nuclear Fukui Function and New Relations for Nuclear Softness and Hardness Kernels. *J. Chem. Phys.* **1998**, *108*, 7549−7554.

(43) Balawender, R.; De, Proft, F.; Geerlings, P. Nuclear Fukui Function and Berlin's Binding Function: Prediction of the Jahn-Teller Distortion. *J. Chem. Phys.* **2001**, *114*, 4441−4449.

(44) Doll, K.; Saunders, V. R.; Harrison, N. M. Analytical Hartree-Fock gradients for periodic systems. *Int. J. Quantum Chem.* **2001**, *82*, 1−13.

CT600202E

JCTC Journal of Chemical Theory and Computation

Catalytic Carbon−Halogen Bond Activation: Trends in Reactivity, Selectivity, and Solvation

G. Theodoor de Jong and F. Matthias Bickelhaupt*

Afdeling Theoretische Chemie, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands

Received November 22, 2006

Abstract: We have theoretically studied the oxidative addition of all halomethanes CH_3X (with X = F, Cl, Br, I, At) to Pd and $PdCl^-$, using both nonrelativistic and zeroth-order-regular-approximation-relativistic density functional theory at BLYP/QZ4P. Our study covers the gas phase as well as the condensed phase (water), where solvent effects are described with the conductor-like screening model. The activation of the C*−X bond may proceed via two stereochemically different pathways: (i) direct oxidative insertion (OxIn) which goes with *retention* of the configuration at C* and (ii) an alternative S_N2 pathway which goes with *inversion* of the configuration at C*. In the gas phase, for Pd, the OxIn pathway has the lowest reaction barrier for all CH_3X's. Anion assistance, that is, going from Pd to $PdCl^-$, changes the preference for all CH_3X's from OxIn to the S_N2 pathway. Gas-phase reaction barriers for both pathways to C−X activation generally decrease as X descends in group 17. Two striking solvent effects are (i) the shift in reactivity of Pd + CH_3X from OxIn to S_N2 in the case of the smaller halogens, F and Cl, and (ii) the shift in reactivity of $PdCl^-$ + CH_3X in the opposite direction, that is, from S_N2 to OxIn, in the case of the *heavier* halogens, I and At. We use the activation strain model to arrive at a qualitative understanding of how the competition between OxIn and S_N2 pathways is determined by the halogen atom in the activated C−X bond, by anion assistance, and by solvation.

1. Introduction

Oxidative addition and reductive elimination (eq 1) are ubiquitous as elementary reaction steps in homogeneous catalysis[1−3] and have been intensively investigated both experimentally[3−5] and theoretically.[5−14]

$$ML_n + R{-}X \underset{\text{reductive elimination}}{\overset{\text{oxidative addition}}{\rightleftharpoons}} R{-}ML_n{-}X \qquad (1)$$

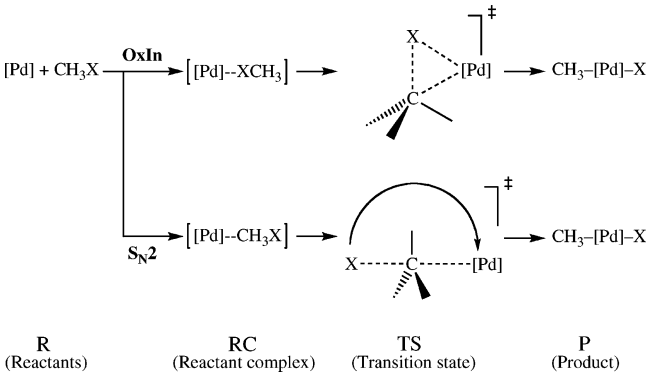
A well-known class of processes involving oxidative addition is catalytic C−X bond activation.[1,15] The catalytically active species in these reactions are generally coordination complexes of palladium or other transition metals. This process is an efficient tool for selectively converting simple educts, via C−C bond formation, into more complex compounds and is therefore of major importance for synthetic chemistry. The most intensively used substrates for such C−C coupling reactions are aryl halides, whereas it is more difficult in this context to exploit alkyl halides.[16]

In the oxidative addition process, the metal increases its formal oxidation state by two units. There has been controversy about the mechanism of this reaction, notably concerning the C−Cl bond.[17] One mechanism that has been proposed requires the concerted transfer of two electrons and involves either a concerted front-side displacement or a concerted nucleophilic displacement (S_N2) proceeding via backside attack of the C−Cl bond by the metal. Theoretical studies on the oxidative addition of the C−Cl bond in chloromethane to the Pd atom in the gas phase show that this process can indeed proceed via direct oxidative insertion of the metal into the C−Cl bond (OxIn) or via S_N2 substitution followed, in a concerted manner, by leaving-group rearrangement (S_N2-

* Corresponding author fax: +31-20-5987629; e-mail: fm.bickelhaupt@few.vu.nl.

10.1021/ct600342j CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/19/2007

Scheme 1. Model Reactions and Nomenclature



ra).[10,14,18] The reaction barrier for OxIn is lower than that for the S_N2 pathway. Interestingly, anion assistance, for example, coordination of a chloride anion to Pd, reverses this order in activation energies and makes S_N2 the preferred pathway. Note that this shift in mechanism also corresponds to a change in stereochemistry at the carbon atom involved, namely, from retention (OxIn) to inversion of the configuration (S_N2). This is of practical relevance for substrates in which the carbon atom, C*, is asymmetric (which is obviously not the case in the simple model substrate chloromethane). The two pathways are schematically summarized in Scheme 1.

In the present study, we aim at obtaining insight into the trends in reactivity of palladium and all possible carbon−halogen bonds, that is, C−X, with X = F, Cl, Br, I, and At. Along this series of halogens, we are particularly interested in the preference for one of the two pathways, OxIn or S_N2, and how this preference is affected by three different aspects, namely, (i) anion assistance, that is, using $PdCl^-$ instead of Pd, (ii) relativistic effects, and (iii) solvent effects, in particular, changing from the gas phase to an aqueous solution. To this end, we have calculated a set of consistent potential energy surfaces (PESs), using both nonrelativistic and relativistic density functional theory (DFT) and, for the solvent effects, using a continuum solvation model, which enables us to infer accurate trends in reactivity for these simple, archetypal oxidative addition reactions.

The differences in reactivity for the various combinations of inserting metal complexes, namely, bare Pd and $PdCl^-$, and substrates CH_3X are analyzed and interpreted in terms of the activation strain model of chemical reactivity.[7,9,10,19] In this model, activation energies $\Delta E^{\neq}$ are decomposed into the activation strain $\Delta E^{\neq}_{\text{strain}}$ of and the stabilizing transition state (TS) interaction $\Delta E^{\neq}_{\text{int}}$ between the reactants in the activated complex: $\Delta E^{\neq} = \Delta E^{\neq}_{\text{strain}} + \Delta E^{\neq}_{\text{int}}$. The activation strain $\Delta E^{\neq}_{\text{strain}}$ depends on the strength of the activated bond and on the extent to which a particular metal expands the bond in the activated complex. The TS interaction $\Delta E^{\neq}_{\text{int}}$ is directly determined by the bonding capabilities and, thus, the frontier orbitals of the reactants. As will emerge from our analyses, much of the trends in reaction characteristics can be traced to the different strengths of the carbon−halogen bonds and the energy levels of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO, respectively) in the substrates.

2. Computational Methods

2.1. DFT Calculations. All DFT[20] calculations have been done with the Amsterdam Density Functional (ADF) program.[21,22] Calculations were done either nonrelativistically or with scalar relativistic effects accounted for using the zeroth-order regular approximation (ZORA).[23] For the dissociation energies of the C−X bonds, where spin−orbit coupling is expected to be important, also single-point calculations using double-group symmetry including spin−orbit effects were done. The BLYP[24] density functional was used, in combination with the QZ4P basis set for all elements except hydrogen, for which the TZ2P basis set was used. The QZ4P basis set is a large uncontracted set of Slater-type orbitals (STOs) containing diffuse functions. It is of quadruple-ζ quality and has been augmented with several sets of polarization functions on each atom: two 3d and two 4f sets on F, three 3d and two 4f sets on Cl, two 4d and three 4f sets on Br, one 5d and three 4f sets on I, one 6d and two 5f sets on At, two 3d and two 4f sets on C, and two 5p and two 4f sets on Pd. The TZ2P basis set, only used for hydrogen, is of triple-ζ quality and has been augmented with two sets of polarization functions: 2p and 3d in the case of H. An auxiliary set of s, p, d, f, and g STOs was used to fit the molecular density and to represent the Coulomb and exchange potentials accurately in each self-consistent field cycle.[21] All electrons were treated variationally (i.e., no frozen-core approximation). The electronic structures of Pd, $PdCl^-$, and CH_3X were analyzed in terms of the quantitative molecular orbital model contained in Kohn−Sham DFT.[25] Recently, it has been shown that our approach is in good agreement with high-level ab initio calculations for oxidative addition reactions of the C−H,[8,11] C−C,[12] C−F,[13] and C−Cl[14] bonds to palladium.

Equilibrium and transition state geometries were fully optimized using analytical gradient techniques. All structures were verified by frequency calculations: for minima, all normal modes have real frequencies, whereas transition states have one normal mode with an imaginary frequency. The character of the normal mode associated with the imaginary frequency was analyzed to ensure that the correct transition state was found.

Solvent effects in water have been estimated using the conductor-like screening model (COSMO),[26] as implemented in the ADF program.[27] The same parameters were used as described on page 49 of the Supporting Information of ref 28. This implies using a solvent-excluding surface with an effective radius for water of 1.9 Å, derived from the macroscopic density and molecular mass, and a relative dielectric constant of 78.4. The empirical parameter in the scaling function in the COSMO equation was chosen to be 0.0. The radii of the atoms were taken to be MM3 radii,[29] divided by 1.2, giving 1.350 Å for H, 1.700 Å for C, 1.425 Å for F, 1.725 Å for Cl, 1.850 Å for Br, 1.967 Å for I, 2.092 Å for At, and 1.975 Å for Pd. Using these radii gave differences of less than 3 kcal/mol between computed and experimental hydration energies of, for example, the chloride anion, tetramethylammonium cation, and tert-butyl cation.[30] Furthermore, the above computational settings were tested on the complexation energy of F^- and C_2H_5F. At ZORA-

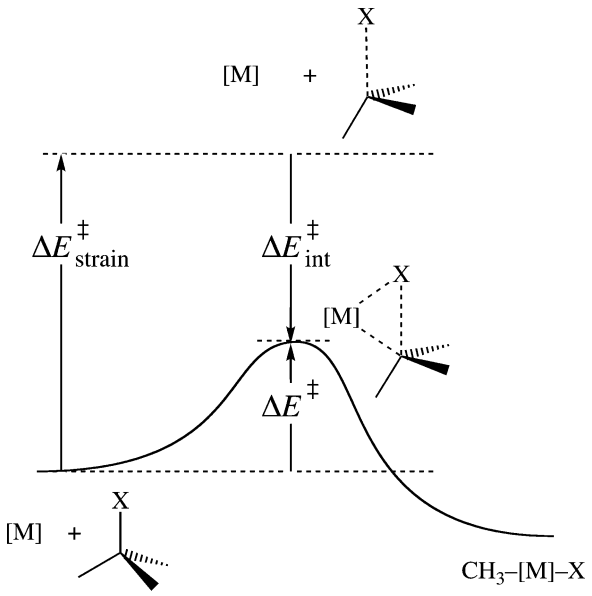


Figure 1. Illustration of the activation strain model in the case of oxidative insertion of a metal complex [M] into a C−X bond. The activation energy $\Delta E^{\neq}$ is decomposed into the activation strain $\Delta E^{\neq}_{\text{strain}}$ of and the stabilizing TS interaction $\Delta E^{\neq}_{\text{int}}$ between the reactants in the transition state.

BLYP/QZ4P, this is −19.0 kcal/mol in the gas phase and a much smaller −0.2 kcal/mol when solvent effects in water are included by COSMO. This is in complete agreement with previous observations.[31]

2.2. Activation Strain Analyses. To gain insight into how the use of different inserting metal complexes, namely, bare Pd and $PdCl^-$, and different substrates, namely, CH_3X, affects the activation barriers of the different oxidative addition reactions, the reactions were analyzed using the activation strain model of chemical reactivity.[7,9,10,19] In this model, the activation energy $\Delta E^{\neq}$ is decomposed into the activation strain $\Delta E^{\neq}_{\text{strain}}$ and the TS interaction $\Delta E^{\neq}_{\text{int}}$ (see eq 2 and Figure 1):

$$\Delta E^{\neq} = \Delta E^{\neq}_{\text{strain}} + \Delta E^{\neq}_{\text{int}} \qquad (2)$$

The activation strain $\Delta E^{\neq}_{\text{strain}}$ is the strain energy associated with deforming the reactants from their equilibrium geometry to the geometry they acquire in the activated complex (Figure 1). The TS interaction $\Delta E^{\neq}_{\text{int}}$ is the actual interaction energy between the deformed reactants in the transition state. In the present study, one of the reactants is either Pd or $PdCl^-$, and the other reactant is one of the substrates CH_3X.

2.3. Ab Initio Calculations. On the basis of the ZORA-BLYP/QZ4P optimized geometries, ab initio dissociation energies of the CH_3-X bonds were calculated at the advanced correlated CCSD(T) level[32] with the program package DIRAC.[33] A full all-electron four-component Dirac−Coulomb approach was used, which allowed nonrelativistic calculations with the Lévy−Leblond approximation,[34] relativistic calculations without spin−orbit coupling using a spin-free Dirac−Coulomb Hamiltonian,[35] and relativistic calculations using the unmodified Dirac−Coulomb Hamiltonian, which includes spin−orbit coupling [DC-CCSD(T)]. The two-electron integrals exclusively over the small components have been neglected and corrected with a simple Coulombic correction, which has been shown to be reliable.[36] The basis sets used for hydrogen, carbon, fluorine, and chlorine were Dunning's correlation consistent augmented triple-ζ (aug-cc-pVTZ) basis sets,[37] and for bromine, iodine, and astatine, Dyall's relativistically optimized triple-ζ basis sets were used.[38]

3. Results and Discussion

3.1. Reaction Profiles and Geometries. In this section, we discuss the fully relativistic PESs in the gas phase of the various oxidative addition reactions as well as the geometries of the stationary points, that is, all computed at the ZORA-BLYP/QZ4P level. In the next section, we examine how relativity has affected the trends in reactivity. Thereafter, we examine how solvent effects in water affect these trends. Finally, we analyze these trends in the framework of the activation strain model. Structural results are summarized in Figure 2 and Table 1 (relativistically) and Table 2 (relativistically in water), and results about reaction profiles are given in Table 3 and in Figures 3 and 4. Structural results for the nonrelativistic calculations, energies with zero-point vibrational energy correction, and enthalpies at 298.15 K can be found in the Supporting Information.

The reactions of Pd + CH_3F, Pd + CH_3Cl, and $PdCl^-$ + CH_3Cl have been reported before.[10,13,14] It was shown that our computational method (ZORA with the BLYP functional) gives results that are in good agreement with high-level relativistic ab initio benchmark calculations. Here, we present a comprehensive overview of all reactions of Pd and $PdCl^-$ with CH_3X.

All reactions proceed from the reactants via two distinct pathways, either via direct oxidative insertion (OxIn) or via S_N2 substitution, to the product, see Figure 2. For the OxIn pathway, the reaction proceeds from the reactants R via the formation of a stable reactant complex RC_{OxIn}, in which the halogen atom coordinates to the palladium atom, to a transition state TS_{OxIn} and, finally, a stable product P (Figure 2). There is one exception, namely, the addition of CH_3F to Pd. Here, the reaction proceeds from the same reactant complex as the alternative S_N2 pathway, because the F−Pd coordination bond is too weak to produce a stable reactant complex (vide infra).

The S_N2 pathway to oxidative addition proceeds via two consecutive stages: first, the actual nucleophilic substitution and followed by a rearrangement of the expelled leaving group toward palladium. The exact nature of this pathway and the shape of the reaction profile depend on the particular model reaction, that is, on the halogen atom in the C−X bond and on whether the metal experiences anion assistance ($PdCl^-$) or not (Pd). For all neutral Pd + CH_3X and for $PdCl^-$ + CH_3F and CH_3Cl, the S_N2 reaction proceeds from the reactants via formation of a stable reactant complex, RC_{S_N2}, that differs from RC_{OxIn}. In the S_N2 reactant complex RC_{S_N2}, CH_3X coordinates either via one hydrogen atom in an η^1 fashion or via two hydrogen atoms in an η^2 fashion to Pd (see Figure 2), completely analogous to the reactant complexes found previously for the reactions of Pd with methane[11] and ethane.[12] The distinction between coordination

Table 1. Geometry Parameters[a] (in Å, Degrees) of Stationary Points on the Potential Energy Surface along the Reaction Coordinates of the OxIn and S_N2-Type Pathways of Pd and $PdCl^-$ Addition to the C−X Bond of CH_3X, with X = F, Cl, Br, I, and At, Computed Relativistically in the Gas Phase[b]

		Pd					$PdCl^-$				
		F	Cl	Br	I	At	F	Cl	Br	I	At
R	C−X	1.413	1.818	1.986	2.189	2.295					
	∠H−C−H	108.5	108.1	107.3	107.1	106.6					
RC_{OxIn}	C−X	1.411[c]	1.853	2.030	2.226	2.336	1.423	1.901	2.112	2.298	2.422
	Pd−X	3.301[c]	2.324	2.425	2.529	2.616	2.429	2.277	2.390	2.515	2.611
	Pd−C	2.372[c]	3.516	3.669	3.817	3.909	3.133	3.582	3.850	4.037	4.228
	∠H−C−H	110.5[c]	111.8	112.5	112.4	112.8	111.3	112.0	112.8	112.4	112.7
TS_{OxIn}	C−X	1.786	2.046	2.194	2.379	2.477	1.733	2.152	2.327	2.491	2.587
	Pd−X	2.304	2.355	2.448	2.554	2.637	2.227	2.315	2.424	2.549	2.639
	Pd−C	2.111	2.506	2.643	2.730	2.807	2.462	2.790	2.902	2.916	2.978
	∠H−C−H	111.1	112.6	113.2	112.9	113.3	113.8	114.4	114.8	114.2	114.3
P	C−X	2.944	3.209	3.319	3.465	3.539	2.892	3.244	3.337	3.502	3.580
	Pd−X	1.938	2.250	2.381	2.543	2.632	2.020	2.352	2.492	2.658	2.751
	Pd−C	1.997	2.001	2.002	2.005	2.006	2.016	2.025	2.026	2.030	2.030
	∠H−C−H	112.5	112.4	112.3	112.1	112.0	111.9	111.9	111.9	111.9	111.8
RC_{S_N2}	C−X	1.411	1.824	2.004	2.222	2.390	1.447	1.899	*d*	*d*	*d*
	Pd−X	3.301	3.768	3.970	4.253	4.567	3.352	3.799	*d*	*d*	*d*
	Pd−C	2.372	2.383	2.366	2.352	2.313	2.308	2.280	*d*	*d*	*d*
	∠H−C−H	110.5	111.2	112.4	113.4	117.0	111.4	112.9	*d*	*d*	*d*
TS_{S_N2}	C−X	1.851[e]	*d*	*d*	*d*	*d*	2.359	2.009	*d*	*d*	*d*
	Pd−X	2.806[e]	*d*	*d*	*d*	*d*	3.843	3.976	*d*	*d*	*d*
	Pd−C	1.926[e]	*d*	*d*	*d*	*d*	1.963	2.214	*d*	*d*	*d*
	∠H−C−H	114.7[e]	*d*	*d*	*d*	*d*	112.2[f]	117.6	*d*	*d*	*d*
IM_{S_N2}	C−X	3.575[e]	*d*	*d*	*d*	*d*	2.856[e]	2.724	2.911	3.173	3.232
	Pd−X	3.166[e]	*d*	*d*	*d*	*d*	3.457[e]	4.546	4.755	5.044	5.113
	Pd−C	1.859[e]	*d*	*d*	*d*	*d*	1.884[e]	1.998	2.002	2.007	2.011
	∠H−C−H	120.0[e]	*d*	*d*	*d*	*d*	107.2[e]	115.0	114.4	113.9	113.9
TS_{S_N2-ra}	C−X	2.566	3.192	3.425	3.641	3.742	2.604	3.354	3.570	3.876	3.981
	Pd−X	2.508	4.022	4.329	4.342	4.506	3.262	4311	4.517	4.819	4.912
	Pd−C	2.019	2.018	2.020	2.007	2.009	1.969	1.997	2.001	2.005	2.007
	∠H−C−H	117.2	114.1	113.8	111.5	111.3	113.5	112.9	112.8	112.7	112.6

[a] See Figure 2; for ∠H−C−H, the average of the three possible ∠H−C−H angles is given. [b] At the ZORA−BLYP/QZ4P level of theory. [c] Same as RC_{S_N2}. [d] Stationary point does not exist. [e] "Anomalous" $PdCH_2\cdots HF$ structure, see text. [f] C in methyl group pointing toward Pd.

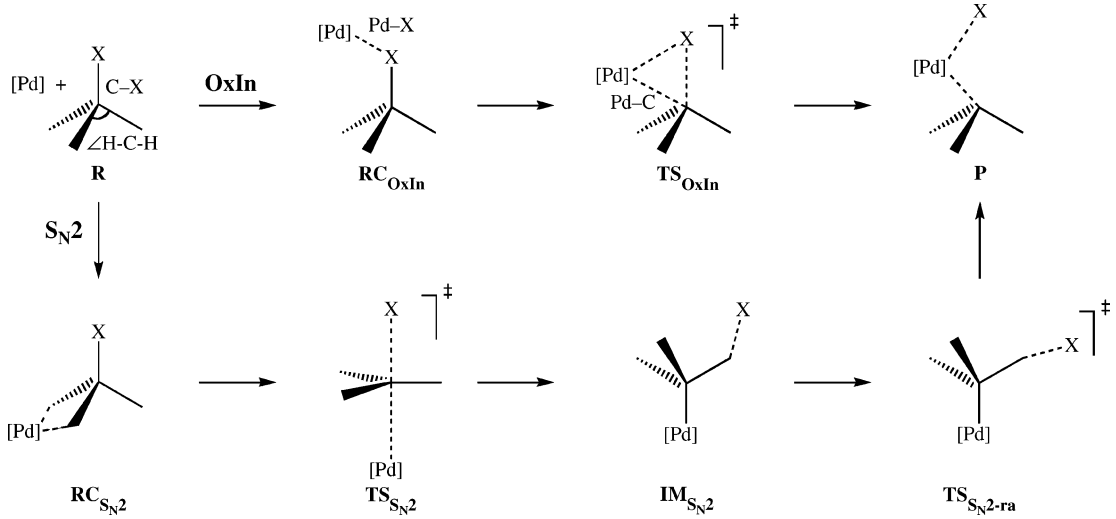


Figure 2. Structures of stationary points along the reaction coordinates of the OxIn- and S_N2-type pathways for oxidative addition of the halomethane C−X bond to [Pd], with [Pd] = Pd and $PdCl^-$ and with X = F, Cl, Br, I, and At. See Tables 1−2 for values of selected geometry parameters, calculated at various levels of theory.

Table 2. Geometry Parameters[a] (in Å, Degrees) of Stationary Points on the Potential Energy Surface along the Reaction Coordinates of the OxIn and S_N2-Type Pathways of Pd and $PdCl^-$ Addition to the C−X Bond of CH_3X, with X = F, Cl, Br, I, and At, Computed Relativistically including Solvent Effects in Water[b]

		Pd					$PdCl^-$				
		F	Cl	Br	I	At	F	Cl	Br	I	At
R	C−X	1.437	1.832	1.997	2.192	2.300					
	∠H−C−H	107.7	107.4	106.7	106.7	106.3					
RC_{OxIn}	C−X	1.436[c]	1.862	2.032	2.218	2.326	1.455	1.863	2.035	2.223	2.332
	Pd−X	3.266[c]	2.312	2.416	2.525	2.611	2.288	2.320	2.428	2.548	2.637
	Pd−C	2.333[c]	3.513	3.658	3.798	3.874	3.302	3.535	3.703	3.864	3.944
	∠H−C−H	111.2[c]	112.3	112.9	112.6	113.0	112.0	112.2	112.8	112.5	112.9
TS_{OxIn}	C−X	1.861	2.083	2.222	2.390	2.484	1.825	2.149	2.276	2.432	2.518
	Pd−X	2.592	2.327	2.430	2.544	2.626	2.157	2.327	2.442	2.570	2.656
	Pd−C	2.020	2.583	2.672	2.718	2.779	2.503	2.622	2.680	2.695	2.743
	∠H−C−H	109.5	113.7	113.9	113.3	113.5	115.3	114.3	114.2	113.5	113.5
P	C−X	2.949	3.203	3.309	3.444	3.517	2.928	3.243	3.320	3.466	3.571
	Pd−X	2.009	2.291	2.413	2.562	2.648	2.054	2.355	2.485	2.635	2.714
	Pd−C	1.993	1.999	2.000	2.004	2.005	2.022	2.031	2.033	2.038	2.037
	∠H−C−H	112.9	112.7	112.6	112.4	112.2	111.9	112.0	111.9	111.8	111.8
RC_{S_N2}	C−X	1.436	1.842	2.022	2.227	*d*	1.440	1.845	2.023	2.221	2.425
	Pd−X	3.266	3.724	3.921	4.201	*d*	3.265	3.687	3.856	4.084	4.453
	Pd−C	2.333	2.343	2.326	2.323	*d*	2.332	2.335	2.319	2.323	2.247
	∠H−C−H	111.2	111.8	113.0	113.7	*d*	111.2	111.6	112.6	112.7	118.4
TS_{S_N2}	C−X	1.838	2.162	2.266	2.447	*d*	1.927	2.173	2.267	2.504	2.557
	Pd−X	4.004	4.270	4.360	4.554	*d*	3.758	4.078	4.198	4.418	4.488
	Pd−C	2.165	2.185	2.196	2.195	*d*	2.044	2.124	2.152	2.135	2.152
	∠H−C−H	119.9	119.6	119.4	119.4	*d*	119.6[e]	119.9	119.8	120.0	119.9
IM_{S_N2}	C−X	2.381	2.879	2.926	3.032	3.038	3.275	5.187	4.318	3.999	3.805
	Pd−X	4.398	4.891	4.946	5.063	5.078	4.132	6.961	6.052	5.976	5.782
	Pd−C	2.017	2.012	2.021	2.031	2.040	2.001	1.999	2.000	2.003	2.009
	∠H−C−H	116.4	115.7	115.9	116.0	116.3	113.1	112.8	112.7	112.8	112.9
TS_{S_N2-ra}	C−X	2.931	3.841	3.990	4.139	4.142	3.241	4.088	4.299	4.543	4.628
	Pd−X	4.611	4.845	5.001	5.138	5.147	4.073	4.775	5.034	5.294	5.329
	Pd−C	1.991	1.987	1.989	1.993	1.998	2.000	2.001	2.001	2.001	2.001
	∠H−C−H	114.3	114.7	114.6	114.3	114.0	113.1	112.8	112.8	112.7	112.7

[a] See Figure 2; for ∠H−C−H, the average of the three possible ∠H−C−H angles is given. [b] At the ZORA−BLYP/QZ4P level of theory, including solvent effects in water by COSMO. [c] Same as RC_{S_N2}. [d] Stationary point does not exist. [e] C in methyl group pointing toward Pd.

to one or two hydrogen atoms is not important from an energetical point of view. Thus, enforcing η^2 coordination in the case of an η^1 equilibrium structure (and vice versa) leads to a destabilization of not more than a few tenths of a kilocalorie per mole. From RC_{S_N2}, the S_N2 substitution can then proceed via transition state TS_{S_N2} to intermediate IM_{S_N2} in which the C−X bond is broken. However, for the anion-assisted reactions of $PdCl^- + CH_3Br$, CH_3I, and CH_3At, the intermediate IM_{S_N2} is rather stable, and it is formed spontaneously, that is, via a barrierless substitution process lacking both a stable encounter complex RC_{S_N2} and a transition state TS_{S_N2}.

On the other hand, for the reactions of Pd + CH_3Cl, CH_3Br, CH_3I and CH_3At, the expulsion of the leaving group goes with (energetically highly unfavorable) charge separation. This causes the intermediate structure IM_{S_N2}, i.e., $PdCH_3^+\cdots X^-$, to become labile (i.e., it is no longer a local minimum) with respect to spontaneous back reaction to the reactant complex RC_{S_N2}. Consequently, the first and only transition state encountered along the S_N2 pathway of Pd + CH_3Cl, CH_3Br, CH_3I and CH_3At is TS_{S_N2-ra} which is associated with the rearrangement of the X^- leaving group from carbon to palladium yielding the product P of oxidative addition.

There are some marked differences between the S_N2 pathways for the addition of CH_3F compared to that of the other substrates (for a detailed discussion, see ref 13). In the first place, the C−F bond is much stronger than the other C−X bonds (vide infra) and activation of the former is associated with significantly higher barriers (via both OxIn and S_N2). Thus, at variance with the other substrates, the minimum energy path for Pd approaching CH_3F from the backside is, in a sense, redirected from straight nucleophilic substitution and proceeds instead via a relatively low-energy transition state for insertion into a C−H bond (not shown here). Furthermore, the much higher basicity of F^- compared to the other X^- causes the former, after its expulsion in the actual S_N2 transition state TS_{S_N2} and on its way toward Pd or $PdCl^-$, to abstract a proton from the methyl moiety, under formation of an "anomalous" structure $PdCH_2\cdots HF$ or $[PdCl]CH_2^-\cdots HF$ for the intermediate complex IM_{S_N2} (i.e., *not* $PdCH_3^+\cdots F^-$ or $[PdCl]CH_3\cdots X^-$). From this intermediate, fluoride migrates via the normal transition state TS_{S_N2-ra} toward Pd or $PdCl^-$ under formation of the product CH_3PdF or $CH_3[PdCl]F^-$.

Regarding the energetics of the reaction, the following trends can be observed. Along X = F, Cl, Br, I, and At, the

Table 3. Energies (in kcal/mol) Relative to Reactants of Stationary Points on the Potential Energy Surface along the Reaction Coordinates of the OxIn- and S_N2-Type Pathways of Pd and $PdCl^-$ Addition to the C−X Bond of CH_3X, with X = F, Cl, Br, I, and At[a]

		RC_{OxIn}	TS_{OxIn}	P	RC_{S_N2}	TS_{S_N2}	IM_{S_N2}	TS_{S_N2-ra}
				Nonrelativistic in Gas Phase				
Pd	F	−3.0[b]	23.2	−1.0	−3.0	39.6	31.5[c]	38.8
	Cl	−7.8	5.4	−17.5	−2.7	*d*	*d*	31.4
	Br	−10.3	0.1	−22.0	−2.8	*d*	*d*	25.5
	I	−16.1	−6.3	−25.7	−3.0	*d*	*d*	23.7
	At	−18.4	−8.9	−27.4	−3.3	*d*	*d*	20.4
$PdCl^-$	F	−12.9[b]	13.1	−29.4	−12.9	5.8	4.0	5.1
	Cl	−13.0	−3.3	−45.1	−14.4	−12.9	−14.9	−11.9
	Br	−16.7	−8.4	−48.7	*d*	*d*	−20.6	−17.2
	I	−23.8	−14.9	−51.1	*d*	*d*	−24.2	−20.8
	At	−26.2	−17.5	−51.9	*d*	*d*	−25.8	−22.2
				Relativistic in Gas Phase				
Pd	F	−5.6[b]	17.2	−16.7	−5.6	25.5	10.9[c]	29.6
	Cl	−13.3	−0.8	−33.4	−5.3	*d*	*d*	22.9
	Br	−16.2	−5.7	−37.7	−5.7	*d*	*d*	16.6
	I	−22.4	−12.2	−41.5	−6.2	*d*	*d*	9.9
	At	−22.6	−13.3	−42.7	−7.2	*d*	*d*	5.8
$PdCl^-$	F	−2.4	9.7	−39.4	−16.0	−1.8	−8.3	−2.2
	Cl	−17.6	−6.9	−54.7	−17.7	−17.2	−21.2	−18.3
	Br	−21.1	−11.6	−57.8	*d*	*d*	−26.7	−23.4
	I	−27.5	−17.2	−59.9	*d*	*d*	−30.6	−27.0
	At	−27.7	−17.9	−59.5	*d*	*d*	−32.2	−28.1
				Relativistic in Water				
Pd	F	−7.6[b]	10.5	−27.5	−7.6	3.6	1.5	3.7
	Cl	−15.7	−2.9	−40.4	−7.3	−3.1	−8.2	−4.5
	Br	−18.7	−7.7	−43.4	−7.8	−6.2	−10.7	−5.7
	I	−25.4	−14.4	−45.7	−8.2	−7.5	−10.9	−3.6
	At	−25.8	−15.6	−46.3	*d*	*d*	−12.1	−3.2
$PdCl^-$	F	−7.2	11.0	−37.4	−11.4	−1.3	−10.9	−10.4
	Cl	−19.1	−3.5	−45.9	−10.8	−7.0	−20.4	−18.8
	Br	−21.1	−7.3	−47.0	−11.2	−9.0	−20.9	−19.6
	I	−26.0	−12.3	−46.1	−10.9	−10.4	−18.3	−16.5
	At	−25.7	−12.5	−45.6	−11.4	−11.3	−17.6	−15.6

[a] Nonrelativistic in Gas Phase: computed fully nonrelativistically at BLYP/QZ4P. Relativistic in Gas Phase: computed fully relativistically at ZORA−BLYP/QZ4P. Relativistic in Water: computed fully relativistically at ZORA−BLYP/QZ4P including solvent effects in water by COSMO. For a definition of stationary points, see Figure 2. [b] Same as RC_{S_N2}. [c] "Anomalous" $PdCH_2\cdots HF$ structure, see text. [d] Stationary point does not exist.

reactant complex for direct insertion, RC_{OxIn}, becomes increasingly stable, from −13.3 kcal/mol for Pd + CH_3Cl to −26.2 kcal/mol for Pd + CH_3At and from −2.4 kcal/mol for $PdCl^-$ + CH_3F to −27.7 kcal/mol for $PdCl^-$ + CH_3At, see Table 3 and Figure 3, upper diagrams. In contrast, the relative energy of the reactant complex for the S_N2 pathway, RC_{S_N2}, does not change much when the halogen in the substrate is changed. This is clearly seen from Figure 3, middle diagrams. In all cases, the reactant complexes for addition to $PdCl^-$ are more stable than for addition to Pd; see again Table 3 and Figure 3.

All reaction barriers become lower along X = F, Cl, Br, I, and At. For example, the relative energy of TS_{OxIn} changes from 17.2 for Pd + CH_3F to −13.3 kcal/mol for Pd + CH_3-At. The relative ordering of barriers for the two pathways does not change along X: for addition to Pd, the relative energy of TS_{OxIn} lies below the relative energy of TS_{S_N2-ra} with a difference of, for example, 12.4 kcal/mol in the case of CH_3F and 19.1 kcal/mol in the case of CH_3At. In contrast, for addition to $PdCl^-$ for all CH_3X's, the relative energy of TS_{S_N2-ra} lies below the relative energy of TS_{OxIn} with a difference of, for example, 11.9 kcal/mol in the case of CH_3F and 10.2 kcal/mol in the case of CH_3At, see Table 3 and Figure 4, middle diagrams. This change of selectivity, from OxIn as the preferred pathway for addition to Pd, to S_N2 as the preferred pathway for addition to $PdCl^-$ has been observed before for the addition of CH_3Cl and can be well-understood within the activation strain model by the stronger, more stabilizing TS interaction $\Delta E_{int}^{\neq}$ in the case of $PdCl^-$, caused by the raise of Pd-4d-derived orbitals in $PdCl^-$, which translates into more stabilizing donor−acceptor orbital interactions between the metal and the substrate, see for a discussion ref 10. It is interesting to note that this same effect extends over all CH_3X additions.

The exothermicity of the reaction, that is, the relative energy of the product, becomes larger along X = F, Cl, Br, I, and At, although there is a certain saturation for the largest halogens. The reaction energy changes for addition to Pd, for example, from −16.7 kcal/mol for CH_3F to −33.4 kcal/mol for CH_3Cl to −37.7 kcal/mol for CH_3Br to −41.5 kcal/mol for CH_3I to −42.7 kcal/mol for CH_3At. The addition to $PdCl^-$ is in all cases more exothermic than that to Pd, see Table 3 and Figure 3, lower diagrams.

In the next sections, the effects of relativity and of changing the environment from the gas phase to water on these trends will be investigated. In the last section, an analysis of the trends in reactivity will be given.

3.2. Relativistic Effects. The use of a relativistic treatment is significant, but it does not change the relative order of reactivity of CH_3X oxidative addition to Pd and $PdCl^-$ along the series of halogens. The effects of relativity can be revealed by comparing the fully relativistic PESs discussed above with the corresponding fully nonrelativistic PESs derived from nonrelativistic energies and nonrelativistic geometries of stationary points. Here, we discuss the relative electronic energy PESs summarized in Table 3 and Figures 3 and 4. Note, however, that the PESs including zero-point vibrational energy corrections and those based on relative enthalpies give rise to the same trends and relativistic effects (see Tables S2 and S3 in the Supporting Information).

Relativity stabilizes the PES of all CH_3X oxidative addition reactions to Pd and $PdCl^-$, as illustrated by Figures 3 (compare dotted and dashed lines) and 4 (compare upper and middle diagrams). Reaction barriers are stabilized by up to 14.6 kcal/mol (for TS_{S_N2-ra} in the case of CH_3At addition to Pd), and reactions become more exothermic by up to 16 kcal/mol (for Pd + CH_3Cl; see Table 3). The effect increases in most cases as one proceeds along the reaction coordinate (see Figure 3; compare the difference between the dotted and dashed lines for both the reactant complexes and the product). Thus, in the case of Pd + CH_3Cl, for example, the RC_{OxIn}, TS_{OxIn}, and P are relativistically stabilized by −5.5, −6.2, and −15.9 kcal/mol (compare

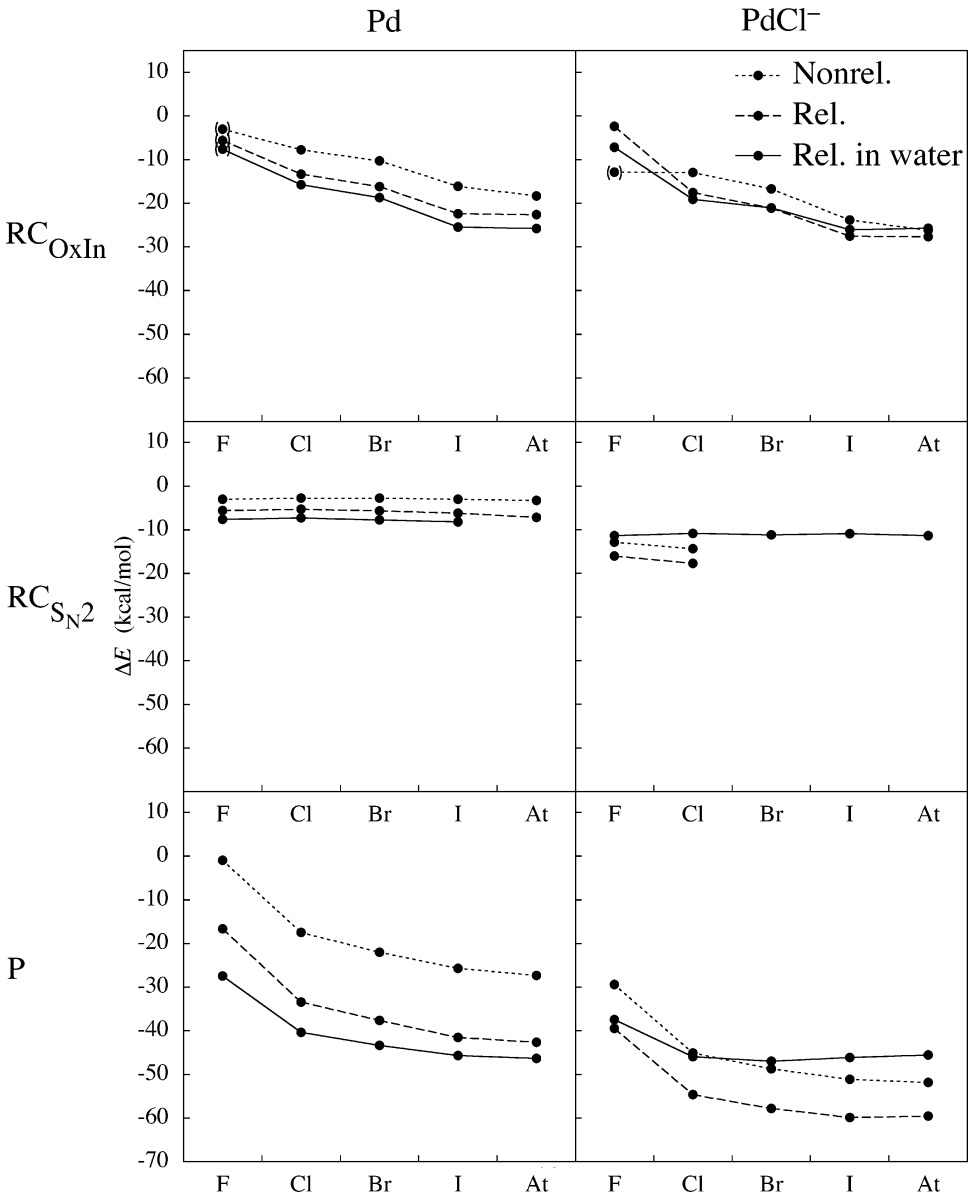


Figure 3. Relative energies (in kcal/mol) of the reactant complexes and the product for the OxIn- and S_N2-type pathways of Pd and $PdCl^-$ addition to the C−X bond of CH_3X, with X = F, Cl, Br, I, and At. Dotted lines: calculated nonrelativistically in the gas phase. Dashed lines: calculated relativistically in the gas phase. Solid lines: calculated relativistically in water.

relativistic with nonrelativistic data in Table 3). One seemingly exceptional case is RC_{OxIn} for the CH_3F addition to $PdCl^-$, which seems to become more stable when relativity is turned off, see Figure 3, upper left diagram. But, in fact, the nonrelativistic RC_{OxIn} is very unstable, such that it does not even exist and changes into the more stable RC_{S_N2}. One might expect relativistic effects to be more pronounced for the reactions involving the heavier halogens. This expectation is born out of the observation that relativistic effects on the CH_3-X bond strengths increase substantially along X = F, Cl, Br, I, and At (vide infra; see also Table 4). Interestingly, however, the relativistic stabilization of the stationary points becomes overall smaller, not stronger, along this series in X, with only two exceptions, namely, RC_{S_N2} and TS_{S_N2-ra}, for addition to Pd (see Table 3).

The relativistic effects originating from the halogen atoms obviously counteract and partially cancel those originating from palladium. Closer inspection of the influence of relativity on the orbital energy levels shows what causes this. The only observable effect in the valence and subvalence orbitals, both on Pd as on the larger halogens, is the relativistic stabilization of the s orbitals. In the case of Pd, the stabilization of the empty 5s orbital (from −3.0 to −3.4 eV) increases its electron-accepting ability. This enhances the Pd−substrate interaction and is thus responsible for the relativistic stabilization of stationary points along the reaction coordinate. In the case of the halogens, the increasing stabilization of the filled valence *n*s orbital along the series of halogens (e.g., for F from −30.4 to −30.5 eV, for Br from −19.7 to −20.4 eV, and for At from −15.1 to −19.4 eV; not shown in diagrams or tables) causes the electron-donating ability to decrease. Thus, the relativistic effects stemming from the halogen atoms have a weakening effect on the Pd−substrate interaction which counteracts (but only

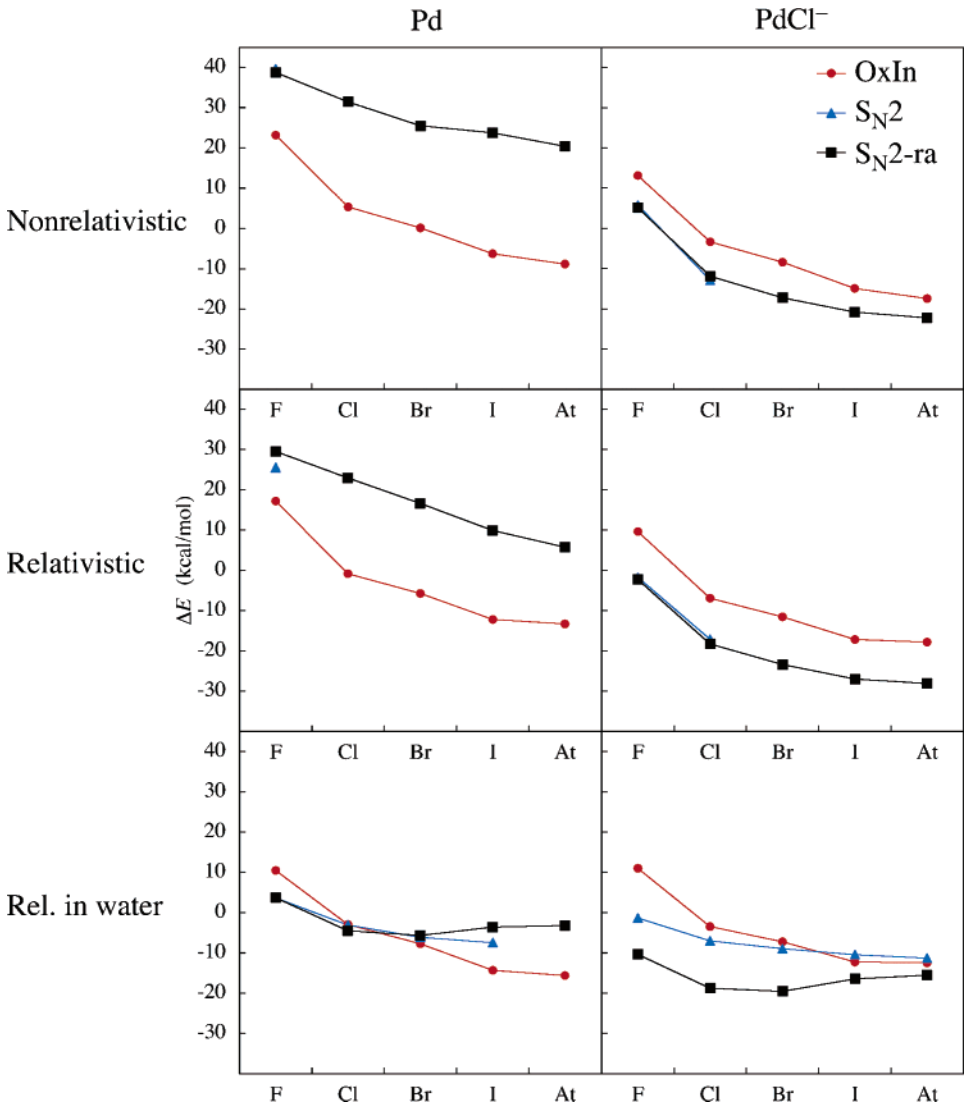


Figure 4. Reaction barriers relative to reactants (in kcal/mol) of the OxIn- and S_N2-type pathways of Pd and $PdCl^-$ addition to the C−X bond of CH_3X, with X = F, Cl, Br, I, and At, calculated at different levels of theory.

partially cancels) the stabilizing relativistic effect originating from palladium.

Relativity also affects the geometries of all species involved in the oxidative addition reactions. The most striking and general effect is a shortening of the Pd−X bond distance (compare Table 2 and Table S1 in the Supporting Information). In a previous study, on the oxidative addition of HX and X_2 to Pd, it was shown that geometry-relaxation processes caused by relativity are typically on the order of 1 kcal/mol or less and, thus, hardly affect the PES along the reaction coordinate.[39] The relativistic stabilization of stationary points along the reaction coordinate can be really directly attributed to the strengthening of the Pd−substrate interaction mentioned above.

3.3. Solvent Effects. Solvent effects, at variance with relativistic effects (vide supra), profoundly affect and qualitatively modify the characteristics of the reactions, in certain instances, to the extent that they change the preference from one to another pathway (see Table 3). Solvent effects are also markedly different for the neutral (Pd + CH_3X) as compared to the anion-assisted model reactions ($PdCl^-$ + CH_3X). In the former, solvation stabilizes all stationary points along the reaction coordinate relative to the reactants. Thus, in water, the reactant complexes of Pd + CH_3X become more strongly bound, activation barriers are reduced, and the reaction becomes more exothermic. This becomes clear, for example, by comparing solid lines (water) with dashed lines (gas phase) in Figure 3. The stabilization relative to reactants that is caused by solvation can be easily understood from electrostatic arguments. The interaction of neutral Pd with neutral CH_3X induces a charge separation $Pd^{\delta+}\cdots CH_3X^{\delta-}$ which is stabilized by the concomitant polarization of the solvent medium. Likewise, the ionic intermediate structure $PdCH_3^+\cdots X^-$ (IM_{S_N2}) which is labile in the gas phase (or, in the case of Pd + CH_3F, exists as $PdCH_2\cdots HX$, vide supra) is particularly stabilized and reappears as a stable intermediate in water (see Table 3).

On the other hand, solvation of the anion-assisted model reactions ($PdCl^-$ + CH_3X) leads in many (but not all) cases to a *destabilization* of stationary points along the reaction coordinate (see Table 3 and Figures 3 and 4). The reason is mainly the strong stabilization of the reactant $PdCl^-$ in which the excess negative charge is highly localized, leading to a strongly stabilizing electrostatic (and orbital) interaction with

Table 4. Homolytic and Heterolytic Dissociation Energies (in kcal/mol) of the CH_3-X Bond for X = F, Cl, Br, I, and At, Computed at Various Levels of DFT and ab Initio Theory[a]

method	F	Cl	Br	I	At
	Homolytic				
BLYP//BLYP[b]	114.7	82.9	72.1	62.1	57.5
BLYP	114.7	82.9	72.1	62.1	57.5
ZORA-BLYP	114.6	82.7	71.7	61.3	55.9
SO-ZORA-BLYP	114.6	82.4	68.3	54.3	41.0
SO-ZORA-BLYP in water[c]	117.9	84.2	70.7	56.1	42.5
CCSD(T)	111.1	83.0	72.6	63.2	58.9
SFDC-CCSD(T)	111.1	82.9	72.2	62.2	56.6
DC-CCSD(T)	110.7	82.0	68.8	55.7	41.7
experiment[d]	109.8	83.8	69.8	56.9	
	Heterolytic				
ZORA-BLYP	262.1	227.5	220.7	214.7	213.5
ZORA-BLYP in water[c]	80.5	72.2	71.7	74.9	75.8

[a] Geometries optimized at ZORA−BLYP/QZ4P, unless stated otherwise. BLYP = nonrelativistic BLYP/QZ4P. ZORA−BLYP = scalar ZORA−relativistic BLYP/QZ4P. SO−ZORA−BLYP = spin−orbit ZORA−relativistic BLYP/QZ4P. CCSD(T) = nonrelativistic CCSD(T). SFDC−CCSD(T) = CCSD(T) with relativistic spin−free Dirac−Coulomb Hamiltonian excluding spin−orbit coupling. DC−CCSD(T) = CCSD(T) with relativistic unmodified Dirac−Coulomb Hamiltonian including spin−orbit coupling. All CCSD(T) values at all levels of theory have been corrected for the basis−set superposition error. [b] Geometries optimized nonrelativistically at BLYP/QZ4P. [c] Solvent effects in water by COSMO; see text. [d] Obtained from corresponding enthalpies of formation at 298 K from ref 41.

the solvent.[31] In the other stationary points (i.e., RC, TS, etc.), the negative charge is delocalized, because of the $PdCl^-$/substrate interaction, over a larger area, leading to a less favorable electrostatic (and orbital) interaction.[31] Note however that the reactant complex structure $PdCl^-\cdots CH_3X$ (RC_{S_N2}), which is labile and thus absent in the gas phase, is stabilized more than TS_{S_N2} and reappears as a stable species in water (see Table 3 and Figure 4). For RC_{S_N2} and P, the relative energy clearly becomes less negative. For example, in the case of $PdCl^- + CH_3Cl$, it changes from −17.7 to −10.8 kcal/mol and from −54.7 to −45.9 kcal/mol, respectively (see Table 3 and Figure 3). For RC_{OxIn}, the effect is somewhat less clear-cut. What is clear, however, is the absence of a pronounced and general stabilization of transition states, at variance with the neutral $Pd + CH_3X$ reactions (see Figure 4).

The most striking phenomenon associated with solvation is, however, the change in preference from one to another pathway. Interestingly, the occurrence of this solvation-induced change in reaction mechanism depends on which C−X bond is activated and also on whether anion assistance is present ($PdCl^-$) or not (Pd). The neutral reactions of $Pd + CH_3X$ shift because of solvation from OxIn (preferred in the gas phase) to S_N2, but this shift occurs only for the smaller halogens, F and Cl (see Figure 4). On the other hand, the anion-assisted reactions of $PdCl^- + CH_3X$ shift from S_N2 (preferred in the gas phase) to OxIn, but this time, the shift happens only for the heavier halogens, I and At (see Figure 4). For example, for $Pd + CH_3F$, the barrier for the OxIn pathway is lowered by solvation from 17.2 to 10.5 kcal/mol, but the barrier for the S_N2 pathway is lowered much more, from 29.6 to 3.7 kcal/mol (see Table 3). On the other hand, for $Pd + CH_3At$, the barrier for the OxIn pathway is lowered by solvation from −13.3 to −15.6 kcal/mol, and the barrier for the S_N2 pathway is lowered more, from 5.8 to −3.2 kcal/mol, but not nearly enough to make the S_N2 pathway the preferred pathway (see Table 3). For comparison, the S_N2 barrier of $PdCl^- + CH_3At$, which is absent in the gas phase, is in water about 1 kcal/mol higher than the corresponding OxIn barrier (see Table 3 and Figure 4). Furthermore, the S_N2-ra barrier of $PdCl^- + CH_3At$ in water is only slightly, that is, 3 kcal/mol, lower than the OxIn barrier, whereas it is more than 10 kcal/mol lower than the OxIn barrier in the gas phase.

Solvation apparently stabilizes the S_N2 transition states of the lighter C−X bonds significantly more efficiently than the corresponding OxIn transition states but also more than the S_N2 transition states of the heavier C−X bonds (see Table 3 and Figure 4). Note that this biased solvation stabilization even breaks the intrinsic (i.e., gas-phase) trend of continuously decreasing S_N2-ra reaction barriers along the halogens: thus, in water, the S_N2-ra barrier still decreases from F to Cl to Br, but thereafter, it increases from Br to I to At (see Figure 4).

3.4. Activation Strain Analysis of Gas-Phase Reactivity. In the following, we aim at understanding the origin of the above trends in reactivity; that is, we wish to understand how the feasibility of reaction mechanisms is exactly determined by the nature of the C−X bond, anion assistance, and solvent effects. We do this using the activation strain model, which, as pointed out earlier, is a fragment approach to understanding activation barriers in terms of properties of the reactants, here, the catalyst [Pd] and the substrate CH_3X. Thus, the activation energy $\Delta E^{\neq}$ is decomposed into the activation strain $\Delta E^{\neq}_{strain}$ associated with deforming the reactants from their equilibrium structures to the geometries they adopt in the TS plus the TS interaction $\Delta E^{\neq}_{int}$, that is, the interaction between the deformed reactants (see eq 2).

First, we explore and analyze the metal−substrate bonding in the reactant complexes and the C−X bond strength in the halomethane substrates. This facilitates interpreting the activation strain analyses of activation barriers $\Delta E^{\neq}$, as the latter are the result of an interplay between these two quantities which appear as the metal−substrate TS interaction $\Delta E^{\neq}_{int}$ and the activation strain $\Delta E^{\neq}_{strain}$ that mainly stems from C−X bond elongation. The metal−substrate interaction between the reactants is provided, among others, by the donor−acceptor orbital interactions between the Pd 4d orbitals (or the Pd 4d-derived orbitals in $PdCl^-$) and the substrate σ^*_{C-X} LUMO. These orbital interactions are strengthened by relativistic effects because, as is well-known, the latter destabilize the Pd 4d orbitals, which leads to a smaller, more favorable HOMO−LUMO gap between Pd and the substrate (the effect is further reinforced by relativistic stabilization of the Pd 5s acceptor orbital).[7] The relativistic strengthening of the metal−substrate interaction in the reactant complexes can be clearly recognized in Figure 3, in which nonrelativistic and relativistic gas-phase bond energies are connected by dotted and dashed lines, respectively.

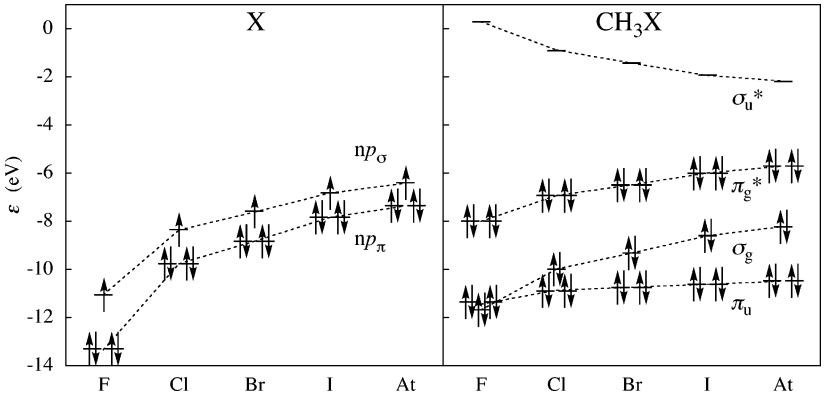


Figure 5. Kohn−Sham orbital energies ϵ (in eV) of the frontier orbitals of X and CH_3X with X = F, Cl, Br, I, and At, at ZORA-BLYP/QZ4P.

The trend of increasing metal−substrate interaction in RC_{OxIn} along X = F, Cl, Br, I, and At (see Figure 3) can be easily understood on the basis of how the electronic structure of the substrate CH_3X depends on the halogen X (see Figure 5). The frontier orbitals of CH_3X are the degenerate π_g* lone-pair HOMOs, which are mainly the halogen np_π atomic orbitals (AOs), and the σ^*_{C-X} LUMO, which is mainly the carbon 2s−halogen np_σ antibonding combination. Along X = F, Cl, Br, I, and At, the halogen np_π AOs rise in energy (the halogen becomes less electronegative) and become more diffuse. Thus, the halogen π_g* lone-pair HOMOs in the substrate also increases in energy (because the halogen np_π AOs rise) while the antibonding σ^*_{C-X} LUMO decreases as the $<2s \mid np_\sigma>$ overlap becomes smaller (because the halogen np_σ AO becomes more diffuse; see Figure 5). This reduces the HOMO−LUMO gaps for donation (substrate π_g* lone-pair to metal 5s) and backdonation (metal 4d to substrate σ^*_{C-X}) between Pd or $PdCl^-$ and CH_3X and thus strengthens the interaction energy along this series (see Figure 3).

In RC_{S_N2}, on the other hand, the metal−substrate interaction depends much less on the halogen atom X in the substrate CH_3X (in Figure 3, it seems to be essentially constant). The reason is that in RC_{S_N2} the metal interacts predominantly through an agostic interaction with C−H bonds of the methyl group in CH_3X and therefore depends much less on X (for a more detailed discussion, see ref 18).

Next, we examine the other player, besides metal−substrate interaction, in the activation strain model, namely, the geometrical rigidity of the reactants. The latter shows up as the activation strain $\Delta E^{\neq}_{strain}$. It is predominantly determined by the C−X bond stretching in the TS and, thus, by the CH_3−X bond strengths D_{C-X} in the halomethane substrates. Thus, we have computed D_{C-X} associated with the reactions $CH_3-X \rightarrow CH_3^\bullet + X^\bullet$ at BLYP/QZ4P (see Table 4). The bond strength D_{C-X} decreases continuously along X = F, Cl, Br, I, and At, from 114.6 (X = F) to 41.0 kcal/mol (X = At), at the spin−orbit corrected relativistic SO-ZORA-BLYP/QZ4P level of theory. This trend is mainly caused by the decreasing charge-stabilization that goes with the decreasing electronegativity difference across the CH_3−X bond when the halogen atom becomes less and less electronegative along the series.[40]

Before continuing the discussion of the interplay between metal−substrate interaction and geometrical deformation or strain energy in the activation strain model, we briefly evaluate the performance of our approach for computing C−X bond strength (see also Table 4). To this end, we have carried out an accurate benchmark study at the CCSD(T) level of ab initio theory in combination with a Gaussian triple-ζ basis set. Relativistic effects have been assessed for both BLYP and CCSD(T), by carrying out the computations nonrelativistically [BLYP or CCSD(T)], scalar relativistically, that is, without spin−orbit coupling [ZORA-BLYP, spin-free Dirac-Coulomb, or SFDC-CCSD(T)], and relativistically including spin−orbit coupling [SO-ZORA-BLYP, Dirac-Coulomb, or DC-CCSD(T)] (see Section 2.3 for methodological details). The BLYP values agree well with those of the CCSD(T) benchmark, and they do so at each level of treating relativistic effects. For example, D_{C-At} amounts to 57.5, 55.9, and 41.0 kcal/mol at BLYP, ZORA-BLYP, and SO-ZORA-BLYP, which agrees within 1.4 kcal/mol with the 58.9, 56.6, and 41.7 kcal/mol obtained at CCSD(T), SFDC−CCSD(T), and DC−CCSD(T), respectively, including corrections for the basis-set superposition error (see Table 4). The DC-CCSD(T) benchmark in turn agrees within a few kilocalories per mole with experimental data[41] where available. Note that relativistic effects on the homolytic D_{C-X} are predominantly caused by spin−orbit coupling. The spin−orbit term stems from the species that have a doublet open-shell configuration, that is, the dissociation products $CH_3^\bullet$ and especially $X^\bullet$. This is in line with earlier work by Ziegler and co-workers.[40] Spin−orbit coupling is a minor term for all closed-shell systems studied, that is, Pd, $PdCl^-$, CH_3X, and the species at the stationary points along the PES of our model reactions.[42] Thus, whereas the computation of reliable homolytic bond dissociation energies requires consideration of spin−orbit effects, the relative energies of stationary points along the PESs of our model reactions (all closed-shell!) can be sufficiently accurately determined through a scalar relativistic approach (i.e., the ZORA-BLYP approach which is used throughout the present work).

Interestingly, the activation strain analyses, to which we now return, reveal that the trend in activation energies $\Delta E^{\neq}$ of our gas-phase model reactions is mainly determined by the trend in C−X bond strength. The results of the analyses are listed in Table 5, both for the gas-phase and the condensed-phase model reactions which are discussed in this

Table 5. Analysis of the Reaction Barriers of the OxIn- and S_N2-Type Pathways of Pd and $PdCl^-$ Addition to the C−X Bond of CH_3X, with X = F, Cl, Br, I, and At, in Terms of the Activation Strain Model[a]

		Pd					$PdCl^-$[b]				
		F	Cl	Br	I	At	F	Cl	Br	I	At
					Relativistic in Gas Phase						
TS_{OxIn}	$\Delta E^{\neq}$	17.2	−0.8	−5.7	−12.2	−13.3	9.7	−6.9	−11.6	−17.2	−17.9
	$\Delta E^{\neq}_{strain}$	37.0	10.0	7.2	5.4	4.6	22.2	14.1	12.5	9.4	8.3
	$\Delta E^{\neq}_{int}$	−19.8	−10.8	−12.9	−17.7	−17.9	−12.5	−21.0	−24.1	−26.6	−26.1
TS_{S_N2}	$\Delta E^{\neq}$	25.5	*c*	*c*	*c*	*c*	−1.8	−17.2	*c*	*c*	*c*
	$\Delta E^{\neq}_{strain}$	103.2	*c*	*c*	*c*	*c*	84.5	9.1	*c*	*c*	*c*
	$\Delta E^{\neq}_{int}$	−77.7	*c*	*c*	*c*	*c*	−86.2	−26.3	*c*	*c*	*c*
TS_{S_N2-ra}	$\Delta E^{\neq}$	29.6	22.9	16.6	9.9	5.8	−2.2	−18.3	−23.4	−27.0	−28.1
	$\Delta E^{\neq}_{strain}$	95.8	82.8	76.4	67.4	62.3	98.1	85.3	78.5	72.5	68.8
	$\Delta E^{\neq}_{int}$	−66.2	−59.8	−59.8	−57.6	−56.5	−100.3	−103.6	−101.8	−99.5	−96.9
					Relativistic in Water						
TS_{OxIn}	$\Delta E^{\neq}$(aq)	10.5	−2.9	−7.7	−14.4	−15.6	11.0	−3.5	−7.3	−12.3	−12.5
	$\Delta E^{\neq}_{strain}$(aq)	47.9	10.0	7.5	5.7	4.8	23.2	13.0	10.0	7.8	6.6
	$\Delta E^{\neq}_{strain,pure}$	53.4	11.4	8.3	5.9	5.0	28.9	14.9	11.2	8.2	6.9
	$\Delta E^{\neq}_{strain,cav}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	$\Delta E^{\neq}_{strain,solv}$	−5.5	−1.4	−0.8	−0.2	−0.2	−5.7	−1.9	−1.2	−0.4	−0.3
	$\Delta E^{\neq}_{int}$(aq)	−37.4	−12.9	−15.2	−20.1	−20.5	−12.1	−16.5	−17.3	−20.0	−19.1
	$\Delta E^{\neq}_{int,desolv}$	0.6	0.2	0.0	−0.3	−0.2	2.2	1.3	1.2	0.9	0.9
	$\Delta E^{\neq}_{int,cav}$	−1.5	−1.5	−1.5	−1.5	−1.5	−1.5	−1.5	−1.5	−1.5	−1.5
	$\Delta E^{\neq}_{int,pure}$	−36.5	−11.6	−13.8	−18.3	−18.8	−12.8	−16.4	−17.0	−19.5	−18.6
	$(\Delta E^{\neq}_{int,gas})$	(−33.7)	(−12.1)	(−14.0)	(−18.2)	(−18.3)	(−19.5)	(−22.5)	(−23.2)	(−25.9)	(−25.2)
TS_{S_N2}	$\Delta E^{\neq}$(aq)	3.6	−3.1	−6.2	−7.5	*c*	−1.3	−7.0	−9.0	−10.4	−11.3
	$\Delta E^{\neq}_{strain}$(aq)	23.4	13.9	10.2	9.0	*c*	35.6	18.2	13.1	14.3	11.5
	$\Delta E^{\neq}_{strain,pure}$	30.0	16.3	11.5	9.6	*c*	43.1	20.7	14.6	15.1	12.0
	$\Delta E^{\neq}_{strain,cav}$	0.0	0.0	0.0	0.0	*c*	0.0	0.0	0.0	0.0	0.0
	$\Delta E^{\neq}_{strain,solv}$	−6.6	−2.4	−1.3	−0.6	*c*	−7.5	−2.5	−1.5	−0.8	−0.5
	$\Delta E^{\neq}_{int}$(aq)	−19.8	−16.9	−16.3	−16.5	*c*	−36.9	−25.2	−22.1	−24.7	−22.8
	$\Delta E^{\neq}_{int,desolv}$	1.8	1.0	0.7	0.4	*c*	2.8	2.1	1.8	1.6	1.4
	$\Delta E^{\neq}_{int,cav}$	−1.4	−1.4	−1.4	−1.4	*c*	−1.5	−1.5	−1.5	−1.5	−1.5
	$\Delta E^{\neq}_{int,pure}$	−20.1	−16.5	−15.6	−15.5	*c*	−38.2	−25.8	−22.5	−24.9	−22.8
	$(\Delta E^{\neq}_{int,gas})$	(−13.2)	(−12.9)	(−13.1)	(−13.9)	*c*	(−49.9)	(−38.8)	(−36.0)	(−39.3)	(−37.4)
TS_{S_N2-ra}	$\Delta E^{\neq}$(aq)	3.7	−4.5	−5.7	−3.6	−3.2	−10.4	−18.8	−19.6	−16.5	−15.6
	$\Delta E^{\neq}_{strain}$(aq)	80.6	75.5	71.7	69.4	66.6	88.6	81.0	77.0	74.0	71.1
	$\Delta E^{\neq}_{strain,pure}$	100.4	86.2	74.8	64.9	59.8	112.9	87.0	76.3	65.9	60.5
	$\Delta E^{\neq}_{strain,cav}$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	$\Delta E^{\neq}_{strain,solv}$	−19.9	−10.8	−3.2	4.4	6.7	−24.4	−6.1	0.6	8.0	10.5
	$\Delta E^{\neq}_{int}$(aq)	−76.9	−79.9	−77.4	−72.9	−69.9	−99.1	−99.8	−96.5	−90.4	−86.6
	$\Delta E^{\neq}_{int,desolv}$	8.0	10.2	7.5	4.1	2.6	13.1	12.4	9.5	5.1	3.7
	$\Delta E^{\neq}_{int,cav}$	−1.4	−1.5	−1.4	−1.4	−1.5	−1.5	−1.5	−1.5	−1.5	−1.5
	$\Delta E^{\neq}_{int,pure}$	−83.5	−88.7	−83.4	−75.6	−71.0	−110.8	−110.8	−104.5	−94.0	−88.8
	$(\Delta E^{\neq}_{int,gas})$	(−57.0)	(−59.7)	(−55.9)	(−53.5)	(−52.0)	(−107.2)	(−102.4)	(−97.2)	(−91.1)	(−87.1)

[a] See footnote *a* of Table 3. [b] Activation strain values include strain in $PdCl^-$ species, which is nowhere larger than 0.6 kcal/mol, but mostly around 0.1 kcal/mol. [c] TS does not exist.

section and the following one, respectively (for the results of a *non*relativistic activation strain analysis, see Table S4 in the Supporting Information). The trend of a decreasing activation energy along X = F, Cl, Br, I, and At derives predominantly from the activation strain $\Delta E^{\neq}_{strain}$ and despite a (in most but not all cases) counteracting trend in the metal−substrate TS interaction $\Delta E^{\neq}_{int}$ (see Table 5, relativistic in gas phase). The activation strain $\Delta E^{\neq}_{strain}$ decreases along X = F, Cl, Br, I, and At for all reaction steps and pathways (i.e., OxIn versus S_N2 and S_N2-ra) of the Pd- and $PdCl^-$-induced reactions. For example, in the case of the OxIn reactions of Pd + CH_3X, $\Delta E^{\neq}_{strain}$ decreases from 37.0 to 10.0 to 7.2 to 5.4 to 4.6 kcal/mol (see Table 5). Similar trends exist for the other model reactions with somewhat lower $\Delta E^{\neq}_{strain}$ values for the $PdCl^-$-induced OxIn reactions (with more eductlike and thus less deformed TSs, vide infra) and higher $\Delta E^{\neq}_{strain}$ values for the S_N2-type reactions (which feature TSs in which the substrate CH_3X is significantly more deformed, vide infra). This trend of decreasing activation strain along X = F, Cl, Br, I, and At and, thus, the trend in activation energies $\Delta E^{\neq}$ is directly inherited from the behavior of the C−X bond strength, which, as discussed

above, also weakens along this series in X. In this context, we note that the activation strain of the Pd-induced reactions arises exclusively from the substrate, in particular, the C−X elongation along the reaction coordinate. That of the $PdCl^-$-induced reactions arises *almost* exclusively from the substrate: the $PdCl^-$ complex never contributes more than a few tenths of a kilocalorie per mole (not shown in Table 5).

Here, we wish to address two phenomena in more detail: (i) the increase in activation energy from OxIn to S_N2 pathways and (ii) the much stronger decrease in activation energies due to anion assistance (i.e., if one goes from Pd to $PdCl^-$) for the S_N2 reactions than that for the OxIn reactions, which causes the preferred pathway to shift from OxIn for Pd + CH_3X to S_N2 for $PdCl^-$ + CH_3X. Note, in connection with the former issue, that the activation strain increases in all cases going from TS_{OxIn} to TS_{S_N2-ra}. We mentioned already that this is caused by the higher extent of deformation that the substrate undergoes in the S_N2 and, especially, the S_N2-ra transition states in which the C−X bond is essentially completely broken (see Figure 2). Interestingly, the activation strain can become even larger than the C−X bond dissociation energy. For example, the activation strain $\Delta E^{\neq}_{strain}$ of 62.3 kcal/mol associated with the S_N2-ra reaction of Pd + CH_3At (see Table 5) is more than 7 kcal/mol higher than the CH_3−At bond dissociation energy of 55.9 kcal/mol (see Table 4). The origin of this phenomenon is that, in the transition state TS_{S_N2-ra}, the methyl group stemming from the substrate remains pyramidal with an average H−C−H angle of 111° (in CH_3At, the H−C−H angle is 107°), while in a straight bond dissociation reaction, the methyl group would adopt its own planar equilibrium geometry with a H−C−H angle of 120°.

Finally, we address the phenomenon that anion assistance (i.e., going from Pd to $PdCl^-$) shifts the preference from the OxIn to the S_N2 mechanism. This happens as anion assistance stabilizes the transition states of both pathways, but it does so significantly more effectively for the latter. What happens is the following: The S_N2 pathway has in all cases a significantly higher, that is, less favorable, activation strain $\Delta E^{\neq}_{strain}$ than the OxIn pathway. The activation strain $\Delta E^{\neq}_{strain}$ is a characteristic of each of the two pathways: for each C−X bond, the activation strain is higher for TS_{S_N2-ra} than for TS_{OxIn} because the former is inherently more distorted than the latter, in which the C−X has to elongate only slightly. Importantly, the activation strain $\Delta E^{\neq}_{strain}$ changes comparatively little if we add a chloride ligand on Pd or solvent on the reaction system (see Table 5). However, the TS interaction $\Delta E^{\neq}_{int}$ does change significantly. Coordination of the chloride anion effectively pushes up the palladium 4d AOs, which leads in most of the reactions to an increase of the metal−substrate TS interaction $\Delta E^{\neq}_{int}$ by roughly a factor 2. In absolute terms, this means a much larger stabilization of TS_{S_N2-ra} because, in this transition state, $\Delta E^{\neq}_{int}$ was already larger. The reason is the higher extent of deformation of the substrate in TS_{S_N2-ra}, in particular, the larger C−X bond expansion (see Table 1). This stabilizes the substrate σ^*_{C-X} LUMO and thus reinforces the metal−substrate interaction in TS_{S_N2-ra} as compared to TS_{OxIn}. Thus, the TS interaction always favors S_N2-ra, but in the Pd-induced reactions, it is too weak to counteract the unfavorable $\Delta E^{\neq}_{strain}$. This changes in the case of the $PdCl^-$-induced reactions in which $\Delta E^{\neq}_{int}$ becomes large enough to overrule the trend in $\Delta E^{\neq}_{strain}$ (which favors OxIn) and to shift the reactivity to S_N2.

In the Pd + CH_3Br reactions, for example, the OxIn barrier (−6 kcal/mol) is lower than the S_N2-ra barrier (17 kcal/mol) because of a significantly lower activation strain for the less distortive OxIn reaction, that is, 7 versus 76 kcal/mol (see Table 5, relativistic in gas phase). The corresponding TS interactions of −13 and −60 kcal/mol, respectively, are too small to change the order set by the activation strain. Switching on anion assistance (i.e., going to $PdCl^-$ + CH_3Br) has little effect on the activation strain values which go from 7 to 13 kcal/mol (OxIn) and from 76 to 79 kcal/mol (S_N2-ra). However, the corresponding TS interactions jump from −13 to −24 kcal/mol (OxIn) and from −60 to −102 kcal/mol (S_N2-ra). The in absolute terms larger stabilization of $\Delta E^{\neq}_{int}$ in TS_{S_N2-ra} causes this transition state to drop below TS_{OxIn}.

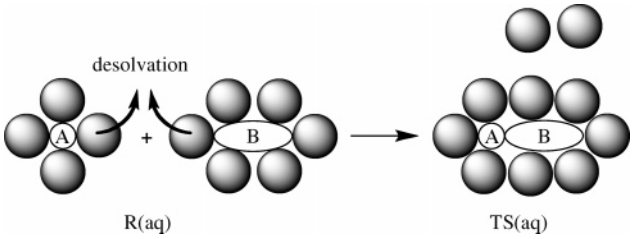
In conclusion, the heights of reaction barriers for C−X bond activation decrease as the C−X bond becomes weaker along X = F, Cl, Br, I, and At, because of the concomitant decrease in activation strain. The latter is furthermore significantly lower for the OxIn pathway, which therefore has a lower barrier and dominates the S_N2 pathway in the Pd-induced reactions. Anion assistance in the $PdCl^-$-induced reactions significantly amplifies the TS interaction which favors the S_N2 pathway. This provides us with a rational approach toward tuning the stereochemistry of the C*−X bond activation process from retention (OxIn) to inversion of configuration (S_N2), simply by increasing the TS interaction with the help of a chloride ligand (see also Scheme 1).

3.5. Activation Strain Model for Condensed-Phase Reactions. Solvation, as pointed out above, pronouncedly changes the intrinsic reactivity trends of C−X bond activation: it causes the neutral reactions of Pd + CH_3X to shift from OxIn to S_N2 (for the *smaller* halogens, F and Cl) and the anion-assisted reactions of $PdCl^-$ + CH_3X from S_N2 to OxIn (for the *heavier* halogens, I and At; see Figure 4). In the following, we aim to understand these solvent effects on the reactivity and stereochemical selectivity of catalytic C−X bond activation, using again the conceptual framework provided by the activation strain model. In this way, the height of reaction barriers is again described and understood in terms of the rigidity and bonding capabilities of the reactants, that is, the reactants *in solution*. This enables a consistent comparison with the activation strain analyses of the gas-phase reactions. Thus, the activation energy in water, $\Delta E^{\neq}(aq)$, is decomposed into the activation strain of the model catalyst and substrate in water, $\Delta E^{\neq}_{strain}(aq)$, plus the corresponding metal−substrate interaction in water, $\Delta E^{\neq}_{int}(aq)$:

$$\Delta E^{\neq}(aq) = \Delta E^{\neq}_{strain}(aq) + \Delta E^{\neq}_{int}(aq) \quad (3)$$

Note that both $\Delta E^{\neq}_{strain}(aq)$ and $\Delta E^{\neq}_{int}(aq)$ contain effects due to solvation. The activation strain in solution, $\Delta E^{\neq}_{strain}(aq)$,

Scheme 2. Desolvation of Fragments A and B When Complex AB Is Formed



is the energy associated with deforming the reactants from their equilibrium structure *in solution* to the geometry they adopt in the TS *in solution*. To reveal which role the solvent plays in this term, the latter is decomposed into three terms (eq 4): (i) the pure deformation energy, that is, the change in energy associated with the geometrical deformation of the reactants in solution *but in the absence of the solvent*, $\Delta E^{\neq}_{\text{strain,pure}}$; (ii) the change in cavitation energy associated with the corresponding deformation of the cavity in the solvent that houses the reactants, $\Delta E^{\neq}_{\text{strain,cav}}$; and (iii) the corresponding change in solute−solvent interaction, that is, in solvation stabilization $\Delta E^{\neq}_{\text{strain,solv}}$:

$$\Delta E^{\neq}_{\text{strain}}(\text{aq}) = \Delta E^{\neq}_{\text{strain,pure}} + \Delta E^{\neq}_{\text{strain,cav}} + \Delta E^{\neq}_{\text{strain,solv}} \quad (4)$$

Thus, the activation strain in water depends on the rigidity of the reactants (mainly the substrate), the extent to which the solvent cavities must change in reaction to the geometrical deformation of the reactants, and the effect of all these geometrical deformations of reactants and solvent on the solute−solvent interaction.

The TS interaction in solution, $\Delta E^{\neq}_{\text{int}}(\text{aq})$, is the energy associated with bringing the deformed reactants *in solution* together in the TS *in solution*. To also uncover how the solvent affects this term, $\Delta E^{\neq}_{\text{int}}(\text{aq})$ is decomposed again into three terms (eq 5): (i) the change in energy associated with the desolvation of those sites on either reactant that in the TS are bonding and/or in steric contact, $\Delta E^{\neq}_{\text{int,desolv}}$, see also Scheme 2; (ii) the change in cavitation energy associated with going from two reactant cavities in the solvent to one cavity that houses the TS, $\Delta E^{\neq}_{\text{int,cav}}$; and (iii) the pure interaction energy between the partially desolvated reactants (i.e., without desolvation and cavitation effects), $\Delta E^{\neq}_{\text{int,pure}}$:

$$\Delta E^{\neq}_{\text{int}}(\text{aq}) = \Delta E^{\neq}_{\text{int,desolv}} + \Delta E^{\neq}_{\text{int,cav}} + \Delta E^{\neq}_{\text{int,pure}} \quad (5)$$

The $\Delta E^{\neq}_{\text{int,desolv}}$ is computed as the change in solute−solvent interaction when each of the deformed reactants is brought from its regular cavity to the TS cavity *in the absence of the other reactant* (i.e., by using ghost atoms for the other reactant). The term $\Delta E^{\neq}_{\text{int,pure}}$ is then computed as $\Delta E^{\neq}_{\text{int}}(\text{aq}) - \Delta E^{\neq}_{\text{int,desolv}} - \Delta E^{\neq}_{\text{int,cav}}$.

3.6. Activation Strain Analysis of Condensed-Phase Reactivity. Solvation shifts the preference of the neutral Pd + CH_3X reactions from OxIn to S_N2 (see Table 5) for two reasons: (i) it weakens the C−X bond regarding heterolytic bond cleavage and thus effectively reduces the rigidity of the substrate, and (ii) it stabilizes the metal−substrate interaction by the concomitant enhancement of the charge separation $C^{\delta+}-X^{\delta-}$ in the substrate. This shows up in a lower activation strain in water, $\Delta E^{\neq}_{\text{strain}}(\text{aq})$, as compared to that in the gas phase, $\Delta E^{\neq}_{\text{strain}}$, as well as in a more stabilizing TS interaction in water, $\Delta E^{\neq}_{\text{int}}(\text{aq})$, as compared to that in the gas phase, $\Delta E^{\neq}_{\text{int}}$ (see Table 5). Importantly, this stabilizing effect is stronger for the more deformed TS_{S_N2-ra} than for TS_{OxIn} as the former has a significantly more elongated and polarized $C^{\delta+}-X^{\delta-}$ bond (vide supra). In the case of Pd + CH_3Cl, for example, hydration stabilizes the activation strain and TS interaction of the OxIn pathway by only 0.0 and −2.1 kcal/mol, respectively, whereas it stabilizes the corresponding terms of the S_N2-ra reaction by a sizable −7.3 and −20.1 kcal/mol, respectively [Table 5: compare $\Delta E^{\neq}_{\text{strain}}(\text{aq})$ with $\Delta E^{\neq}_{\text{strain}}$ and $\Delta E^{\neq}_{\text{int}}(\text{aq})$ with $\Delta E^{\neq}_{\text{int}}$]. The solvent-induced stabilization of the activation strain and TS interaction furthermore increases along At, I, Br, Cl, and F, that is, if one goes to the smaller and more electronegative halogens. Thus, for Pd + CH_3I, the hydration-induced stabilization of the activation strain and TS interaction of the S_N2-ra reaction has been reduced to +2 (i.e., a slight *destabilization*) and −15.3 kcal/mol, respectively. As a result, the barriers of the S_N2 pathway approach those of the OxIn pathway, and for F and Cl, TS_{S_N2-ra} becomes lower in energy than TS_{OxIn}.

Further analyses show that the hydration-induced stabilization of the activation strain in the case of the smaller halogens is indeed caused by a better solvation term, $\Delta E^{\neq}_{\text{strain,solv}}$. In the case of the S_N2-ra reaction, the latter amounts to −19.9 (F), −10.8 (Cl), −3.2 (Br), +4.4 (I), and +6.7 kcal/mol (At) (see Table 5). The cavitation term, $\Delta E^{\neq}_{\text{strain,cav}}$, is nowhere larger than a virtually negligible 0.1 kcal/mol. The reason for the large solvent stabilization in the case of TS_{S_N2-ra} is that solvation in water stabilizes charge separation and opens the possibility for heterolytic bond dissociation of the CH_3-X bond, that is, dissociation into the ionic fragments CH_3^+ and X^-. This is nicely illustrated by the data in Table 4, which shows homolytic and heterolytic C−X bond dissociation energies for both the gas phase and the water phase. In the gas phase, the homolytic C−X dissociation is always more favorable than the heterolytic one. But, in water, heterolytic dissociation is significantly more strongly stabilized. In the case of F and Cl, heterolytic dissociation becomes even more favorable than homolytic dissociation. In the cases of Br, I, and At, the selective stabilization of the heterolytic C−X dissociation is not strong enough to open this ionic dissociation mode as a more favorable alternative to homolytic dissociation.

The above effects of hydration on the activation strain can be easily understood with the classical electrostatic Born model of a spherical ion in a dielectric continuum (eq 6):[43]

$$\Delta E_{\text{solv}} = -\frac{q^2}{8\pi\epsilon_0 a}\left(1 - \frac{1}{\epsilon_r}\right) \quad (6)$$

In this equation, ϵ_0 is the dielectric constant in a vacuum and ϵ_r is the relative dielectric constant of the solvent (i.e., 78.4 for water). The charge q is −1 for the X^- ion. The appearance of the radius a of the ion in the denominator leads to smaller solvation energies for larger ions. On the

basis of this simple model, it is immediately clear why the dissociation of CH_3X into CH_3^+ and X^- in water is more favored for the smaller halogens: the large halogenide anions are in a sense too large to be well-stabilized by solvation.

Likewise, further analyses show that the hydration-induced stabilization of the TS interaction in the case of the smaller halogens is indeed caused by a reinforcement of the pure interaction between the solvated reactants, $\Delta E^{\neq}_{\text{int,pure}}$, which dominates all other effects (Table 5: compare this term directly with $\Delta E^{\neq}_{\text{int}}$ or $\Delta E^{\neq}_{\text{int,gas}}$). The contribution of the changing cavitation energy $\Delta E^{\neq}_{\text{int,cav}}$ is in all cases small, that is, around −1.5 kcal/mol (Table 5). This decrease in cavitation energy is caused by the fact that the deformed reactants occupy less space when they form the complex than when they are apart from each other. The partial desolvation energy $\Delta E^{\neq}_{\text{int,desolv}}$ is (naturally) always destabilizing.

Finally, we extend our analysis to the anion-assisted $PdCl^- + CH_3X$ reactions in water to understand why here solvation shifts the reactivity from S_N2 to OxIn, that is, in the opposite direction as compared to the effect of solvation on the neutral reactions. The main effect of solvation appears to be the *weakening* of the metal−substrate TS interaction, at variance with the strengthening in the case of the neutral reactions $Pd + CH_3X$ (vide infra). On the other hand, the activation strain behaves quite similar in the anion-assisted and neutral reactions, since the substrates are identical and the strain in Pd and $PdCl^-$ is zero versus negligible.

The TS interaction, as discussed above, is more stabilizing for the S_N2 pathway and therefore favors this mechanism over OxIn. In the gas phase, anion assistance amplifies the TS interaction term, which causes the reactivity to switch from OxIn, for $Pd + CH_3X$, to S_N2, for $PdCl^- + CH_3X$ (see Figure 4 and Table 5). Now, if we go from the gas phase to the water phase, the TS interaction becomes again weaker, an effect that favors the OxIn transition states because it destabilizes the barrier of the anion-assisted TS_{S_N2-ra} more than that of the corresponding TS_{OxIn}. For example, for the neutral $Pd + CH_3At$ reaction, solvation *strengthens* the TS interaction of TS_{OxIn} and TS_{S_N2-ra} by −2.6 and −13.4 kcal/mol, respectively, while for the anion-assisted $PdCl^- + CH_3$-At reaction, solvation *weakens* the TS interaction of TS_{OxIn} and TS_{S_N2-ra} by +7.0 and +10.3 kcal/mol, respectively (Table 5: compare $\Delta E^{\neq}_{\text{int}}$(aq) with $\Delta E^{\neq}_{\text{int}}$ values).

Similar solvent effects occur for the barrier associated with TS_{S_N2} which, in the case of X = I and At, becomes even higher in energy than that with TS_{OxIn} (see Figure 4 and Table 5). But the similarity between the solvent effects in the precise electronic mechanism through which solvation destabilizes both TS_{S_N2} and TS_{S_N2-ra} is masked by the relatively large concomitant shift, along the reaction coordinate of the former, toward the product side. As explained in detail in ref 10, this brings the solvated TS_{S_N2-ra} to a position on the reaction coordinate at which interactions are much stronger, and this, in turn, masks the solvation-induced weakening of the TS interaction. In addition, a direct comparison between the gas-phase and water-phase TS_{S_N2} is further complicated by the fact that for $PdCl^- + CH_3X$ this transition state exists only for X = F and Cl but not for the other halogens. To nevertheless enable a systematic gas-phase versus condensed-phase comparison, we have also computed the TS interaction in all the *condensed-phase* TS_{S_N2} geometries, *however, in the absence of solvent*. The metal−substrate interaction associated with this structure is designated $\Delta E^{\neq}_{\text{int,gas}}$ and serves as a measure for the gas-phase TS interaction $\Delta E^{\neq}_{\text{int}}$. Moreover, $\Delta E^{\neq}_{\text{int,gas}}$ serves as a point of reference to which the condensed-phase TS interaction $\Delta E^{\neq}_{\text{int}}$(aq) can be compared in a consistent fashion, that is, without strong geometry effects that mask the intrinsic change in the interaction. Indeed, the values of the condensed-phase TS interaction $\Delta E^{\neq}_{\text{int}}$(aq) of −22 to −37 kcal/mol are substantially less stabilizing than those of $\Delta E^{\neq}_{\text{int,gas}}$ of −36 to −50 kcal/mol (see Table 5).

Further analyses show that the hydration-induced weakening of the TS interaction of the anion-assisted $PdCl^- + CH_3X$ reactions is, in most cases, mainly caused by a weakening of the pure interaction between the solvated reactants, $\Delta E^{\neq}_{\text{int,pure}}$, which dominates all other effects (Table 5: compare this term directly with $\Delta E^{\neq}_{\text{int}}$ or $\Delta E^{\neq}_{\text{int,gas}}$). The contribution of the changing cavitation energy $\Delta E^{\neq}_{\text{int,cav}}$ is in all cases small, that is, around −1.5 kcal/mol (Table 5). This decrease in cavitation energy is caused by the fact that the deformed reactants occupy less space when they form the complex than when they are apart from each other. The partial desolvation energy $\Delta E^{\neq}_{\text{int,desolv}}$ is (naturally) always destabilizing but relatively unimportant, that is, about 1 order of magnitude smaller than $\Delta E_{\text{int,pure}}$ (see Table 5). The hydration-induced weakening of the pure metal−substrate interaction is ascribed to the polarization of the negative charge on $PdCl^-$ (away from the desolvated binding site) and the stabilization of the highest occupied molecular orbitals on $PdCl^-$, which reduces the capability of these orbitals to participate in donor−acceptor orbital interactions. Similar effects have been analyzed in detail for the E2 and S_N2 reactions of $F^- + CH_3CH_2F$.[31]

4. Conclusions

Palladium-catalyzed C*−X bond activation in halomethanes can proceed through two stereochemically different pathways: direct oxidative insertion (OxIn, which goes with the retention of configuration at C*) and S_N2 substitution (which goes with the inversion of configuration at C*). Using the activation strain model of chemical reactivity, we have shown that the barriers of all pathways decrease along X = F, Cl, Br, I, and At because the C−X bond becomes less stable and therefore less rigid along this series. Relativistic effects substantially stabilize the stationary points along the PES, but they do not change the relative order in barrier heights and reaction energies along the various model reactions.

Interestingly, our activation strain analyses provide a very simple and transparent picture of how anion assistance (i.e., going from the model catalyst Pd to $PdCl^-$) and solvation (water, described through the COSMO model) affect the overall reactivity and the selectivity between the OxIn (retention of configuration) and S_N2 pathways (inversion of configuration). The latter pathway is inherently connected with a higher extent of deformation of the substrate in the TS, which leads to a higher activation strain $\Delta E^{\neq}_{\text{strain}}$ for S_N2 than for OxIn. Through the relationship $\Delta E^{\neq} = \Delta E^{\neq}_{\text{strain}} +$

$\Delta E_{int}^{\neq}$, this contributes to a higher reaction barrier $\Delta E^{\neq}$ for S_N2 than for OxIn.

This situation can now be modulated by the TS interaction $\Delta E_{int}^{\neq}$ which *favors* the S_N2 pathway because the more deformed substrate in the latter is also a better partner in electrostatic and donor−acceptor orbital interactions. Thus, whenever the TS interaction $\Delta E_{int}^{\neq}$ is small, the trend in selectivity is determined more by the activation strain $\Delta E_{strain}^{\neq}$, and vice versa.

Thus, anion assistance, which increases the bonding capabilities of the model catalyst, favors the S_N2 pathway. On the other hand, solvation on top of anion assistance diminishes the bonding capabilities of the model catalyst and therefore favors again the OxIn pathway.

Acknowledgment. We thank the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO-CW and NWO-NCF) for financial support.

Supporting Information Available: Energies with zero-point vibrational energy corrections and enthalpies at 298.15 K of all stationary points at all levels of theory and Cartesian coordinates of all species occurring in this study. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Collman, J. P.; Hegedus, L. S.; Norton, J. R.; Finke, R. G. *Principles and Applications of Organotransition Metal Chemistry*; University Science Books: Mill Valley, California, 1987. Elschenbroich, Ch. A. *Organometallics*; Wiley-VCH: Weinheim, Germany, 2006.

(2) Grushin, V. V.; Alper, H. *Chem. Rev.* **1994**, *94*, 1047. Amatore, C.; Jutand, A. *Acc. Chem. Res.* **2000**, *33*, 314.

(3) Luh, T.-Y.; Leung, M.-k.; Wong, K. T. *Chem. Rev.* **2000**, *100*, 3187.

(4) Ritter, D.; Weisshaar, J. C. *J. Am. Chem. Soc.* **1990**, *112*, 6425. Fayet, P.; Kaldor, A.; Cox, D. M. *J. Chem. Phys.* **1990**, *92*, 254. Carroll, J. J.; Weisshaar, J. C. *J. Am. Chem. Soc.* **1993**, *115*, 800. Casado, A. L.; Espinet, P. *Organometallics* **1998**, *17*, 954. Stürmer, R. *Angew. Chem.* **1999**, *111*, 3509. Hinrichs, R. Z.; Schroden, J. J.; Davis, H. F. *J. Am. Chem. Soc.* **2003**, *125*, 860. de Pater, B. C.; Zijp, E. J.; Frühauf, H.-W.; Ernsting, J. M.; Elsevier, C. J.; Vrieze, K. *Organometallics* **2004**, *23*, 269. Espinet, P.; Echavarren, A. *Angew. Chem.* **2004**, *116*, 4808. Wang, G.; Chen, M.; Zhou, M. *Chem. Phys. Lett.* **2005**, *412*, 46. Lersch, M.; Tilset, M. *Chem. Rev.* **2005**, *105*, 2471.

(5) Weisshaar, J. C. *Acc. Chem. Res.* **1993**, *26*, 213. Carroll, J. J.; Haug, K. L.; Weisshaar, J. C.; Blomberg, M. R. A.; Siegbahn, P. E. M.; Svensson, M. *J. Phys. Chem.* **1995**, *99*, 13955. Porembski, M.; Weisshaar, J. C. *J. Phys. Chem. A* **2000**, *104*, 1524. Haynes, A.; Maitlis, P. M.; Morris, G. E.; Sunley, G. J.; Adams, H.; Badger, P. W.; Bowers, C. M.; Cook, D. B.; Elliott, P. I. P.; Ghaffar, T.; Green, H.; Griffin, T. R.; Payne, M.; Pearson, J. M.; Taylor, M. J.; Vickers, P. W.; Watt, R. J. *J. Am. Chem. Soc.* **2004**, *126*, 2847.

(6) Siegbahn, P. E. M.; Blomberg, M. R. A.; Svensson, M. *J. Am. Chem. Soc.* **1993**, *115*, 1952. Griffin, T. R.; Cook, D. B.; Haynes, A.; Pearson, J. M.; Monti, D.; Morris, G. E. *J. Am. Chem. Soc.* **1996**, *118*, 3029. Siegbahn, P. E. M. *J. Am. Chem. Soc.* **1996**, *118*, 1487. Wittborn, A. M. C.; Costas, M.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Chem. Phys.* **1997**, *107*, 4318. Cui, Q.; Musaev, D. G.; Morokuma, K. *J. Chem. Phys.* **1998**, *108*, 8418. Torrent, M.; Solà, M.; Frenking, G. *Chem. Rev.* **2000**, *100*, 439. Senn, H. M.; Ziegler, T. *Organometallics* **2004**, *23*, 2980. Gooβen, L. J.; Koley, D.; Hermann, H.; Thiel, W. *Chem. Commun.* **2004**, *2141*. Kozuch, S.; Shaik, S.; Jutand, A.; Amatore, C. *Chem.—Eur. J.* **2004**, *10*, 3072. Diefenbach, A.; Bickelhaupt, F. M. *J. Organomet. Chem.* **2005**, *690*, 2191. de Jong, G. Th.; Visser, R.; Bickelhaupt, F. M. *J. Organomet. Chem.* **2006**, *691*, 4341.

(7) Diefenbach, A.; Bickelhaupt, F. M. *J. Chem. Phys.* **2001**, *115*, 4030.

(8) de Jong, G. Th.; Solà, M.; Visscher, L.; Bickelhaupt, F. M. *J. Chem. Phys.* **2004**, *121*, 9982.

(9) Diefenbach, A.; de Jong, G. Th.; Bickelhaupt, F. M. *Mol. Phys.* **2005**, *103*, 995.

(10) Diefenbach, A.; de Jong, G. Th.; Bickelhaupt, F. M. *J. Chem. Theory Comput.* **2005**, *1*, 286.

(11) de Jong, G. Th.; Geerke, D. P.; Diefenbach, A.; Bickelhaupt, F. M. *Chem. Phys.* **2005**, *313*, 261.

(12) de Jong, G. Th.; Geerke, D. P.; Diefenbach, A.; Solà, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2005**, *26*, 1006.

(13) de Jong, G. Th.; Bickelhaupt, F. M. *J. Phys. Chem. A* **2005**, *109*, 9685.

(14) de Jong, G. Th.; Bickelhaupt, F. M. *J. Chem. Theory Comput.* **2006**, *2*, 322.

(15) March, J. *Advanced Organic Chemistry*; McGraw-Hill: Tokyo, 1977. Dedieu, A. *Chem. Rev.* **2000**, *100*, 543. Nicolaou, K. C.; Bulger, P. G.; Sarlah, D. *Angew. Chem.* **2005**, *117*, 4516.

(16) Frisch, A. C.; Beller, M. *Angew. Chem.* **2005**, *117*, 680.

(17) Kiplinger, J. L.; Richmond, Th. G.; Osterberg, C. E. *Chem. Rev.* **1994**, *94*, 373.

(18) Bickelhaupt, F. M.; Ziegler, T.; von Ragué Schleyer, P. *Organometallics* **1995**, *14*, 2288.

(19) Bickelhaupt, F. M. *J. Comput. Chem.* **1999**, *20*, 114.

(20) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864. Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133. Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(21) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41.

(22) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931. Baerends, E. J.; Autschbach, J.; Bérces, A.; Bo, C.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; van Faassen, M.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Harris, F. E.; van den Hoek, P.; Jacobsen, H.; Jensen, L.; van Kessel, G.; Kootstra, F.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Osinga, V. P.; Patchkovskii, S.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Ros, P.; Schipper, P. R. T.; Schreckenbach, G.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visser, O.; Wang, F.; van Wezenbeek, E.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; Yakovlev, A. L.; Ziegler, T. *ADF2005.01*; SCM, Theoretical Chemistry: Vrije Universiteit, Amsterdam, The Netherlands, 2005.

(23) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. *J. Chem. Phys.* **1994**, *101*, 9783.

(24) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(25) Bickelhaupt, F. M.; Baerends, E. J. Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers Inc.: New York, 2000; Vol. 15; p 1.

(26) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799. Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224.

(27) Pye, C. C.; Ziegler, T. *Theor. Chem. Acc.* **1999**, *101*, 396.

(28) Bon, R. S.; van Vliet, B.; Sprenkels, N. E.; Schmitz, R. F.; de Kanter, F. J. J.; Stevens, Chr. V.; Swart, M.; Bickelhaupt, F. M.; Groen, M. B.; Orru, R. V. A. *J. Org. Chem.* **2005**, *70*, 3542.

(29) Allinger, N. L.; Zhou, X.; Bergsma, J. *THEOCHEM* **1994**, *312*, 69.

(30) van Duijnen, P. T.; Grozema, F. C.; Swart, M. *THEOCHEM* **1999**, *464*, 191.

(31) Bickelhaupt, F. M.; Baerends, E. J.; Nibbering, N. M. M. *Chem.—Eur. J.* **1996**, *2*, 196.

(32) Cizek, J. *J. Chem. Phys.* **1966**, *45*, 4256. Purvis, G. D., III; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910. Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

(33) Visscher, L.; Lee, T. J.; Dyall, K. G. *J. Chem. Phys.* **1996**, *105*, 8769. Jensen, H. J. Aa.; Saue, T.; Visscher, L. *DIRAC*, version 4.0; Syddansk Universitet: Odense, Denmark, 2004.

(34) Lévy-Leblond, J. M. *Commun. Math. Phys.* **1967**, *6*, 286.

(35) Dyall, K. G. *J. Chem. Phys.* **1994**, *100*, 2118.

(36) Visscher, L. *Theor. Chem. Acc.* **1997**, *98*, 68.

(37) Kendall, R. A.; Dunning, Th. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(38) Dyall, K. *Theor. Chem. Acc.* **2002**, *108*, 335.

(39) de Jong, G. Th.; Kovács, A.; Bickelhaupt, F. M. *J. Phys. Chem. A* **2006**, *110*, 7943.

(40) Deng, L.; Branchadell, V.; Ziegler, T. *J. Am. Chem. Soc.* **1994**, *116*, 10645.

(41) Linstrom, J.; Mallard, W. G. P. *NIST Standard Reference Database Number 69*, June 2005 Release. http://webbook.nist.gov/chemistry (accessed Jul 2006).

(42) For example, the relativistic energy *including* spin−orbit coupling of the oxidative-addition product P of Pd + CH_3-At relative to R is −45.1 kcal/mol (not shown in a table), which differs only 2.4 kcal/mol from the scalar relativistic value of −42.7 kcal/mol (see Table 3).

(43) Born, M. *Z. Phys.* **1920**, *1*, 45. Atkins, P. W. *Physical Chemistry*; Oxford University Press: Oxford, United Kingdom, 1994.

CT600342J

JCTC Journal of Chemical Theory and Computation

Modeling Aromatic Liquids: Toluene, Phenol, and Pyridine

Christopher M. Baker[†,‡] and Guy H. Grant*[,‡]

Department of Chemistry, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford, OX1 3QZ, United Kingdom, and Unilever Centre for Molecular Informatics, The University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

Received July 4, 2006

Abstract: Aromatic groups are now acknowledged to play an important role in many systems of interest. However, existing molecular mechanics methods provide a poor representation of these groups. In a previous paper, we have shown that the molecular mechanics treatment of benzene can be improved by the incorporation of an explicit representation of the aromatic π electrons. Here, we develop this concept further, developing charge-separation models for toluene, phenol, and pyridine. Monte Carlo simulations are used to parametrize the models, via the reproduction of experimental thermodynamic data, and our models are shown to outperform an existing atom-centered model. The models are then used to make predictions about the structures of the liquids at the molecular level and are tested further through their application to the modeling of gas-phase dimers and cation$-\pi$ interactions.

Introduction

Interactions involving aromatic groups are now recognized as playing a significant role in many important biomolecular processes. Aromatic−aromatic interactions are thought to be important in protein−ligand interactions,[1] DNA structure,[2] and protein structure.[3,4] Computational study of such phenomena therefore requires potential models that can accurately treat interactions involving aromatic groups. Unfortunately, conventional all-atom force fields provide a rather poor description of these interactions. The aromatic groups possess a substantial amount of charge density associated with the aromatic π system that lies above and below the plane of the ring. This charge distribution gives rise to a quadrupole moment perpendicular to the plane of the ring, which cannot be adequately reproduced using a simple atom-centered model. A variety of approaches have been attempted to remedy this problem, including central multipoles,[5] distributed multipoles,[5] and off-atom charges, as suggested by Hunter and Sanders.[6] In a recent paper,[7] the charge-separation model of Hunter and Sanders was combined with an atom-centered model to develop a new force field, based on the OPLSAA force field[8] and named OPLS-CS, for the study of liquid benzene. This new model offered an improved representation of the structure of the liquid relative to the existing all-atom model.

The work on liquid benzene demonstrated the utility of the methodology, but the ultimate aim of this research is to establish a new force field for the modeling of aromatic interactions in biological systems. For such a model to be successful, it is necessary that it be able to treat a variety of aromatic groups, with a variety of different substitution patterns. Here, the method is extended to develop new force field parameters for the benzene derivatives toluene, phenol, and pyridine and to gain new insights into their properties in the liquid phase. As a test of the transferability of these models, they will also be used to investigate the cation$-\pi$ interaction. Cation$-\pi$ interactions are believed to be of fundamental importance in many biological problems and will provide a first test of whether the OPLS-CS model is likely to be of use in the study of biological problems.

Although they have received comparatively little experimental and theoretical attention relative to that of the

* Corresponding author e-mail: ghg24@cam.ac.uk.
† Physical and Theoretical Chemistry Laboratory.
‡ Unilever Centre for Molecular Informatics.

10.1021/ct600218f CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/14/2007

archetypical aromatic molecule, benzene, the molecules toluene, phenol, and pyridine play numerous important roles in chemistry. For example, toluene and phenol have been used as analogues for the aromatic amino acids phenylethylamine[9] and tyrosine,[10,11] respectively, in studies on the role of $\pi-\pi$ interactions in proteins,[9] the role of NH$-\pi$ interactions in drug recognition,[10] and the ability of the aromatic amino acids to form different types of hydrogen bonds.[11] In addition, all three chemicals are used in the manufacture of a wide range of products including aviation fuel, paints and dyes, explosives, insecticides, and drugs.

The study of these substituted benzenes is complicated relative to that of benzene itself by the presence of the substituents, which in all cases affect the steric bulk of the molecule and also give rise to a dipole moment where previously none existed. In phenol and pyridine, the OH and N groups can also participate in hydrogen-bonding interactions, increasing the range of structural properties; it has also been shown that methyl groups can act as hydrogen-bond donors to aromatic rings,[12] which may provide a source of structural stabilization in toluene.

Toluene. As with the study of the benzene dimer,[13] there is some controversy over the exact nature of the global energy minimum for the toluene dimer. Gervasio et al.[14] performed molecular mechanics and ab initio calculations which suggested that no "T-shaped" minimum is observed but that, instead, only two stacked minima exist. The work of Chipot et al.,[9] using the AMBER force field,[15] also came to the conclusion that the stacked structure was preferred over the T-shaped, despite failing to consider any displaced structures. Simulations incorporating quadrupolar interactions have identified the T-shaped dimer as an energy minimum[16] but still concluded that stacked structures lie lower in energy.

Spectroscopic results are also ambiguous, with hole-burning experiments[17] having identified the presence of two different dimer structures. These were attributed, on the basis of Schauer and Bernstein's results,[16] to sandwich-shaped and T-shaped structures. Using optical absorption spectroscopy, Law et al.[18] also identified the presence of two structures, which they assigned, by analogy to the case of the benzene−toluene dimer, to parallel displaced and T-shaped arrangements. The most recent spectroscopic work[19] suggests that there might in fact be only one structure present, with the second signal being due to a hot band. This study also proceeded to explore larger clusters of toluene molecules and concluded that they are built from the same basic structure as the dimer, suggesting that an antiparallel stacked arrangement is the most likely.

From a theoretical viewpoint, it is apparent that so-called sandwich structures, which are disfavored for the benzene dimer,[20] should be more stable in all of the aromatic molecules considered in this study because they allow for the possibility of favorable dipole−dipole interactions which would result from antiparallel stacking. Furthermore, in the benzene dimer, the exact energetic preference is known to depend on a subtle balance between the electrostatic and dispersion interactions.[21] The larger size of toluene and phenol would be expected to increase the dispersion interaction in the stacked arrangements.[22] The most accurate theoretical calculations to date on the toluene dimer, at the CCSD(T) level of theory, indeed predict that a face-to-face arrangement of molecules is considerably more favorable in toluene than in benzene.[22,23] They also predict that binding in the T-shaped structure is slightly less favorable in toluene than in benzene, though it still remains as the energetic minimum by 0.28 kcal mol^{-1} (cf. benzene where the difference is 0.82 kcal mol^{-1}).[23]

In the liquid phase, there is almost as little consensus. There are no available experimental studies on liquid toluene, but molecular dynamics (MD)[24,25] have suggested that the structure in the liquid is largely perpendicular. One set of Monte Carlo[26] calculations concluded that liquid toluene was broadly similar to liquid benzene, "with edge-to-face and offset-stacked arrangements predominating". A second Monte Carlo study, however, concluded that stacking interactions occur between nearest neighbors[27] and that the predominant orientational distribution could be described as being "V-shaped". There was also disagreement from Fioroni and Vogt,[28] who used MD simulation with a reparameterization of the GROMOS96 force field[29] to study liquid toluene, concluding that the structure contains "a prevalence of the parallel stacking between dimers of toluene molecules".

The crystal structures of two phases of solid toluene have been determined by X-ray diffraction,[30,31] and in both cases the nearest neighbor arrangement is found to be perpendicular.

Phenol. The study of phenol has received a relatively small amount of interest compared to toluene, but high-level ab initio calculations[32] reveal the presence of a minimum energy structure for the dimer in which the arrangement of molecules is perpendicular, but with one molecule donating an OH−O bond rather than an OH$-\pi$ bond. This structure was also predicted by Sagarik and Asawakun,[33] using a potential based on the test particle model,[34] and is also in good agreement with experimental results.[35−37] A hydrogen-bonded minimum in which both molecules lie approximately in the same plane has also been predicted,[38] though it is the result of Hartree−Fock (HF) calculations which neglect the important role of dispersion in the interactions of aromatic molecules.[21] As in toluene, stacked dimers of phenol are found to be more favorable than those of benzene.[22]

The perpendicular structure also seems to have translated well into the simulation of the liquid phase. Molecular dynamics simulations[39] have found that the structure is largely perpendicular but that molecules "stack over the OH group". Work in the group of Jorgensen using Monte Carlo simulations[26] focused on compounds that are liquid at 298 K and so did not consider phenol, but in the closely related *m*-cresol, they found that a perpendicular structure is preferred but that there is competition between ring stacking and hydrogen bonding.

These results could be called in to question by the work of Thornton et al.,[40] which suggests that all-atom models are insufficient for the modeling of the interactions of phenol. In their work on the interactions between phenol and a carboxylate group, they found that a model constructed using distributed multipole analysis[41] outperformed the CHARMM[42] force field, suggesting that higher-order multipole moments

are required in order to model accurately even those aromatic compounds that possess a dipole moment.

Experimental work has been performed to determine the distribution of tyrosine residues in proteins[43] and found that, while stacked residues were preferred at short separations, as the inter-residue distances increased a T-shaped arrangement became more favorable. It is important to note here that the distribution of aromatic groups in proteins need not necessarily correspond to the minimum energy arrangement. Overstabilization of a protein structure could lead to adverse effects, and for this reason, the structure may not be that which minimizes all possible interactions. In spite of this, it is instructive to consider the biological distribution of the groups as it provides an important insight into the types of interaction that a force field must be able to successfully treat.

The crystal structure of solid phenol has also been determined by X-ray diffraction,[44] and under ambient conditions, it is characterized by "infinite hydrogen-bonded chains, where the molecules are aligned in approximately threefold helices".[45]

Pyridine. As a heteroaromatic molecule, pyridine might be expected to show somewhat different behavior to either toluene or phenol. The presence of the N atom introduces a dipole moment which might be expected to favor the parallel arrangements of the molecules, but the potential dispersion interaction will be smaller than in either toluene or phenol. Mishra and Sathyamurthy[46] have used the MP2 method to investigate the dimer and trimer of pyridine, finding that in both cases an antiparallel displaced arrangement is the most favorable, with the T-shaped structures lying comparatively high in energy. The authors also found that Hartree−Fock and density functional methods predict the dimer and trimer to be unstable, indicating the importance of accounting for the dispersion interaction in these complexes. Piacenza and Grimme[47] attempted to account for the failings of density functional theory methods by introducing an empirical dispersion correction. With this method, they also investigated the structures of the dimer and trimer of pyridine, finding again that the parallel structures are preferred over the T-shaped structure. This result was attributed to the presence of the dipole moment, which was calculated to contribute "about 1 kcal mol^{-1}" to the binding. The pyridine dimer has also been studied by Megiel et al.,[48] using the Roothaan−Hartree−Fock (RHF) method. They identified five dimer minima, all of which are hydrogen-bonded with the global minimum being a planar structure forming a cyclic hydrogen-bonding pattern. However, the fact that RHF neglects dispersion means that this result should be treated cautiously.

The same work by Megiel et al.[48] went on to consider the structure of the liquid via molecular dynamics calculations with the AMBER force field.[15] The authors found that the liquid is dominated by the formation of dimers but that the planar structure identified as the gas-phase energy minimum is now only infrequently observed with the structure instead dominated by three other dimers having hydrogen bonds between the nitrogen atom of the first molecule and either the β or γ hydrogen of the second molecule. The results of these calculations are supported by experimental data from ^{14}N nuclear magnetic resonance. Sagarik and Spohr[49] have studied liquid pyridine using a test particle model[34] in molecular dynamics and Monte Carlo simulations, also concluding that the structure is dominated by the formation of dimers. Monte Carlo simulations using the OPLS all-atom potential[8] have been used by Jorgensen and McDonald[50] to study liquid pyridine. The OPLS model was found to perform well in terms of reproducing thermodynamic data and for the structure of the liquid to predict a largely antiparallel arrangement of the molecules. A molecular dynamics simulation with a simpler potential function (that of Williams and Weller[51]) was performed by Gamba and Klein,[52] who concluded that such a potential was inadequate for the task and decided that the inclusion of the N lone pair is essential for accurate results.

Unlike the other aromatic molecules that have been considered, the solid-phase behavior of pyridine has been difficult to determine. Until 2002, only one crystal structure had been identified,[53] and this was found to be unusually complex. Anghel et al.[54] then used a model derived from distributed multipole analysis[41] to identify 17 structures that are thermodynamically more stable than the observed structure. This result prompted an experimental search for new polymorphs of pyridine, which discovered one new structure, but it was not one of the 17 predicted structures. Five of these 17 structures can be discounted on the grounds of poor mechanical properties, but the reason why the other 12 have not been observed has not been satisfactorily explained.

From the preceding discussion it is clear that all three of the molecules considered in this work have already received a good deal of attention. It is also clear that this attention has thus far failed to yield any unanimous conclusions on the structure of the liquid phase of toluene, phenol, or pyridine. The need for more accurate model potentials to address these issues, as well as many other issues related to the properties of aromatic groups, is pressing.

In this work, we will present new force field models for each of these molecules based on the charge-separation model of Hunter and Sanders.[6] We will parametrize the models by fitting to the available experimental thermodynamic data and then use the resulting models to make predictions on the structure of the liquids, comparing our results to those obtained using a conventional all-atom force field. We will then go on to consider the possibility of using these same parameters to investigate the role of aromatic groups in biological systems. Unfortunately, little experimental data is available for any of these molecules, and in its absence, the OPLSAA model will be used as a reference source where appropriate. OPLSAA is a potential developed around 10 years ago with the goal of being able to model a variety of situations including the liquid phase, gas phase, and aqueous solution. Our objectives here are different: to develop models for the pure liquids, and as such, a failure to reproduce the thermodynamic results obtained with the OPLSAA model would indicate either inappropriate parameter choice or an inappropriate potential model.

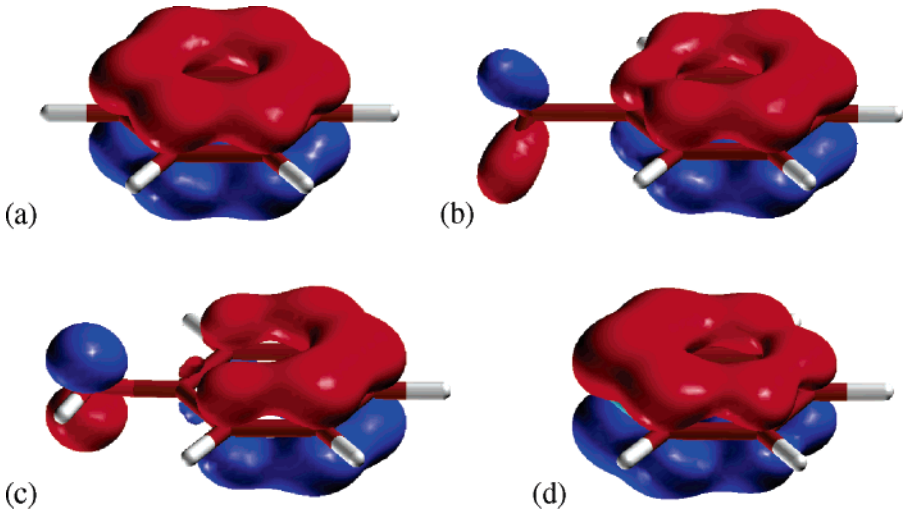


Figure 1. π orbitals for aromatic molecules, calculated at the MP2/6-31+G*//MP2/6-31+G* level of theory. In each case, the lowest-energy orbital in which the aromatic ring lies on the nodal plane is shown. (a) Benzene, (b) toluene, (c) phenol, and (d) pyridine. A contour value of 0.06 is used in all cases.

Cation−π Interactions. Of the many interactions involving aromatic groups that play a role in biology, perhaps the most important is the cation−π interaction, one of the strongest known noncovalent interactions. Energetically, it is on a par with the hydrogen bond,[55] and its role in biology is often just as significant. Cation−π interactions are known to be of importance in determining the structure of small molecules,[56] protein structure,[57,58] protein−protein binding,[59] protein−DNA binding,[60] protein−ligand binding,[61] and protein−lipid interactions.[62]

If a force field aspires to provide an accurate representation of aromatic groups, it is essential that the cation−π interaction be correctly treated. Unfortunately, existing molecular mechanics methodologies are known to provide a rather poor representation of cation−π interactions,[63] the major component of which is electrostatic.[64] As has already been discussed, the electrostatic interactions of aromatic rings are dominated by the quadrupole moment that arises from the aromatic π electrons, which atom-centered force fields fail to account for. As a first step toward the goal of modeling biologically significant aromatic groups, the ability of the OPLS-CS model to treat the cation−π interaction will be investigated.

Many previous studies have employed high levels of theory to identify the preferred binding modes and energetics of cation−π interactions.[65] Unsurprisingly, benzene has been favored, and its interactions with metal ions[66−68] and organic cations[69−72] have been considered. Toluene[73,74] and phenol[75,76] have also been studied in this context, and much is known about the nature of these interactions. The objective of this work is not to gain further insight into the interactions themselves but rather to test whether the OPLS-CS force field is capable of reproducing the results of high-level ab initio calculations which, in the absence of experimental data, can be considered as a benchmark for the calculations.

Methods

Constructing the Potential Model. Conventional all-atom force fields represent the electrostatic interactions of molecules by a series of point charges centered on the nuclear positions. Hunter and Sanders[6] suggested that the modeling of aromatic interactions could be improved by incorporating a number of non-atom-centered point charges into the model, representing the π-electron clouds of the molecule. These point charges lie directly above and below the C atoms, forming two planes that lie above and below the plane of the aromatic ring. This enables the reproduction of the quadrupole moment perpendicular to the plane of the ring, which is absent from an all-atom force field representation. In addition to our work on liquid benzene,[7] such an approach has also been successful in the modeling of a variety of other systems, including porphyrin rings[6] and the aromatic amino acids.[77] In parametrizing their model, Hunter and Sanders assumed that each C atom contributed one π electron to the π system, resulting in each of the 12 π-electron points having a charge of $-0.50e$. These points were then placed at a distance of 0.47 Å from the benzene ring so as to reproduce the component of the gas-phase quadrupole moment in the direction perpendicular to the plane of the benzene molecule.[78]

When considering liquid benzene, this model tended to overestimate the degree of perpendicular ordering in the liquid, and it was concluded that reparameterization of the model was necessary to obtain an accurate description of the liquid.[7] Initial attempts to treat the substituted aromatics using these benzene parameters yielded poor results, indicating that the transferability of these parameters was low. Inspection of the lowest-lying π orbitals of these molecules (Figure 1) suggested that this lack of transferability was not unreasonable; the exact nature of the π orbitals differs considerably throughout the series (as does the component of the quadrupole moment perpendicular to the ring, a key parameter used in fitting of the benzene model). An analysis of the π-electron density (Table 1) also supported this conclusion. Accordingly, it was necessary to develop a new set of parameters for each model.

Computational Details. As in the benzene OPLS-CS simulations,[7] the new models were parametrized by fitting

Table 1. π-Electron Density at Each of the Atomic Sites in the Aromatic Rings, Calculated at the MP2/6-31+G*//MP2/6-31+G* Level of Theory[a]

	C1/N	C2	C2′	C3	C3′	C4
benzene	0.448	0.448	0.448	0.448	0.448	0.448
toluene	0.397	0.449	0.448	0.437	0.435	0.458
phenol	0.291	0.468	0.493	0.418	0.420	0.482
pyridine	0.600	0.406	0.406	0.480	0.480	0.398

[a] The π-electron density is calculated, in each case, by considering the contribution from each of the three molecular orbitals in which the aromatic ring lies on the nodal plane.

Table 2. Calculated Values of Θ_{zz} for Benzene, Evaluated Using a Variety of Different ab Initio Techniques and Basis Sets[a]

	6-31+G**	6-31+G*	6-31G*	6-31G
MP2	−7.05	−7.12	−6.40	−6.73
B3LYP	−6.02	−6.12	−5.34	−5.46
HF	−6.85	−6.89	−6.21	−6.29

[a] All values are in units of ea_0^2.

Table 3. Calculated Values of Θ_{zz} for Pyridine, Evaluated Using a Variety of Different Ab Initio Techniques and Basis Sets

	6-31+G**	6-31+G*	6-31G*	6-31G
MP2	−4.57	−4.61	−4.18	−4.28
B3LYP	−3.64	−3.71	−3.30	−3.20
HF	−4.32	−4.34	−3.95	−3.82

[a] All values are in units of ea_0^2.

the models to reproduce experimental thermodynamic data of the molecular liquids, subject to the constraint that the model must also reproduce the experimental quadrupole moment of the molecule. For toluene and phenol, no experimental quadrupole moment was available, and instead it was necessary to use calculated values for the quadrupole moment. The values used in the parametrization process were calculated ab initio using the GDMA program[80] with the MP2/6-31+G**//MP2/6-31+G** wavefunction. The choice of wavefunction here is arbitrary, and it is necessary to consider whether such a choice is likely to have a significant effect on the calculated quadrupole moment. To assess this, quadrupole moments have been calculated using a variety of different techniques for two aromatic molecules, benzene and pyridine, in which the quadrupole moment is known experimentally (Tables 2 and 3).

For benzene, experimental values for Θ_{zz} range from -6.02 ± 0.25 to -7.42 ± 0.47 ea_0^2.[79] While the values calculated using B3LYP are generally rather low, all of the values calculated with either the MP2 or HF wavefunctions (Table 2) fall comfortably within this range, and there is nothing to suggest that calculating the quadrupole moment at the MP2/6-31+G**//MP2/6-31+G** level of theory would be inappropriate. For pyridine, fewer experimental values of Θ_{zz} are available, but they still range from -4.01 ± 0.45 to -5.13 ± 1.64 ea_0^2.[79] Here, the values calculated using HF methods tend to be slightly low, but all those calculated using MP2 are within the range of experimental values. In both pyridine and benzene, the value calculated using the MP2/6-31+G**//MP2/6-31+G** level of theory provides a reasonable approximation of the experimentally determined quadrupole moment. There is no reason to suppose that, for toluene and phenol, where experimental quadrupole moments are not available, the MP2/6-31+G**//MP2/6-31+G** calculated value of Θ_{zz} will not provide a reasonable approximation to the true value of Θ_{zz}.

In accordance with Hunter and Sanders' model, two new parameters are introduced for each model, q_π, the charge on the π points, and δ, the separation between the nucleus and the π points. Following initial difficulties in parametrizing a model for pyridine, distributed multipole analysis[41] using the program GDMA[80] revealed that the π-electron density is unevenly distributed throughout the molecule, locating preferentially at the nitrogen atom. For this reason, in the pyridine model, an additional parameter, $q_{\pi N}$, was introduced, being the charge on the N π point, and defined such that $q_{\pi N} = 1.75 q_\pi$. In all cases, the bond lengths and angles used are taken directly from the OPLSAA force field.[8]

Because OPLS is an effective potential, interactions that are not explicitly included in the potential will have been "mixed-in" to the model. This means that simply adding charge points to reproduce the quadrupole moment is not an acceptable solution—it would almost certainly have adverse effects on the properties that have been "mixed-in" to the potential. Accordingly, it was deemed necessary to re-examine all of the nonbonding parameters present within the OPLS model. The parametrization scheme followed for toluene and phenol was identical to that used in our previous work on benzene.[7] To the bare molecular frame are added a series of charge and van der Waals parameters. The charge parameters were subject to the constraints that the individual molecules must be charge-neutral, and that $q_\pi = -q_H$, where q_π and q_H are the charges on the π-electron points and hydrogen atoms, respectively. A better fit to the experimental data could probably be obtained by relaxing this constraint and, indeed, by allowing for charge redistribution within the ring. However, such an approach would greatly increase the risk of overfitting,[81] especially given the limited amount of experimental data available, and for this reason was avoided. q_π and δ were then varied systematically along with the van der Waals parameters, σ and ϵ, until the models gave the minimum deviation from experimental thermodynamic results.

Pyridine proved more complicated to parametrize. OPLS assigns different charge values to each of the three different types of C atoms in pyridine, and it was felt necessary to keep this scheme in place when parametrizing the OPLS-CS model. The total charge on each CH unit was determined using the MP2/6-31+G*//MP2/6-31+G* method, and this value was then kept constant throughout the parametrization process. Again, the charge parameters were subject to the constraints that the individual molecules must be charge-neutral, and that $q_\pi = -q_H$, where q_π and q_H are the charges on the π-electron points and hydrogen atoms, respectively. q_π and δ were then varied systematically along with the van

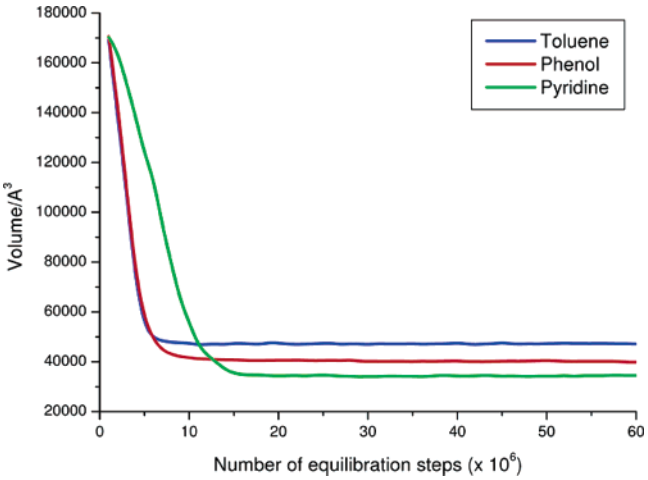


Figure 2. Plot monitoring the volume of the OPLS-CS parametrization simulations over the course of the equilibration period.

der Waals parameters, σ and ϵ, until the models gave the minimum deviation from experimental thermodynamic results.

As noted above, the development of the OPLS-CS models has required wholesale reparameterization of the original OPLS models. If the OPLS-CS models do show an improvement over the OPLS models, the question that then arises is: is the improvement due to the additional point charges or the reparameterization? This is in fact part of a larger issue of how sensitive the obtained results are to the choice of parameters, and whether this is more or less important than the inclusion of the additional point charges. For example, what would happen if the OPLS model compounds had larger bond dipoles? To investigate this issue further, an additional model has been parametrized for phenol. In this model, termed OPLS-Q, the atom-centered representation is retained, but the atomic charge on the aromatic C and H atoms is increased, from $\pm$ 0.115e in the OPLS model to $\pm$ 0.200e in the OPLS-Q model. With these charge values in place, the values of the nonbonding parameters σ and ϵ were then varied until the best set of thermodynamic results were obtained. From the results of the OPLS-CS calculations, it was noted that the calculated values of ΔH_{vap} were somewhat large compared to the experimental values. To investigate the significance of this effect, the OPLS-Q model was parametrized to reproduce the OPLS-CS thermodynamics values rather than the experimental thermodynamic values.

For the purpose of parametrization, a series of Monte Carlo simulations was performed for each molecule. Each simulation included 267 molecules in the constant-temperature, constant pressure (NPT) ensemble, with $P = 1$ atm. For toluene and pyridine, a temperature $T = 298$ K was used, and for phenol, which is solid at 298 K, $T = 333$ K was used. Each simulation consisted of 6.0×10^7 steps of equilibration, sufficient for convergence to have been achieved (Figure 2), followed by 6.0×10^7 steps of averaging. The new parameters obtained from these simulations are given in Table 4. In the simulations of liquid benzene, the molecule was held rigid, and in this case, the rings were also fixed, but the methyl group of toluene and the hydroxyl group of phenol were allowed to rotate. The rotational profile for each of these groups was, however, affected by the addition of the π-electron points, and it was necessary to reparameterize these two dihedral angle terms.

Table 4. New Parameters Obtained for OPLS-CS Simulations and the Corresponding Parameters Used in the OPLS Simulations[a]

molecule	parameter	OPLS-CS	OPLS
toluene	δ/Å	1.600	n/a
	q_{CA}/e	0.0455	−0.1150
	q_{HA}/e	0.0455	0.1150
	q_{π}/e	−0.0455	n/a
	q_{CT}/e	−0.1345	−0.0650
	q_{HC}/e	0.0600	0.0600
	σ_{CA}/Å	3.5750	3.5500
	σ_{HA}/Å	2.4370	2.4200
	σ_{CT}/Å	3.5000	3.5000
	σ_{HC}/Å	2.5000	2.5000
	ϵ_{CA}/kcal mol^{-1}	0.0630	0.0700
	ϵ_{HA}/kcal mol^{-1}	0.0270	0.0300
	ϵ_{CT}/kcal mol^{-1}	0.0594	0.0660
	ϵ_{HC}/kcal mol^{-1}	0.0270	0.0270
	$\psi_{CACACTHC}$/deg	0.000, −0.435, 0.000, −0.085	0.000, 0.000, 0.000, 0.000
Phenol	δ/Å	1.100	n/a
	q_{C1}/e	0.3480	0.1500
	q_{CA}/e	0.0990	−0.1150
	q_{HA}/e	0.0990	0.1150
	q_{π}/e	−0.0990	n/a
	q_{OH}/e	−0.5850	−0.5850
	q_{HO}/e	0.4350	0.4350
	σ_{C1}/Å	3.5500	3.5500
	σ_{CA}/Å	3.6200	3.5500
	σ_{HA}/Å	2.4677	2.4200
	σ_{OH}/Å	3.0700	3.0700
	σ_{OH}/Å	0.0000	0.0000
	ϵ_{C1}/kcal mol^{-1}	0.0700	0.0700
	ϵ_{CA}/kcal mol^{-1}	0.0700	0.0700
	ϵ_{HA}/kcal mol^{-1}	0.0300	0.0300
	ϵ_{OH}/kcal mol^{-1}	0.1700	0.1700
	ϵ_{HO}/kcal mol^{-1}	0.0000	0.0000
	$\psi_{CACAOHHO}$/deg	0.000, 1.845, 0.000, 0.000	0.000, 1.682, 0.000, 0.000
pyridine	δ/Å	0.600	n/a
	q_N/e	0.0990	−0.6780
	q_{C1}/e	0.1170	0.4730
	q_{C2}/e	0.2750	−0.4470
	q_{C3}/e	−0.0330	0.2270
	q_{H1}/e	0.1000	0.0120
	q_{H2}/e	0.1000	0.1550
	q_{H3}/e	0.1000	0.0650
	q_{π}/e	−0.1000	n/a
	$q_{\pi N}/e$	−0.1750	n/a
	σ_N/Å	3.1585	3.2500
	σ_{CA}/Å	3.4500	3.5500
	σ_{HA}/Å	2.3518	2.4200
	ϵ_N/kcal mol^{-1}	0.1637	0.1700
	ϵ_{CA}/kcal mol^{-1}	0.0674	0.0700
	ϵ_{HA}/kcal mol^{-1}	0.0289	0.0300

[a] Dihedral-angle parameters show the four Fourier coefficients used for energy evaluation. A full list of parameters is given in the Supporting Information.

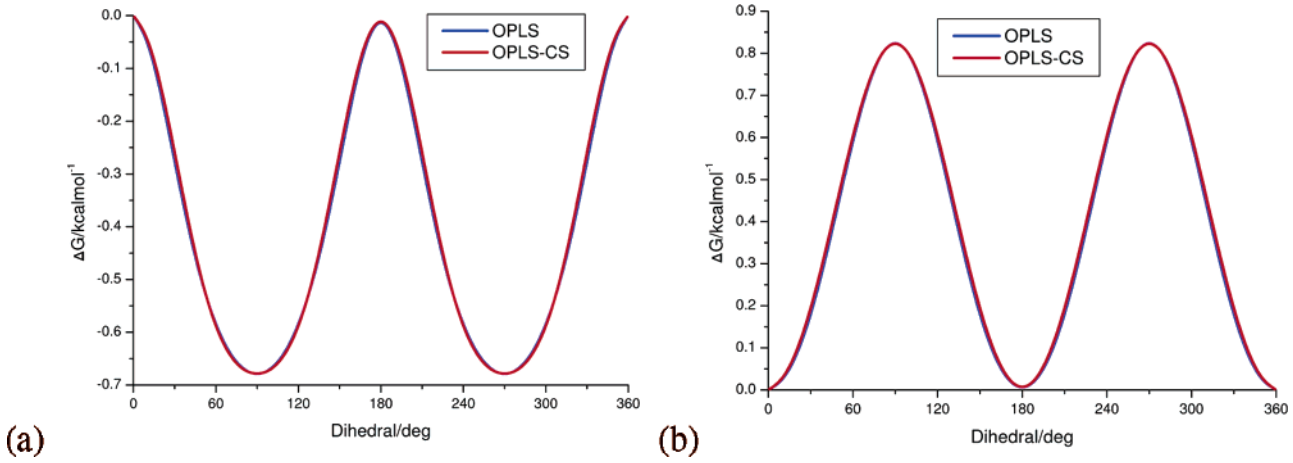


Figure 3. Fittings to OPLS rotational profiles. (a) Toluene and (b) phenol.

This was achieved by fitting to the gas-phase rotational profiles generated using the OPLS potential: for each dihedral angle, the Fourier coefficients were modified until they gave the best fit to the original OPLS-calculated rotational profile. This fitting was performed in the gas phase using simulations at $T = 298$ K with 4×10^5 steps of equilibration and 8×10^5 steps of averaging. The resulting parameters are given in Table 2, and the free-energy profiles are shown in Figure 3.

With the necessary parameters in place, two Monte Carlo simulations were performed for each of toluene and pyridine, and three were performed for phenol. In all cases, the first simulation treated the aromatic molecules using the OPLS all-atom potential[8] (denoted OPLS), the second the OPLS all-atom potential modified in such a way as to incorporate the π-electron points of the charge-separation model (denoted OPLS-CS). In phenol, the third simulation employed the OPLS-Q potential described above. All simulations were performed using BOSS version 4.2;[82] in the OPLS simulations, the standard OPLSAA parameters were used; in the OPLS-CS simulations, the OPLSAA parameters were modified so as to incorporate the π parameters described above. In all cases, the molecules were regarded as being rigid with the exception of the C−O and C−Me bonds in phenol and toluene, respectively, which were allowed to rotate subject to the torsional parameters of the OPLSAA force field in the OPLS and OPLS-Q simulations and the modified OPLSAA parameters in the OPLS-CS simulation. A system consisting of 267 molecules was used in simulations that were run in the NPT ensemble with $P = 1.0$ atm. For toluene and pyridine, a value of $T = 298$ K was used, but for phenol, in which the melting point occurs at 313 K, a value of $T = 333$ K was used. The simulations were begun from a configuration in which all of the molecules were arranged in a parallel fashion, and in all simulations, 4×10^6 equilibration steps were performed followed by 2.5×10^8 steps of averaging.

All molecular graphics were produced using the PyMOL program,[83] except those in Figure 1, which were produced using MOLDEN,[84] and Figures 7b, 10b, and 14b, which were produced using VMD.[85] Where values of the quadrupole moments are given for the force field models, they were evaluated using the BOSS program.[82]

Gas-Phase Dimers. As a first test of the transferability of the OPLS-CS models derived from the liquid-phase calculations, their ability to reproduce the binding energies of the gas-phase dimers of toluene, phenol, and pyridine has been investigated. Dimer structures for each of the molecules were first identified using ab initio theory. Dimer structures obtained from the literature, as discussed above, were constructed and minimized at the MP2/6-31+G* level of theory. The energies of the resulting complexes were then evaluated using the same level of theory with basis set superposition error[86] corrected using the counterpoise method.[87] In this way, a total of 15 minimum energy structures were identified: four for toluene, three for phenol, and eight for pyridine. The energy of each of the dimers was then evaluated using the OPLS and OPLS-CS models (and, in the case of phenol, the OPLS-Q model). Comparison of the energies obtained from the force field and ab initio calculations can give us an insight into whether the charge-separation models are providing a good reproduction of the intermolecular interactions between the molecules.

Modeling Cation−π Interactions. To examine the cation−π interaction, the phenol OPLS-CS model was used to construct an OPLS-CS model of *p*-methylphenol, an analogue of tyrosine (henceforth referred to as Tyr). The toluene OPLS-CS model previously obtained was taken as an analogue of phenylalanine (henceforth referred to as Phe). These models were then used to investigate the interactions of the two molecules with the archetypical organic cation—ammonium. For comparison, equivalent models for Phe and Tyr were constructed using the OPLS models of toluene and phenol. As a reference, ab initio calculations were performed at the MP2/6-311+G** level of theory, with the basis set superposition error[86] corrected using the counterpoise method.[87]

Two sets of calculations were performed. First, an ammonium molecule was placed in a monodentate fashion above the center of the aromatic ring lying in the *xy* plane. The ammonium molecule was then moved in increments of 0.1 Å along the *z* axis, with the energy evaluated via a single point calculation at each increment. This procedure was performed using the ab initio method as well as the OPLS-CS and OPLS models. Second, the separation of the molecules in the *z* direction was fixed at 3 Å while the ammonium molecule was scanned, again with a monodentate

Table 5. Thermodynamic Properties of Liquid Toluene

	OPLS	OPLS-CS	experiment[26]
θ_{zz}/ea$_o^2$	0.45	−6.34	−6.64[a]
dipole/ea$_o$	0.20	0.22	0.14
density/g cm^{-3}	0.857	0.863	0.862
ΔH_{vap}/kcal mol^{-1}	8.89	9.09	9.08
molecular volume/Å^3	178.6	177.3	177.4

[a] Experimental quadrupole moment values for phenol and toluene are not available. The values used in the parametrization process were calculated ab initio using the GDMA program[80] with the MP2/6-31+G**//MP2/6-31+G** wavefunction.

Table 6. Thermodynamic Properties of Liquid Phenol

	OPLS	OPLS-Q	OPLS-CS	experiment[39]
θ_{zz}/ea$_o^2$	0.00	0.00	−6.90	−6.93[a]
dipole/ea$_o$	0.91	0.99	0.79	0.57
density/g cm^{-3}	1.03	1.04	1.04	1.04
ΔH_{vap}/kcal mol^{-1}	12.81	1.49	13.44	13.19

[a] Experimental quadrupole moment values for phenol and toluene are not available. The values used in the parametrization process were calculated ab initio using the GDMA program[80] with the MP2/6-31+G**//MP2/6-31+G** wavefunction.

Table 7. Thermodynamic Properties of Liquid Pyridine

	OPLS	OPLS-CS	experiment[50]
θ_{zz}/ea$_o^2$	0.00	−4.61	−4.61
dipole/ea$_o$	0.94	0.89	0.86
density/g cm^{-3}	0.97	0.97	0.98
ΔH_{vap}/kcal mol^{-1}	9.58	9.66	9.61

geometry, in the x and y directions over the aromatic molecule. The energy was evaluated at 0.2 Å increments in both directions with the same calculations performed using the ab initio, OPLS, and OPLS-CS models. All ab initio calculations were performed using the program Gaussian 98.[88]

Results and Discussion

Aromatic Liquids. From the initial simulations performed, a set of parameters was determined as the "best" OPLS-CS model for each of the aromatic molecules (full details of the parameters are given in the Supporting Information). The comparison with experimental thermodynamic data is shown in Tables 5−7.

Although it should be noted that the model has been parametrized well within experimental uncertainty, the reproduction of the experimental thermodynamic data by the OPLS-CS models is generally good. It offers an improvement over OPLS in all areas, with a significant improvement in the reproduction of the quadrupole moment perpendicular to the plane of the ring in all cases. It is possible to learn something about the behavior of these liquids by consideration of their thermodynamic properties. At constant pressure, the entropy change of vaporization of a system is related to its enthalpy of vaporization via eq 1.

$$\Delta S_{vap} = \frac{\Delta H_{vap}}{T_b} \quad (1)$$

Trouton's rule[89] then tells us that molecular liquids have a standard entropy of vaporization of approximately 85 J K^{-1} mol^{-1} (20 cal K^{-1} mol^{-1}). This observation is based on the fact that the evaporation of any liquid into the gas phase will generate a similar amount of disorder. As a first measure of the level of structuring in the liquids in this study, we can consider the relevant ΔS_{vap} values, which are (calculated via eq 1) 23.68, 29.05, and 24.75 cal K^{-1} mol^{-1} for toluene, phenol, and pyridine, respectively.

It follows that phenol, with the largest deviation from Trouton's rule, is the most structured of the liquids, and toluene the least. This result is unsurprising. Phenol, which contains both hydrogen-bond donors and acceptors, would be expected to be highly hydrogen-bonded. Pyridine will also form hydrogen bonds, but they will be weaker because the only possible donors are the C−H bonds. Toluene, with no scope for hydrogen bonding, will be forced to rely on weaker interactions. Although this is much as would be expected intuitively, it is encouraging. Because the OPLS-CS model has performed better than OPLS in reproducing the experimental ΔH_{vap} values, and these values are an indirect measure of structure in the liquid, it suggests that OPLS-CS should give a better account of liquid structure than does OPLS.

It is also possible to learn something about the strength of the interactions between the models. By comparing the ΔH_{vap} value determined experimentally to the ΔH_{vap} value predicted by Trouton's rule (via eq 2), we can make an estimate of the strength of these ordering interactions.

$$\Delta\Delta H = \Delta H_{vap}^{exp} - \Delta H_{vap}^{Trouton} \quad (2)$$

In benzene, $\Delta\Delta H = 0.92$ kcal mol^{-1}, suggesting that the intermolecular interactions are small. In toluene, $\Delta\Delta H = 1.30$ kcal mol^{-1}; in phenol, $\Delta\Delta H = 3.98$ kcal mol^{-1}, and in pyridine, $\Delta\Delta H = 1.73$ kcal mol^{-1}. Again, the relative size of these interactions is much as would be expected, and again the fact that OPLS-CS has reproduced the ΔH_{vap} values more accurately than OPLS suggests that the model should be able to give a better description of the intermolecular interactions within the liquid.

The OPLS-CS model also outperforms the OPLS model in reproducing the density of all three of the liquids. This can again be seen as an encouraging sign since the liquid density provides a fundamental insight into the packing of the molecules in the liquid and, hence, the structure of the liquid. It follows that, with a better reproduction of the liquid densities, we would expect the OPLS-CS model to provide a better representation of the liquid structures.

Structural Information. Toluene. An analysis of $g_{CC}(r)$ (Figure 4a) reveals that the two simulations performed for liquid toluene, using the OPLS and OPLS-CS potentials, give almost identical results. Two small peaks, at around 5.2 Å and 6.2 Å, are identified in each case, with a slight shoulder in the distribution at around 4.3 Å. In the OPLS-CS simulations of liquid benzene,[7] this shoulder appeared as a clear peak and was taken as evidence of a largely perpendicular structure in the liquid. The absence of such a peak in this case indicates that the perpendicular arrangement is not nearly so dominant in toluene. An additional difference between $g_{CC}(r)$ for toluene and benzene is the presence of a small shoulder in the distribution at around 7.5 Å, which is

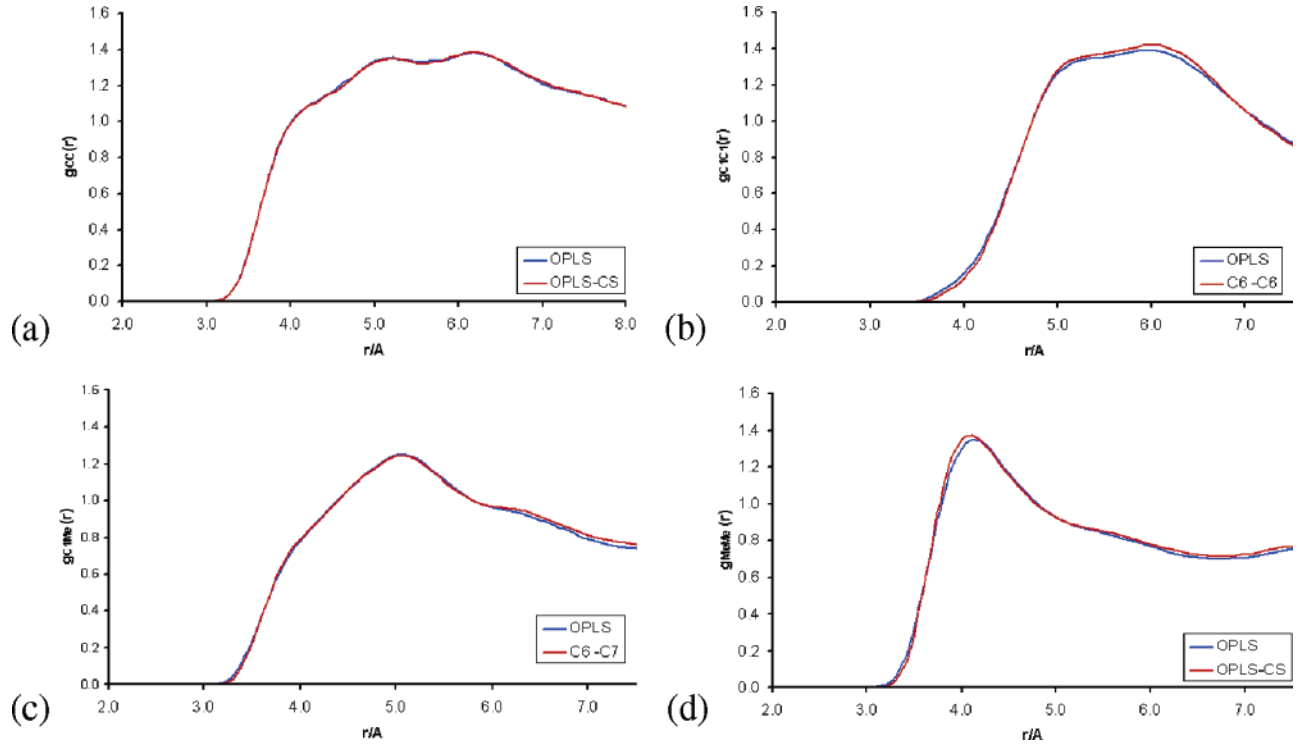


Figure 4. Radial distribution functions for liquid toluene. (a) $g_{CC}(r)$, (b) $g_{C1C1}(r)$, (c) $g_{C1Me}(r)$, and (d) $g_{MeMe}(r)$.

not observed in liquid benzene. Kim and Lee[25] also observed this feature and attributed it to the presence of the methyl group attached to the phenyl ring in a perpendicular arrangement. It is worth noting that the simulation of Kim and Lee in the NPT ensemble[25] heavily overestimated the density of the liquid, and as a result, all four features in $g_{CC}(r)$ were considerably larger than in the current results. A simulation by the same group in the canonical ensemble[24] gave results in much better agreement with our results.

In addition to $g_{CC}(r)$, several other radial distribution functions have been considered (Figure 4). In general, the rather broad and unstructured peaks are similar to those that have been previously observed.[26] The only exception to this rule is in $g_{C1C1}(r)$ (Figure 4b). Previous simulations using the OPLS potential have identified this as a single, broad peak,[26] whereas in the OPLS-CS simulation results, a slight structure is obtained in the form of a splitting into two peaks, the first at around 5.3 Å with a second, larger peak at 6.1 Å. These two distances fit well with two of the minimum energy structures of the toluene dimer (Figure 5), a stacked structure stabilized by a methyl$-\pi$ hydrogen bond[12] and a T-shaped structure. That the second peak, corresponding to a T-shaped structure, is more prominent in the OPLS-CS simulation would be expected; the inclusion of the quadrupole moment arising from the π electrons would be expected to favor the T-shaped orientations. This effect appears to be the only real difference between the simulations performed using the OPLS and OPLS-CS potentials.

The slight preference for the T-shaped structure shown in $g_{C1C1}(r)$ confirms what has been learned from $g_{CC}(r)$: in liquid toluene, a perpendicular arrangement of the molecules is present, but not to such an extent as in benzene, and both parallel displaced and perpendicular structures are well-represented in the liquid phase. This idea is confirmed by

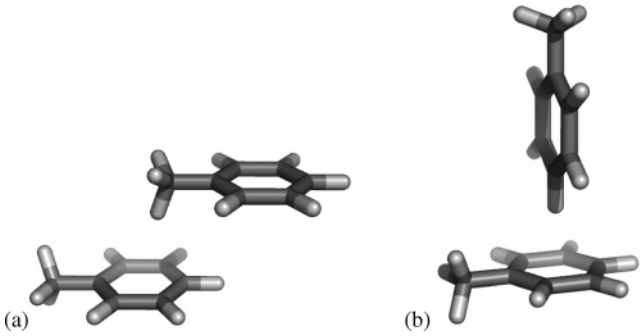


Figure 5. Minimum-energy structures for the toluene dimer. (a) $R_{C1C1} = 5.3$ Å and (b) $R_{C1C1} = 6.1$ Å.

an analysis of the angular distribution function $g(r,\theta)$ (Figure 6), which considers both the relative orientations of the molecules [measured in terms of the angle between the normals to the rings (Figure 6a)] and their separation. The first peak in $g(r,\theta)$ corresponds to the first solvation shell of liquid toluene, and it is clear that both the OPLS and OPLS-CS simulations predict very similar structures; in both cases, the parallel and perpendicular arrangements are approximately equally well-populated.

As has been noted, the molecules in liquid benzene showed a clear preference for orienting in a perpendicular arrangement, while those in liquid toluene show no such preference. This difference must be attributable to the presence of the methyl group in toluene. The methyl group introduces a dipole moment into the molecules, which would be expected to favor parallel (head-to-tail) orientations. In addition, the presence of the methyl group stabilizes parallel displaced interactions such as that shown in Figure 5a by stacking over the center of the ring in an energetically favorable arrangement. The lack of preference for any one orientation can also be seen in the structure of the first solvation shell of a

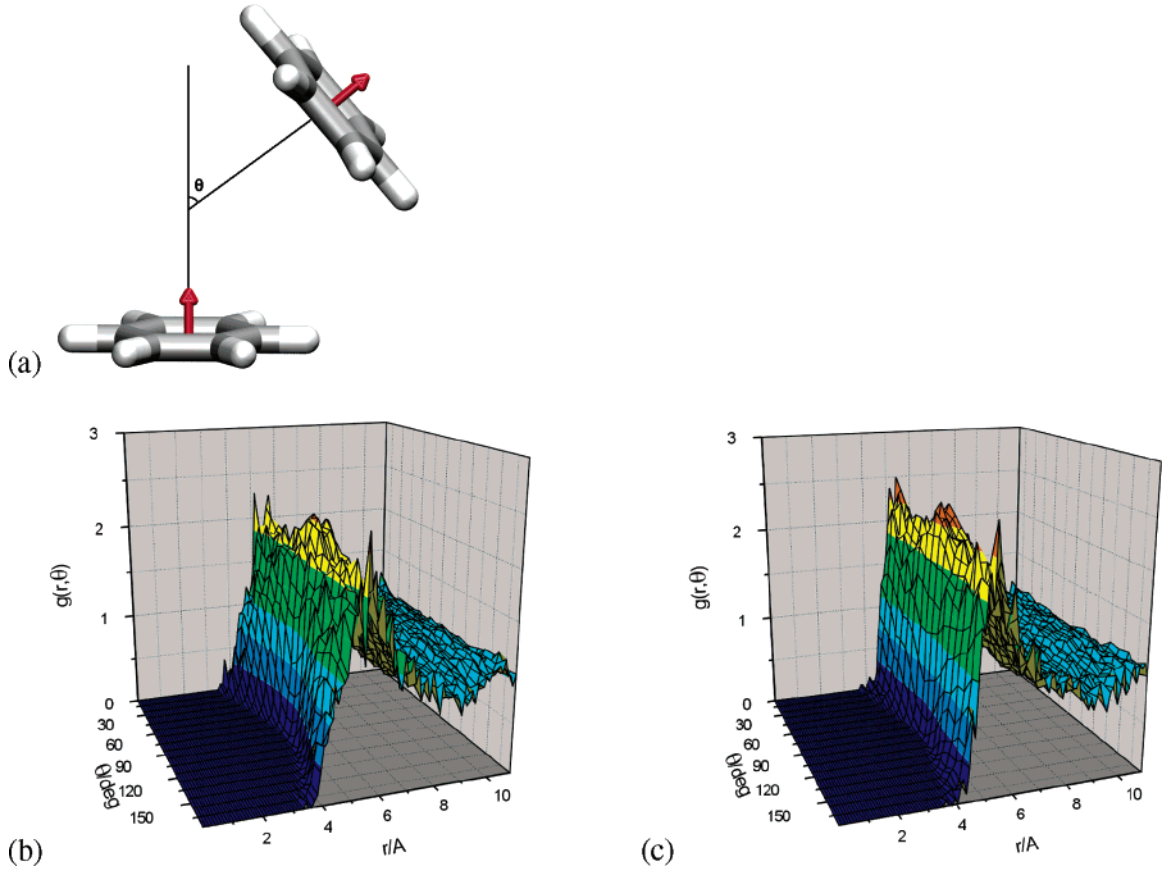


Figure 6. Angular distribution functions for liquid toluene. (a) Definition of θ. (b) $g(r,\theta)$ calculated using OPLS. (c) $g(r,\theta)$ calculated using OPLS-CS.

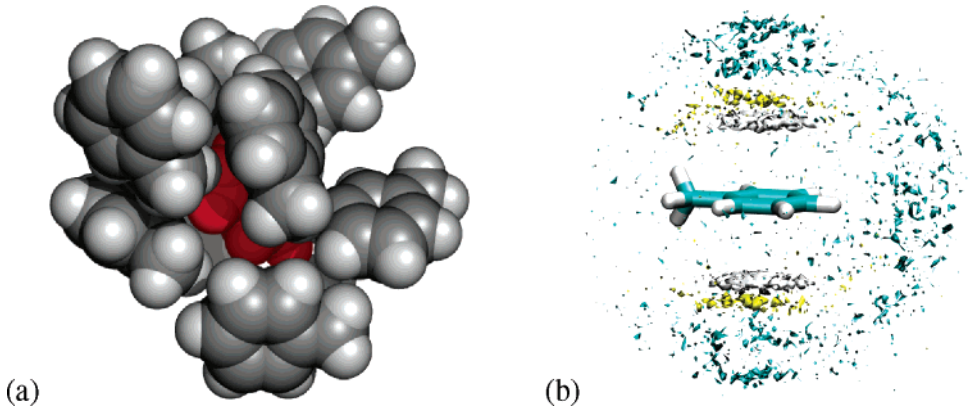


Figure 7. Distribution of molecules in liquid toluene. (a) The first solvation shell around a toluene molecule (in red) taken from the OPLS-CS simulation of liquid toluene. (b) The average atomic density around toluene molecules in the OPLS-CS simulation (cyan = aromatic carbon; yellow = methyl carbon; gray = hydrogen).

toluene molecule (Figure 7a). Examining the average atomic density around toluene molecules (Figure 7b) also leads to the same conclusion. The only real preference we see is for the methyl group to locate above the aromatic ring. Such an arrangement would be expected in either the T-shaped or parallel minima.

Phenol. The use of the OPLS-CS potential had only a small effect on the predicted properties of liquid toluene, but for phenol, its effect is far larger. In the OPLS simulations, $g_{CC}(r)$ appears as a single broad peak, with no discernible structure. In the OPLS-CS simulations, however, the formation of three distinct features within that single peak is observed. Differences between the force fields are even more apparent in the individual $g(r)$ (Figure 8). All of the $g(r)$'s show approximately the same form in both the OPLS and OPLS-CS simulations, but in all cases, the OPLS-CS calculated $g(r)$'s contain smaller peaks. The sizes of these peaks, which focus on the interactions involving the phenol hydroxyl group, can be considered as a measure of the hydrogen bonding in the liquid, and it follows that in the OPLS-CS model it is predicted that less hydrogen bonding occurs. Evidence for hydrogen bonding also comes from the C(O)−H(O) distribution (Figure 8c), which shows the double-peaked structure that is characteristics of alcohols.[90]

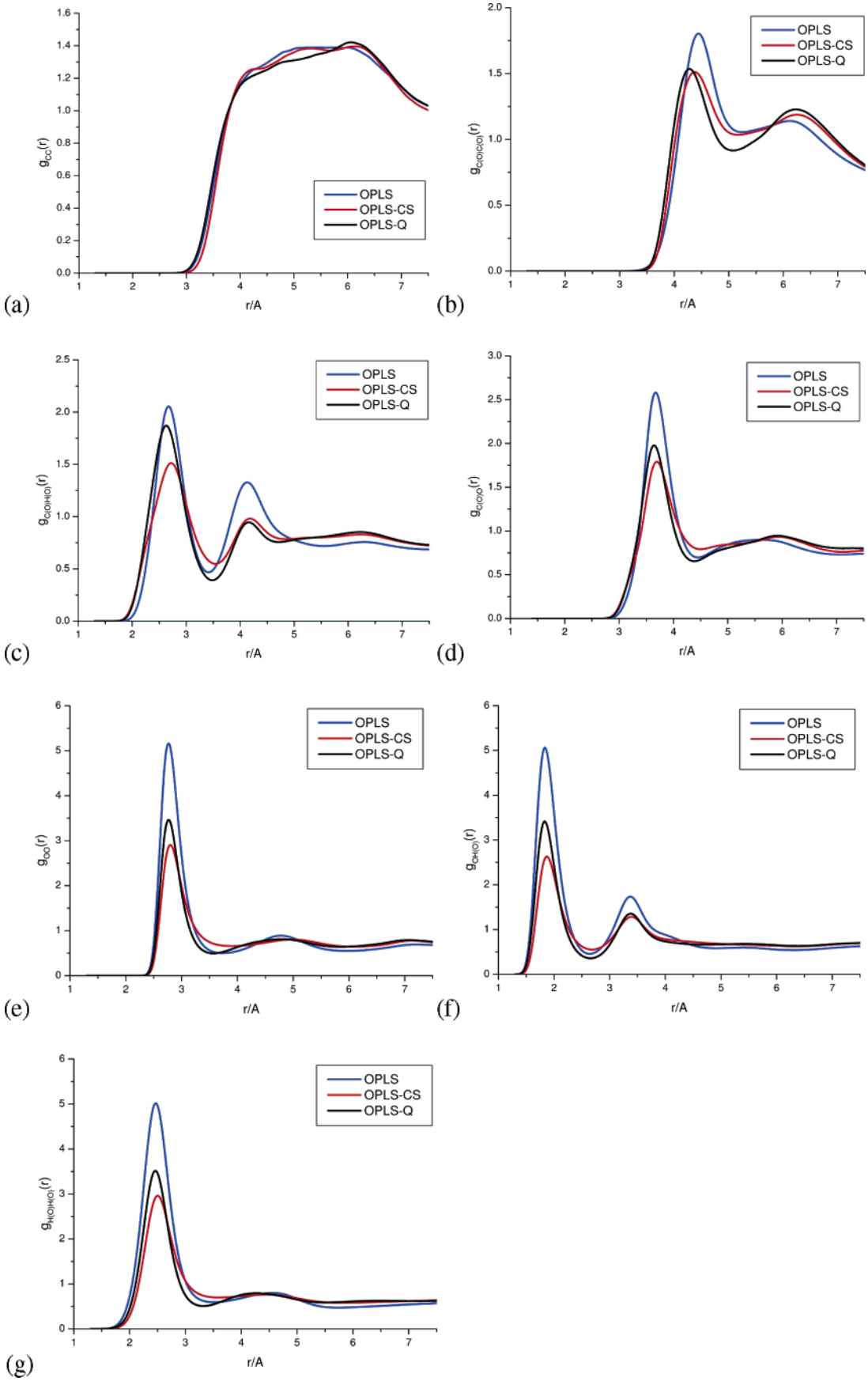


Figure 8. Radial distribution functions for liquid phenol. (a) $g_{CC}(r)$, (b) $g_{C(O)C(O)}(r)$, (c) $g_{C(O)H(O)}(r)$, (d) $g_{C(O)O}(r)$, (e) $g_{OO}(r)$, (f) $g_{OH(O)}(r)$, and (g) $g_{H(O)H(O)}(r)$.

By calculating the areas under the peaks in $g(r)$, we can make an estimate of the average number of hydrogen bonds formed per molecule. Integration of the first peak in $g_{OO}(r)$ to the minimum at 3.7 Å gives a value of 2.6 in the OPLS

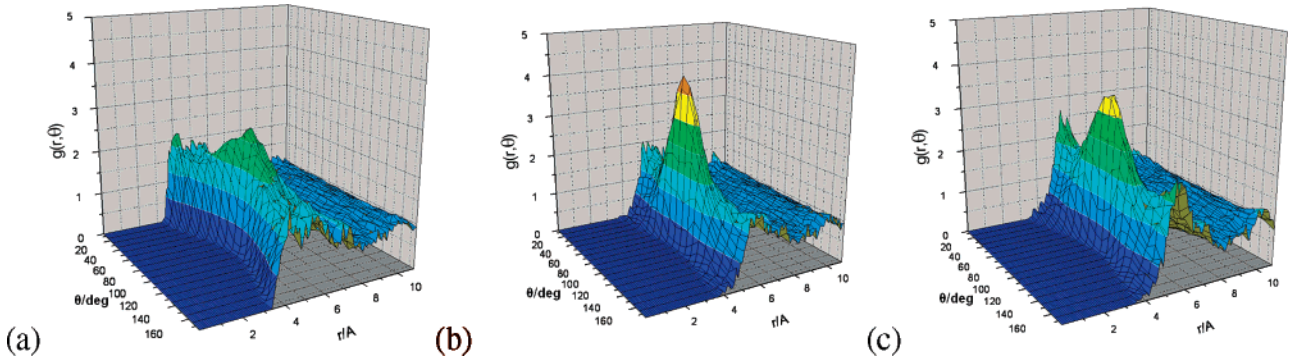


Figure 9. Angular distribution functions for liquid phenol. Calculated using (a) OPLS, (b) OPLS-Q, and (c) OPLS-CS.

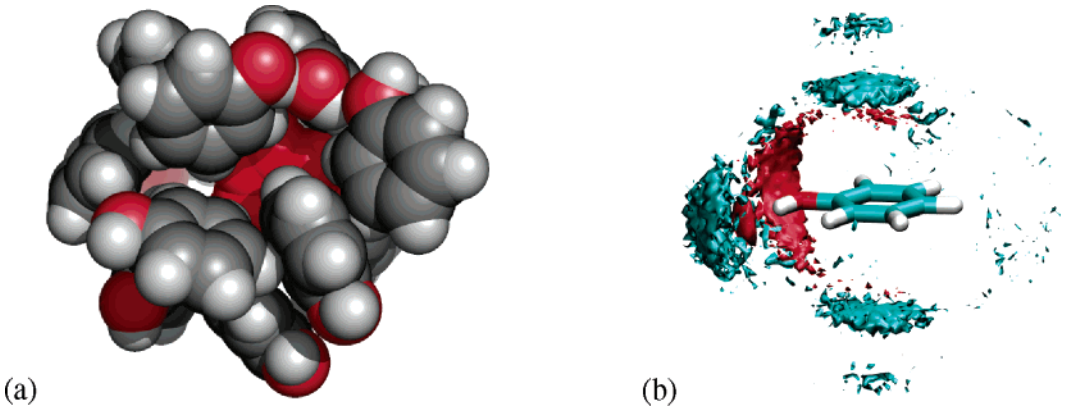


Figure 10. Distribution of molecules in liquid phenol. (a) The first solvation shell around a phenol molecule (in red) taken from the OPLS-CS simulation of liquid phenol. (b) The average atomic density around phenol molecules in the OPLS-CS simulation (cyan = carbon; red = oxygen).

simulations and 1.8 in the OPLS-CS simulations. The OPLS simulation is, therefore, predicting that a molecule forms, on average, 2.5 hydrogen bonds, with the OPLS-CS simulation predicting 1.8. This value is in close agreement with the value of 1.6 predicted by Jorgensen et al.[26] in their simulations of liquid *m*-cresol.

For phenol, a third potential model, OPLS-Q, has also been considered. It is a reparameterization of OPLS in which the OPLS-CS density and ΔH_{vap} are reproduced and the C−H bond dipole is increased. The radial distribution functions obtained using OPLS-Q are shown in Figure 8. In general, the OPLS-Q model seems to be somewhere in between the OPLS-CS and OPLS models, which suggests that a simple reparameterization of the OPLS model might be able to yield a model that is comparable to the OPLS-CS model. Such a conclusion would indicate that the charge-separation strategy might be an overcomplicated method for the development of force field parameters in general, and this idea will be investigated in more detail in the relevant sections of this paper.

Further insight into the structure of the liquid comes from consideration of the angular distribution function $g(r,\theta)$ (Figure 9). Now, we can see that the OPLS-CS simulation predicts a far more perpendicular structure than the OPLS simulation. What is more, the OPLS-CS simulation suggests that the ordering in liquid phenol is far more perpendicular than in liquid toluene, and almost as perpendicular as liquid benzene.[7] Although we have no experimental structural data for liquid phenol, this result seems to be in good agreement with experimental and theoretical studies on the phenol dimer where the minimum energy structure is found to be T-shaped, but with hydrogen bonding occurring between the hydroxyl groups. Examples of such structures can clearly be seen in the first solvation shell of a phenol molecule taken from the OPLS-CS simulation (Figure 10a). Moreover, by considering the average phenol atomic density around phenol molecules in the simulation (Figure 10b), it can clearly be seen that the preferred position for an oxygen atom is adjacent to the oxygen atom of another phenol molecule, either donating or accepting a hydrogen bond. Comparatively little oxygen density is seen above or below the ring, indicating that the formation of OH−π hydrogen bonds is rare, although the carbon density reveals that T-shaped stacking over the ring center does still occur. The angular distribution function obtained from the OPLS-Q simulation (Figure 9b) is surprising. The radial distribution functions suggested that OPLS-Q was somehow intermediate between the OPLS and OPLS-CS simulations. The angular distribution obtained using OPLS-Q favors the perpendicular arrangement even more strongly than the OPLS-CS model does. It is clear from this that the results of the calculations are highly dependent on the choice of parameters. They, along with the type of model to be used, must be carefully considered if a useful potential is to be obtained.

Pyridine. A first measure of the structure of liquid pyridine comes from $g_{NN}(r)$ (Figure 11), which shows two distinct peaks in both the OPLS and OPLS-CS simulations. Jorgensen and McDonald have previously studied liquid pyridine using

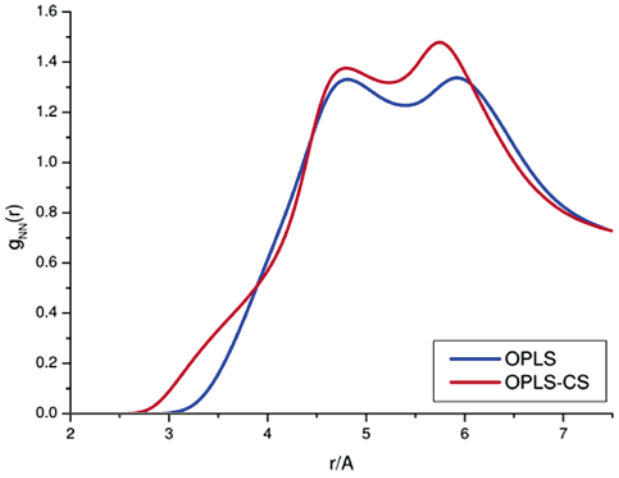


Figure 11. Radial distribution function for liquid pyridine $g_{NN}(r)$.

the OPLS force field[50] and obtained the same $g_{NN}(r)$. They have attributed the two peaks to two different antiparallel arrangements (which correspond to the OPLS calculated dimer minimum), one directly stacked and the other parallelly displaced. However, an analysis of $g(r,\theta)$ (Figure 12) suggests that even with the OPLS potential a perpendicular arrangement of the molecules is slightly favored. In the OPLS-CS simulation, the preference for perpendicular arrangements is even more pronounced.

Inspection of the structures produced in the OPLS-CS simulation suggests that we are not in fact seeing the formation of antiparallel stacked structures, but rather the formation of hydrogen-bonded structures, in which the molecules then adopt an orthogonal arrangement. The peak in $g_{NN}(r)$ at 4.9 Å results from the formation of a structure in which one pyridine molecule donates a hydrogen bond through the hydrogen in the ortho position, with the peak at 5.9 Å arising from a structure in which a pyridine molecule donates a hydrogen bond through the hydrogen in the meta position (Figure 13). The preference for a perpendicular arrangement can also be seen in the first solvation shell of a molecule taken from the OPLS-CS simulation (Figure 14a). Examining the average atomic density around molecules in liquid pyridine (Figure 14b) supports the view that the interactions in the liquid are dominated by the formation of CH−N hydrogen bonds. Nitrogen atom density occurs only in the regions around the hydrogen atoms. Carbon density, however, is seen additionally above the aromatic ring, arising from dimer structures that are stabilized by the formation of a CH−π interaction.

Computational Cost. An important consideration when running any molecular simulation is the relative cost, in terms of computer time, of the calculation. Typically, the need to achieve an acceptable level of accuracy within an acceptable time frame is the determining factor in the choice of the force field model used. With the addition of extra point charges, one would expect the OPLS-CS simulations to take longer than the OPLS simulations, and this is indeed found to be the case. Typically, running on equivalent processors, the OPLS-CS calculations take around 3 times as long as the analogous OPLS calculations. While this represents a significant increase in computational resources, it is still considerably faster than simulations based on distributed multipole models, which have been found to take around 8 times as long as their atom-centered equivalents.[91]

Gas-Phase Dimers. The gas-phase dimer minimum energy structures identified via ab initio calculation are shown in Figure 15, and the corresponding interaction energies are given in Tables 8−10.

For toluene, the OPLS model does well in reproducing the energetic ordering of the dimers, even if the energetic separation between them is underestimated. The OPLS-CS model performs similarly, underestimating the relative stability of tol_b relative to the T-shaped structures tol_c and tol_d but correctly identifying the minimum energy structure tol_a. This information is consistent with that obtained from the simulations on liquid toluene, in which both the OPLS and OPLS-CS models are found to predict very similar structures, but with the OPLS-CS model showing a slight preference for a perpendicular arrangement of molecules.

Interpretation of the results for phenol is hampered by the fact that no perpendicular structures have been identified as minima by the ab initio minimization process. Of the three structures identified as minima, OPLS fails to correctly identify the lowest-energy structure. The reparameterized OPLS model, OPLS-Q, performs better, giving the correct ordering of the minima but heavily underestimating the energy separation between them. OPLS-CS correctly identifies the minimum energy structure and gives the best estimate of its interaction energy. However, it completely fails in its treatment of phe_c, not even identifying it as a minimum.

The relatively simple structure of pyridine means that it has been possible to characterize more fully the minimum energy dimer structures, and it is this molecule that should give the best indication of the ability of the force field models to reproduce dimer interaction energies.

Both the OPLS and OPLS-CS models fail to predict correctly the minimum energy structure of the pyridine dimer. Where the ab initio calculations prefer the formation of a cyclic hydrogen-bonding pattern (pyri_h), OPLS prefers an offset stacked arrangement and OPLS-CS prefers CH−N hydrogen-bonded perpendicular arrangements. Again, these results are consistent with the analysis of the liquid pyridine simulations. For structure pyri_c, the OPLS-CS model performs very poorly, predicting the interaction to be strongly repulsive.

Overall, the results of the calculations on the gas-phase dimers suggest that the OPLS-CS models do not offer a general improvement over the OPLS models. In fact, there are several flaws that can be identified in the OPLS-CS representations. The repulsion in complexes where the molecules stack directly above each other is strongly overestimated, and the relative stability of the T-shaped structures is slightly overestimated relative to the parallel structures. These observations would lead us to suggest that the perpendicular arrangement of molecules is probably overrepresented in the liquid structure predicted using the OPLS-CS models. The absence of perpendicular dimer structures for phenol makes it impossible to truly assess the quality of the OPLS-Q model. If the OPLS-CS model is overestimating

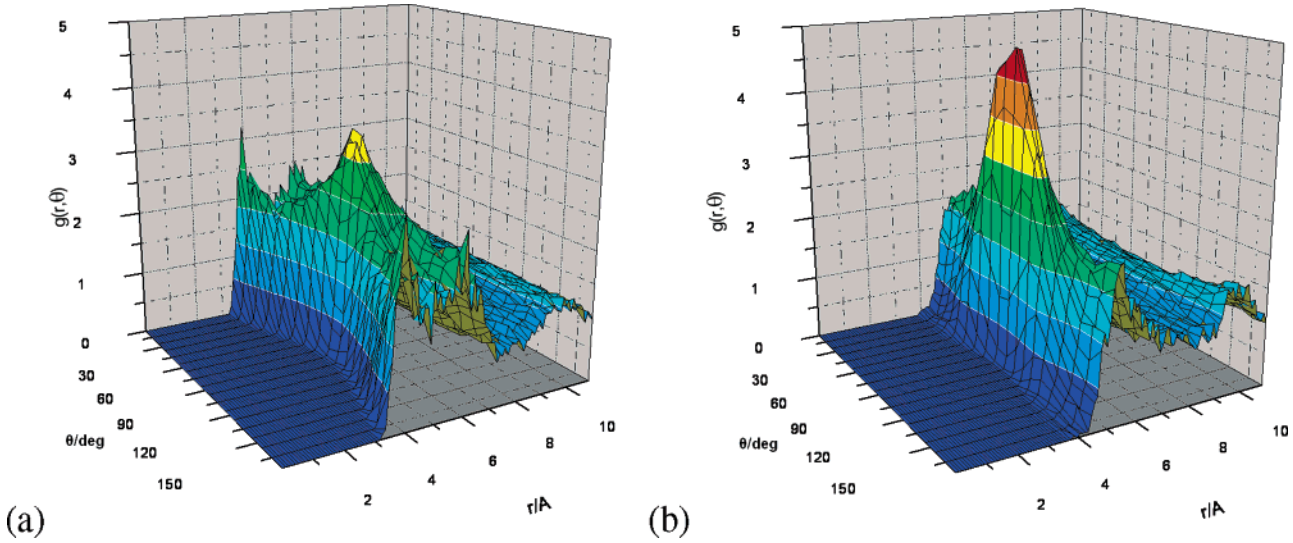


Figure 12. Angular distribution functions for liquid pyridine. Calculated using (a) OPLS and (b) OPLS-CS.

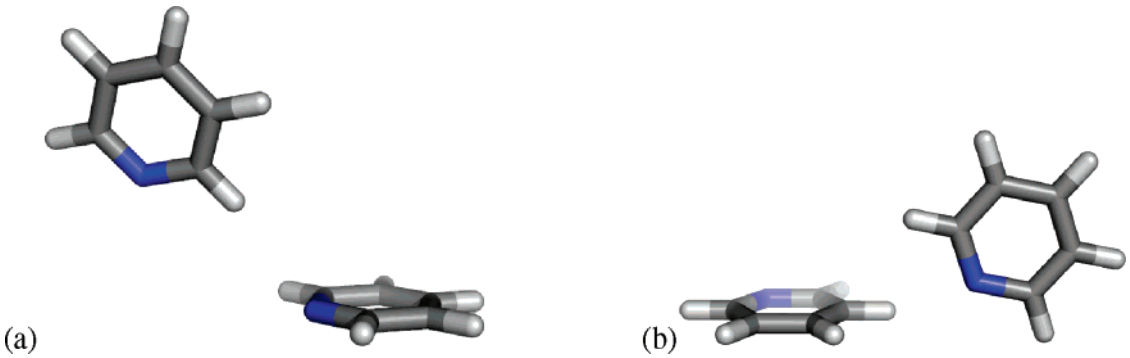


Figure 13. Pyridine−pyridine interactions taken from the OPLS-CS simulation of liquid pyridine. (a) R_{NN} = 4.9 Å and (b) R_{NN} = 5.9 A.

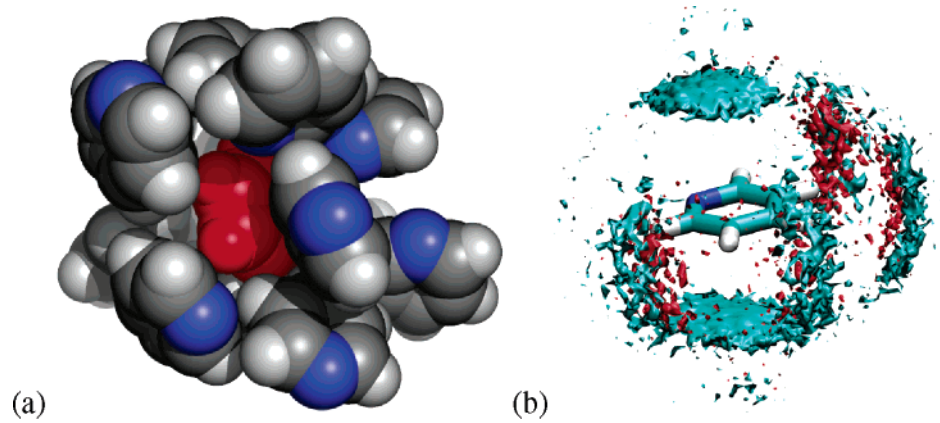


Figure 14. Distribution of molecules in liquid pyridine. (a) The first solvation shell around a pyridine molecule (in red) taken from the OPLS-CS simulation of liquid pyridine. (b) The average atomic density around pyridine molecules in the OPLS-CS simulation (cyan = carbon; red = nitrogen).

perpendicular structure, then the OPLS-Q model is overestimating it even more.

Cation−π Interactions. The models developed thus far have been shown to be successful in the modeling of aromatic liquids. However, the main goal of this work is the development of a force field capable of accurately modeling aromatic groups in biological contexts. The models are not yet sufficiently well-developed that we might be able to consider modeling aromatic−aromatic interactions in proteins, but we can begin to consider isolated examples of the types of interaction that are important in these cases. Such investigations will allow us to test whether the OPLS-CS model is likely to be useful in the study of biological problems and will give an indication of the transferability of the models to situations for which they have not been explicitly parametrized.

As a first test case, the interactions between the amino acid side chains from Phe and Tyr and the ammonium cation

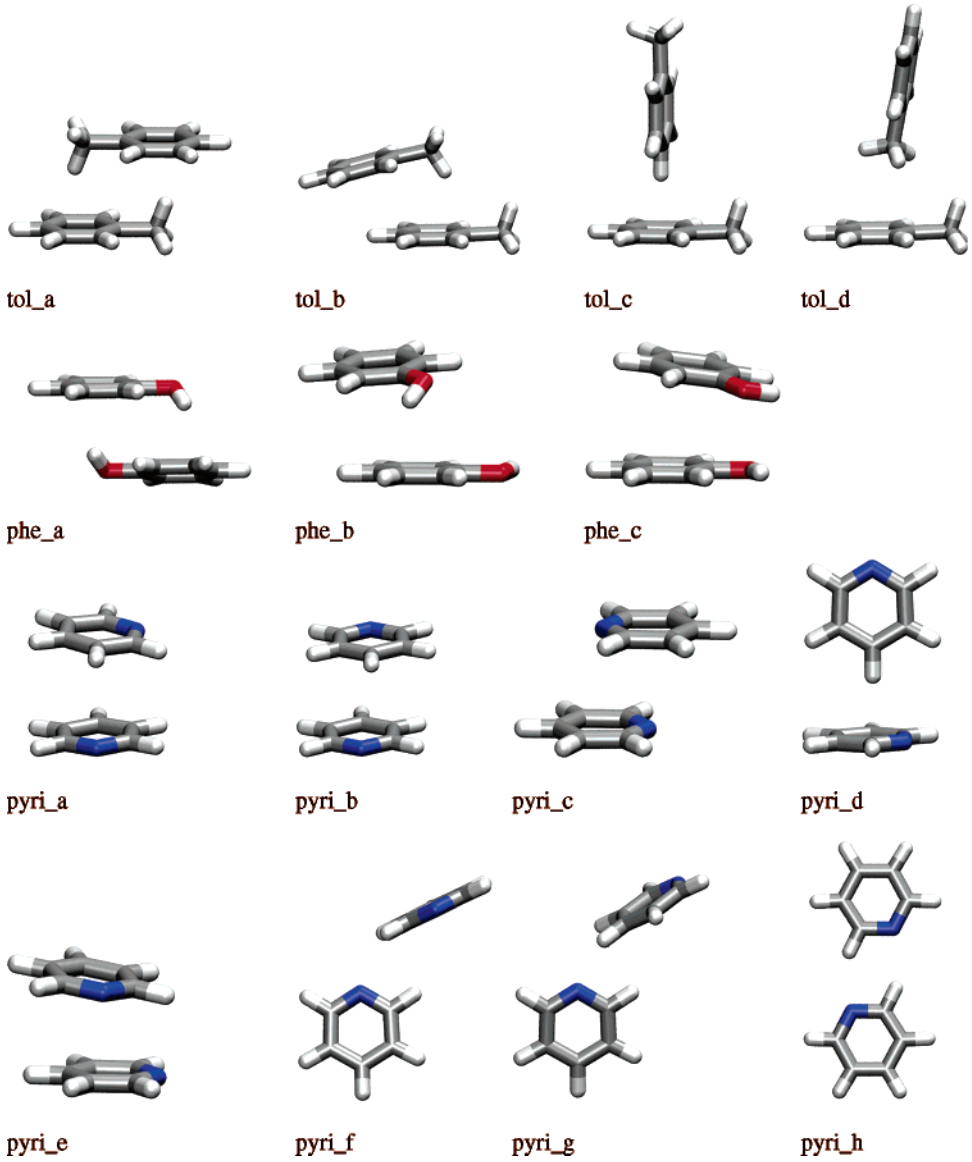


Figure 15. Gas-phase dimer structures for toluene, phenol, and pyridine. Calculated at the MP2/6-31+G*//MP2/6-31+G* level of theory.

Table 8. Toluene Dimer Interaction Energies (in kcal mol^{-1})

	tol_a	tol_b	tol_c	tol_d
MP2/6-31+G*	−3.69	−2.90	−2.02	−1.29
OPLS	−3.14	−2.24	−2.16	−1.95
OPLS-CS	−2.44	−1.28	−1.96	−1.71

Table 9. Phenol Dimer Interaction Energies (in kcal mol^{-1})

	phe_a	phe_b	phe_c
MP2/6-31+G*	−5.33	−3.48	−3.59
OPLS	−2.19	−2.65	−3.11
OPLS-Q	−2.41	−2.23	−2.24
OPLS-CS	−6.26	−5.25	0.91

have been investigated. The results of scanning the ammonium molecule along the *z* axis above the aromatic ring center (Figure 16) reveal that both the OPLS and OPLS-CS models perform reasonably well in reproducing the general shape of the ab initio calculated curves. For the Phe−ammonium interaction, OPLS-CS does slightly better in reproducing the position of the energy minimum while OPLS does better in reproducing the well depth. For Tyr−ammonium, the situation is more clear-cut; OPLS-CS performs much better than OPLS in reproducing both the position and depth of the curve.

While the *z*-scan results do not definitively demonstrate that OPLS-CS outperforms OPLS, the results obtained by scanning the ammonium molecules in the *xy* plane above the aromatic ring (while keeping *z* fixed at 3 Å) are far more conclusive (Figures 17 and 18). For both Phe and Tyr, the OPLS model fails completely to reproduce the general shape of the potential energy surface, severely overestimating the repulsion that occurs when the cation is above the atomic sites. The OPLS-CS models give a much better reproduction of the ab initio data. The well depth is exaggerated, but the shape of the potential energy surface is close to that obtained from the ab initio calculations. The same set of calculations has also been performed using the OPLS-Q model (results

Table 10. Pyridine Dimer Interaction Energies (in kcal mol^{-1})

	pyri_a	pyri_b	pyri_c	pyri_d	pyri_e	pyri_f	pyri_g	pyri_h
MP2/6-31+G*	−3.02	−2.94	−1.85	−1.58	−2.93	−2.57	−2.89	−3.31
OPLS	−3.02	−3.33	−1.84	−2.08	−2.99	−2.62	−2.77	−2.27
OPLS-CS	−1.22	−0.09	4.41	−2.76	−1.43	−2.72	−2.93	−1.00

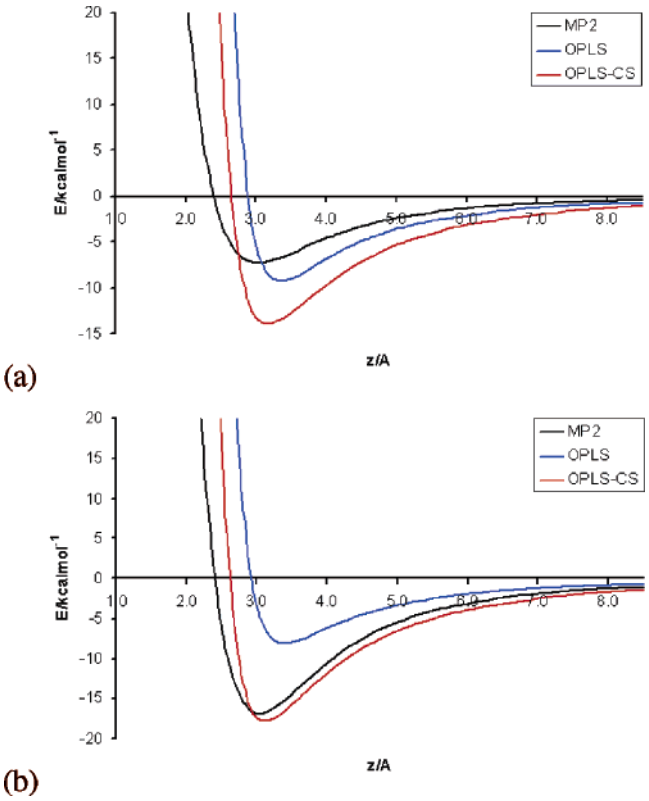


Figure 16. Energy profiles obtained from scanning an ammonium cation in the *z* direction above the aromatic ring center of (a) Phe and (b) Tyr.

not shown). In this case, the results are found to be similar to those obtained using the OPLS-CS model. As was found to be true in the discussion of the liquid-phase results, it seems likely that results equivalent to those obtained using a charge-separation model could be obtained using a standard OPLS model with appropriate reparameterization.

An accurate reproduction of the potential energy surface for cation−π interactions is likely to be essential for the accurate modeling of a variety of biological systems. For example, Thornton and Singh have performed a detailed analysis of protein side-chain interactions,[92] finding that both Arg and Lys are observed experimentally interacting with both Phe and Tyr in a way that places the charged groups in the region above the ring atomic sites rather than above the ring center. Where the OPLS model would fail to find such interactions, predicting them to be repulsive, the OPLS-CS model would be expected to perform much better.

The lack of quantitative agreement between the OPLS-CS and ab initio results might be considered a problem. Indeed, a force field that aspires to provide the most accurate possible account of aromatic interactions must reproduce accurately the true interaction energies. It must be acknowledged that there is still a considerable amount of work to be done before the OPLS-CS model can treat perfectly cation−π interactions. Learning from other cases, it is highly likely that a "correct" representation of these interactions will not be approached without the use of more sophisticated potential functions incorporating, for example, explicit polarization. This work can be seen as a step in the right direction rather than the ultimate solution to the problem of modeling cation−π interactions.

In spite of these caveats, it is possible to say that the OPLS-CS potential, previously derived for the modeling of aromatic liquids, offers a considerable improvement over the OPLS model for the modeling of cation−π interactions.

Conclusions

Monte Carlo calculations have been used to parametrize new OPLS-CS force field models for the aromatic molecules toluene, phenol, and pyridine. Using these models, an accurate reproduction of the available experimental data has been obtained.

The models developed have then been used to examine the structure of these aromatic liquids at the molecular level. Predictions of the structures have been made, but unfortunately, the lack of experimental structural data for these systems makes corroboration of these predictions impossible. Comparison with an existing all-atom potential (OPLS) reveals both similarities and differences. In toluene, both models predict an approximately equal distribution of parallel and perpendicular arrangements of molecules, the parallel structures probably stabilized by a combination of large dispersion interactions and the presence of a dipole moment.

In phenol and pyridine, the OPLS-CS model predicts a far more perpendicular structure than does the OPLS model. Both liquids are characterized by the formation of hydrogen bonds.

To test the transferability of the models developed, they have been used to evaluate the energies of a number of dimeric structures for toluene, phenol, and pyridine. It is found that the OPLS-CS model often overestimates the stability of the perpendicular dimer structures relative to the parallel dimer structures. If this energetic preference is carried over into the liquid, it is likely that the OPLS-CS model is overestimating the extent of perpendicular structure.

As a first test of the ability of the OPLS-CS model to model aromatic interactions in biology, the OPLS-CS models of toluene and phenol have been used to construct models of the aromatic amino acids phenylalanine and tyrosine, respectively. The interactions between these amino acids and the ammonium cation have then been investigated, and it has been shown that the OPLS-CS model considerably outperforms the OPLS model in reproducing the ab initio calculated potential energy surface for this interaction, most significantly in regions that are known experimentally to be

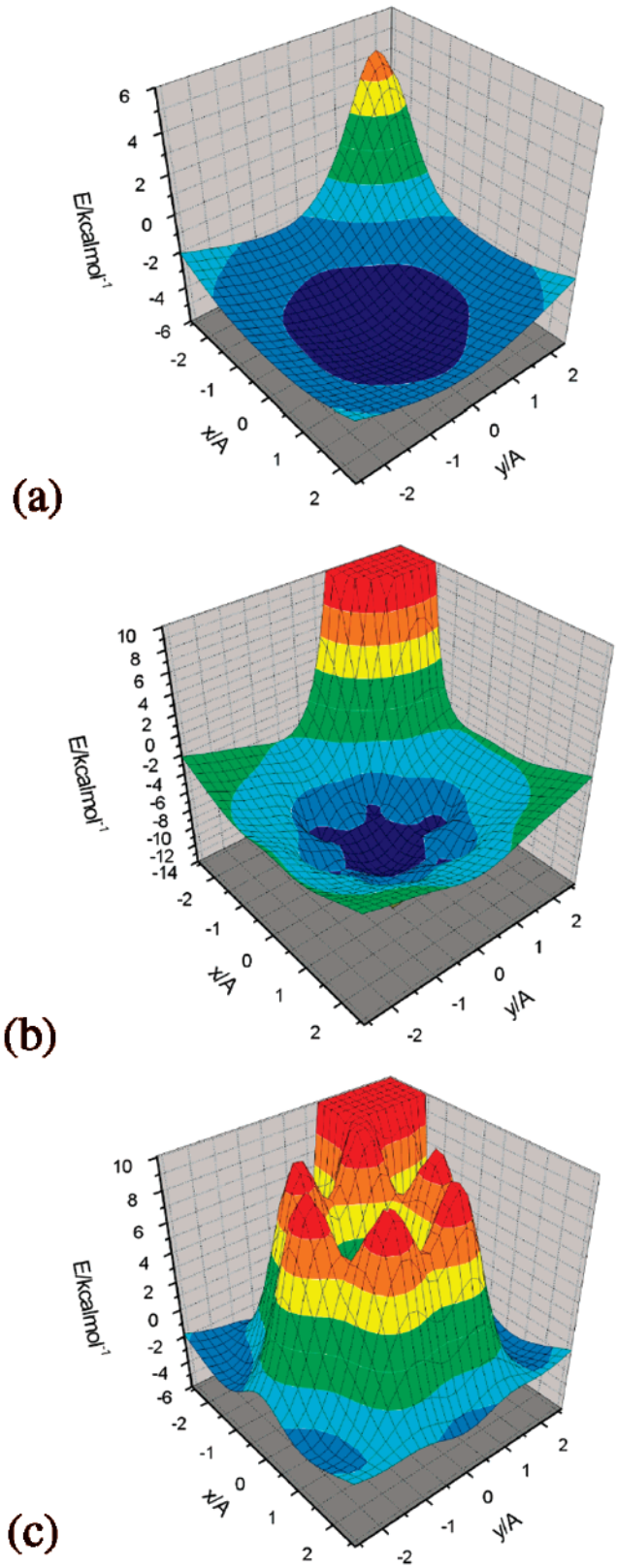


Figure 17. Potential-energy surface for Phe−ammonium interactions calculated using (a) MP2/6-311+G**, (b) the OPLS-CS model, and (c) the OPLS model. In all cases, $z =$ 3.0 Å.

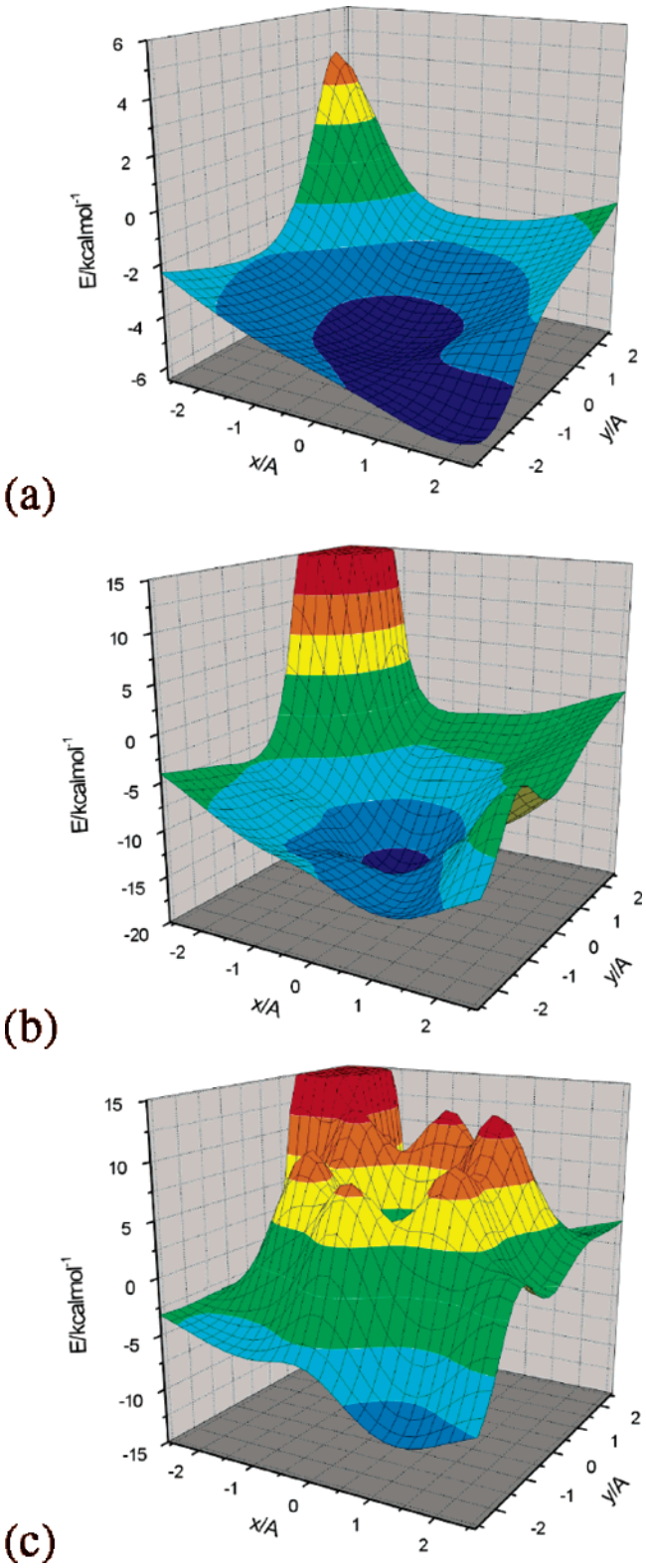


Figure 18. Potential-energy surface for Tyr−ammonium interactions calculated using (a) MP2/6-311+G**, (b) the OPLS-CS model, and (c) the OPLS model. In all cases, $z =$ 3.0 Å.

of biological significance. That the OPLS-CS model outperforms the atom-centered method in this case suggests that it might well find use in the modeling of aromatic interactions in biology, though further development will be necessary before this goal can be realized.

In tandem with the development of the OPLS-CS potential for phenol, an additional model has been developed as a reparameterization of the basic OPLS model. When this

reparameterized model is used, it is possible to obtain results similar to those obtained using the OPLS model—it is likely that the quality of the parametrization process is as important as the model used.

Overall, it has been shown that there are situations in which the OPLS-CS models can outperform the OPLS models. However, this is not always the case, and as with all empirical force fields, it is essential to choose the appropriate model to tackle the problem in question.

Acknowledgment. The authors thank Prof. W. L. Jorgensen for his generous provision of the BOSS program. C.M.B. thanks the National Foundation for Cancer Research for funding, and G.H.G. thanks Boehringer Ingelheim for a fellowship.

Supporting Information Available: Force field parameters for the OPLS-CS models of toluene, phenol, and pyridine. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. *Science* **1991**, *253*, 872.

(2) Hunter, C. A. *Philos. Trans. R. Soc. London, Ser. A* **1993**, *345*, 77.

(3) Vondrášek, J.; Bendová, L.; Klusák, V.; Hobza, P. *J. Am. Chem. Soc.* **2005**, *127*, 2615.

(4) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23.

(5) Fowler, P. W.; Buckingham, A. D. *Chem. Phys. Lett.* **1991**, *176*, 11.

(6) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.

(7) Baker, C. M.; Grant, G. H. *J. Chem. Theory Comput.* **2006**, *2*, 947.

(8) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(9) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 11217.

(10) Cheney, J.; Cheney, B. V.; Richards, W. G. *Biochim. Biophys. Acta* **1988**, *954*, 137.

(11) Scheiner, S.; Kar, T.; Pattanayak, J. *J. Am. Chem. Soc.* **2002**, *124*, 13257.

(12) Hobza, P.; Špirko, V.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem. A* **1998**, *102*, 2501.

(13) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.

(14) Gervasio, F. L.; Chelli, R.; Procacci, P.; Schettino, V. *J. Phys. Chem. A* **2002**, *106*, 2945.

(15) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. C.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(16) Schauer, M.; Bernstein, E. R. *J. Chem. Phys.* **1985**, *82*, 3722.

(17) Ishikawa, S.; Ebata, T.; Ishikawa, H.; Inoue, T.; Mikami, N. *J. Phys. Chem.* **1996**, *100*, 10531.

(18) Law, K. S.; Schauer, M.; Bernstein, E. R. *J. Chem. Phys.* **1984**, *81*, 4871.

(19) Musgrave, A.; Wright, T. G. *J. Chem. Phys.* **2005**, *122*, 074312.

(20) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 4209.

(21) Tsuzuki, S.; Honda, K.; Uchimara, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104.

(22) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*, 8377.

(23) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690.

(24) Kim, J. H.; Lee, S. H. *Bull. Korean Chem. Soc.* **2002**, *23*, 441.

(25) Kim, J. H.; Lee, S. H. *Bull. Korean Chem. Soc.* **2002**, *23*, 447.

(26) Jorgensen, W. L.; Laird, E. R.; Nguyen, T. B.; Tirado-Rives, J. *J. Comput. Chem.* **1993**, *14*, 206.

(27) Kumar, A. V. A.; Yashonath, S.; Chaplot, S. L. *J. Chem. Phys.* **2000**, *113*, 8070.

(28) Fioroni, M.; Vogt, D. *J. Phys. Chem. B* **2004**, *108*, 11774.

(29) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596.

(30) Andre, D.; Fourme, R.; Bruneaux-Poulle, J.; Bosio, L. *J. Mol. Struct.* **1982**, *81*, 253.

(31) Anderson, M.; Bosio, L.; Bruneaux-Poulle, J.; Fourme, R. *J. Chim. Phys. Phys. Chim. Biol.* **1977**, *74*, 68.

(32) Hobza, P.; Riehn, C.; Weichert, A.; Brutschy, B. *Chem. Phys.* **2002**, *283*, 331.

(33) Sagarik, K.; Asawakun, P. *Chem. Phys.* **1997**, *219*, 173.

(34) Böhm, H.-J.; Ahlrichs, R. *J. Chem. Phys.* **1982**, *77*, 2028.

(35) Hartland, G. V.; Henson, B. F.; Venturo, V. A.; Felker, P. M. *J. Phys. Chem.* **1992**, *96*, 1164.

(36) Ebata, T.; Watanabe, T.; Mikami, N. *J. Phys. Chem.* **1995**, *99*, 5761.

(37) Connell, L. L.; Ohline, S. M.; Joireman, P. W.; Corcoran, T. C.; Fleker, P. M. *J. Chem. Phys.* **1992**, *96*, 2585.

(38) Parthasarathi, R.; Subramanian, V.; Sathyamurthy, N. *J. Phys. Chem. A* **2005**, *109*, 843.

(39) Mooney, D. A.; Müller-Plathe, F.; Kremer, K. *Chem. Phys. Lett.* **1998**, *294*, 135.

(40) Thornton, J. M.; MacArthur, M. W.; McDonald, I. K.; Jones, D. T.; Mitchell, J. B. O.; Nandi, C. L.; Price, S. L.; Zvelebil, M. J. J. M. *Philos. Trans. R. Soc. London, Ser. A* **1993**, *345*, 113.

(41) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(42) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(43) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *J. Am. Chem. Soc.* **2002**, *124*, 6133.

(44) Zavodnik, V. E.; Bel'skii, V. K.; Zorkii, P. M. *J. Struct. Chem.* **1987**, *28*, 793.

(45) Allan, D. R.; Clark, S. J.; Dawson, A.; McGregor, P. A.; Parsons, S. *Acta. Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **2002**, *58*, 1018.

(46) Mishra, B. K.; Sathyamurthy, N. *J. Phys. Chem. A* **2005**, *109*, 6.

(47) Piacenza, M.; Grimme, S. *Chem. Phys. Chem.* **2005**, *6*, 1554.

(48) Megiel, E.; Kasprzycka-Guttman, T.; Jagielska, A.; Wróblewska, L. *J. Mol. Struct.* **2001**, *569*, 111.

(49) Sagarik, K.; Spohr, E. *Chem. Phys.* **1995**, *199*, 73.

(50) Jorgensen, W. L.; McDonald, N. A. *THEOCHEM* **1998**, *424*, 145.

(51) Williams, D. E.; Weller, R. R. *J. Am. Chem. Soc.* **1983**, *105*, 4143.

(52) Gamba, Z.; Klein, M. L. *Chem. Phys.* **1989**, *130*, 15.

(53) Mootz, D.; Wussow, H.-G. *J. Chem. Phys.* **1981**, *75*, 1517.

(54) Anghel, A. T.; Day, G. M.; Price, S. L. *CrystEngComm* **2002**, *4*, 348.

(55) Dougherty, D. A. *Science* **1996**, *271*, 163.

(56) Urban, J. J.; Cramer, C. J.; Famini, G. R. *J. Am. Chem. Soc.* **1992**, *114*, 8226.

(57) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *17*, 9459.

(58) Tsou, L. K.; Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2002**, *124*, 14917.

(59) Scrutton, N. S.; Raine, A. R. C. *Biochem. J.* **1996**, *319*, 1.

(60) Wintjens, R.; Lievin, J.; Rooman, M.; Buisine, E. *J. Mol. Biol.* **2000**, *302*, 395.

(61) Zacharias, N.; Dougherty, D. A. *Trends Pharmacol. Sci.* **2002**, *23*, 281.

(62) Petersen, F. N. R.; Jensen, M. O.; Nielsen, C. H. *Biophys. J.* **2005**, *89*, 3985.

(63) Caldwell, J. W.; Kollmann, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4177.

(64) Mecozzi, S.; West, A. P., Jr.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10566.

(65) Mo, Y.; Subramanian, G.; Gao, J.; Ferguson, D. M. *J. Am. Chem. Soc.* **2002**, *124*, 4832.

(66) Tsuzuki, S.; Uchimaru, T.; Mikami, M. *J. Phys. Chem. A* **2003**, *107*, 10414.

(67) Alberti, M.; Castro, A.; Lagana, A.; Moix, M.; Pirani, F.; Cappelletti, D.; Liuti, G. *J. Phys. Chem. A* **2005**, *109*, 2906.

(68) Rocha-Rinza, T.; Hernandez-Trujillo, J. *Chem. Phys. Lett.* **2006**, *422*, 36.

(69) Zhu, W.-L.; Tan, X.-J.; Puah, C. M.; Gu, J.-D.; Jiang, H.-L.; Chen, K.-X.; Felder, C. E.; Silman, I.; Sussman, J. L. *J. Phys. Chem. A* **2000**, *104*, 9573.

(70) Liu, T.; Zhu, W.; Gu, J.; Shen, J.; Luo, X.; Chen, G.; Puah, C. M.; Silman, I.; Chen, K.; Sussman, J. L.; Jiang, H. *J. Phys. Chem. A* **2004**, *108*, 9400.

(71) Xu, Y.; Shen, J.; Zhu, W.; Luo, X.; Chen, K.; Jiang, H. *J. Phys. Chem. B* **2005**, *109*, 5945.

(72) Lee, J. Y.; Lee, S. J.; Choi, H. S.; Cho, S. J.; Kim, K. S.; Ha, T.-K. *Chem. Phys. Lett.* **1995**, *232*, 67.

(73) Reddy, A. S.; Sastry, G. N. *J. Phys. Chem. A* **2005**, *109*, 8893.

(74) Chipot, C.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 2998.

(75) Tan, X. J.; Jiang, H. L.; Zhu, W. L.; Chen, K. X.; Ji, R. Y. *J. Chem. Soc., Perkin Trans. 2* **1999**, 107.

(76) Vaden, T. D.; Lisy, J. M. *J. Chem. Phys.* **2005**, *123*, 074302.

(77) Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837.

(78) Price, S. L.; Stone, A. J. *J. Chem. Phys.* **1987**, *86*, 2859.

(79) Doerksen, R. J.; Thakkar, A. J. *J. Phys. Chem. A* **1999**, *103*, 10009.

(80) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128.

(81) Hawkins, D. M. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1.

(82) Jorgensen, W. L. *BOSS 4.2*; Yale University: New Haven, CT, 2001.

(83) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.

(84) Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 123.

(85) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.

(86) Ransil, B. J. *J. Chem. Phys.* **1961**, *34*, 2109.

(87) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(88) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Rev. A.7*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(89) Atkins, P. W. *Physical Chemistry*, 6th ed.; Oxford University Press: Oxford, U. K., 1998; p 106.

(90) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276.

(91) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.

(92) Singh, J.; Thornton, J. M. *Atlas of Protein Side-Chain Interaction*; IRL Press: Oxford, United Kingdom, 1992; pp 94−105, 494−505.

CT600218F

JCTC Journal of Chemical Theory and Computation

Spin−Spin Coupling Constants for Iminoboranes RBNH, HBNR, and RBNR and Comparisons with Corresponding Isoelectronic Acetylenes RCCH and RCCR, for R = H, CH_3, NH_2, OH, and F

Janet E. Del Bene,*,† José Elguero,‡ Ibon Alkorta,‡ Manuel Yáñez,§ and Otilia Mó§

Department of Chemistry, Youngstown State University, Youngstown, Ohio 44555, Instituto de Química Médica, Consejo Superior de Investigaciones Científicas, Juan de la Cierva, 3, E-28006 Madrid, Spain, and Departamento de Química, C-9, Universidad Autónoma de Madrid, Cantoblanco, E-28049 Madrid, Spain

Received December 5, 2006

Abstract: Ab initio equation-of-motion coupled-cluster singles and doubles method calculations have been performed on iminoboranes RBNH, HBNR, and RBNR, for R = H, CH_3, NH_2, OH, and F, to evaluate substituent effects on one- and two-bond $^{15}N-^{11}B$, $^{11}B-^{1}H$, and $^{15}N-^{1}H$ spin−spin coupling constants. For comparison purposes, $^{13}C-^{13}C$ coupling constants were evaluated for corresponding isoelectronic molecules RCCH and RCCR. The absolute values of $^{1}J(N-B)$ and $^{1}J(C-C)$ increase in the series HBNR and RCCH as the σ-electron-withdrawing ability of R increases. In contrast, NH_2 substitution at B leads to a decrease in the absolute value of $^{1}J(N-B)$, but OH and F substitution increase $^{1}J(N-B)$. Disubstitution has dramatically different effects on coupling constants for RCCR and RBNR. In the former, $^{1}J(C-C)$ more than doubles relative to the corresponding RCCH and HCCH; in the latter, disubstitution of OH and F decreases $^{1}J(N-B)$ relative to the corresponding monosubstituted derivatives, while NH_2 substitution increases $^{1}J(N-B)$. Changes in one- and two-bond B−H and N−H coupling constants upon substitution are similar to changes observed for $^{1}J(N-B)$ in the corresponding monosubstituted derivatives RBNH and HBNR.

Introduction

The examination of substituent effects on molecular properties has proven to be a very fruitful approach in a wide range of studies of organic molecules.[1−5] In a previous study,[6] we used this approach to examine the relative stabilities of corresponding isomers in two series of iminoboranes RBNH and HBNR, where R is H, CH_3, NH_2, OH, and F. We investigated ground-state bonding and selected IR spectroscopic properties of these molecules, as well as the bonding characteristics and IR properties of their disubstituted counterparts, RBNR. The present paper reports the results of an extension of our investigation of iminoboranes to include substituent effects on $^{15}N-^{11}B$ spin−spin coupling constants in molecules RBNH, HBNR, and RBNR, as well as one- and two-bond $^{15}N-^{1}H$ and $^{11}B-^{1}H$ spin−spin coupling constants in the monosubstituted derivatives. We also present $^{13}C-^{13}C$ coupling constants for RCCH and RCCR, molecules that are isoelectronic with corresponding RBNH and HBNR, and RBNR, respectively. In this paper, we compare and contrast substituent effects of one-bond B−N coupling constants in mono- and disubstituted iminoborazines and compare them to substituent effects on C−C coupling constants in the corresponding mono- and disubstituted acetylenes.

The present study is also an extension of our previous investigation of B−N coupling constants in borazine and its fluoro- and lithio-substituted derivatives.[7] In that study, we

* Corresponding author e-mail: jedelbene@ysu.edu.
† Youngstown State University.
‡ Consejo Superior de Investigaciones Científicas.
§ Universidad Autónoma de Madrid.

10.1021/ct600351p CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/09/2007

examined the sensitivity of ^1J(N−B) to the site of substitution and the number of substituents and also observed the opposing effects of σ-electron-donating and -electron-withdrawing substituents, since F substitution at either B or N increases ^1J(N−B), whereas Li substitution decreases ^1J(N−B). These substituent effects in borazine are similar to those observed for benzene and are consistent with available experimental data.[7] Recently, we have also reported one-bond B−H and B−Li coupling constants for the five-membered boron-containing rings diazaborole and Li-diazaborole, as well as B−N coupling constants for these molecules and the corresponding diazaborole anion,[8] species which served as models for newly synthesized five-membered nucleophilic boron-containing rings.[9] NMR coupling constants and chemical shifts were used experimentally to identify these species. It is anticipated that the newly developed synthetic techniques which produced nucleophilic boron moieties will lead to the synthesis of a new class of molecules containing B−N bonds and that NMR spin−spin coupling constants will be an important tool for the identification and characterization of these molecules. Hence, a systematic study of substituent effects on B−N coupling constants is both warranted and timely.

Methods

The geometries of the molecules RBNH, HBNR, and RBNR were optimized at second-order Møller−Plesset theory[10−13] with the 6-311++G(d,p) basis set[14−16] and have been reported previously.[6] Coupling constants were computed for these molecules using the ab initio equation-of-motion coupled-cluster singles and doubles method (EOM-CCSD) in the configuration-interaction- (CI-)like approximation,[17−20] with all electrons correlated. The Ahlrichs qzp basis set[21] was used on C, N, O, and F atoms, and the qz2p basis set was used for H atoms bonded to either B or N. The Dunning cc-pVDZ[22,23] basis set was placed on all other H atoms. Since an Ahlrichs qzp basis is not available for B, a new "hybrid" basis had been constructed for this atom, using the boron cc-pV5Z basis for 1s orbitals, cc-pVQZ for the valence s and p orbitals, and one set of d orbitals. This basis set has the same number of contracted functions (6s, 4p, and 1d) as the Ahlrichs qzp basis used for C, N, O, and F and was used previously in studies of B−N, B−H, and B−Li coupling constants.[7,8]

In the nonrelativistic approximation, the nuclear spin−spin coupling constant is composed of four terms: the paramagnetic spin−orbit (PSO), diamagnetic spin−orbit (DSO), Fermi-contact (FC), and spin-dipole (SD).[24] All terms have been computed for all molecules. The EOM-CCSD calculations were carried out using the ACES II[25] program on the Itanium Cluster at the Ohio Supercomputer Center.

Results and Discussion

Substitution Effects on ^1J(C−C): RCCH and RCCR. Before analyzing the B−N coupling constants as a function of the substituent, it is advantageous to first examine C−C coupling constants for molecules RCCH, which are isoelectronic with corresponding molecules RBNH and HBNR, and for RCCR, which are isoelectronic with the corresponding RBNR. The individual contributions of the PSO, DSO, FC, and SD terms to ^1J(C−C) can be found in Table S1 of the Supporting Information. The FC terms, total ^1J(C−C), and C−C distances are reported in Table 1. The Fermi-contact term is the dominant term, although contributions from the PSO terms (ranging from 6 to 15 Hz) and from the SD terms (8 to 10 Hz) cannot be neglected. Moreover, although the data of Table 1 indicate that ^1J(C−C) has its maximum values in the RCCH and RCCR series when R is F and the C−C distances are shortest, ^1J(C−C) does not correlate with this distance. For example, although the C−C distance in H_2NCCH is intermediate between that in HCCH and that in H_3CCCH, differing from both by only 0.001 Å, ^1J(C−C) for H_2NCCH is 34 Hz greater than ^1J(C−C) for HCCH and H_3CCCH.

Table 1. C−C Fermi-Contact Terms and Coupling Constants [^1J(C−C), Hz], C−C Distances (Å), and Symmetries for Molecules RCCH and RCCR

Coupling Constants				
	RCCH		RCCR	
molecule	FC	^1J(C−C)	FC	^1J(C−C)
R = H	178.1	192.3	178.1	192.3
CH_3	178.5	192.7	180.9	195.2
NH_2	210.8	226.3	253.7	270.1
OH	232.8	250.2	313.6	334.0
F	258.8	278.7	394.1	418.8

C−C Distances and Molecular Symmetry		
molecule	RCCH	RCCR
R = H	1.216 ($C_{\infty v}$)	1.216 ($D_{\infty h}$)
CH_3	1.218 (C_{3v})	1.220 (D_{3h})
NH_2	1.217 (C_s)	1.217 (C_2)
OH	1.214 (C_s)	1.208 (C_2)
F	1.208 ($C_{\infty v}$)	1.197 ($D_{\infty h}$)

The only experimental coupling constants available for the molecules investigated in this study are for acetylene (170.6 Hz)[26] and methylacetylene (175 Hz).[27] From Table 1, it can be seen that the computed EOM-CCSD values of 192.3 and 192.7 Hz, respectively, overestimate the experimental values by about 10%. This may be due in part to the sensitivity of ^1J(C−C) to the length of the C−C bond, as demonstrated by Wigglesworth et. al,[28] and also to the neglect of zero-point and thermal vibrational effects.[29−32] However, our emphasis in this study is on changes in coupling constants due to the substitution of σ-electron-withdrawing groups, and the results reported below for substituted acetylenes are given only for comparison with substituent effects in iminoboranes. Nevertheless, as will be evident below, our results for substituent effects on ^1J(C−C) are in agreement with results reported by Kamienska-Trela and co-workers, who have carried out extensive studies of substituent effects on C−C coupling constants.[27,33]

What happens to ^1J(C−C) upon the substitution of H by the ground-state σ-electron-withdrawing substituents CH_3, NH_2, OH, and F? As evident from Table 1, ^1J(C−C) values for HCCH, H_3CCCH, and H_3CCCCH_3 are similar. However, as the σ-electron-withdrawing ability of the substituents

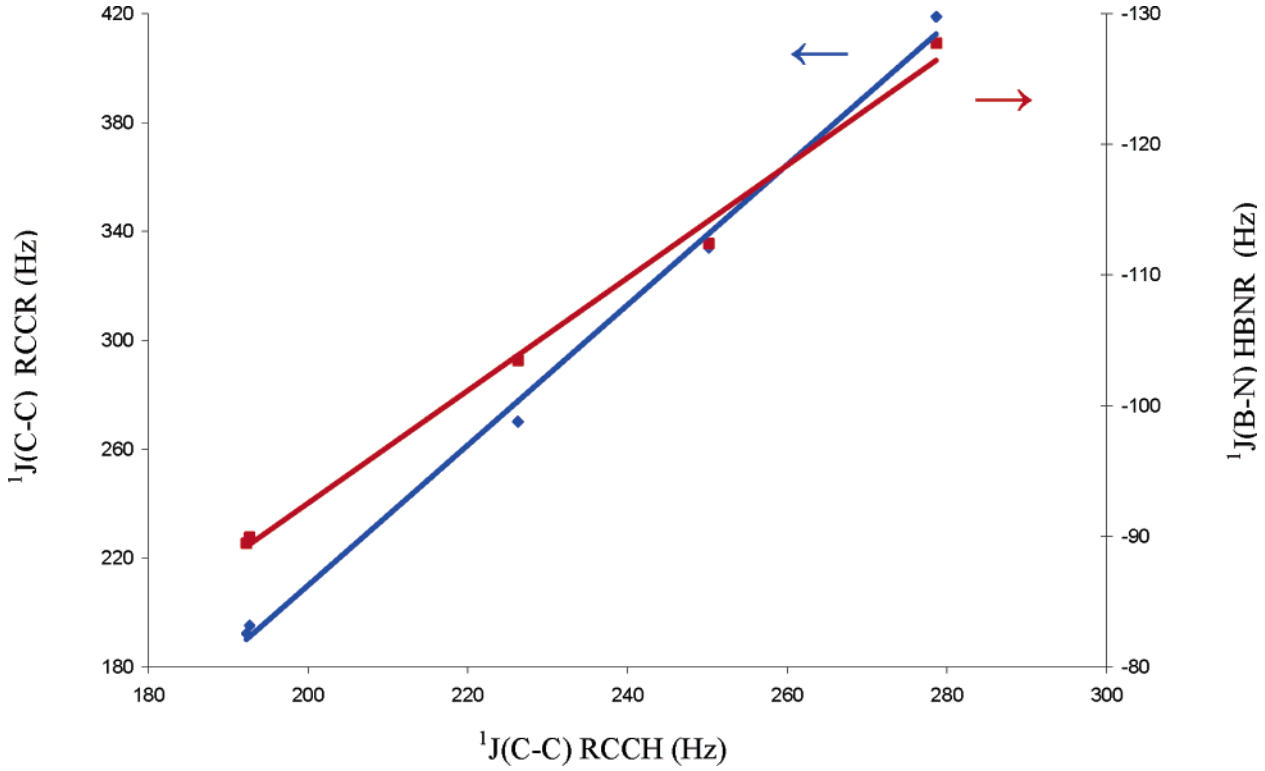


Figure 1. ^1J(C−C) for RCCR (blue ◆, left *y* axis) and ^1J(B−N) for HBNR (red ■, right *y* axis) vs ^1J(C−C) for RCCH.

increases in the order CH_3 < NH_2 < OH < F, ^1J(C−C) increases in the same order in both the mono- and disubstituted derivatives. Figure 1 graphically illustrates the linear correlation between ^1J(C−C) for RCCR and ^1J(C−C) for RCCH. As the σ-electron-withdrawing ability of the substituent increases, ^1J(C−C) increases, primarily because of an increase in the FC term. A second very interesting observation is the slope of 2.6 for the trend line shown in Figure 1, which indicates that disubstitution is nonadditive in a positive sense, insofar as it more than doubles ^1J(C−C) in RCCR relative to the corresponding RCCH and HCCH. These observations indicate that ^1J(C−C) increases systematically as the σ-electron-withdrawing ability of the R group increases.

How can these results be explained? Because ^1J(C−C) is dominated by the FC term, some insights can be obtained by considering the nature of this term and the nature of the excited triplet-state wavefunctions which couple to the ground state through the FC operator. First, it should be recalled that the FC operator does not contain a specific distance-dependent term. Second, if the function of the R groups is to withdraw σ-electron density from the C−C bond in the ground state, then the increase in ^1J(C−C) as the σ-electron-withdrawing ability of the substituent increases suggests that the excited states which dominate must have increased σ-electron densities on the two carbon atoms. Alternatively, in going from NH_2 to OH to F, the number of nonbonding pairs of electrons on the substituent increases. These lone pairs can interact with the C−C π bonds and, from a sum-over-states perspective, influence both the sign and magnitude of the contributions to ^1J(C−C) from the various excited states that couple to the ground state. Since the FC term dominates and the sign of ^1J(C−C) is positive, triplet excited states with an odd number of nodes intersecting the C−C bond assume increased importance.[34] The effect of σ-electron-withdrawing substituents on ^1J(C−C) in the substituted acetylenes is consistent with the effect of F substitution on one-bond C−C coupling constants in fluorobenzene, which also increase when coupling involves the fluoro-substituted C atom.[7]

^1J(N−B) and B−N Distances. Molecular symmetries, B−N distances, Fermi-contact terms, and one-bond B−N coupling constants [^1J(N−B)] for the entire set of molecules RBNH, HBNR, and RBNR are reported in Table 2. The individual contributions of PSO, DSO, FC, and SD terms can be found in Table S2 of the Supporting Information. As observed previously for borazine,[7] the FC term dominates ^1J(N−B), accounting for more than 90% of the total coupling constant in all cases. The data of Table 2 also indicate that there is little correlation between one-bond B−N coupling constants and B−N distances. For example, among the disubstituted iminoboranes, the B−N distances are shortest in HBNH and H_3CBNCH_3, but these two molecules have the smallest B−N coupling constants. Thus, it can be concluded that, in general, changes in B−N coupling constants in substituted iminoboranes do not depend simply on changes in B−N distances. This observation is consistent with previous theoretical and experimental data on coupling constants and distances for molecules[35,36] and is really not surprising in view of the absence of a distance-dependent term in the Fermi-contact operator.

Substituent Effects on ^1J(N−B). Having examined ^1J(C−C) as a function of the substituent for RCCH and RCCR, it is appropriate to now return to ^1J(N−B) for RBNH, HBNR, and RBNR. The B−N coupling constants for these molecules are reported in Table 2. As observed for CH_3 substitution

Table 2. B−N Fermi-Contact Terms and Coupling Constants [^1J(N−B), Hz]; B−N Distances (Å); and Symmetries for Molecules RBNH, HBNR, and RBNR

Coupling Constants						
	RBNH		HBNR		RBNR	
molecule	FC	^1J(N−B)	FC	^1J(N−B)	FC	^1J(N−B)
R = H	−84.9	−89.4	−84.9	−89.4	−84.9	−89.4
CH_3	−84.1	−88.2	−84.9	−89.9	−84.3	−89.0
NH_2	−83.4	−86.4	−96.9	−103.4	−110.1	−115.6
OH	−108.5	−112.6	−104.5	−112.4	−91.0	−96.1
F	−118.3	−123.1	−118.1	−127.7	−99.5	−106.6

B−N Distances and Molecular Symmetry			
molecule	RBNH	HBNR	RBNR
R = H	1.247 ($C_{\infty v}$)	1.247 ($C_{\infty v}$)	1.247 ($C_{\infty v}$)
CH_3	1.250 (C_{3v})	1.250 (C_{3v})	1.252 (C_{3v})
NH_2	1.264 (C_s)	1.248 (C_s)	1.254 (C_1)
OH	1.250 (C_s)	1.246 (C_s)	1.268 (C_1)
F	1.245 ($C_{\infty v}$)	1.241 ($C_{\infty v}$)	1.260 (C_s)

on ^1J(C−C), substitution of one CH_3 group at either B or N, or two CH_3 groups, one at B and one at N, has little effect on ^1J(N−B). Moreover, in the series HBNR, substitution of the stronger σ-electron-withdrawing substituents NH_2, OH, and F increases the absolute value of ^1J(N−B) in the order $CH_3 < NH_2 < OH < F$, consistent with the RCCH series. The correlation between substituent effects on ^1J(N−B) in the series HBNR and ^1J(C−C) for RCCH is also illustrated in Figure 1. [It should be noted that, since the magnetogyric ratio of ^{15}N is negative, while that of ^{11}B is positive, the one-bond B−N coupling constant ^1J(N−B) is negative, but the reduced one-bond coupling constant ^1K(N−B) is positive, as is ^1K(C−C).]

In contrast, a correlation between ^1J(C−C) and ^1J(N−B) is not observed for substitution at B in the series RBNH. In particular, NH_2 substitution at B leads to a decrease of ^1J(N−B) relative to HBNH, while OH and F substitution increase ^1J(N−B). [Changes in ^1J(N−B) will be discussed in terms of absolute values.] Thus, ^1J(N−B) is extremely sensitive to the site of substitution when the substituent is NH_2 but is relatively insensitive to which site is substituted when R is OH or F. This is illustrated graphically in Figure 2, a plot of ^1J(N−B) for HBNR versus ^1J(N−B) for RBNH. The reference line shown passes through the origin and has a slope of 1. Relative to the other molecules in the two series of monosubstituted derivatives, ^1J(N−B) for H_2NBNH appears to be too small; that is, NH_2 bonded to B acts more like H and CH_3 rather than the stronger σ-electron-withdrawing substituents OH and F. Since ^1J(N−B) for HBNR is linearly correlated with ^1J(C−C), it is apparent that no such correlation exists for coupling constants between RBNH and RCCH.

Why does the substitution of NH_2 at B fail to produce the expected increase in ^1J(N−B)? Some insight into the answer to this question can be gained by considering the nature of the BN bond in H_2NBNH. At the outset, it must be stated that coupling constants are second-order properties that depend on electron densities in both the ground state and the excited states to which they couple, and any explanation of changes in coupling constants based on ground-state properties must be incomplete. However, it is not unreasonable to suggest that, in a series of related molecules, differences in ground-state bonding characteristics may be manifested by differences in coupling constants. In our previous study of the ground-state bonding characteristics of substituted iminoboranes,[6] we concluded that, while substitution at N only slightly perturbs the BN triple bond, NH_2 substitution at B has a rather dramatic effect on this bond. In H_2N-BNH, there is a strong contribution to the BN bond from a canonical form of the type $H_2N{=}B{=}N-H$, which indicates that the BN bond loses triple bond character. This is associated with a significant increase in the B−N distance relative to all other monosubstituted derivatives and a value of ^1J(N−B) that is much less than expected, as clearly illustrated in Figure 2.

The effect of disubstitution of NH_2, OH, and F on ^1J(N−B) is dramatically different from the effect of monosubstitution. In particular, disubstitution of NH_2 increases ^1J(N−B) relative to both H_2NBNH and $HBNNH_2$, but disubstitution of OH or F decreases ^1J(N−B) relative to the corresponding monosubstituted derivatives RBNH and HBNR. Specifically, ^1J(N−B) decreases from −112 Hz in HOBNH and HBNOH to −96 Hz in HOBNOH, and from −123 and −128 Hz in FBNH and HBNF, respectively, to −107 Hz in FBNF, while ^1J(N−B) increases from −86 and −103 Hz in H_2NBNH and $HBNNH_2$, respectively, to −116 Hz in H_2NBNNH_2. These differences are also shown graphically in Figure 2, where ^1J(N−B) for RBNR is plotted against ^1J(N−B) for RBNH. Once again, the reference line passes through the origin and has a slope of 1. The corresponding plot for RBNR versus HBNR has a similar appearance. It is apparent that disubstitution of NH_2 leads to a dramatic increase in ^1J(N−B) relative to $HBNNH_2$ and H_2NBNH, whereas disubstitution of OH or F actually decreases ^1J(N−B) relative to the corresponding monosubstituted derivatives.

Some insight into at least one factor which influences this rather strange behavior can be obtained by comparing ^1J(N−B) for RBNR and ^1J(C−C) for RCCR. Figure 3 indicates

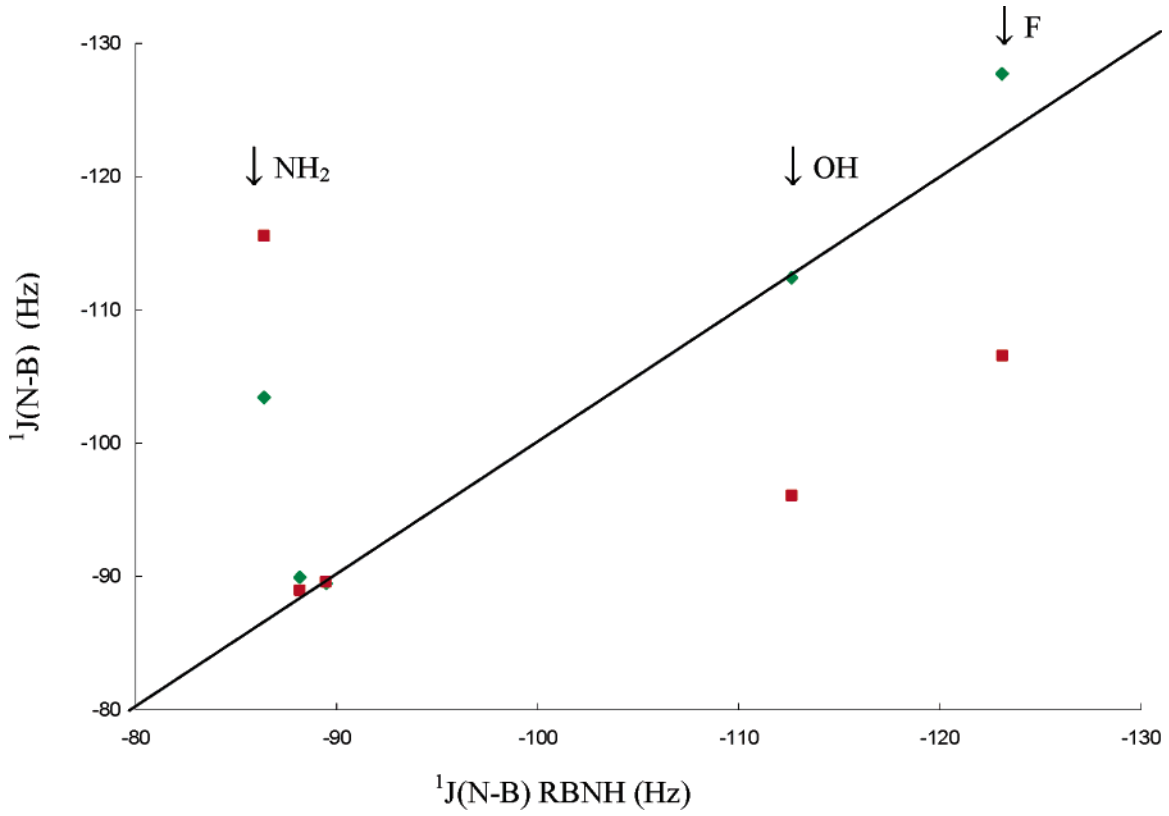


Figure 2. ^1J(N−B) for HBNR (green ◆) and ^1J(N−B) for RBNR (red ■) vs ^1J(N−B) for RBNH. The reference line has a slope of 1 and passes through the origin. The points corresponding to NH_2, OH, and F as substituents are marked by arrows. The points for H and CH_3 are at the bottom left of the graph.

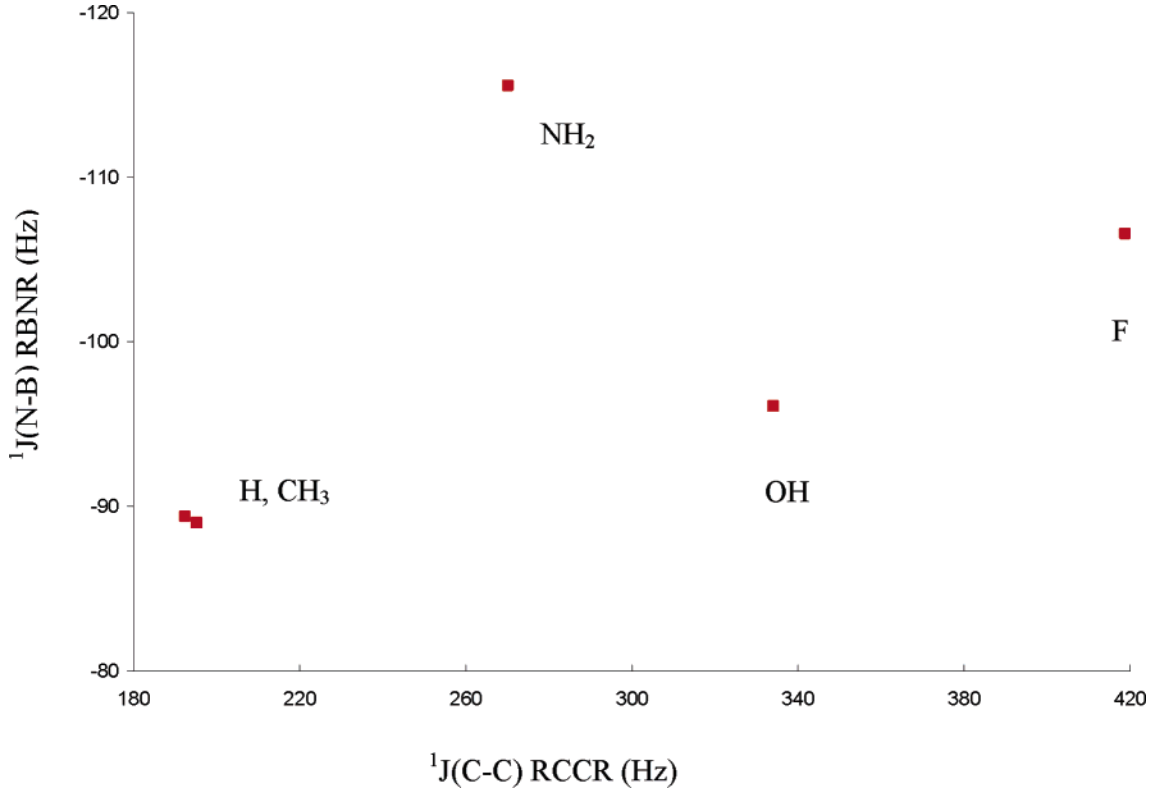


Figure 3. ^1J(N−B) for RBNR vs ^1J(C−C) for RCCR.

that, while disubstitution of CH_3 has very little effect on these two one-bond coupling constants, there is no correlation between ^1J(N−B) for RBNR and ^1J(C−C) for RCCR when R is NH_2, OH, and F. From Figure 1, it is apparent that disubstitution of OH and F in the acetylenes significantly increases ^1J(C−C) for HOCCOH and FCCF relative to

Table 3. One- and Two-Bond B−H and N−H Coupling Constants and Fermi-Contact Terms (Hz) and B−H and N−H Distances (Å) for Molecules RBNH and HBNR

Coupling Constants				
RBNH	FC	^1J(N−H)	FC	^2J(B−H)
R = H	110.3	−111.9	54.8	57.6
CH_3	−110.3	−112.1	54.5	57.2
NH_2	−102.1	−104.4	46.0	47.6
OH	−116.9	−119.1	66.0	68.5
F	−119.8	−121.8	70.9	73.6
HBNR	FC	^1J(B−H)	FC	^2J(N−H)
R = H	230.7	231.0	−13.9	−15.5
CH_3	229.2	229.5	−13.2	−15.0
NH_2	239.4	239.6	−14.4	−16.3
OH	249.8	249.9	−15.2	−17.2
F	265.0	265.1	−16.9	−18.9

Distances				
	RBNH		HBNR	
molecule	N−H	B−H	B−H	N−H
R = H	0.996	2.243	1.168	2.415
CH_3	0.995	2.245	1.169	2.418
NH_2	0.997	2.195	1.168	2.415
OH	0.993	2.244	1.167	2.412
F	0.993	2.239	1.166	2.407

HOCCH and FCCH, respectively. In contrast, disubstitution of OH and F in the iminoboranes actually decreases ^1J(N−B) relative to either of the corresponding monosubstituted derivatives. The opposite effects of these two substituents on ^1J(N−B) and ^1J(C−C) are consistent with our earlier observations that disubstitution of OH and F have opposite effects on C−C and B−N bonds.[6] Disubstitution of these two groups appears to strengthen and shorten the C−C bond relative to the corresponding monosubstituted derivatives but lengthen the B−N bond relative to either of the corresponding monosubstituted derivatives, as evident from Tables 1 and 2, respectively. What remains unclear is why substitution of two NH_2 groups in iminoborane should increase ^1J(N−B) relative to either monosubstituted derivative, thereby having a similar effect to mono- and disubstitution of NH_2 in acetylene.

Substitution Effects on B−H and N−H Coupling Constants. Table 3 presents one- and two-bond B−H and N−H Fermi-contact terms and coupling constants and corresponding distances for the monosubstituted derivatives RBNH and HBNR. The contributions of all terms to these coupling constants are reported in Table S2 of the Supporting Information. Once again, the FC term dominates, contributing more than 90% of the coupling constant except for two-bond N−H coupling in HBNR molecules, in which case ^2J(N−H) is relatively small, ranging from −15 to −19 Hz. One-bond N−H distances in RBNH and B−H distances in HBNR show little sensitivity to substitution, varying by only 0.004 and 0.003 Å, respectively, in these two series. Despite the small variation in N−H and B−H distances, ^1J(N−H) varies from −104 to −122 Hz in the series RBNH, while ^1J(B−H) varies from 230 to 265 Hz in the series HBNR.

Patterns similar to those observed for B−N coupling constants in the two series RBNH and HBNR are also observed for one- and two-bond B−H and N−H coupling constants. Both B−H and N−H coupling constants are relatively insensitive to the replacement of H by CH_3. With

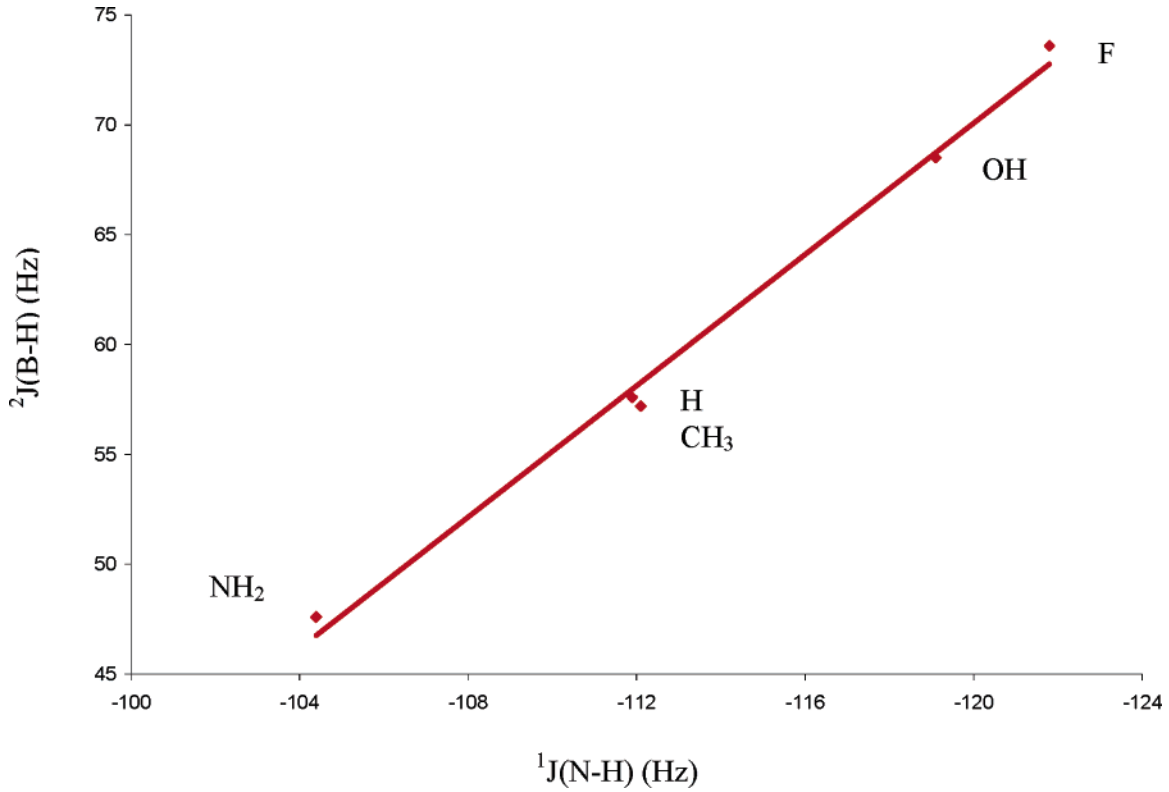


Figure 4. ^2J(B−H) vs ^1J(N−H) for molecules RBNH.

respect to the substituents, changes in ^1J(N−H) and ^2J(B−H) in the series RBNH, and ^1J(B−H) and ^2J(N−H) in the series HBNR, are similar to changes in ^1J(N−B) in the corresponding series. Figure 4 shows a linear correlation between ^2J(B−H) and ^1J(N−H) for iminoboranes RBNH. Just as NH_2 substitution at B decreases ^1J(N−B) relative to its value for HBNH, NH_2 substitution also decreases ^1J(N−H) and ^2J(B−H) relative to HBNH. OH and F substitution increases ^1J(N−H), ^2J(B−H), and ^1J(N−B) relative to the corresponding coupling constants for the parent molecule HBNH. B−H and N−H coupling constants in molecules HBNR are also linearly correlated.

All of the reduced one-bond B−N, B−H, and N−H coupling constants are positive in RBNH, HBNR, and RBNR, in agreement with the Dirac vector model,[37] which states that reduced one-bond coupling constants are positive, two-bond negative, three-bond positive, and so on. However, ^2J(N−H) values are negative for HBNR, and ^2J(B−H) values are positive for RBNH. Thus, the reduced two-bond coupling constants ^2K(N−H) and ^2K(B−H) in these series of molecules are both positive and, as such, are exceptions to this generalization.

Conclusions

In this study, ab initio EOM-CCSD calculations were carried out on iminoboranes RBNH, HBNR, and RBNR, for R = H, CH_3, NH_2, OH, and F, to evaluate substituent effects on one- and two-bond ^{15}N−^{11}B, ^{11}B−^{1}H, and ^{15}N−^{1}H coupling constants. For comparison purposes, calculations were also performed on corresponding isoelectronic acetylene derivatives RCCH and RCCR. The following statements are supported by the results of these calculations:

1. Mono- or disubstitution of CH_3 in iminoborane or acetylene has little effect on B−N or C−C coupling constants.

2. As the σ-electron-withdrawing ability of the substituent increases in the order CH_3 < NH_2 < OH < F, ^1J(C−C) increases in both series RCCH and RCCR. The increase in ^1J(C−C) upon disubstitution is more than double the increase upon monosubstitution.

3. For the monosubstituted iminoborane derivatives HBNR, ^1J(N−B) also increases as the σ-electron-withdrawing ability of the substituent increases in the order CH_3 < NH_2 < OH < F, as observed for ^1J(C−C) in the corresponding isoelectronic series RCCH.

4. ^1J(N−B) does not vary systematically with the σ-electron-withdrawing ability of the substituent in the series RBNH. NH_2 substitution decreases ^1J(N−B) relative to HBNH, while substitution of OH or F leads to similar increases for ^1J(N−B) in RBNH and HBNR. The sensitivity of ^1J(N−B) to the site of NH_2 substitution may be related to the significant loss of BN triple-bond character in H_2NBNH.

5. The effect of disubstitution in the series RBNR is dramatically different from the effect of monosubstitution. Disubstitution of NH_2 increases ^1J(N−B) relative to either monosubstituted derivative, while disubstitution of OH or F decreases ^1J(N−B) relative to the corresponding monosubstituted derivatives. Disubstitution of OH or F has the opposite effect on coupling constants in RBNR and RCCR.

6. Changes in B−N coupling constants in iminoboranes do not correlate with changes in B−N distances.

7. Changes in ^1J(N−H) and ^2J(B−H) for molecules RBNH, and ^2J(N−H) and ^1J(B−H) for HBNR, are similar to changes in ^1J(N−B) for RBNH and HBNR, respectively. Changes in N−H and B−H coupling constants in each series are linearly correlated.

Acknowledgment. This work was supported by the United States National Science Foundation through Grant CHE-9873815; the DGI Project Nos. BQU-2003-00894, BQU2003-06553, and BQU-2003-01251; and the Project MADRISOLAR, ref S-0505/PPQ/0225 of the Comunidad Autónoma de Madrid. The continuing support of the Ohio Supercomputer Center is also gratefully acknowledged.

Supporting Information Available: Tables reporting the PSO, DSO, FC, and SD components of one- and two-bond coupling constants for acetylenes and iminoboranes and the complete author list for ref 25, are included as Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96.

(2) Taft, R. L.; Lewis, I. C. *J. Am. Chem. Soc.* **1959**, *81*, 5343.

(3) Exner, O. *Collect. Czech. Chem. Commun.* **1966**, *31*, 65.

(4) Peyerimhoff, S. D. *Interactions in Molecules: Electronic and Steric Effects*; Wiley- VCH: Weinheim, Germany, 2003.

(5) Selassie, C. D. The History of Quantitative Structure Activity Relationships. In *Burger's Medicinal Chemistry and Drug Discovery*, 6th ed.; Abraham, D. J., Ed.; John Wiley and Sons Publishers: New York, 2003; Vol. 1, pp 1−48.

(6) Mó, O.; Yañez, M.; Martin-Pendás, A.; Del Bene, J. E.; Alkorta, I.; Elguero, J. *Phys. Chem. Chem. Phys.* (submitted).

(7) Del Bene, J. E.; Elguero, J.; Alkorta, I.; Yañez, M.; Mó, O. *J. Phys. Chem. A* **2006**, *110*, 9959.

(8) Del Bene, J. E.; Alkorta, I.; Elguero, J.; Yañez, M.; Mó, O. *J. Phys. Chem. A* **2007**, *111*, 419.

(9) Segawa, Y.; Yamashita, M.; Nozaki, K. *Science* **2006**, *134*, 113.

(10) Pople, J. A.; Binkley J. S.; Seeger, R. *Int. J. Quantum Chem., Quantum Chem. Symp*. **1976**, *10*, 1.

(11) Krishnan, R.; Pople, J. A. *Int. J. Quantum Chem.* **1978**, *14*, 91.

(12) Bartlett, R. J.; Silver. D. M. *J. Chem. Phys*. **1975**, *62*, 3258.

(13) Bartlett, R. J.; Purvis, G. D. *Int. J. Quantum Chem.* **1978**, *14*, 561.

(14) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(15) Spitznagel, G. W.; Clark, T.; Chandrasekhar, J.; Schleyer, P. v. R. *J. Comput. Chem.* **1982**, *3*, 3633.

(16) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.

(17) Perera, S. A.; Sekino, H.; Bartlett, R. J. *J. Chem. Phys.* **1994**, *101*, 2186.

(18) Perera, S. A.; Nooijen, M.; Bartlett, R. J. *J. Chem. Phys.* **1996**, *104*, 3290.

(19) Perera, S. A.; Bartlett, R. J. *J. Am. Chem. Soc.* **1995**, *117*, 8476.

(20) Perera, S. A.; Bartlett, R. J. *J. Am. Chem. Soc.* **1996**, *118*, 7849.

(21) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.

(22) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(23) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1995**, *103*, 4572.

(24) Kirpekar, S.; Jensen, H. J. A.; Oddershede, J. *Chem. Phys.* **1994**, *188*, 171.

(25) Stanton, J. F. *ACES II*; University of Florida: Gainesville, FL.

(26) Stothers, J. B. *Carbon-13 NMR Spectroscopy*; Academic Press: New York, 1972; p 370.

(27) Kamieńska-Trela, K. *J. Mol. Struct.* **1982**, *32*, 143.

(28) Wigglesworth, R. D.; Raynes, W. T.; Kirpekar, S.; Oddershede, J.; Sauer, S. P. A. *J. Chem. Phys.* **2000**, *112*, 3735.

(29) Kirpekar, S.; Enevoldsen, T.; Oddershede, J.; Raynes, W. T. *Mol. Phys.* **1997**, *91*, 897.

(30) Ruden, T. A.; Lutnæs, O. B.; Helgaker, T. *J. Chem. Phys.* **2003**, *118*, 9572.

(31) Contreras, R. H.; Barone, V.; Facelli, J. C.; Peralta, J. E. *Annu. Rep. NMR Spectrosc.* **2003**, *51*, 167−260.

(32) Ruden, T. A.; Ruud, K. Ro-Vibrational Corrections to NMR Parameters. In *Calculation of NMR and EPR Parameters*; Kaupp, M., Bühl, M., Malkin, V., Eds.; Wiley-VCH: Weinheim, Germany, 2004; pp 153−173.

(33) Kamieńska-Trela, K.; Kania, L.; Schilf, W.; Balova, I. *Spectrochim. Acta A* **1999**, 55, 817. Dąbrowski, A.; Kamieńska-Trela, K; Wójcik, J. *Spectrochim. Acta A* **2000**, *56*, 91. Kamieńska-Trela, K.; Kania, L.; Bernatowicz, P.; Bechcicka, M.; Kaczmarek, L.; Wójcik, J. *Spectrochim. Acta A* **2000**, *56*, 2079. Kamieńska-Trela, K. *Annu. Rep. NMR Spectrosc.* **1995**, *30*, 131.

(34) Del Bene, J. E.; Elguero, J. *Chem. Phys. Lett.* **2003**, *382*, 100.

(35) Del Bene, J. E.; Elguero, J.; Alkorta, I. *J. Phys. Chem. A* **2004**, *108*, 3662.

(36) Del Bene, J. E.; Elguero, J. *Magn. Reson. Chem.* **2006**, *44*, 784.

(37) Lynden-Bell, R. M.; Harris, R. K. *Nuclear Magnetic Resonance Spectroscopy*; Appleton Century Crofts: New York, 1969.

CT600351P

JCTC Journal of Chemical Theory and Computation

Improving Convergence of Replica-Exchange Simulations through Coupling to a High-Temperature Structure Reservoir

Asim Okur,[†] Daniel R. Roe,[†] Guanglei Cui,[†] Viktor Hornak,[#] and Carlos Simmerling*[,†,#,‡]

Department of Chemistry and Center for Structural Biology, Stony Brook University, Stony Brook, New York 11794, and Computational Science Center, Brookhaven National Laboratory, Upton, New York 11973

Received August 10, 2006

Abstract: Parallel tempering or replica-exchange molecular dynamics (REMD) significantly increases the efficiency of conformational sampling for complex molecular systems. However, obtaining converged data with REMD remains challenging, especially for large systems with complex topologies. We propose a new variant to REMD where the replicas are also permitted to exchange with an ensemble of structures that have been generated in advance using high-temperature MD simulations, similar in spirit to J-walking methods. We tested this approach on two non-trivial model systems, a β-hairpin and a 3-stranded β-sheet and compared the results to those obtained from very long (>100 ns) standard REMD simulations. The resulting ensembles were indistinguishable, including relative populations of different conformations on the unfolded state. The use of the reservoir is shown to significantly reduce the time required for convergence.

Introduction

Conformational sampling remains one of the largest challenges in simulating biologically relevant events in atomic detail. Even when a sufficiently accurate Hamiltonian of the system is used, the rugged and complex potential-energy surfaces usually result in simulations being trapped, prohibiting complete exploration of conformational space. Thus, significant effort has been put into devising efficient simulation strategies that locate low-energy minima for these complex systems. The challenges of conformational sampling have been discussed in several reviews.[1,2]

One major problem for molecular simulations is quasi-ergodicity where simulations may appear converged when observing some simulation parameters, but in reality large energy barriers may prevent them from sampling important regions of the energy landscape. Another simulation initiated in a different conformation may look converged as well, but comparison may show that only partial equilibration was achieved. An example of this behavior has been demonstrated by Smith et al. who reported that MD simulations of short peptides starting from different initial conformations were in poor agreement despite apparent convergence in some measured properties.[3]

One popular approach to overcoming quasi-ergodicity in a biomolecular simulation is the replica-exchange method .[4−8] In replica-exchange molecular dynamics (REMD)[9] (also known as parallel tempering[6]), a series of molecular dynamics simulations (replicas) are performed for the system of interest. In the original form of REMD, each replica is an independent realization of the system, coupled to a thermostat at a different temperature. The temperatures of the replicas span a range from low values of interest (experimentally accessible temperatures such as 280 or 300 K) up to high values (such as 600 K) at which the system is expected to rapidly overcome potential energy barriers that would otherwise impede conformational transitions on a computationally affordable time scale.

* To whom correspondence should be addressed. E-mail: carlos.simmerling@stonybrook.edu.
[†] Department of Chemistry, Stony Brook University.
[#] Center for Structural Biology, Stony Brook University.
[‡] Computational Science Center, Brookhaven National Laboratory.

10.1021/ct600263e CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/10/2007

At intervals during the otherwise standard simulations, conformations of the system being sampled at different temperatures are exchanged based on a Metropolis-type criterion[10] that considers the probability of sampling each conformation at the alternate temperature (further details are discussed in Methods). In this manner, REMD is hampered to a lesser degree by the local-minima problem, since simulations at low temperatures can escape kinetic traps by "jumping" directly to alternate minima being sampled at higher temperatures. Recently, it was shown that temperature distributions can be optimized to maximize the rate at which replicas traverse the temperature span. Moreover, the transition probability is constructed such that the canonical ensemble properties are maintained during each simulation, thus providing potentially useful information about conformational probabilities as a function of temperature. Because of these advantages, REMD has been widely applied to studies of peptide and small protein folding.[6,9,11−19]

For large systems, REMD can become intractable since the number of replicas needed to span a given temperature range increases with the square root of the number of degrees of freedom in the system.[20−23] Since the number of accessible conformations also typically increases with system size, the current computational cost for REMD simulations of large systems limits the simulation lengths to tens of nanoseconds per replica, which limits the ability to obtain converged ensembles for large systems. Several promising techniques have been proposed[20,24−27] to deal with this apparent disadvantage of REMD. To our knowledge, converged REMD simulations in explicit solvent from independent starting conformations have been reported only for short helical or unstructured peptides.[27−29]

Several studies have compared the sampling efficiencies of standard MD and REMD. Sanbonmatsu and Garcia reported a five-fold increase in sampled conformations using REMD over MD in the five-residue Met-enkephalin peptide in explicit solvent.[30] Zhang et al. showed that REMD enhances sampling over conventional MD by 15−70 times at different temperatures for the 21-residue Fs peptide in continuum solvent.[31] A recent study by Zuckerman and Lyman investigated the sampling efficiency of REMD through consideration of the rate acceleration afforded by increased temperature.[32] For slower converging systems (such as β-hairpins or more complex topologies where folding time is on the order of microseconds), REMD simulations are typically initiated from the native conformation (see recent example by Zhang et al.[33]) where unfolding through high-temperature replicas is obtained and temperature-dependent properties are calculated from the resulting structures.

REMD simulations increase conformational sampling over standard MD simulations, but reliable results for nontrivial systems are still challenging to obtain. It is possible that REMD does not provide even greater efficiency gains for peptides and proteins because the temperature dependence of the folding rate tends to be more weakly temperature dependent than the unfolding rate, as has been shown experimentally[34−38] and computationally.[39,40] When starting from non-native conformations, high-temperature replicas give limited advantage for finding native states since more minima on the free-energy landscape become accessible at higher temperatures, further complicating the search. Furthermore, when a high-temperature REMD replica locates a favorable low-energy basin (such as the native structure), this conformation is exchanged to lower temperature, and the high-temperature replica needs to repeat the search process. Importantly, during the search by the high-temperature replicas, all replicas continue to be simulated. Thus a very large set of simulations, all of which are long enough for the high-temperature replicas to sample multiple folding events, can be required to achieve correct Boltzmann-weighted ensembles across the range of replicas. From another perspective, REMD drives the generation of correct equilibrium ensembles of structures by employing an exchange criterion that explicitly assumes that structures being considered for exchange have Boltzmann-weighted probability of being sampled (see Methods for details). However, this assumption is only true after the generalized ensemble has already reached convergence and is typically incorrect at the start of the REMD simulation. Thus until all temperatures sample an equilibrium ensemble, none of the temperatures would be expected to have correct distributions because of coupling of replicas through an incorrect exchange probability.

An approach to the reduction of quasi-ergodicity that is conceptually similar to REMD was reported by Frantz et al. for Monte Carlo (MC) simulations of atomic clusters.[41] In their approach, called jump-walking (or J-walking), they coupled one MC simulation to another at higher temperature. Somewhat analogously to REMD, the low-temperature simulation was used to sample local minima and provide thermodynamic ensemble data at the temperature of interest, while the high-temperature simulation was used to facilitate barrier crossing. Periodically the low-temperature structures escape local minima by "jumping" to basins sampled at high temperatures. The Boltzmann distribution generated by the high-temperature walker becomes the sampling distribution for attempted jumps by the low-temperature walkers. One drawback is that too large a temperature difference results in poor acceptance probabilities for the jump, comparable to the need to optimize the spacing between REMD temperatures. Variations of the J-walking scheme were tested by employing high-temperature simulations on a different time scale than the low-temperature simulation or using multiple high-temperature simulations. They determined that the most efficient method is running the high-temperature walker to obtain an adequate distribution and using the stored conformations for jumps in a MC run at slightly lower temperature. The results of this lower-temperature run were then used as the seed set for a new J-walking run at even lower temperature. They validated this approach using simple double-well potentials where comparison to analytical results was possible and in simulations of Argon clusters of various sizes. Similar approaches to J-walking have been developed, such as smart walking (S-walking),[42] smart darting,[43] and cool walking .[44] The J-walking scheme has been adapted to REMD simulations that employ a resolution exchange scheme, where replicas were run using a coarse-grained

model to obtain conformations to be subsequently sampled by an all-atom model.[45,46]

Here we introduce a variant to REMD where we draw upon the strengths of the J-walking approach to overcome the slow convergence and high computational expense of REMD. Similar to J-walking, an ensemble of structures is generated using standard MD simulation at high temperature. Instead of the reduction of the temperature stepwise and re-equilibratation of the ensemble in stages, an REMD run is used to link, in a single step, the high-temperature ensemble to low temperatures of interest. Periodic exchanges are made between randomly chosen conformations from the reservoir set and the highest-temperature replica. This process formally provides correct ensembles at lower temperature with free energies that reflect the proper relative populations of minima. Importantly, the convergence speed of the REMD run is greatly enhanced since exchanges are attempted from an already converged Boltzmann ensemble and thus the exchange probabilities are correct at the start of the REMD run. We call this method reservoir REMD (R-REMD) because REMD is coupled to a high-temperature reservoir.

One major advantage of the reservoir approach with REMD is that a converged ensemble of conformations has to be generated only once and only for one temperature. After extensive conformational search at one temperature, the remaining temperatures can sample from and anneal these structures to rapidly construct equilibrium distributions consistent with their thermostat temperature. This is in contrast to the typical REMD approach where all replicas are run simultaneously, and the computational expense for running long simulations must be paid for each of the replicas even though only a few high-temperature ones may be contributing to the sampling of new basins. Another advantage is that the exchanges with the reservoir need not be time correlated with the replica simulations. Folding events sampled during reservoir generation can provide multiple native structures for the other replicas, in contrast to standard REMD where an independent folding event is required for each temperature that will have a substantial native population. Overall, the convergence rate for the set of replicas is greatly enhanced by exchanging with a previously converged ensemble.

We have implemented the reservoir REMD approach in the Amber[47] simulation package and have tested it on two models peptides, the trpzip2 β-hairpin[48] and the dPdP[49] three-stranded antiparallel β-sheet. These systems were selected because of the complexity and slow folding of β-sheets and hairpins as compared to that of α-helices, which fold rapidly enough that the performance advantage of R-REMD may not be apparent. For both systems, reservoir ensembles were generated at 400 K using the generalized Born[50] (GB) implicit solvent model using multiple simulations with different initial conditions. Subsequent R-REMD simulations were compared to standard REMD calculations with the same temperature ranges. In all cases, simulations were extended until close agreement was obtained between results obtained from independent runs with different initial structure ensembles (folded and unfolded). For both peptides, the use of reservoir structures is shown to provide the same structure ensembles and thermal melting profiles as standard REMD, with a reduction in overall computational cost of 5 to 20 times, including generation of the reservoir.

Methods

Replica Exchange Molecular Dynamics (REMD). We briefly summarize the key aspects of REMD as they relate to the present study. In standard parallel tempering or replica-exchange molecular dynamics,[6,9] the simulated system consists of M noninteracting copies (replicas) at M different temperatures. The positions, momenta and temperature for each replica are denoted by $q^{[i]}$, $p^{[i]}$, and T_m, $i = 1,..., M$ and $m = 1, ..., M$. The equilibrium probability for this generalized ensemble is

$$W(p^{[i]},q^{[i]},T_m) = \exp\left\{-\sum_{i=1}^{M}\frac{1}{k_B T_m}H(p^{[i]},q^{[i]})\right\} \quad (1)$$

where the Hamiltonian $H(p^{[i]},q^{[i]})$ is the sum of kinetic energy $K(p^{[i]})$ and potential energy $E(q^{[i]})$. For convenience, we denote $\{p^{[i]},q^{[i]}\}$ at temperature T_m with $x_m^{[i]}$ and further define $X = \{x_1^{[i(1)]}, ..., x_M^{[i(M)]}\}$ as one state of the generalized ensemble. We now consider exchanging a pair of replicas. Suppose we exchange replicas i and j, which are at temperatures T_m and T_n, respectively,

$$X = \{ ...; x_m^{[i]}; ...; x_n^{[j]}; ...\} \rightarrow X' = \{...; x_m^{[j]}; ...; x_n^{[i]}; ...\} \quad (2)$$

To maintain detailed balance of the generalized system, microscopic reversibility has to be satisfied, thus giving

$$W(X)\rho(X \rightarrow X') = W(X')\rho(X' \rightarrow X) \quad (3)$$

where $\rho(X \rightarrow X')$ is the exchange probability between two states X and X'.

A key step in the derivation of the exchange criterion[9] is the substitution of the Boltzmann factor for the weight of each conformation into eq 3, yielding eq 4. We note that this is not strictly correct until equilibrium has been reached, at which point the structures are actually considered for exchange with this probability.

$$\exp\left\{-\frac{1}{k_B T_m}H(p^{[i]}, q^{[i]}) - \frac{1}{k_B T_n}H(p^{[j]}, q^{[j]})\right\}\cdot\rho(X \rightarrow X') = \exp\left\{-\frac{1}{k_B T_m}H(p^{[j]}, q^{[j]}) - \frac{1}{k_B T_n}H(p^{[i]}, q^{[i]})\right\}\cdot\rho(X' \rightarrow X) \quad (4)$$

In the canonical ensemble, the potential energy, E, rather than the total Hamiltonian, H, can be used because the momentum can be integrated out .[9] By rearranging eq 4, one obtains the following Metropolis exchange probability

$$\rho = \min\left(1, \exp\left\{\left(\frac{1}{k_B T_m} - \frac{1}{k_B T_n}\right)(E(q^{[i]}) - E(q^{[j]}))\right\}\right) \quad (5)$$

It is important to reiterate that eq 4 is valid only for equilibrated ensembles that follow Boltzmann distributions. This assumption is true at the end of the simulation, and use of this exchange probability drives each replica toward adoption of the correct ensemble.

In standard REMD, several replicas at different temperatures are simulated simultaneously and independently for a chosen number of MD steps. Exchange between a pair of replicas is then attempted with a probability of success calculated from eq 5. If the exchange is accepted, the bath temperatures of these replicas will be swapped, and the velocities will be scaled accordingly. Otherwise, if the exchange is rejected, each replica will continue on its current trajectory with the same thermostat temperature.

Reservoir REMD (R-REMD). Reservoir REMD simulations (R-REMD) were run using same simulation parameters as standard REMD simulations. The only difference is that the highest temperature replica is replaced with a previously generated structure reservoir (replica R^N). Standard replicas (MD simulations) were used for each of the lower temperatures (replicas R^1 to R^{N-1}). Exchanges are attempted on the basis of the same criterion as used for standard REMD (eq 4). During the exchange attempts for replicas between R^1 and R^{N-1}, the exchange calculation is performed using current simulation coordinates. The only difference between R-REMD and REMD is when an exchange is attempted between replica R^{N-1} and the reservoir set R^N. The exchange attempt is made between the current structure of R^{N-1} and a randomly selected structure from the reservoir. If the exchange is accepted, the coordinates and velocities from R^N are sent to replica R^{N-1}. Formally the coordinates from replica R^{N-1} would be placed into the reservoir; however, for computational convenience, it was discarded since we assume that the reservoir constitutes a complete representation of the ensemble and that the inclusion of the new coordinates will have a negligible effect on the reservoir. Similarly, the chosen reservoir structure is left in the reservoir. Since the reservoir implemented for these tests is finite in size (10 000 structures), although the method is rigorous the present *implementation* does not formally obey detailed balance. For all practical purposes, the effect is minimal and such exchanges could readily be incorporated into the code.

Model Systems. The first model system chosen was the tryptophan zipper (trpzip) developed by Starovasnik and coworkers.[48] This β-hairpin structural motif is stabilized through cross-strand tryptophan pairs. Trpzip2 (SWTWENGKWTWK, with a type I′ β-turn at NG) has the most cooperative melting curve and highest stability (~90% at 300 K) among the trpzips and was selected for use in this study. Thermodynamic properties for this peptide have been determined by NMR and CD spectroscopy, and a family of structures was refined using restraints from NMR experiments[48] (PDB code 1LE1). The N-terminal of the peptide was acetylated and the C-terminal was amidated, in accordance with the experiments.

The second model system was created from the sequence of DPDP (VFITSdPGKTYTEVdPGOKILQ, dP = D-proline, O = ornithine), except that lysine was substituted for the ornithine. The replacement of ornithine with lysine in a related peptide analogous to the C-terminal hairpin of DPDP caused no detectable effect on the structure .[51] The termini were amidated and acetylated in accordance with experiments. DPDP was designed with a net charge of +2 to prevent aggregation, and our model retains this net charge.[49]

REMD Simulations. For both systems, standard REMD simulations were carried out with Amber, version 8.[47] For trpzip2, all covalent bonds were constrained using SHAKE .[52] For dPdP, only the bonds involving hydrogen atoms were constrained. A 2fs time step was used, and temperatures were maintained using weak coupling[53] to a bath with a time constant of 0.5 ps^{-1}. All nonbonded interactions were calculated at each time step (i.e., no cutoff was used). To permit comparison to our previously published data, both peptides were simulated with the Amber ff99 force field with modified backbone parameters .[54] Steepest-descent energy minimization was performed for both systems for 500 steps prior to REMD simulations. Both systems were simulated with the generalized Born solvation model[50] with GB^{HCT} [55] implementation in Amber. Scaling factors were taken from the TINKER modeling package .[56]

Standard REMD simulations were performed for both systems using 14 replicas for trpzip2 and 12 replicas for dPdP, covering a temperature range of ~260−570 K with an expected exchange probability of 15%. For trpzip2, additional replicas were manually placed between 300 and 370 K to increase statistics around the experimentally observed melting transition. Exchanges between neighboring replicas were attempted at 1ps intervals.

For both systems, two independent replica exchange simulations were run. For trpzip2, one simulation initiated all replicas in the published native conformation. The other simulation started from a compact non-native conformation where no hairpin backbone hydrogen bonds were present. Both REMD simulations were run to 155 000 exchange attempts (155 ns per replica). dPdP simulations were run as explained in our previous work,[17] with a simulation starting with all replicas in a fully extended structure and another with all replicas in a compact non-native structure. dPdP simulations were carried out for 170 000 exchange attempts (170 ns per replica).

Generation of Reservoir Structures. The reservoir structures were generated through molecular dynamics at 400 K with the same simulation parameters as used for REMD simulations. For trpzip2, four independent MD simulations each of ~38 ns in length were run starting from an extended conformation. Multiple folding and unfolding transitions were observed for each trajectory. For dPdP, a single long trajectory of 260 ns was generated. Multiple folding and unfolding transitions were observed. For both systems, velocities and coordinates were saved at 1 ps intervals. In the present implementation of R-REMD in Amber, coordinates and velocities for the reservoir were loaded into memory at the start of the R-REMD run. To minimize memory requirements, the reservoir ensembles were reduced to 10 000 structures by selecting equidistant snapshots from the trajectories.

Reservoir REMD Simulations. For trpzip2, four replicas were used below the 400 K reservoir with temperatures of 300, 323, 350, and 373 K. No additional replicas were used since these four replicas were sufficient to provide a 25−30% exchange ratio. Two sets of R-REMD simulations were

each run for 50 000 exchange attempts, starting from the same native or unfolded initial conformations as used for the standard REMD calculations.

Since dPdP is a larger system, the R-REMD simulations used 6 replicas below 400 K with the same temperature distribution as the standard REMD reported by Roe et al.[17] One R-REMD simulation starting from extended conformations was run for 50 000 exchange attempts.

Analysis. The trajectories obtained from standard and reservoir REMD simulations were analyzed using the Amber ptraj module. Trpzip2 simulations were compared to the experimentally determined native structure[48] (model 1 of PDB code 1LE1), where backbone rmsd's were calculated for residues 2−11. Terminal residues were omitted to remove the effects of fluctuations. An rmsd cutoff of 1.7 Å was used to determine native structures on the basis of the free-energy profile along rmsd where the native minimum reached up to 1.7 Å (data not shown). For dPdP, the fraction of native contacts were calculated, and the native population was calculated using a cutoff of 0.50 for both hairpin1 and hairpin2 contacts (as described in Roe et al.).[17]

Melting curves were generated by calculating the average population of native structures at each temperature. For trpzip2 simulations, data from the first 55 000 exchange attempts were discarded for each standard REMD simulation to remove initial structure bias. For dPdP, REMD simulations data from the first 20 000 exchange attempts were discarded.

Native fractions as a function of time were calculated by averaging the native population up to that time point for both systems using their respective criteria. For all systems, the rate of convergence was observed by comparing populations starting from different initial conformations. When both simulations show similar observables and a flat profile is obtained for all temperatures, the simulations are classified as converged.

Cluster analysis was performed as described previously[27] using the Moil-View program.[57] The trajectories from standard and reservoir REMD simulations were combined. Cluster analysis was performed on the combined set, and then normalized populations for each cluster type were calculated for each of the original simulations. This process permits direct comparison of the populations since the structure families are defined using the combined trajectories.

Results and Discussion

We apply the R-REMD method to two model systems (trpzip2 and dPdP) that we have studied previously using standard REMD. To validate the R-REMD approach, we first compare the resulting structure ensembles to those obtained with standard REMD to show that R-REMD provides accurate results. Next, we examine whether R-REMD provides these results more efficiently than standard REMD.

Trpzip2 REMD Simulations. We performed 2 independent REMD simulations of the trpzip2 peptide, one starting with all replicas in the published NMR structure (native) and one from a compact non-native structure. Both simulations were run ∼155 000 exchange attempts (equivalent to 155 ns per replica) where 14 replicas were used to cover a temperature range of 260−570K.

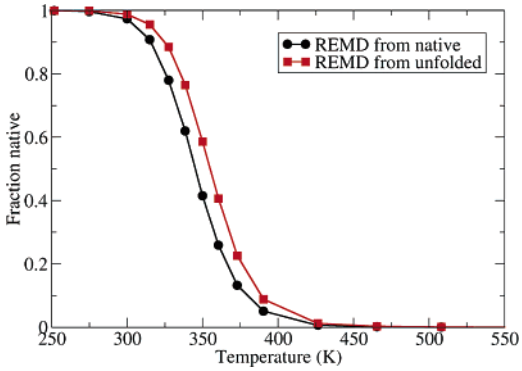


Figure 1. Melting curves for trpzip2 REMD simulations starting from native and unfolded conformations. Symbols represent temperatures at which simulation data is obtained. The similar profiles suggests that the data is reasonably well converged. Simulations show melting temperatures of 342.3 and 352.4 K, in excellent agreement with the experimentally measured value of 345 K.

Even though trpzip2 is a small system, long simulations were required to obtain good agreement between simulations with different initial conformations. Throughout the simulations the melting profiles were monitored and compared. After ∼150 ns, both REMD simulations showed similar melting profiles which no longer changed with increasing simulation times. The convergence rates of each simulation will be discussed later in this section. Since significant time was required to overcome the bias from initial conformations, data from the first 55 000 exchange attempts (55 ns) were discarded for constructing the melting curves (Figure 1). It should be noted that the amount discarded is larger than the total simulation time of most current published REMD studies. Simulations starting from unfolded conformations show slightly higher stability than those initiated with the native state, suggesting that these differences involve fluctuations in the data and do not reflect initial structure bias. As determined by fitting of the native fractions to the Gibbs−Helmholz equation, both simulations show comparable thermodynamic properties with melting temperatures of 342.4 and 352.4 K and ΔH_m of −15.90 and −16.46 kcal mol^{-1}. These values are in excellent agreement with the experimental melting temperature of 345 K and ΔH_m of −16.8 kcal mol^{-1}.[48] While the accuracy of the force field is not the subject of this study, it indicates that we are evaluating the performance of the R-REMD method under conditions that are relevant to experimental observations.

Testing the Accuracy of R-REMD. After the benchmark results were obtained using converged standard REMD simulations, we generated the high-temperature reservoir ensemble at 400 K. We chose 400 K because it is high enough to allow rapid conformational transitions, and it is well above the T_m, thus requiring R-REMD to significantly transform the reservoir ensemble to obtain accurate ensembles at lower temperatures.

Four standard molecular dynamics simulations were performed at 400 K using conditions identical to those for standard REMD simulations. Each simulation was run for ∼38 ns with a cumulative simulation time of ∼152 ns, where multiple folding and unfolding transitions were observed for

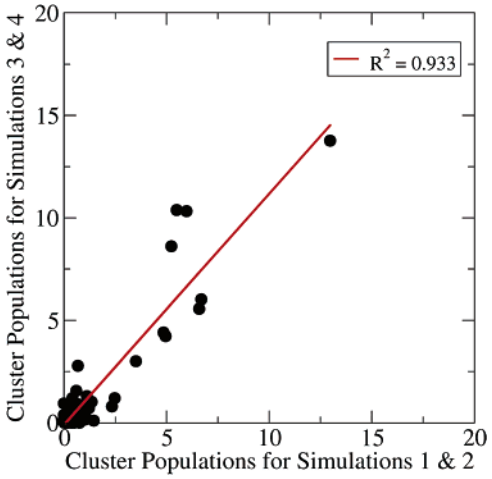


Figure 2. Populations of different trpzip2 structure clusters sampled by standard MD simulations. Populations of the first two trajectories are compared to populations of the same clusters in the remaining two trajectories. All clusters with large populations in runs 1 and 2 are also present with similar populations in runs 3 and 4, suggesting good convergence.

each trajectory (Figure S1). The presence of reversible folding transitions during standard MD is a reasonable indicator that the ensemble is fairly well converged (discussed in more detail below). Because of the elevated temperature, rapid unfolding takes place after each folding event, and the native population for each simulation is between 1 and 5%, in good agreement with the melting curves shown in Figure 1 (3.3 and 6.1% native populations at 400 K, calculated using the Gibbs−Helmholz equation and native fractions at the other temperatures).

Following our previously published work,[27] we evaluate the population of each cluster to determine whether independent simulations provide the same ensembles. This analysis is more rigorous than that comparing only native populations. Cluster analysis resulted in 136 structure families. In Figure 2, we compare the populations of each family sampled in the first two trajectories to the populations from the other two trajectories. A good correlation is observed, suggesting that the simulations not only sample the same types of structures, but that the relative population of each structure family is similar. While the composition of the unfolded ensemble will be discussed elsewhere, it is important to note that the most populated clusters (10−15% of the ensemble) are non-native at this elevated temperature, with a native population of only ∼3%.

This pool of 10 000 structures (coordinates and velocities) was used as the reservoir set for the R-REMD simulations. Four replicas were used with temperatures 300, 323, 350, and 373 K, where the 373 K replica periodically attempted to exchange with the 400 K reservoir as described in Methods. Two sets of R-REMD simulations with different initial structures were run for 50 000 exchange attempts. During the simulations, 25−30% exchange ratios are observed between replicas, and a 30% ratio is observed between the 373 K replica and the 400 K reservoir. Potential-energy overlaps were adequate for all temperature pairs (Figure S2).

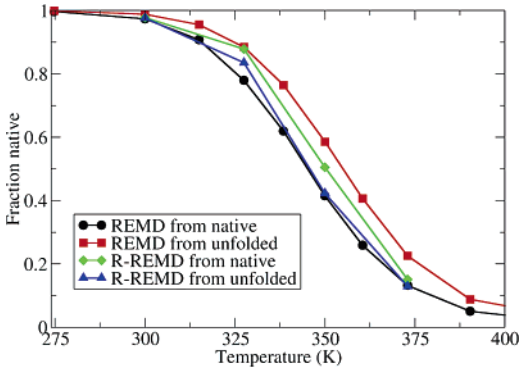


Figure 3. Thermal melting profiles for trpzip2 obtained from standard REMD (black and red) and R-REMD simulations (blue and green). Symbols represent temperatures at which simulation data is obtained. Standard REMD simulations are shown in black and red and R-REMD results are shown in green and blue. For easier comparison, only temperatures below 400 K are shown. Both R-REMD simulations are in good agreement with each other and lie fully within the precision range defined by the standard REMD results.

We next evaluated whether the use of the reservoir had any negative impact on the accuracy of the simulations. We calculated the thermal melting profiles for ensembles from the R-REMD simulations using the same procedure that was used for the standard REMD data. In Figure 3, we show the comparison of these melting curves to those from standard REMD. Excellent agreement is observed; the melting curves from the two R-REMD simulations lay within the bounds defined by the curves obtained from the two standard REMD simulations. Importantly, the R-REMD ensembles at low temperature are nearly fully native despite the low (3%) native population in the reservoir; thus, the REMD replicas are capable of accurately transforming the ensemble in the reservoir to what should be sampled at alternate temperatures. This result also suggests that it is possible to use this method for structure prediction, since the native conformation at low temperature is correctly identified despite the fact that it is not the most populated structure type in the high-temperature reservoir (Figure 2).

Figure 3 shows a striking agreement between the melting profiles obtained using standard REMD and reservoir REMD simulations. As we noted above, however, analysis of only native populations gives an incomplete view of the composition of an ensemble of structures. To be able to more fully evaluate the ensembles provided by R-REMD, one must compare populations not only of the native conformation but of all accessible states. We selected the ensemble at 350 K for this analysis; the proximity to the T_m makes this an excellent temperature to characterize the ensemble under conditions where native and non-native conformations are well populated. Cluster analysis on the combined set of structures sampled at 350 K in all REMD and R-REMD simulations resulted in 63 clusters with the native conformation being the highest-populated cluster in each simulation (Figure 4). We note that the most populated cluster is different at these temperatures (native at 350 K and non-native at 400 K).

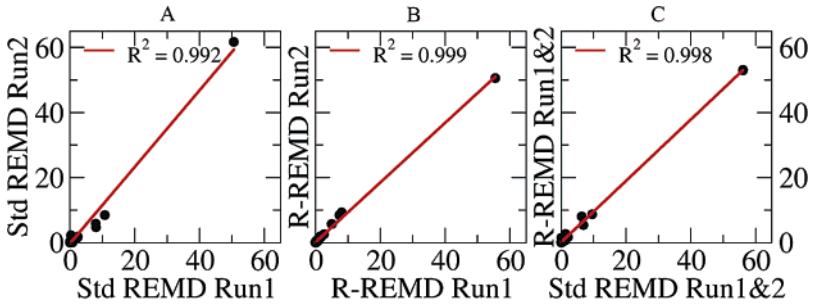


Figure 4. Comparison of the populations of a set of trpzip2 structure types sampled in different simulations. Structure families are defined using the combined set of structures, permitting direct comparison of populations between trajectories. (A) comparison of standard REMD from native vs standard REMD from unfolded. (B) Comparison of R-REMD from native vs R-REMD from unfolded. (C) Comparison of the combined data from standard REMD and the combined data from R-REMD. High correlations were observed in each case ($R^2 \approx 0.99$), and the most populated cluster is the same in all runs. Regression analysis after discarding the most populated cluster results in a similar level of agreement.

Standard REMD simulations starting from different initial conformations show high correlation between cluster populations ($R^2 > 0.99$), suggesting that the ensembles are well converged and the data are suitable as a reference to evaluate R-REMD results (Figure 4A). Similarly both R-REMD simulations starting from different conformations are in excellent agreement with $R^2 > 0.99$ (Figure 4B). Having thus validated the precision of the results from each method, we compare the populations of different structures in the ensemble obtained from standard REMD to that from R-REMD (Figure 4C). The agreement between the two data sets is impressive, with $R^2 = 0.998$ and a slope of 0.932.

The regression data obtained using all clusters may be biased by a single cluster with large population (native). We repeated the regression analysis for the data shown in Figure 4 after removal of this data point, thus comparing the preference to sample the various weakly populated structures in the unfolded state. For all cases, the resulting fit is similar to the original, with correlation coefficients of 0.974 (A), 0.997 (B), and 0.966 (C) between the unfolded ensembles sampled in the REMD and R-REMD simulations. Thus, we conclude that the ensemble obtained from R-REMD is essentially indistinguishable from that obtained using standard REMD, including the relative populations of the various conformations that make up the unfolded state.

Testing the Efficiency of R-REMD. We have demonstrated that R-REMD can produce the same ensembles of structures as standard REMD, validating the accuracy of the approach. We next investigate whether R-REMD offers any advantage over standard REMD in terms of computational cost. To analyze the rate of convergence, the population of native conformation with respect to simulation time was calculated for each simulation and temperature.

Figure 5A shows the native populations vs time for several temperatures in the two independent standard REMD simulations. As expected the values undergo very large fluctuations at the beginning of the REMD run and slowly approach their equilibrium values (obtained by combining the two data sets and discarding a significant amount of data to remove bias from initial conditions as described for Figure 3). After 155 000 exchange attempts (155 ns per replica), populations near the melting temperatures still fluctuate and do not show a flat profile with increasing simulation time. It is interesting

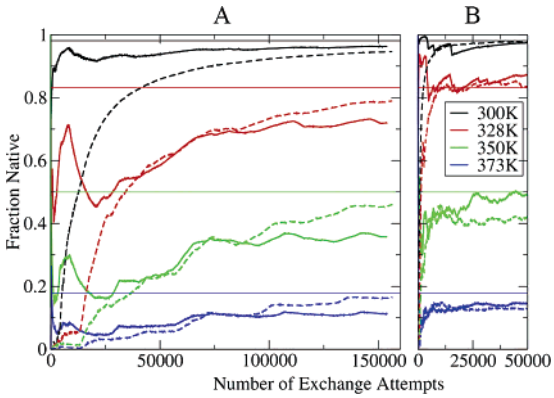


Figure 5. Convergence of native population in standard REMD runs (left) and R-REMD runs (right) vs number of exchange attempts. Solid lines represent simulations starting from native conformation and dashed lines represent simulations starting from unfolded conformations. Thin lines on both graphs represent the average equilibrium values obtained from the standard melting curves (Figure 3). For both graphs, the *x* axis is on the same scale. For standard REMD (left), the results fluctuate at the beginning of the simulations and slowly converge to their equilibrium values. Even though the simulations were extended to 155 000 exchange attempts, the average native populations show about 10% deviation between the two runs at multiple temperatures and plateau values have not been reached. R-REMD simulations (right) converge much faster (~5000−10 000 exchange attempts).

to note that the simulation initiated with all replicas in the native conformation still underestimates the equilibrium native population. Data near the thermal melting transition (where native and non-native conformations are both sampled) is critically important for characterizing the folding landscape. Even at 100 ns per replica, the population values differ significantly from the final values. Importantly, the populations from the two independent simulations provide similar values (i.e., good precision) at times where the population value is dramatically different from the final value (poor accuracy), indicating that precise results for the native population are not a reliable indicator of the overall convergence of the data. As an example, if we perform cluster analysis on the ensembles sampled up to 50 000 exchange attempts, the native population in both simulations

is similar (55% and 60%). However, the correlation coefficient for the populations of unfolded conformations is only 0.796, showing that, even though the largest cluster populations agree with each other, the overall sampling is not complete and the unfolded state and folding landscape may be poorly converged.

In marked contrast to the slow convergence obtained with standard REMD, both R-REMD simulations reach their equilibrium values after only 10 000 exchange attempts and fluctuate around this value for each temperature (Figure 5B). As observed in the melting curves, good agreement between the two methods over the temperature range is observed. The results seem to differ about 7% at 350 K, which is reasonable since the melting transition is sharp around this temperature, and this small difference corresponds to only ~0.16 kcal/mol difference in free energy (49.0 vs 42.3%). Overall, the R-REMD simulations converge to their equilibrium values much faster than standard REMD simulations. Standard REMD simulations have not reached their equilibrium values even at 150 000 exchange attempts (150 ns per replica). In contrast, R-REMD simulations reach their equilibrium values in ~5000−10 000 exchange attempts and remain near these values throughout the remainder of the simulations. This represents an improvement of over an order of magnitude in efficiency with R-REMD as compared to standard REMD.

Up to this point, the R-REMD simulations were compared to standard REMD simulations that employed a much larger temperature range (up to 570 K). For examination of computational efficiency, however, a more direct comparison between standard and R-REMD would involve using the same number of replicas and temperatures for each method. To test this, a new REMD simulation was prepared starting from the same unfolded conformation used for R-REMD but with 5 replicas: 4 matching the temperatures used in the R-REMD run and an additional replica at 400 K. The only difference between this REMD run and the previous R-REMD run is that the 400 K trajectory is a continuous simulation with exchanges that are synchronized with the other replicas instead of being chosen randomly from the pre-generated 400 K structure reservoir used for R-REMD. The native populations versus time for this REMD simulation are shown in Figure S3. These standard REMD simulations with a highest temperature of 400 K converge much more slowly than the original standard REMD which used more replicas covering a wider temperature range. After 180 000 exchange attempts (180 ns per replica), the replicas still did not reach the equilibrium values determined from standard REMD runs, and they also show relatively little progress toward these values. This slow convergence is somewhat unexpected since this REMD simulation was run longer than the cumulative simulation time of our standard MD simulations at 400 K (180 ns per replica vs 152 ns of standard MD), and these standard simulations were shown to be reasonably well converged (Figure 2 and S1). We believe that this difference in convergence between high-temperature MD and REMD demonstrates the effect of "scavenging" of low-energy structures sampled at the highest temperature by the lower temperatures, slowing the convergence of the high-temperature REMD ensemble. This interpretation is consistent with the observation that the lowest temperature converges within ~50 ns to nearly fully native ensemble; once this structure is located at higher *T* and exchanged to the lowest *T*, it becomes trapped and no further exchanges take place (and will not until other low-energy basins are located at higher temperatures). Thus the rapid convergence of this low temperature is not an adequate measure for simulation convergence at higher temperatures. As discussed above, temperatures such as 350 K where the native state does not fully dominate the ensemble are likely to be much more useful in characterizing the folding landscape and composition of the unfolded state. The poor convergence of standard REMD at these temperatures and the rapid convergence of R-REMD under otherwise identical conditions confirm that using an equilibrated structure reservoir instead of a synchronous high-temperature replica significantly increases the rate of convergence of REMD simulations.

One remaining question with the R-REMD simulations is how much the convergence rate and final results depend on the composition of the reservoir set. We tested this dependence by repeating the R-REMD run from an initial unfolded ensemble, using only the first half of the original structure reservoir (corresponding to two of the four MD trajectories at 400 K). The resulting pool of 5000 structures had a native population of ~1.5%.The resulting R-REMD melting curve is shown in Figure S4, along with those obtained from standard REMD and R-REMD with the larger reservoir. The thermal stability of trpzip2 in the R-REMD run with the smaller pool is somewhat lower, with ~15 K reduction in the midpoint of the melting transition. This likely results from a lower population of native conformations in the smaller reservoir (~1.5 vs 3%). Even with this much smaller native population in the reservoir, the R-REMD run shows good agreement at the lowest temperatures away from the reservoir, and the native population at higher temperatures is reduced accordingly. The simulations still converge as fast as the R-REMD simulations using the full structure pool (data not shown), suggesting that repeating R-REMD simulations with independent reservoirs would be an excellent approach to validating data convergence.

Testing R-REMD Performance with an Antiparallel β-sheet. To test the efficiency of the R-REMD method on a different and more challenging system, we simulated the peptide dPdP, which has been shown to adopt a 3-stranded antiparallel β-sheet.[49,51] We previously reported results from independent standard REMD simulations starting from fully extended and compact initial conformations.[17] Here, we compare those results to data from new simulations performed using R-REMD, starting from a fully extended conformation.

We employed a single long MD simulation of dPdP to generate the structure reservoir (260 ns, with 5 folding transitions observed). The reservoir was again generated at 400 K, and the native content in the resulting ensemble was 7.7%, in reasonable agreement with data at 399 K in our standard REMD simulations (4.5 and 4.7% in the independent REMD runs). Once again 10 000 structures were selected at equal intervals for use as the structure reservoir. Since dPdP is a larger system than trpzip2, six replicas were

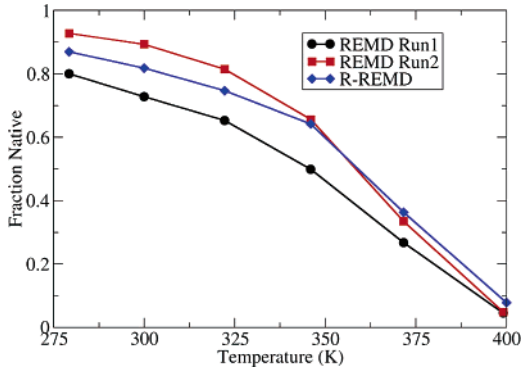


Figure 6. Comparison of dPdP melting curves from standard REMD simulations (black and red) and R-REMD simulation (blue). For standard REMD simulations, data from the first 20 000 exchange attempts were discarded to remove bias introduced by initial conformations. For the R-REMD simulation, the 400 K population reflects the reservoir ensemble.

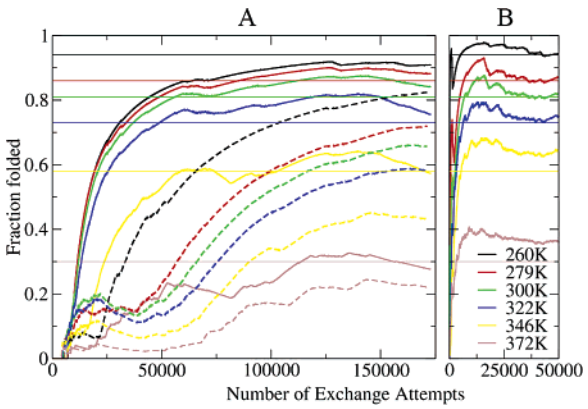


Figure 7. Native fraction vs number of exchange attempts for standard REMD simulation (A) and R-REMD simulation (B). Solid lines in panel A represent simulations starting from compact non-native structure, and dashed lines represent simulations starting from extended conformation. Even after 170 000 exchange attempts, plateau values have not been reached. During R-REMD simulations (B), all replicas converge to their equilibrium values after ~10 000 exchange attempts and show a flat profile thereafter.

used with the same temperature distribution that we employed in the standard REMD simulations resulting in 15−20% exchange ratios between replicas and 14% between highest replica and the 400 K reservoir. No data were discarded, since within 28 exchange attempts every one of the initial fully extended conformations had been exchanged with the reservoir, as expected since the fully extended conformation is energetically less favorable than the MD-generated conformations in the reservoir.

We compare the dPdP melting curves from our standard REMD simulations with the R-REMD results in Figure 6. As we observed with trpzip2, good agreement is obtained between REMD and R-REMD, with the R-REMD melting profile falling within the precision bounds obtained from the two independent standard REMD runs. As we reported previously,[17] these values are in good agreement with experimental observations.

Having confirmed that R-REMD is once again able to accurately reproduce the thermal melting profiles obtained using standard REMD, we evaluated how long it took each simulation to reach these equilibrium values (Figure 7). Even after 170 000 exchange attempts (~2 μs per run) for standard REMD simulations it was not possible to conclude that the simulations were well converged since the populations at some temperatures varied more than 10% and, in many cases, a plateau had not yet been reached.

In contrast with the standard REMD results, dPdP R-REMD simulations reach their equilibrium values within ~10 000 exchange attempts and show an essentially flat profile after that point. Simulations were continued up to 50 000 exchange attempts, with no significant changes for any of the simulated temperatures. The total simulation time used to obtain fully converged ensembles using R-REMD, including reservoir generation (260 ns), was ~320 ns, although we did not test whether a shorter reservoir generation simulation would have been sufficient. On the basis of this time, we estimate that R-REMD is at least 6.4 times more efficient than standard REMD.

Conclusions

We introduced a new variant of the replica exchange method where slow convergence and the high computational cost of REMD have been greatly improved by coupling of the REMD replicas to an ensemble of conformations that is generated in advance, similar in spirit to J-walking schemes. This approach builds on the hypothesis that the main contribution to sampling efficiency during REMD is obtained from the replicas exploring the free-energy landscape at high temperatures. Rather than simulating all replicas during this search process, R-REMD performs the search for alternate local minima *in advance* and subsequently uses a relatively short REMD run to generate accurate Boltzmann-weighted ensembles at other temperatures. An important advantage is that exchanges with the reservoir need not be time-correlated with the replica simulations, permitting REMD replicas to obtain many low-energy (such as native) conformations from a smaller number of folding events; this is not possible with standard REMD, which may a contributing factor in slow convergence.

In our current implementation in Amber, we employed a relatively small ensemble of 10 000 structures in the reservoir. Although successful exchanges would, in principle, involve placing the structure from the MD replica into the reservoir and removing the previous reservoir structure, we made the approximation that the reservoir is large enough that these changes would not affect the reservoir ensemble, and therefore, reservoir was not modified for computational convenience. As a result, the present implementation in Amber does not strictly obey detailed balance conditions; however, the resulting ensembles are in very good agreement with the standard REMD data. Furthermore, the use of a larger reservoir or true exchange of conformations could be readily accommodated in future implementations, and we note that the method itself is rigorous.

We tested R-REMD by comparing to standard REMD results for two systems, a β-hairpin and a three-stranded

β-sheet, under conditions in which the standard REMD data were in good agreement with experimental observations. We find that the thermal melting profiles obtained from R-REMD simulations were highly accurate compared to those of standard REMD, as expected because of the lack of any approximations in development of the method. Furthermore, excellent agreement was noted between the compositions of the structure ensembles obtained from standard REMD and R-REMD, including very high correlations between the two methods for the populations of native and non-native conformational families.

To summarize the efficiency comparisons, standard REMD simulations with 14 trpzip2 replicas were run for 155 ns per replica from two initial conformations resulting in a cumulative simulation time of ~4.3 μs, and they still did not fully converge. The R-REMD simulations were run using 4 replicas and two initial conformations, and both runs reached their equilibrium values in under 10 ns per replica (40 ns total). Generation of the reservoir does require additional computational effort that must be included in the comparison. In the present case, four simulations of ~40 ns were employed (152 ns, almost as long as each replica in the REMD simulations but importantly only 1 temperature is needed for R-REMD). The ability to use multiple simulations provides the reservoir generation with parallel efficiency comparable to the REMD simulations. Thus the cumulative simulation time for R-REMD *including the reservoir generation* is about 232 ns, approximately 19 times more efficient than the less well converged 4.3 μs standard REMD simulation. These values result from comparison to a primitive standard REMD implementation in Amber. Future studies will compare the efficiency of R-REMD to other, more recent REMD variants, as well as investigate how the R-REMD approach can be further optimized

Acknowledgment. C.S. thanks Adrian Roitberg for helpful discussions. C.S. gratefully acknowledges financial support from the NIH (GM6167803) and DOE (DE-AC02-98CH10886) and significant supercomputer time at NCSA (NPACI MCA02N028).

Supporting Information Available: Figures S1−S4. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tai, K. Conformational sampling for the impatient. *Biophys. Chem.* **2004**, *107* (3), 213−220.

(2) Roitberg, A.; Simmerling, C. Special issue: Conformational sampling. *J. Mol. Graphics Modell.* **2004**, *22* (5), 317−317.

(3) Smith, L. J.; Daura, X.; van Gunsteren, W. F. Assessing equilibration and convergence in biomolecular simluations. *Proteins: Struct., Funct., Genet.* **2002**, *48* (3), 487−496.

(4) Geyer, C. J.; Thompson, E. A. Annealing Markov-chain Monte-Carlo with applications to ancestral inference. *J. Am. Stat. Assoc.* **1995**, *90* (431), 909−920.

(5) Hukushima, K.; Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **1996**, *65* (6), 1604−1608.

(6) Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281* (1−3), 140−150.

(7) Swendsen, R. H.; Wang, J. S. Replica Monte-Carlo simulation of spin-glasses. *Phys. Rev. Lett.* **1986**, *57* (21), 2607−2609.

(8) Tesi, M. C.; van Rensburg, E. J. J.; Orlandini, E.; Whittington, S. G. Monte Carlo study of the interacting self-avoiding walk model in three dimensions. *J. Stat. Phys.* **1996**, *82* (1−2), 155−181.

(9) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141−151.

(10) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21* (6), 1087−1092.

(11) Feig, M.; Karanicolas, J.; Brooks, C. L. MMTSB tool set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **2004**, *22* (5), 377−395.

(12) Garcia, A. E.; Sanbonmatsu, K. Y. Exploring the energy landscape of a β hairpin in explicit solvent. *Proteins: Struct., Funct., Genet.* **2001**, *42* (3), 345−354.

(13) Garcia, A. E.; Sanbonmatsu, K. Y. α-Helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (5), 2782−2787.

(14) Karanicolas, J.; Brooks, C. L. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: Lessons for protein design? *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (7), 3954−3959.

(15) Kinnear, B. S.; Jarrold, M. F.; Hansmann, U. H. E. All-atom generalized-ensemble simulations of small proteins. *J. Mol. Graphics Modell.* **2004**, *22* (5), 397−403.

(16) Pitera, J. W.; Swope, W. Understanding folding and design: Replica-exchange simulations of "Trp-cage" miniproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (13), 7587−7592.

(17) Roe, D. R.; Hornak, V.; Simmerling, C. Folding cooperativity in a three-stranded β-sheet model. *J. Mol. Biol.* **2005**, *352* (2), 370−381.

(18) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113* (15), 6042−6051.

(19) Zhou, R. H.; Berne, B. J.; Germain, R. The free energy landscape for β hairpin folding in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (26), 14931−14936.

(20) Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. Modified Replica Exchange Simulation Methods for Local Structure Refinement. *J. Phys. Chem. B* **2005**, *109* (16), 8220−8230.

(21) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116* (20), 9058−9067.

(22) Kofke, D. A. On the acceptance probability of replica-exchange Monte Carlo trials. *J. Chem. Phys.* **2002**, *117* (15), 6911−6914.

(23) Rathore, N.; Chopra, M.; de Pablo, J. J. Optimal allocation of replicas in parallel tempering simulations. *J. Chem. Phys.* **2005**, *122* (2), 024111.

(24) Jang, S. M.; Shin, S.; Pak, Y. Replica-exchange method using the generalized effective potential. *Phys. Rev. Lett.* **2003**, *91* (5), 058305.

(25) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test. *J. Chem. Phys.* **2003**, *118* (14), 6664−6675.

(26) Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.* **2000**, *329* (3−4), 261−270.

(27) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *J. Chem. Theory Comput.* **2006**, *2* (2), 420−433.

(28) Nymeyer, H.; Garcia, A. E. Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (24), 13934−13939.

(29) Wickstrom, L.; Okur, A.; Song, K.; Hornak, V.; Raleigh, D. P.; Simmerling, C. L. The unfolded state of the villin headpiece helical subdomain: Computational studies of the role of locally stabilized structure. *J. Mol. Biol.* **2006**, *360* (5), 1094−1107.

(30) Sanbonmatsu, K. Y.; Garcia, A. E. Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins: Struct., Funct., Genet.* **2002**, *46* (2), 225−234.

(31) Zhang, W.; Wu, C.; Duan, Y. Convergence of replica exchange molecular dynamics. *J. Chem. Phys.* **2005**, *123*, 154105.

(32) Zuckerman, D. M.; Lyman, E. A second look at canonical sampling of biomolecules using replica exchange simulation. *J. Chem. Theory Comput.* **2006**, *2*, 1200−1202.

(33) Zhang, J.; Qin, M.; Wang, W. Folding mechanism of β-hairpins studied by replica exchange molecular simulations. *Proteins: Struct. Funct. Bioinformatics* **2006**, *62* (3), 672−685.

(34) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. α-Helix peptide folding and unfolding activation barriers: A nanosecond UV resonance Raman study. *J. Am. Chem. Soc.* **1999**, *121* (35), 8074−8086.

(35) Matagne, A.; Jamin, M.; Chung, E. W.; Robinson, C. V.; Radford, S. E.; Dobson, C. M. Thermal unfolding of an intermediate is associated with non-Arrhenius kinetics in the folding of hen lysozyme. *J. Mol. Biol.* **2000**, *297* (1), 193−210.

(36) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding dynamics and mechanism of β-hairpin formation. *Nature* **1997**, *390* (6656), 196−199.

(37) Oliveberg, M.; Tan, Y. J.; Fersht, A. R. Negative Activation Enthalpies in the Kinetics of Protein-Folding. *Proc. Natl. Acad. Sci.* U.S.A. **1995**, *92* (19), 8926−8929.

(38) Segawa, S. I.; Sugihara, M. Characterization of the transition-state of lysozyme unfolding. 1. Effect of protein solvent interactions on the transition-state. *Biopolymers* **1984**, *23* (11), 2473−2488.

(39) Cavalli, A.; Ferrara, P.; Caflisch, A. Weak temperature dependence of the free energy surface and folding pathways of structured peptides. *Proteins: Struct., Funct., Genet.* **2002**, *47* (3), 305−314.

(40) Ferrara, P.; Apostolakis, J.; Caflisch, A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. B* **2000**, *104* (20), 5000−5010.

(41) Frantz, D. D.; Freeman, D. L.; Doll, J. D. Reducing quasi-ergodic behavior in Monte-Carlo simulations by J-walking−Applications to atomic clusters. *J. Chem. Phys.* **1990**, *93* (4), 2769−2784.

(42) Zhou, R. H.; Berne, B. J. Smart walking: A new method for Boltzmann sampling of protein conformations. *J. Chem. Phys.* **1997**, *107* (21), 9185−9196.

(43) Andricioaei, I.; Straub, J. E.; Voter, A. F. Smart darting Monte Carlo. *J. Chem. Phys.* **2001**, *114* (16), 6994−7000.

(44) Brown, S.; Head-Gordon, T. Cool walking: A new Markov chain Monte Carlo sampling method. *J. Comput. Chem.* **2003**, *24* (1), 68−76.

(45) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. Resolution exchange simulation. *Phys. Rev. Lett.* **2006**, *96* (2).

(46) Lyman, E.; Zuckerman, D. M. Resolution exchange simulation with incremental coarsening. *J. Chem. Theory Comput.* **2006**, *2* (3), 656−666.

(47) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. A.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.

(48) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci.* U.S.A. **2001**, *98* (10), 5578−5583.

(49) Schenck, H. L.; Gellman, S. H. Use of a designed triple-stranded antiparallel β-sheet to probe β-sheet cooperativity in aqueous solution. *J. Am. Chem. Soc.* **1998**, *120* (19), 4869−4870.

(50) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127−6129.

(51) Syud, F. A.; Espinosa, J. F.; Gellman, S. H. NMR-based quantification of β-sheet populations in aqueous solution through use of reference peptides for the folded and unfolded states. *J. Am. Chem. Soc.* **1999**, *121* (49), 11577−11578.

(52) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-integration of Cartesian equations of motion of a system with constraints−Molecular-dynamics of *N*-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.

(53) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.

(54) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *J. Comput. Chem.* **2003**, *24* (1), 21−31.

(55) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246* (1−2), 122−129.

(56) Ponder, J. W.; Richards, F. M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **1987**, *8* (7), 1016−1024.

(57) Simmerling, C.; Elber, R.; Zhang, J. MOIL-View, a program for visualization of structure and dynamics of biomolecules, and STO, a program for computing stochastic paths. In *Modelling of Biomolecular Structures and Mechanisms*, Proceedings of the 27th Jerusalem Symposium on Quantum Chemistry and Biochemistry, Jerusalem, Israel, May 23−26, 1994; Pullman A., Jortner, J., Pullams, B., Ed.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995; pp 241−265.

CT600263E

JCTC Journal of Chemical Theory and Computation

Representative Benchmark Suites for Barrier Heights of Diverse Reaction Types and Assessment of Electronic Structure Methods for Thermochemical Kinetics

Jingjing Zheng, Yan Zhao, and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received September 8, 2006

Abstract: We propose three small sets of barrier heights for heavy-atom transfer, nucleophilic substitution, and unimolecular and association reactions as benchmarks for comparing and developing theoretical methods. We chose the data sets to be statistically representative subsets of the NHTBH38/04 database. Each data set consists of 6 barrier heights; we call these small benchmark suites HATBH6, NSBH6, and UABH6. Benchmark values are tabulated for 205 combinations of theory level and basis set. The theory levels studied include single-level wave function theory like Hartree−Fock, Møller−Plesset perturbation theory, quadratic configuration interaction, and coupled cluster theory; they also include multicoefficient correlation methods, local and hybrid density functional theory, and semiempirical molecular orbital methods. The three new representative data sets are combined with a previous representative data set for hydrogen-transfer reactions to form a new compact but diverse and representative data set called DBH24. Comparison of a large number of methods for their performance on DBH24 leads us to recommend the following methods for barrier height calculations, in order of decreasing cost: G3SX, BMC-CCSD, PWB6K, BB1K, M06-L, MPW1K, HF/MIDI!, and PM3.

1. Introduction

Data sets composed of experimental and high-level electronic structure results are very useful for assessing the performance of new theoretical methods. Examples of such data sets are G2/97,[1−3] G3/99,[2−4] Database/3,[5] and NHTBH38/04.[6] These databases usually contain a large number of data to ensure diversity. But a disadvantage of using them to test new methods is that it requires a burdensomely large number of calculations, which is not always affordable. Hence, it is useful to develop a smaller set of data that is representative of the larger database and can be used more conveniently as a benchmark. With this motivation, our group has developed small representative benchmarks suites[7] for the atomization energies and hydrogen-transfer (HT) reactions in Database/3. These benchmark suites have been widely used because of their efficiency. In this paper we develop new small benchmark suites for non-hydrogen-transfer (non-HT) reactions.

The non-HT database NHTBH38/04 contains 38 barrier heights (BH) for heavy-atom transfer (HAT), nucleophilic substitution (NS), and unimolecular and association (UA) reactions. Because it can be expensive to use this suite, we present here a subset of data that has statistical errors close to those exhibited by the entire set, in particular, the small set adequately reproduces the mean signed error (MSE), mean unsigned error (MUE), and root-mean-square error (RMSE) of the full set.

A modern theorist has many quantum chemistry methods available for calculating thermochemical kinetics. In general, high level ab initio wave function theory (WFT) methods can give very accurate results, but they are very time-consuming. Semiempirical molecular orbital methods are computationally inexpensive but less accurate and less reliable. Density-functional theory (DFT) has proven to be very efficient with an excellent performance-to-cost ratio. Multilevel methods can also achieve high accuracy with low computational cost by extrapolating the results of some low-

* Corresponding author e-mail: truhlar@umn.edu.

10.1021/ct600281g CCC: $37.00 © 2007 American Chemical Society
Published on Web 12/24/2006

level methods. In the present work a wide variety of methods covering various kinds of theory and basis sets has been tested against barrier heights for both the full data sets and the representative subsets for both non-HT and HT reactions. These methods are assessed according to their mean errors for calculating barrier heights and their computational scaling and costs in order to allow one to choose the theory level and basis set according to the required performance-to-cost ratio for applications of thermochemical kinetics. In addition, the mean errors presented here can be used to estimate the reliability limits of various methods of calculation.

Section 2 describes the methods and the database used in the present work, section 3 describes the strategies for selecting a representative subset, section 4 gives the results and discussion, and section 5 gives our conclusions.

2. Methods and Databases

2.1. Theoretical Methods. In order to develop the new test sets, we used 205 methods (where a method is a combination of theory level and basis set) to calculate the barrier heights in the NHTBH38/04[6] database. For DFT, theory "level" means choice of density functional.

The single-level methods include local DFT (BB95,[8] BLYP,[9,10] BP86,[9,11] G96LYP, [0,12,13] HCTH,[14] M06-L,[15] mPWKCIS,[16−21] mPWLYP,[10,16] mPWPW,[16] OLYP,[10,22] PBE,[23] PBE1W,[23,24] PW91,[25−29] TPSS,[20,21] TPSSKCIS,[17−21] SPL,[30] SPWL,[31,32] VSXC,[33] and τHCTH[34]), hybrid DFT (B1B95,[8,9] B1LYP,[8,36] B3LYP,[8,35,37] B3LYP*, [38] B97-1,[39] B97-2,[39] B97-3,[40] B98,[41] BB1K,[42] BHandHLYP,[9,10,35] BMK,[43] HFLYP,[44] M05,[45] M05-2X,[45] M06-HF,[46] MPW1B95,[47] MPW1-KCIS,[6,8,17−19] mPW1PW,[16] MPW1K,[16,17,48] MPW3LYP,[10,16,47] MPWB1K, [47] MPWKCIS1K, [6,8,17−19] O3LYP, [10,22,49] PBE1-KCIS,[17−19,23,50] PBE1PBE,[23] PW6B95,[51] PWB6K,[51] TPSS1-KCIS,[17−21,52] TPSSh,[20,21] X3LYP,[9,10,26,53] and τHCTHh[34]), and single-level WFT (Hartree−Fock (HF), Møller−Plesset second-, third-, and fourth-order perturbation theory (MP2,[54] MP2(full), MP3,[55] MP4[55]), Møller−Plesset fourth-order perturbation theory without triple excitations[55] (MP4SDQ), coupled cluster theory (CCSD,[56] CCSD(T)[57]), and quadratic configuration interaction (QCISD[58] and QCISD(T)[58]). Note that the local DFT methods include strictly local ones (SPWL and SPL), GGAs (BP86, BLYP, HCTH, G96LYP, mPWLYP, mPWPW, OLYP, PBE, PBE1W, and PW91), and meta GGAs (BB95, M06-L, mPWKCIS, TPSS, TPSSKCIS, VSXC, and τHCTH), where GGA denotes generalized gradient approximation. In single-level WFT methods, core electrons are uncorrelated (i.e., doubly occupied in all configurations, sometimes called frozen) except for MP2(full) calculations.

We also tested some semiempirical molecular orbital methods such as AM1,[59] PM3,[60] PDDG/PM3,[62] PM6,[61] RM1,[63] and SCC-DFTB.[64] The first five of these methods are based on neglect of diatomic differential overlap (NDDO), whereas SCC-DFTB is a self-consistent-field nonorthogonal tight-binding method.

In addition to the above single-level methods, we also used WFT-based multilevel methods (CBS-4M,[65,66] CBS-Q,[65] CBS-q,[67] CBS-QB3,[66,68] G3SX,[74] G3SX(MP3),[74] MCCM/3,[81] BMC-CCSD,[69] MLSEn+d ($n = 1-4$),[78] and the scaling all correlation method[79−81] (SAC-MP2 including SAC/3)). Within the MCCM/3 suite, we considered MCG3/3, MC-QCISD/3, MC-UT/3, and MC-CO/3. Finally we studied multicoefficient extrapolated DFT methods (MC3BB,[82] MC3MPW,[82] MC3MPWB,[52] MC3TS,[52] MCG3-MPW, -MPWB and -TS,[52] MCCO-MPW, -MPWB and -TS,[52] MCUT-MPW, -MPWB and -TS,[52] MCQCISD-MPW, and MCQ-CISD-MPWB and -TS).[52] Note that multicoefficient extrapolated DFT methods are examples of fifth-rung[83] DFT.

2.2. Basis Sets. The theoretical levels are combined with one or more of the following basis sets: 6-31B(d),[69] 6-31+B-(d,p),[69] 6-31G(d),[55] 6-31G(d,p),[55] 6-31G(2df,p),[55] 6-311G-(2d,p),[55] 6-311G(2df,2p),[55] 6-31+G(d),[55] 6-31+G(d,p),[55] 6-31+G(d,2p),[55] 6-31+G(2df,p),[55] 6-31++G(d,p),[55] 6-31++G(2df,2pd),[55] 6-311++G(d,p),[55] 6-311++G(2df,-2pd),[55] aug-cc-pV(D+d)Z,[70] aug-cc-pVTZ,[71] aug-pc1,[72] aug-pc2,[72] cc-pV(D+d)Z,[70] cc-pV(T+d)Z,[70] cc-pVTZ,[70] G3Large,[73] G3XLarge,[74] MG3,[75,76] MG3S,[77] MG3T,[77] MIDI!,[84] MIDIX+,[84,85] and MIDIY.[84,85] Note that 6-31+G(d,p) is abbreviated to DIDZ in some of our previous papers. We also note that MG3 is also known as G3LargeMP2 and identical to 6-311++G(3d2f,2df,2p) for H through Si and very similar to 6-311+G(3d2f) for P, S, and Cl. The MG3S basis set is identical to MG3 except that diffuse functions on hydrogen have been removed. The MG3T basis set removes all diffuse functions from the MG3 basis set.

2.3. Geometries, Spin−Orbit Coupling, and Software. All calculations in the present work used structures optimized at the QCISD/MG3 level with the spin-restricted formalism for closed-shell and the spin-unrestricted formalism for open-shell systems. The effect of spin−orbit coupling was added to the energies of the Cl, F, and the OH radicals, which lowers their energies by 0.84, 0.38, and 0.20 kcal/mol, respectively.[80]

All single-level and CBS electronic structure calculations in this work were performed using the *Gaussian03* package[86] except that B97-3, M05, M05-2X, M06-HF, M06-L, PW6B95, and PWK6B were carried out with a locally modified version of *Gaussian03*, four of the semiempirical molecular orbital methods (AM1, PM3, PDDG/PM3, RM1) were calculated by the MOPAC 5.012mn[87] program, which is available from the Truhlar group's Web page,[88] the PM6 method was calculated by the MOPAC 7.2[61] program, SCC-DFTB calculations were performed by the DFTB[64,89] program, and single-level CCSD(T) calculations were carried out by the MOLPRO[90] program. All multilevel methods except the CBS ones were carried out with the MLGAUSS[91] program in conjunction with *Gaussian03* except that the MLSE multilevel methods were calculated with *Gaussian03* and a Fortran code. For all single-point energies except CCSD(T), we used the spin-restricted formalism for closed-shell calculations and the spin-unrestricted formalism for open-shell calculations. For CCSD(T), we used a spin-restricted calculation to obtain the orbitals but a spin-unrestricted correlated calculation.

2.4. NHTBH38/04 Database. The NHTBH38/04[6] database for non-hydrogen-transfer reactions consists of 6 heavy-atom transfer reactions, 4 bimolecular nucleophilic substitution reactions, 4 unimolecular nucleophilic substitution reactions of ion-dipole complexes that produce product ion-dipole complexes, and 5 non-nucleophilic-substitution uni-

molecular or association reactions. This database is repeated in the Supporting Information. The details of the database are presented in ref 6. As explained in more detail in ref 6, three of the 19 forward barrier heights are semiexperimental (meaning that they are derived by correcting calculated rate constants to agree with experimentally observed ones), and the other 16 are based on Weizmann-1[6,92] and Weizmann-2[92,93] calculations. The reverse barrier heights are obtained from the forward barrier height and the energy of reaction. All barrier heights, in both directions, are zero-point-exclusive.

3. Strategies for Selecting a Representative Subset

Since the NHTBH38/04 database contains three different types of reactions, we should select a subset that can represent all the reaction types, e.g., the reactions of the subset should not come from only one or two of these types. Here, we used two strategies to select the representative subset.

3.1. Strategy A. In this strategy, we selected a small subset that represents the *whole* database with the minimum root-mean-square deviation (RMSD) between three error measures, mean signed error, mean unsigned error, and root-mean-square error (MSE, MUE, and RMSE) calculated using NHTBH38/04 and the same error measures calculated using the subset (SS), e.g., the deviation between the MSE using NHTBH38/04 (MSE(NH38)) and the MSE using the small subset (MSE(SS)). In this strategy, all the subsets must contain at least one reaction from each reaction type. This restriction guarantees that the representative subset is meaningful for whole database. The rmsd is calculated by using eq 1

$$\mathrm{RMSD} = \left[\frac{1}{3n}\sum_{i=1}\Big[(\mathrm{MSE}_i(\mathrm{NH38}) - \mathrm{MSE}_i(\mathrm{SS}))^2 = (\mathrm{MUE}_i(\mathrm{NH38}) - \mathrm{MUE}_i(\mathrm{SS}))^2 + (\mathrm{RMSE}_i(\mathrm{NH38}) - \mathrm{RMSE}_i(\mathrm{SS})^2)\Big]\right]^{1/2} \quad (1)$$

where n is the number of methods used to generate representative test sets, in particularly $n = 154$. The mean error (ME) is then defined by

$$\mathrm{ME} = \frac{1}{3n}\sum_{i=1}^{n}(|\mathrm{MSE}_i(\mathrm{NH38})| + \mathrm{MUE}_i(\mathrm{NH38}) + \mathrm{RMSE}(\mathrm{NH38})) \quad (2)$$

which is calculated only once for whole database. Finally, we calculated the percentage error in representation (PEIR) using

$$\mathrm{PEIR} = 100\% \times \frac{\mathrm{RMSD}}{\mathrm{ME}} \quad (3)$$

Note that although the present study includes 204 methods, we used $n = 154$ in the above statistical analysis, because we excluded the six semiempirical molecular orbital methods at this stage because they increase PEIR significantly, and the other 44 methods were only used to calculate the representative subsets.

This choice of 154 methods yields ME = 5.26 kcal/mol. The value of RMSD and PEIR were calculated for each possible subset of reactions in the NHTBH38/04 database, and the subset for a given number of reactions is the one with the smallest PEIR.

3.2. Strategy B. An alternative strategy of generating a representative subset is to consider the three types of reactions (HAT, NS, and UA) as *independent* data sets, calculate the corresponding subset's rmsd and PEIR for each reaction type (HAT, NS, and UA) by using equations analogous to eqs 1−3, respectively, and then select the most representative subset for each reaction type.

4. Results and Discussion

4.1. The Large Non-Hydrogen-Transfer Database. Table 1 contains mean signed and unsigned errors for NHTBH38/04 for 160 methods.

4.2. Representative Databases. In order to select a representative subset of the NHTBH38/04 database, at least three reactions are required, which represent the three reaction types in the database. Hence, we calculated a series of the most representative subsets consisting of 3−9 reactions according to strategy A (see Figure 1(a)). The 9-reaction representative subset gives the lowest PEIR, namely 12.8%, and it consists of 3 reactions for each reaction type. Encouragingly, strategy B also selected the same reactions for the representative subset, so the results of strategy A and B are consistent with each other. The 3-reaction subsets for HAT, NS, and UA reactions according to strategy B have similar PEIR, which are 12.4%, 9.2%, and 11.1%, respectively (see Figure 1(b)). We call these representative subsets HATBH6, NSBH6, and UABH6, respectively. In a previous paper,[7] we have identified a representative data set for hydrogen-transfer reactions, namely BH6. We may label this BH6 database more explicitly as HTBH6 in the present paper. The combined set of these four representative databases is called DBH24, where D denotes diverse, and it is given in Table 2. As explained more fully in refs 5 and 6, four of the 12 forward barrier heights in DBH24 are semiexperimental (defined in section 2.4), six are from Weizmann-1 calculations, and two are from Weizmann-2 calculations. The reverse barrier heights are obtained from the forward barrier height and the energy of reaction. All barrier heights in the database, in both directions, are zero-point-exclusive.

The HATBH6 data set is based on three reactions involving nitrogen, fluorine, and chlorine transfer. It represents the six heavy-atom transfer reactions in the NHTBH38/04 database not only by having the smallest PEIR but also by covering all the heavy-atom transfer types in the whole database. NSBH6 is also appealing in that not only is it statistically representative but also it involves all three anions, OH^-, F^-, and Cl^-, that appear in NHTBH38/04. UABH6 includes two open-shell systems and one closed-shell system. Fortunately, the most time-consuming reaction, $CH_3 + C_2H_4 \rightarrow CH_3CH_2CH_2$, is not involved in this subset.

The subsets of non-hydrogen-transfer and hydrogen-transfer barrier heights yield errors that are representative

Table 1. Calculated Errors and Costs of Various Methods for Non-Hydrogen-Transfer Reactions in NHTBH38/04

	HAT(12)		NS(16)		UA(10)			
methods	MSE	MUE	MSE	MUE	MSE	MUE	MMUE[a]	cost[b]
			N^7 Methods					
G3SX	−0.59	0.74	−0.50	0.50	−0.16	0.56	0.60	246
CCSD(T)/aug-cc-pVTZ	−0.34	0.71	-0.71	0.71	-0.18	0.41	0.61	12898
G3SX(MP3)	−0.47	0.76	0.13	0.88	−0.21	0.85	0.75	175
MCG3-MPW	−0.82	1.16	−0.40	0.58	−0.22	0.73	0.82	113
MCG3-MPWB	−0.94	1.15	−0.16	0.65	−0.26	0.76	0.85	114
QCISD(T)/MG3S	1.04	1.21	−0.62	1.08	0.30	0.53	0.94	5076
MCG3-TS	−0.74	1.35	−0.91	0.91	−0.04	0.59	0.95	108
MCG3/3	−0.61	1.17	−0.35	0.94	−0.48	0.91	1.01	102
MLSE4+d	−0.44	1.04	−3.64	4.47	−0.36	0.55	2.02	186
MLSE3+d	0.05	1.09	−3.80	4.64	−0.20	0.45	2.06	186
MLSE2+d	−0.35	1.02	−4.37	5.65	−0.38	0.64	2.44	186
MLSE1+d	−0.15	1.22	−4.77	5.90	−0.33	0.53	2.55	186
QCISD(T)/cc-pV(D+d)Z	2.14	3.04	−4.32	7.87	0.00	1.63	4.18	160
QCISD(T)/6-31G(d)	1.84	5.66	−3.30	7.77	0.96	2.95	5.46	63
MP4/6-31+G(d)	8.53	9.14	−0.95	2.72	3.56	5.03	5.63	84
MP4/6-31G(2df,p)	8.01	8.01	−5.03	9.13	3.74	4.18	7.11	699
MP4/6-31G(d)	8.64	9.24	−3.57	8.08	3.70	5.54	7.62	37
			N^6 Methods					
BMC-CCSD	−0.27	1.14	−0.02	0.42	−0.39	0.69	0.75	27
MC-QCISD/3	1.26	1.53	−0.15	0.53	0.44	0.61	0.89	23
MCQCISD-MPW	−1.16	1.38	−0.17	0.51	−0.11	0.85	0.91	34
MCQCISD-MPWB	−0.94	1.27	0.07	0.78	−0.09	0.85	0.97	35
MCQCISD-TS	−1.37	1.51	−0.98	0.98	−0.02	0.70	1.06	31
MCUT-MPWB	1.36	2.15	−0.07	0.58	1.04	1.31	1.35	34
CCSD/aug-cc-Pvtz	2.44	2.44	1.70	1.70	0.76	0.80	1.65	3868
MCUT-MPW	1.31	2.72	−0.74	0.75	1.16	1.48	1.65	33
MCUT-TS	0.90	2.71	−1.43	1.43	1.12	1.24	1.79	28
QCISD/MG3S	3.43	3.43	1.26	1.32	1.04	1.08	1.94	152
CCSD/6-31+G(d,p)	4.31	4.59	1.88	1.92	2.03	2.03	2.85	12
CCSD/6-31+B(d,p)	4.53	6.48	0.68	1.97	2.19	2.19	3.55	11
MC-UT/3	7.47	7.47	0.23	0.33	3.02	3.02	3.61	22
MP4SDQ/aug-cc-pVTZ	7.24	7.24	1.07	1.07	2.62	2.62	3.64	673
MP4SDQ/MG3S	8.60	8.60	1.42	1.44	3.08	3.12	4.39	95
CCSD/6-31B(d)	4.49	7.09	−0.39	4.70	1.98	2.73	4.84	2.3
MP4SDQ/6-31+G(p,d)	9.55	9.55	1.18	1.86	4.05	4.05	5.15	3.4
MP3/aug-cc-pVTZ	9.24	9.24	3.09	3.09	3.69	3.69	5.34	530
MP4SDQ/6-31+B(d,p)	9.36	9.78	0.07	2.34	4.16	4.16	5.43	3.3
QCISD/6-31G(d)	3.43	5.93	−2.37	7.53	1.53	2.89	5.45	1.7
MP3/MG3S	10.59	10.59	3.62	3.62	4.14	4.14	6.12	71
MP3/6-31+B(d,p)	11.35	11.50	2.54	2.86	5.21	5.21	6.52	2.8
MP4SDQ/cc-pV(D+d)Z	9.23	9.23	−2.44	6.95	2.81	3.79	6.66	2.8
MP3/6-31+G(d)	11.05	11.05	3.55	3.55	4.78	5.37	6.66	1.4
MP3/6-31+G(d,p)	11.49	11.49	3.66	3.66	5.08	5.08	6.74	2.7
MP4SDQ/6-31G(2df,p)	8.68	8.68	−2.73	7.73	3.86	3.98	6.80	14
MP4DQ/6-31B(d)	11.20	11.48	1.14	4.98	4.92	5.55	7.34	0.83
MP4SDQ/6-31G(d)	9.07	9.34	−1.94	7.58	3.86	5.23	7.38	0.87
MP3/6-31G(2df,p)	10.64	10.64	−0.81	6.90	4.91	4.91	7.48	12
MP3/6-31G(d)	10.97	10.97	0.02	6.93	4.89	5.81	7.90	0.73
			N^5 Methods					
MCCO-MPWB	0.60	1.86	1.29	1.60	0.95	1.14	1.53	33
MC3BB	1.18	2.44	0.55	0.66	1.53	1.53	1.54	14
MC3MPWB	1.23	2.55	0.18	0.67	1.57	1.57	1.60	14
MCCO-TS	0.78	2.75	−0.92	1.32	1.34	1.52	1.86	27
MC3MPW	1.50	2.69	0.17	0.89	2.14	2.20	1.93	13
MC3TS	1.35	2.82	−0.90	1.05	2.09	2.16	2.01	7.4
MCCO-MPW	2.13	2.97	0.63	1.48	1.75	1.85	2.10	32
MP2/aug-cc-pVTZ	10.46	10.46	0.37	0.67	4.29	4.88	5.34	140

Table 1 (Continued)

	HAT(12)		NS(16)		UA(10)			
methods	MSE	MUE	MSE	MUE	MSE	MUE	MMUE[a]	cost[b]
			N^5 Methods					
MC-CO/3	10.61	10.61	0.08	1.37	4.04	4.04	5.34	21
MP2/aug-cc-pV(D+d)Z	10.28	10.28	−0.95	1.27	3.80	4.74	5.43	4.3
MP2/6-31++G(2df,2pd)	11.07	11.07	0.55	0.65	4.64	5.15	5.62	13
MP2/6-311++G(2df,2pd)	11.07	11.07	0.55	0.65	4.64	5.15	5.62	18
MP2/MG3	11.46	11.46	0.70	0.73	4.65	5.40	5.86	14
MP2(full)/6-31G(2df,p)	11.41	11.41	0.91	0.91	4.72	5.46	5.93	3.6
MP2/MG3S	11.76	11.76	0.74	0.74	4.71	5.44	5.98	13
MP2/6-31 ++ G(d,p)	11.19	11.19	1.12	2.18	5.28	5.87	6.41	1.1
MP2/6-31 + G(d,2p)	12.04	12.04	0.97	2.21	4.96	5.31	6.52	1.5
SAC/3[c]	11.55	11.55	0.26	2.84	5.27	5.81	6.73	2.5
MP2/6-31 + B(d,p)	12.15	12.15	−0.22	2.62	5.57	5.76	6.84	1.0
MP2/6-31 + G(d,p)	12.56	12.56	1.14	2.21	5.49	5.93	6.90	1.0
SAC-MP2/6-31 + G(d,p)	12.06	12.06	0.32	2.95	5.93	6.44	7.15	1.7
MP2/cc-pV(T + d)Z	11.75	11.75	−2.14	4.42	4.49	5.30	7.16	15
MP2/6-311 ++ G(d,p)	12.57	12.57	3.28	3.28	4.71	5.89	7.25	2.0
MP2/6-311 + G(d,p)	12.57	12.57	3.28	3.28	4.71	5.89	7.25	1.7
MP2/6-31 + G(d)	12.14	12.54	1.08	2.32	5.31	7.07	7.31	0.60
MP2/6-31B (d)	12.00	12.76	−1.45	4.99	5.57	7.10	8.28	0.41
MP2/cc-pV(D+d)Z	12.09	12.09	−2.75	7.29	4.23	5.95	8.44	1.4
MP2/cc-pVDZ	12.20	12.20	−2.97	7.45	4.23	5.95	8.53	0.92
MP2/6-31G(2df,p)	11.63	11.63	−3.47	8.28	5.34	6.23	8.71	3.0
MP2/6-31G(d)	12.00	12.43	−2.22	8.02	5.37	7.54	9.33	0.41
SAC-MP2/6-31G(d)	11.51	13.32	−3.64	8.85	6.11	8.69	10.29	0.56
			N^4 Methods					
BMK/MG3S	−1.21	1.49	0.75	0.91	0.80	1.58	1.33	13
PWB6K/MG3S	−0.24	1.61	0.94	1.10	0.65	1.53	1.41	12
BB1K/MG3S	−0.69	1.58	1.23	1.30	0.53	1.44	1.44	12
MPWB1K/MG3S	−0.77	1.69	1.08	1.19	0.52	1.61	1.50	12
PWB6K/aug-pc2	−0.68	1.59	1.48	1.48	0.53	1.51	1.53	64
MPWB1K/aug-pc2	−1.19	1.81	1.66	1.66	0.40	1.58	1.68	64
BB1K/6-31 + G(d,p)	−0.68	1.86	1.02	1.17	0.84	2.10	1.71	2.0
M05-2X/MG3S	1.15	2.00	−0.79	1.48	0.91	1.77	1.75	13
MPW1K/MG3S	−0.83	1.89	1.12	1.28	0.96	2.42	1.86	12
MPWB1K/aug-pc1	−1.78	2.92	0.20	1.15	0.02	1.95	2.01	3.7
M05-2X/6-31 + G(d,p)	1.06	2.53	−0.63	1.68	1.00	1.90	2.04	2.2
BHandHLYP/MG3S	0.07	3.04	0.95	1.39	0.76	1.98	2.14	10
M05/MG3S	−2.84	3.79	0.00	0.80	0.69	2.24	2.28	13
B97-2/MG3S	−3.13	3.52	−1.43	1.47	0.62	1.91	2.30	11
B1B95/MG3S	−4.73	4.73	−0.95	1.08	−0.58	1.21	2.34	12
MPW1B95/MG3S	−4.62	4.62	−0.81	1.21	−0.52	1.31	2.38	12
MPWKCIS1K/MG3S	−1.94	2.82	1.69	1.69	0.95	2.97	2.49	13
M05/6-31 + G(d,p)	−2.73	4.29	−0.74	1.14	0.81	3.06	2.83	2.1
PW6B95/MG3S	−5.36	5.36	−2.05	2.05	−0.76	1.43	2.95	12
mPW1PW/MG3S	−5.99	5.99	−1.81	1.94	−0.38	2.00	3.31	11
B98/6-311 + G(3df,2p)	−5.19	5.19	−2.93	2.93	−0.39	1.96	3.36	10
B98/MG3S	−5.18	5.18	−2.96	2.96	−0.31	1.97	3.37	11
B97-1/MG3S	−5.18	5.18	−3.21	3.21	−0.23	1.83	3.41	11
PBE1PBE/MG3S	−6.62	6.62	−1.87	2.05	−0.58	2.16	3.61	10
PBE1KCIS/MG3	−8.56	8.56	−1.77	1.88	−0.86	2.64	4.36	12
X3LYP/MG3S	−8.48	8.48	−2.89	2.90	−1.43	2.06	4.48	11
B3LYP/MG3S	−8.49	8.49	−3.25	3.25	−1.42	2.02	4.59	9.4
tHCTHh/MG3S	−6.73	6.73	−4.53	4.53	−0.23	2.65	4.64	11
MPWB1K/cc-pVDZ	−2.06	3.14	−5.79	9.40	0.01	1.74	4.76	1.6
O3LYP/MG3S	−8.27	8.27	2.61	4.42	−1.02	2.27	4.99	11
B3LYP/6-31 + G(d,p)	−8.79	8.79	−3.59	3.59	−1.27	2.65	5.01	1.4
MPW3LYP/MG3S	−9.29	9.29	−4.29	4.29	−1.61	2.21	5.26	11
MPW1KCIS/MG3S	−9.89	9.89	−3.46	3.46	−1.27	2.61	5.32	13

Table 1 (Continued)

methods	HAT(12) MSE	HAT(12) MUE	NS(16) MSE	NS(16) MUE	UA(10) MSE	UA(10) MUE	MMUE[a]	cost[b]
				N^4 Methods				
TPSS1KCIS/MG3S	−9.26	9.89	−4.88	4.88	−1.39	2.12	5.42	13
B3LYP*/MG3T	−6.02	6.02	−6.92	8.40	1.45	4.00	6.14	7.5
B97-2/6-31G(d)	−4.25	5.85	−7.33	11.03	0.95	2.16	6.35	0.89
TPSSh/MG3S	−11.51	11.51	−5.78	5.78	−2.94	3.23	6.84	13
B3LYP/6-311G(2df,2p)	−8.92	8.92	−8.97	11.05	−1.71	1.98	7.32	5.2
B3LYP/6-311G(2d,p)	−8.89	8.89	−9.68	11.65	−1.85	1.98	7.51	1.5
BB1K/MIDIY	−6.80	6.90	−11.71	16.40	0.03	1.83	8.38	1.4
HF/aug-cc-pV(D+d)Z	14.23	16.34	5.23	5.23	2.35	3.77	8.45	0.55
B3LYP/6-31G(d,p)	−9.99	9.99	−9.71	12.92	−1.51	2.49	8.47	0.92
HF/cc-pV(T+d)Z	15.03	16.97	4.18	5.04	2.49	3.75	8.59	1.9
HF/cc-pV(D+d)Z	14.19	16.50	1.20	5.64	2.22	3.80	8.65	0.45
HF/6-31+B(d,p)	14.55	17.73	4.15	4.34	3.23	3.99	8.69	0.87
HF/cc-pVDZ	14.25	16.57	0.93	5.80	2.22	3.80	8.72	0.28
HF/6-31G(d)	13.63	15.47	2.52	6.74	2.95	4.23	8.81	0.15
HF/aug-cc-pVTZ	14.43	16.39	6.76	6.76	2.53	3.72	8.96	45
HF/6-31B(d)	13.93	17.34	3.22	5.16	3.43	4.40	8.97	0.17
HF/6-31+G(d,2p)	15.33	17.44	5.66	5.66	2.86	3.91	9.00	0.97
HF/6-31+G(d,p)	15.29	17.38	5.72	5.72	3.03	3.95	9.02	0.67
HF/MG3	14.64	16.67	6.67	6.67	2.67	3.79	9.04	9.2
HF/G3XLarge	14.62	16.64	6.73	6.73	2.67	3.79	9.05	18
HF/G3Large	14.64	16.66	6.72	6.72	2.67	3.79	9.06	13
HF/MG3S	14.86	16.87	6.67	6.67	2.70	3.82	9.12	7.9
M06-L/MG3S	−5.68	5.93	−3.56	3.56	0.27	2.35	3.95	5.7/8.9[d]
M06-L/6-31+G(d,p)	−5.36	6.22	−4.55	4.55	0.16	2.27	4.35	2.1/2.2
HCTH/MG3S	−8.84	8.84	−2.71	2.71	−0.75	2.20	4.58	3.9/6.7
VSXC/MG3S	−7.44	7.44	−5.30	5.30	−0.91	2.40	5.05	5.2/8.8
VSXC/6-31+G(d,p)	−7.58	7.58	−5.52	5.52	−0.67	2.90	5.33	1.9/2.0
OLYP/MG3S	−11.23	11.23	−2.73	2.73	−1.92	2.53	5.50	3.7/6.8
τHCTH/MG3S	−9.21	9.21	−5.71	5.71	−1.04	2.82	5.91	5.4/8.8
G96LYP/MG3S	−13.03	13.03	−5.80	5.80	−2.86	3.04	7.29	3.8/7.3
mPWKCIS/MG3S	−13.65	13.65	−6.66	6.66	−2.67	3.07	7.79	6.1/9.7
				N^3 Methods				
BB95/MG3S	−13.88	13.88	−6.36	6.36	−3.22	3.40	7.88	5.4/8.7
TPSSKCIS/MG3S	−13.37	13.37	−7.64	7.64	−2.56	2.98	8.00	5.7/9.3
MPWPW/MG3S	−14.10	14.10	−7.45	7.45	−2.67	3.10	8.22	3.8/7.1
PBE1W/MG3S	−14.51	14.51	−7.06	7.06	−2.69	3.13	8.23	3.8/7.4
PBE/MG3S	−14.93	14.93	−6.97	6.97	−2.94	3.35	8.42	3.7/7.3
BP86/MG3S	−15.51	15.51	−6.91	6.91	−3.41	3.87	8.76	3.9/6.9
PW91/MG3S	−15.42	15.42	−7.75	7.75	−2.78	3.22	8.80	3.8/6.9
TPSS/MG3S	−14.65	14.65	−7.75	7.75	−3.84	4.04	8.81	5.5/8.8
BLYP/MG3S	−14.66	14.66	−8.40	8.40	−3.38	3.51	8.86	3.8/7.3
mPWLYP/MG3S	−15.76	15.76	−8.14	8.14	−3.64	3.79	9.23	3.8/7.3
SPL/MG3S	−23.36	23.36	−8.58	8.58	−5.08	5.82	12.59	2.5/5.8
SPWL/MG3S	−23.48	23.48	−8.50	8.50	−5.17	5.90	12.63	3.5/5.9
PM6	−18.83	19.37	1.76	6.54	9.61	15.66	13.86	5×10^{-5}
AM1	−8.99	10.60	16.02	19.08	9.04	15.01	14.90	5×10^{-5}
PM3	−13.97	15.88	19.91	20.24	3.53	11.25	15.79	5×10^{-5}
RM1	−14.80	18.34	6.59	18.43	6.72	14.28	17.02	5×10^{-5}
PDDG/PM3	−21.19	21.90	20.53	20.53	3.01	10.52	17.65	5×10^{-5}
SCC-DFTB	−23.02[e]	23.02[e]			−3.47	9.94	16.48[f]	4×10^{-4}

[a] MMUE is defined as eq 4. [b] The cost for each method is measured by the computer time for an energy gradient calculation of phosphinomethanol divided by the computer time for an MP2/6-31 + G(d,p) energy gradient calculation with the same software (except semiempirical molecular orbital methods, see text) on the same computer. Although such costs depend to some extent (for example, 15%) on the machine, the program, and the computer load, they still provide a useful indication of computer resource demand. [c] Also called SAC-MP2/6-31 + G(d,2p). [d] The double entries for local DFT methods correspond to timings with/without density fitting. [e] The forward and reverse barriers of the reaction $H + N_2O \rightarrow OH + N_2$. [f] Mean value of MUE for the forward and reverse barriers of the reaction $H + N_2O \rightarrow OH + N_2$ and the MUE for UA.[10]

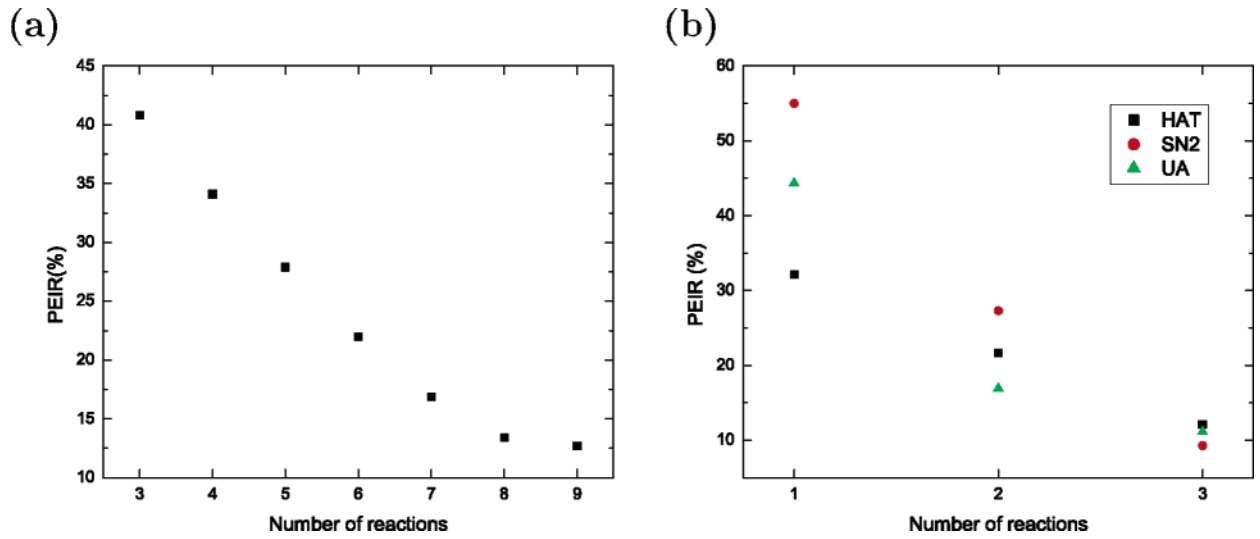


Figure 1. Percentage error in representation vs number of reactions: (a) is for strategy A and (b) is for strategy B.

Table 2. Representative Barrier Heights Database DBH24

database	reaction	$V_f^{\neq}$ a	$V_r^{\neq}$ a
HATBH6	$H + N_2O \rightarrow OH + N_2$	18.14	83.22
	$H + ClH \rightarrow HCl + H$	18.00	18.00
	$CH_3 + FCl \rightarrow CH_3F + Cl$	7.43	61.01
NSBH6	$Cl^-\cdots CH_3Cl \rightarrow ClCH_3\cdots Cl^-$	13.61	13.61
	$F^-\cdots CH_3Cl \rightarrow FCH_3\cdots Cl^-$	2.89	29.62
	$OH^- + CH_3F \rightarrow HOCH_3 + F^-$	−2.78	17.33
UABH6	$H + N_2 \rightarrow HN_2$	14.69	10.72
	$H + C_2H_4 \rightarrow CH_3CH_2$	1.72	41.75
	$HCN \rightarrow HNC$	48.16	33.11
HTBH6	$OH + CH_4 \rightarrow CH_3 + H_2O$	6.7	19.6
	$H + OH \rightarrow O + H_2$	10.7	13.1
	$H + H_2S \rightarrow H_2 + HS$	3.6	17.3

[a] $V_f^{\neq}$ denotes forward BH, and $V_r^{\neq}$ denotes reverse BH (in kcal/mol).

of their respective errors using NHTBH38/04 and Database/3, respectively. Table 3 gives the errors calculated using the HATBH6, NSBH6, and UABH6 subsets. The size of the representative subsets is less than half of the whole NHTBH38/04, and using these subsets can significantly reduce the computational cost for testing and developing new theoretical methods.

4.3. Assessment of Methods. We have tested 160 methods against the full NHTBH38/04 database and 153 methods against both the non-hydrogen-transfer and hydrogen-transfer representative barrier height databases. All the mean errors (MSE and MUE) by these methods are given in Table 1 for the non-hydrogen-transfer full data set and in Table 3 for both non-hydrogen-transfer and hydrogen-transfer representative subsets. We also tabulated a value for MMUE (mean MUE) as defined

$$\mathrm{MMUE} = \frac{1}{3}[\mathrm{MUE(HAT)} + \mathrm{MUE(NS)} + \mathrm{MUE(UA)}] \quad (4)$$

in Table 1 or

$$\mathrm{MMUE} = \frac{1}{4}[\mathrm{MUE(HAT)} + \mathrm{MUE(NS)} + \mathrm{MUE(UA)} + \mathrm{MUE(HT)}] \quad (5)$$

in Table 3. In the two tables, all methods are sorted by MMUE and the scaling order σ, where a method's computational cost, in the limit of a large number N of atoms, scales[94] as N^{σ}.

The computational costs in Table 1 and 3 are the computer time for calculating an energy gradient of the molecule phosphinomethanol divided by the time for an MP2/6-31+G(d,p) energy gradient calculation with the same computer program on the same computer. For calculations where the energy was calculated with *Gaussian03* or MOLPRO we calculated the MP2/6-31+G(d,p) gradient with the same program. For semiempirical molecular orbital methods calculated with MOPAC or DFTB, we take the ratio to the MP2/6-31+G(d,p) computer time with *Gaussian03* on the same computer. We used a gradient calculation to illustrate the cost because gradients are important for geometry optimization and dynamics calculations. When analytic gradients are available in the computer programs specified in section 2.3, we used them. Otherwise, we used numerical gradients. In *Gaussian03* the cost for a numerical gradient of phosphinomethanol is 49 times the cost of a single-point energy, whereas in MOLPRO it is 19 times the cost of a single-point energy. For local DFT methods, we give two costs corresponding to carrying out the calculation with and without density fitting. The density fitting method[95,96] (sometimes called resolution of the identity) employs auxiliary basis functions to represent the electron density; in this way, the four-center two-electron repulsion integrals are decomposed into three- and two-center integrals, reducing the formal scaling from N^4 to N^3.

4.3.1. Non-Hydrogen-Transfer Reactions. Among the tested multilevel methods, G3SX gives the lowest mean errors. The MMUEs of G3SX(MP3) and BMC-CCSD are identical and a little bit higher than that of G3SX, and the MMUE of BMC-CCSD is only 13% higher than that of G3SX(MP3). But the computational cost of BMC-CCSD is about 6 times smaller than that of G3SX(MP3) and 9 times smaller than that of G3SX. The MLSEn+d methods give good performance for neutral systems but are not as good

Table 3. Calculated Mean Errors of Various Methods for Non-Hydrogen-Transfer and Hydrogen-Transfer Representative Databases and for the Diverse Barrier Heights Database DBH24

	HATBH6		NSBH6		UABH6		HTBH6			
methods	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	DBH24 MMUE[a]	cost[b]
					N^7 Methods					
G3SX	−1.02	1.20	−0.54	0.54	−0.06	0.27	0.31	0.53	0.64	246
CCSD(T)/aug-cc-pVTZ	−0.11	0.84	−0.67	0.68	−0.06	0.40	−0.11	0.64	0.64	12898
G3SX(MP3)	−0.98	1.18	0.00	0.73	−0.15	0.40	0.49	0.51	0.71	175
MCG3-MPW	−1.31	1.40	−0.50	0.67	−0.26	0.71	−0.15	0.36	0.79	113
MCG3-TS	−1.14	1.16	−0.99	0.99	−0.02	0.99	−0.30	0.46	0.80	108
MCG3-MPWB	−1.31	1.35	−0.26	0.79	−0.29	0.72	−0.39	0.52	0.85	114
MCG3/3	−1.16	1.65	0.43	0.90	−0.44	0.68	0.58	0.76	1.00	102
QCISD(T)/MG3S	0.94	1.29	−0.59	1.14	−0.59	0.52	1.02	1.12	1.02	5076
CBS-QB3	−1.26	1.76	−0.51	0.59	−2.21	2.47	−1.42	1.42	1.56	363[c]
CBS-Q	−2.52	2.52	−1.43	1.48	−2.61	2.67	−0.87	0.88	1.89	369[c]
CCSD(T)/cc-pV(T+d)Z	1.12	1.12	−2.91	4.93	0.18	0.49	0.68	1.26	1.95	1761
CCSD(T)/cc-pVTZ	1.26	1.35	−3.05	5.10	0.18	0.49	0.80	1.30	2.06	1660
CBS-q	−1.95	2.72	−0.32	1.58	−3.25	3.68	−1.00	1.18	2.29	180[c]
CBS-4M	0.50	3.59	1.94	2.48	−2.38	2.85	−0.40	0.57	2.37	173[c]
CCSD(T)/6-311G(2df,2p)	1.24	2.16	−3.81	7.98	0.43	0.66	1.15	1.71	3.13	1080
QCISD(T)/6-311G(2df,2p)	1.13	2.25	−4.13	8.20	0.42	0.65	1.08	1.62	3.18	0.65
QCISD(T)/6-31G(d)	3.61	4.68	−2.63	7.89	1.31	2.94	4.68	5.53	5.26	63
MP4/6-311G(2df,2p)	7.36	7.36	−4.48	8.47	3.07	3.72	2.27	2.36	5.48	1608
MP4/6-311G(2d,p)	8.49	8.49	−4.90	8.71	2.94	4.00	2.90	3.39	6.15	458
MP4/6-31G(d)	10.09	10.09	−2.78	8.27	4.08	5.31	5.94	6.38	7.51	37
					N^6 Methods					
BMC-CCSD	−0.12	1.36	0.09	0.54	−0.38	0.40	0.00	0.57	0.72	27
MC-QCISD/3	0.84	1.22	−0.10	0.46	0.36	0.61	0.81	0.91	0.80	23
MCQCISD-MPW	−1.46	1.46	−0.24	0.70	−0.28	0.99	−0.50	0.50	0.91	34
MCQCISD-MPWB	−1.14	1.25	0.02	0.96	−0.22	0.95	−0.61	0.64	0.95	35
MCUT-MPWB	0.89	1.03	−0.11	0.81	0.96	1.42	−0.60	0.66	0.98	34
MCQCISD-TS	−1.45	1.45	−1.02	1.02	−0.07	0.70	−0.73	0.73	0.98	31
MCUT-MPW	0.74	1.43	−0.79	0.83	1.03	1.53	−0.62	0.62	1.10	33
MCUT-TS	0.54	1.35	−1.45	1.45	1.14	1.22	−0.67	0.67	1.17	28
QCISD/MG3S	3.76	3.76	1.29	1.29	1.17	1.17	2.27	2.27	2.12	152
CCSD/MG3S	3.54	3.54	2.03	2.03	1.48	1.51	2.50	2.50	2.40	242
MC-UT/3	6.38	6.38	0.36	0.36	2.98	2.98	2.25	2.25	2.99	22
MP4SDQ/MG3S	8.31	8.31	1.51	1.51	3.24	3.24	3.48	3.48	4.14	95
QCISD/6-31B(d)	5.76	8.63	−0.62	5.01	2.13	2.31	4.78	5.09	5.26	1.7
CCSD/6-31B(d)	6.10	8.76	0.42	5.26	2.43	2.58	4.91	5.11	5.43	2.3
QCISD/6-31G(d)	5.54	5.84	−1.65	7.78	1.84	2.61	5.36	5.81	5.44	1.7
MP3/MG3S	10.06	10.06	3.44	3.44	4.31	4.31	4.21	4.21	5.51	71
MP4SDQ/6-31G(2df,p)	8.41	8.41	−2.02	7.92	4.19	4.19	4.62	4.62	6.29	14
MP3/6-31+G(d)	11.72	11.72	3.76	3.76	5.09	5.09	7.05	7.05	6.91	1.4
MP4SDQ/6-31G(d)	10.53	10.53	−1.19	7.91	4.24	4.97	6.54	6.55	7.49	0.87
MP4DQ/6-31B(d)	12.20	12.39	1.96	6.17	5.49	5.49	6.81	6.81	7.72	0.83
MP3/6-31G(d)	12.21	12.21	0.62	7.29	5.29	5.54	7.26	7.26	8.08	0.73
					N^5 Methods					
MC3BB	1.05	2.25	0.52	0.59	1.47	1.47	−0.51	0.72	1.26	14
MC3MPWB	1.14	2.42	0.25	0.63	1.58	1.58	−0.61	0.84	1.37	14
MC3TS	1.38	2.10	−0.80	0.82	2.15	2.15	−0.30	0.44	1.38	7.4
MCCO-MPWB	0.46	1.39	1.09	1.63	0.98	1.16	−1.35	1.35	1.38	33
MC3MPW	1.27	2.24	0.16	0.78	2.06	2.06	−0.54	0.72	1.45	13
MCCO-TS	0.52	1.84	−1.05	1.46	1.49	1.49	−1.00	1.00	1.45	27
MCCO-MPW	1.52	2.38	0.38	1.84	1.76	1.76	−0.73	1.18	1.79	32
MC-CO/3	8.80	8.80	−0.01	1.24	4.66	4.66	2.58	2.67	4.34	21
MP2/aug-cc-pV(D+d)Z	9.73	9.73	−0.48	1.00	4.06	4.96	2.74	3.27	4.74	4.3
MP2/MG3	10.79	10.29	0.72	0.80	5.15	5.65	3.88	3.88	5.28	14
MP2/MG3S	10.95	10.95	0.80	0.81	5.12	6.00	3.91	3.91	5.42	13
MP2(full)/6-31G(2df,p)	10.55	10.55	0.94	0.94	5.11	6.04	4.74	4.74	5.57	3.6
SAC/3	10.53	10.53	0.62	2.99	5.36	6.25	2.62	2.62	5.60	2.5

Table 3 (Continued)

	HATBH6		NSBH6		UABH6		HTBH6			
methods	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	DBH24 MMUE[a]	cost[b]
					N^5 Methods					
MP2/6-31+G(d,2p)	11.44	11.44	1.27	2.52	5.14	5.56	3.98	3.98	5.88	1.5
SAC-MP2/6-31+G(d,p)	11.35	11.35	0.75	3.09	6.13	6.09	4.13	4.13	6.17	1.7
MP2/6-31+G(d,p)	11.44	11.44	1.27	2.52	5.14	5.56	5.49	5.49	6.25	1.0
MP2/6-31++G(d,p)	11.87	11.87	1.46	2.53	5.53	6.22	5.42	5.42	6.51	1.1
MP2/6-31G(2df,p)	10.87	10.87	−2.78	8.29	5.91	6.80	4.98	4.98	7.74	3.0
MP2/6-31B(d)	12.47	12.58	−0.94	5.03	6.20	7.40	6.14	6.14	7.79	0.41
MP2/6-31G(d)	13.22	13.22	−1.48	8.31	6.05	7.83	6.81	6.81	9.04	0.41
					N^4 Methods					
PWB6K/MG3S	−0.06	1.05	0.96	0.96	0.71	1.59	−0.99	1.22	1.21	12
BB1K/MG3S	−0.60	1.09	1.07	1.17	0.42	1.57	−1.03	1.14	1.24	12
MPWB1K/MG3S	−0.69	1.16	1.01	1.01	0.46	1.63	−1.32	1.32	1.28	12
BMK/MG3S	−1.58	1.58	0.72	0.86	0.85	2.06	−1.20	1.20	1.43	13
MPWB1K/aug-pc2	−0.93	1.17	1.57	1.57	0.28	1.61	−1.37	1.37	1.43	64
MPW1K/MG3S	−0.79	1.36	0.91	1.15	0.82	2.42	−1.14	1.40	1.58	11
MPWB1K/MG3	−1.15	1.56	0.97	1.97	0.78	2.46	−1.34	1.34	1.58	13
M05-2X/MG3S	0.72	1.96	−0.94	1.48	1.25	1.60	−0.36	1.40	1.61	13
M05-2X/MG3	0.33	1.58	−0.96	1.50	1.50	2.25	−0.36	1.42	1.69	14
BB1K/6-31+G(d,p)	−0.95	1.92	0.92	1.26	0.70	2.20	−1.03	1.42	1.70	2.0
PWB6K/6-31+G(d,p)	−0.45	2.05	0.79	1.23	0.99	2.30	−1.02	1.35	1.73	2.0
MPWB1K/6-31+G(d,p)	−1.05	1.99	0.86	1.24	0.75	2.37	−1.33	1.43	1.76	2.0
MPW1K/6-31+G(d,p)	−1.14	1.55	0.60	1.50	1.15	3.21	−1.05	1.42	1.92	1.4
M05-2X/6-31+G(d,p)	0.07	2.44	−0.80	1.72	1.32	1.90	−0.64	1.60	1.92	2.1
B97-3/MG3S	−2.93	2.93	−0.39	1.07	0.57	1.63	−2.22	2.29	1.98	11
MPWB1K/aug-pc1	−1.27	2.60	0.36	1.32	−0.31	1.92	−2.17	2.17	2.00	3.7
BHandHLYP/MG3S	1.06	2.79	0.81	1.50	0.88	2.35	0.10	1.98	2.16	10
MPWKCIS1K/MG3S	−1.93	2.20	1.48	1.48	0.81	3.35	−1.97	1.97	2.25	13
BB1K/6-31+B(d,p)	−2.13	4.06	0.13	1.48	0.91	2.23	−1.29	1.29	2.27	2.3
M06-HF/MG3S	3.69	4.11	−0.79	1.74	0.98	1.69	1.14	1.95	2.37	16
M06-HF/6-31+G(d,p)	2.93	4.49	−0.09	1.56	1.25	1.56	0.82	2.34	2.49	2.7
B1B95/MG3S	−4.46	4.46	−1.12	1.22	−0.69	1.12	−3.14	3.14	2.49	12
M05/MG3S	−3.89	5.09	0.13	1.00	0.87	2.48	−0.76	1.64	2.55	13
MPW1B95/MG3S	−4.37	4.37	−0.56	1.26	−0.57	1.23	−3.38	3.38	2.56	12
B97-2/MG3S	−3.68	4.46	−1.63	1.63	0.79	1.81	−2.88	3.21	2.78	11
PW6B95/MG3S	−4.92	4.92	−2.08	2.08	−0.68	1.17	−3.46	3.46	2.91	12
M05/6-31+G(d,p)	−4.37	5.59	−0.47	1.02	0.75	3.28	−0.91	2.09	3.00	2.1
MPW1K/MIDIX+	−3.86	3.86	−2.30	2.30	1.72	3.16	−2.73	3.09	3.10	1.0
MPWB1K/MG3T	−1.52	1.65	−3.68	7.09	0.69	2.27	−1.64	1.64	3.16	8.8
mPW1PW/MG3S	−5.73	5.73	−2.00	2.00	−0.53	1.93	−3.95	3.95	3.40	11
M05-2X/MG3T	0.00	1.79	−5.56	8.53	1.41	2.04	−0.54	1.60	3.49	11
BB1K/6-31B(d)	−0.91	5.20	−1.97	4.67	1.11	1.90	−0.90	2.35	3.53	1.3
B98/MG3S	−5.39	5.39	−3.05	3.05	−0.05	1.84	−4.00	4.00	3.57	11
B97-1/MG3S	−5.45	5.45	−3.21	3.21	0.05	1.68	−4.14	4.14	3.62	11
MPW1K/6-31B(d)	−0.96	4.75	−2.22	4.55	1.64	3.21	−0.92	2.13	3.66	0.80
PBE1PBE/MG3S	−6.45	6.45	−1.99	1.99	−0.77	1.96	−4.62	4.62	3.76	10
mPW1PW/6-31+G(d,p)	−6.20	6.20	−2.17	2.19	−0.30	2.76	−3.94	3.94	3.77	1.4
MPWB1K/6-31G(d,p)	−1.32	2.17	−3.63	9.35	0.71	2.09	−1.55	2.13	3.94	1.3
B1LYP/6-31+G(d,p)	−6.28	6.28	−3.03	3.03	−0.92	2.49	−3.79	4.13	3.98	1.2
M05-2X/6-31G(d,p)	−0.21	2.08	−5.34	10.88	1.38	1.75	−0.79	2.06	4.19	1.5
PBE1PBE/6-31+G(d,p)	−6.90	6.90	−2.16	2.16	−0.53	2.75	−4.61	4.61	4.11	1.4
MPWB1K/cc-pVDZ	−1.93	3.14	−5.07	8.86	−0.53	1.58	−2.84	2.84	4.11	1.6
X3LYP/MG3S	−7.36	7.36	−2.85	2.85	−1.33	1.75	−4.91	4.91	4.22	11
MPW1K/6-31G(d)	−1.24	2.15	−3.54	9.11	1.50	3.11	−0.66	2.70	4.27	0.80
B3LYP/MG3S	−7.38	7.38	−3.44	3.44	−1.35	1.69	−4.73	4.73	4.31	9.4
M05-2X/6-31G(d)	−0.02	2.89	−5.26	10.84	1.77	1.77	−0.28	2.74	4.56	1.5
B97-3/6-31G(d)	−3.53	4.43	−5.51	10.29	1.08	1.89	−1.82	2.44	4.76	0.89
B3LYP*/6-31+G(d,p)	−6.52	6.52	−2.44	2.85	1.72	4.84	−3.81	4.17	4.60	1.4

Table 3 (Continued)

	HATBH6		NSBH6		UABH6		HTBH6			
methods	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE	DBH24 MMUE[a]	cost[b]
					N^4 Methods					
PBE1KCIS/MG3S	−8.21	8.21	−1.90	1.90	−1.01	2.80	−5.71	5.71	4.66	12
B3LYP/6-31+G(d, p)	−8.08	8.08	−3.73	3.73	−1.30	2.59	−4.99	5.03	4.86	1.4
O3LYP/MG3S	−7.98	7.98	3.41	5.14	−1.42	2.19	−4.45	4.45	4.94	11
MPW3LYP/MG3S	−8.17	8.17	−4.55	4.55	−1.48	1.81	−5.27	5.27	4.95	11
TPSS1KCIS/MG3S	−8.45	8.45	−4.95	4.95	−1.29	1.66	−4.99	4.99	5.01	13
B97-2/6-31G(d)	−4.27	6.27	−6.71	11.02	1.25	2.05	−2.40	3.30	5.66	0.89
MPW1KCIS/MG3S	−9.45	9.45	−4.44	4.44	−1.38	2.61	−6.36	6.36	5.72	13
HFLYP/MG3S	11.17	11.17	5.28	5.28	3.51	4.20	5.44	5.44	6.52	10
TPSSh/MG3S	−10.75	10.75	−5.82	5.82	−2.94	2.94	−6.72	6.72	6.56	13
M05-2X/MIDI!	−5.80	6.22	−12.34	18.14	1.26	2.75	−3.44	3.85	7.74	1.20
MPW1K/MIDI!	−7.44	7.44	−11.08	16.24	0.90	3.26	−3.92	4.93	7.97	0.66
HF/MIDI!	11.10	11.10	−2.15	10.28	3.39	4.84	8.90	9.69	8.98	0.10
HF/6-31G(d)	16.98	16.98	2.78	6.17	4.18	4.18	12.20	12.20	9.88	0.15
HF/6-31+G(d,p)	17.50	17.50	5.63	5.63	3.97	3.97	12.42	12.42	9.88	0.67
HF/G3Large	17.50	17.50	6.41	6.41	3.48	3.57	12.33	12.33	9.95	13
HF/MG3S	17.65	17.65	6.38	6.38	3.52	3.62	12.34	12.34	10.00	9.2
M06-L/MG3S	−6.72	7.22	−3.25	3.25	0.78	2.58	−4.21	4.32	4.34	5.7/8.9
M06-L/6-31+G(d,p)	−6.90	7.62	−4.13	4.13	0.29	2.36	−4.09	4.09	4.55	2.1/2.2
VSXC/MG3S	−7.53	7.53	−4.90	4.90	−0.19	1.49	−4.98	4.98	4.73	5.2/8.8
HCTH/MG3S	−9.46	9.46	−2.78	2.78	−0.79	1.74	−5.25	5.25	4.81	3.9/6.7
VSXC/6-31+G(d,p)	−8.06	8.06	−4.52	4.52	−0.12	2.16	−5.05	5.05	4.95	2.0/1.9
OLYP/MG3S	−10.81	10.81	−3.10	3.10	−2.35	2.35	−5.88	5.88	5.54	3.7.8/6
τHCTH/MG3S	−9.83	9.83	−6.01	6.01	−0.71	2.09	−6.21	6.21	6.04	5.4/8.8
M06-L/6-31G(d)	−6.95	8.86	−8.41	11.84	0.57	1.84	−3.21	3.33	6.47	1.7/1.6
G96LYP/MG3S	−11.57	11.57	−6.25	6.25	−2.88	2.88	−6.60	6.60	6.83	3.8/7.3
TPSSKCIS/MG3S	−12.27	12.27	−7.56	7.56	−2.36	2.36	−7.08	7.08	7.32	5.7/9.3
mPWKCIS/MG3S	−12.62	12.62	−6.69	6.69	−2.59	2.59	−7.56	7.56	7.37	6.1/9.7
BB95/MG3S	−13.21	13.21	−6.49	6.49	−3.20	3.20	−8.02	8.02	7.73	5.4/8.7
mPWPW/MG3S	−13.35	13.35	−7.32	7.32	−2.72	2.72	−8.47	8.47	7.97	3.8/7.1
					N^3 Methods					
BLYP/MG3S	−13.01	13.01	−8.64	8.64	−3.19	3.19	−7.83	7.83	8.17	3.8/7.3
TPSS/MG3S	−13.67	13.67	−7.43	7.43	−3.76	3.76	−8.30	8.30	8.29	5.5/8.8
PBE/MG3S	−14.25	14.25	−6.90	6.90	−3.01	3.01	−9.33	9.33	8.37	3.7/7.3
mPWLYP/MG3S	−14.09	14.09	−8.08	8.08	−3.37	3.37	−8.85	8.85	8.60	3.8/7.3
BLYP/6-31+G(d, p)	−13.88	13.88	−7.54	7.54	3.31	−3.31	−8.26	8.26	8.25	1.4/1.3
BP86/MG3S	−14.64	14.64	−7.02	7.02	−3.54	3.54	−9.29	9.29	8.62	3.9/6.9
M06-L/MIDI!	−11.88	13.07	−16.46	18.81	0.62	2.87	−6.20	6.20	10.24	1.6/1.4
PM3	−13.45	16.86	13.93	14.80	5.97	13.92	−3.63	5.82	12.85	5×10^{-5}
AM1	−9.08	11.83	10.53	15.56	13.06	18.88	−0.22	5.25	12.88	5×10^{-5}
SPL/MG3S	−23.05	23.05	−8.33	8.33	−5.21	5.21	−17.75	17.75	13.59	2.5/5.8
SPWL/MG3S	−23.16	23.16	−8.25	8.25	−5.34	5.34	−17.97	17.97	13.68	3.5/5.9
PDDG/PM3	−16.86	18.32	15.56	15.56	5.25	13.54	−4.00	12.47	14.97	5×10^{-5}
RM1	−20.05	20.93	0.27	15.46	10.47	19.83	−5.79	7.35	15.89	5×10^{-5}
PM6	−21.65	22.73	−0.79	4.13	13.97	22.04	−9.10	18.57	16.87	5×0^{-5}
SCC-DFTB	−23.02[d]	23.02[d]			1.00	9.78	−30.53	30.53	19.34[e]	4×10^{-4}

[a] MMUE is defined as eq 5. [b] The cost for each method is measured by the computer time for an energy gradient calculation of phosphinomethanol divided by the computer time for an MP2/6-31+G(d,p) energy gradient calculation with the same software (except semiempirical molecular orbital methods, see text) on the same computer. The double entries of local DFT methods correspond to with/without density fitting. [c] Although CBS methods are defined to use a lower-level geometry and are not normally employed in gradient calculations, we include a timing here so that the reader can judge their approximate cost on the same basis as the other methods. This timing is estimated as 49 times the cost of a single-point energy, since that is the cost of a numerical gradient for phosphinomethanol when using the *Gaussian03* program. [d] These values are only for reaction $H + N_2O \rightarrow OH + N_2$ and its reverse. [e] It is the mean value of MUEs for HAT ($H + N_2O \rightarrow OH + N_2$ and its reverse), UABH6, and HTBH6.

for general charged systems since they lack diffuse functions in their basis sets.

Among the ab initio WFT methods in Table 1, CCSD(T)/aug-cc-pVTZ gives a very high accuracy with 0.61 kcal/mol MMUE and underestimates non-hydrogen-transfer BHs slightly, but it is very time-consuming. The QCISD(T)/MG3S method has a slightly higher MMUE for non-hydrogen-transfer reactions but with smaller cost than the CCSD(T)/

aug-cc-pVTZ method. MP4, MP4SDQ, MP3, MP2, and HF systematically overestimate the BHs of HAT and UA reactions as shown by their high positive MSE.

Among the hybrid DFT methods in Table 1, BMK, PWB6K, BB1K, and MPWB1K are the best performing methods with regards to their MMUE, and they give even higher accuracy than some multilevel methods for calculating BHs of non-HT reactions. M05 gives the lowest MUE for NS reactions, and B1B95 gives the lowest MUE for UA reactions. The B3LYP* functional improves upon the B3LYP functional (see Table 3) by reducing HF exchange energy from 20% to 15%.

Among the local DFT methods in Table 1, our recently developed M06-L functional shows very good performance even with a double-ζ basis set. All the local DFT methods systematically underestimate the BHs of non-HT reactions as shown by their negative MSE, except that M06-L overestimates barrier heights of UA reactions with a small positive MSE.

The five NDDO semiempirical methods tested, AM1, PM3, PM6, RM1, and PDDG/PM3, all underestimate the BHs of HAT reactions and significantly overestimate the BHs of NS and UA reactions. But they are faster than MP2/6-31+G(d,p) by about 5 orders of magnitude for phosphinomethanol gradients and by even more for a larger system.

4.3.2. The Diverse DBH24 Database. For 108 of the 160 methods in Table 1 plus 44 new methods, we list mean unsigned errors for DBH24 in Table 3. The mean errors of HATBH6, NSBH6, and UABH6 shown in Table 3 are consistent with those of the full data set shown in Table 3. This illustrates the representative quality of these subsets. Table 3 also gives the mean errors of hydrogen-transfer reactions with the HTBH6 data set. Table 3 shows that the quality of all the tested methods in Table 3 for calculating hydrogen-transfer barrier heights correlates well with their quality for calculating non-hydrogen-transfer barrier heights. Mean errors of non-hydrogen-transfer representative data sets for the methods shown in Table 1 but not in Table 3 are given in Table S2 of the Supporting Information.

In Table 3 we compare several basis sets for DFT methods. We found that the MG3S basis set is both less expensive and more accurate than the aug-pc2 basis set for DFT methods, although aug-pc2 was specially designed[72] for use with DFT methods. The aug-pc1 basis set is 16 times less expensive than aug-pc2, but its MMUE is higher than that of aug-pc2 by only about 0.6 kcal/mol (2.00 vs 1.43 kcal/mol). Hence aug-pc1 has a better performance-to-cost ratio than aug-pc2. But it is still less accurate than the 6-31+G(d,p) basis set, at least for the MPWB1K functional. We also tested the 6-31+G(d,p) basis set with several other functionals, and we found that it is competitive with MG3S because, although its computational costs are about 7−10 times faster, its MMUEs are slightly higher than those of MG3S. But 6-31G(d,p) gives much higher MMUEs, in particular for NS reactions, due to its lack of diffuse functions, a subject we have discussed previously.[77] According to our calculations, MG3S gives the best performance for barrier heights among the tested basis sets for DFT methods.

Table 4. Best Methods for Barrier Height Calculations as Judged by Performance for a Given Cost and Scaling

method	scaling	cost	MMUE/DBH24 (kcal/mol)
G3SX	7	246	0.64
BMC-CCSD	6	27	0.72
PWB6K/MG3S	4	12	1.21
BB1K/6-31+G(d,p)	4	2.0	1.70
MPW1K/6-31+G(d,p)	4	1.4	1.92
MPW1K/MIDIX+	4	1.0	3.10
MPW1K/6-31B(d)	4	0.80	3.66
MPW1K/MIDI!	4	0.66	7.97
HF/MIDI!	4	0.10	8.98
M06-L/MG3S	3	5.7	4.34
M06-L/6-31+G(d,p)	3	2.1	4.55
M06-L/6-31G(d)	3	1.6	6.47
BLYP/6-31+G(d,p)	3	1.3	8.25
PM3	3	5×10^{-5}	12.85

Since the SCC-DFTB method lacks parameters for halogens, we only calculated barrier heights for UA reactions, HTBH6 reactions, and two HAT reactions ($H + N_2O \rightarrow OH + N_2$ and its reverse) with this method. SCC-DFTB gives better results than the NDDO semiempirical molecular orbital methods for UA reactions with a 9.78 kcal/mol MUE, and it overestimates the barrier heights for UA reactions on average by only 1.00 kcal/mol. But the MSE of full UA reaction barrier heights (see Table 1) becomes −3.47 kcal/mol. The MSEs of full data set and representative subset are not consistent with each other because DBH24 does not have the representative quality for semiempirical methods since these methods were not used to select the representative subset as we discussed in section 3.1. Nevertheless the small test sets are instructive even for semiempirical methods. The SCC-DFTB MUEs for hydrogen-transfer barrier heights and for the one heavy-atom-transfer reaction are 30.53 and 23.02 kcal/mol, respectively, which are much larger than those of the NDDO semiempirical molecular orbital methods. Since the SCC-DFTB method is parametrized against the B3LYP method, it is interesting to compare them. If we take B3LYP/MG3S barrier heights as the standards, the MUE of SCC-DFTB is improved to 8.37 kcal/mol for UA reactions, and the MSE relative to B3LYP/MG3 is 2.35 kcal/mol. We conclude that SCC-DFTB might be improved if it is reparameterized against more accurate functionals.

Table 4 summarizes the best methods when methods are judged solely on the basis of the accuracy of the predicted barrier heights, the cost of the method, and its scaling. Any method not in Table 4 performs less well for DBH24 than at least one method in Table 4 that has the same or lower cost or better scaling. This table, though, is not a replacement for careful consideration when choosing a basis set. For example, BB1K is better than MPWB1K and M05-2X, on average, for barrier heights, but only slightly so, but MPWB1K and M05-2X are better as all-purpose functionals, especially for noncovalent interactions.[52,97] Therefore Table 5 lists some additional recommended methods with performance for barrier heights almost as good as those in Table 4 and with good overall quality on a performance-for-a-given-cost basis.

Table 5. Other Methods To Be Considered for Reaction Dynamics Calculations Based on Overall Quality

method	scaling	cost	MMUE/DBH24 (kcal/mol)
G3SX(MP3)	7	175	0.71
MCG3-MPW	7	113	0.79
MCQCISD-MPWB	6	35	0.95
MC3MPWB	5	14	1.37
M05-2X/MG3S	4	13	1.61
M05-2X/6-31+G(d,p)	4	2.1	1.92
PWB6K/6-31+G(d,p)	4	2.0	1.73
AM1	3	5×10^{-5}	12.88

5. Conclusions

Three small but representative data sets, HATBH6, NSBH6, and UABH6, are identified for the barrier heights of heavy-atom transfer, nucleophilic substitution, and unimolecular and association reactions, respectively. They are representative of the full data set within 12% (HAT), 9% (NS), and 11% (UA), respectively. We combine these data sets with a previous small representative data set for hydrogen-transfer reactions to create a diverse representative data set of zero-point-exclusive barrier heights called DBH24. Assessment of methods with DBH24 shows that DFT and multilevel methods have much better performance-to-cost ratios than single-level WFT methods. The best N^6 method is BMC-CCSD, and its cost is an order of magnitude smaller than the best N^7 methods, although it is almost as accurate. The two best N^4 methods, PWB6K and BB1K, outperform the best N^5 method MC3BB. The best local DFT method is M06-L. The MG3S basis set gives the best performance for DFT methods among the tested basis sets.

Acknowledgment. The authors are grateful to Jingzhi Pu for helpful discussions. This work was supported in part by the Air Force Office of Scientific Research (AFOSR) by a Small Business Technology Transfer (STTR) grant to Scientific Applications & Research Assoc., Inc. and by the U.S. Department of Energy, Office of Basic Energy Sciences, under Grant No. DE-FG02-86ER13579.

Supporting Information Available: Tables of the NHTBH38/04 data base, mean unsigned and signed errors of HATBH6, NSBH6, and UABH6 for 52 methods that are not included in Table 3, geometries of species in the DBH24 database and phosphinomethanol, and T1[98] diagnostic values of species in the DBH24 database. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople. J. A. *J. Chem. Phys.* **1991**, *94*, 7221.

(2) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople. J. A. *J. Chem. Phys.* **1997**, *106*, 1063.

(3) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople. J. A. *J. Chem. Phys.* **1998**, *109*, 42.

(4) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople. J. A. *J. Chem. Phys.* **2000**, *112*, 7374.

(5) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 3898.

(6) Zhao, Y.; González-García, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012.

(7) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 8996.

(8) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.

(9) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(10) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(11) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.

(12) Gill, P. M. W. *Mol. Phys.* **1996**, *89*, 433.

(13) Adamo, C; Barone, V. *J. Comput. Chem.* **1998**, *19*, 419.

(14) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264.

(15) Zhao, Y. Truhlar, D. G. *J. Chem. Phys.* In press.

(16) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.

(17) Rey, J.; Savin, A. *Int. J. Quantum Chem.* **1998**, *69*, 581.

(18) Krieger, J. B.; Chen, J.; Iafrate, G. J. Savin, A. In *Electron Correlations and Materials Properties;* Gonis, A., Kioussis, N., Eds.; Plenum: New York, 1999; p 463.

(19) Toulouse, J.; Savin, A.; Adamo, C. *J. Chem. Phys.* **2002**, *117*, 10465.

(20) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129.

(21) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(22) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403.

(23) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

(24) Dahlke, E. E.; Truhlar, D. G. *J. Phys. Chem. B* **2005**, *109*, 15677.

(25) Burke, K.; Perdew, J. P.; Wang, Y. In *Electronic Density Functional Theory: Recent Progress and New Directions;* Dobson, J. F., Vignale, G., Das, M. P., Eds.; Plenum: New York, 1998.

(26) Perdew, J. P. In *Electronic Structure of Solids;* Ziesche, P., Eschrig, H., Eds.; Kademie Verlag: Berlin, 1991; p 11.

(27) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.

(28) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1993**, *48*, 4978.

(29) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533.

(30) Seidl, M.; Perdew, J. P.; Levy, M. *Phys. Rev. A* **1999**, *59*, 51.

(31) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

(32) Slater, J. C. In *Quantum Theory of Molecular and Solids. Vol 4: The Self-Consistent Field for Molecular and Solids;* McGraw-Hill: New York, 1974.

(33) Voorhis, T. V.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.

(34) Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2002**, *116*, 9559.

(35) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(36) Adamo, C,; Barone, V. *Chem. Phys. Lett.* **1997** *274*, 242.

(37) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.

(38) Reiher, M.; Salomon, O.; Hess, B. A. *Theor. Chem. Acc.* **2001**, *107*, 48.

(39) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264.

(40) Keal, T. W.; Tozer, D. J. *J. Chem. Phys.* **2005**, *123*, 121103

(41) Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, *108*, 9624.

(42) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715.

(43) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.

(44) Valentin, C. D.; Pacchioni, G.; Bredow, T.; Dominguez-Ariza, D.; Illas, F. *J. Chem. Phys.* **2002**, *117*, 2299.

(45) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

(46) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* In press.

(47) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.

(48) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.

(49) Hoe, W. M.; Cohen, A. J.; Handy, N. C. *Chem. Phys. Lett.* **2001**, *341*, 319.

(50) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.

(51) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656.

(52) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.

(53) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673.

(54) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(55) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory;* Wiley: New York, 1986.

(56) Cizek, J. *J. Chem. Phys.* **1966**, *45*, 4256.

(57) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

(58) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(59) (a) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902. (b) Holder, A. J.; Dennington, R. D.; Jie, C.; Yu, G. *Tetrahedron* **1994**, *50*, 627. (c) Dewar, M. J. S.; Jie, C.; Yu, G. *Tetrahedron* **1993**, *23*, 5003.

(60) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

(61) Stewart, J. J. P. http://openmopac.net/ (accessed Oct 25, 2006).

(62) (a) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601. (b) Tubert-Brohman, I.; Guimarâes, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 138. (c) Tubert-Brohman, I.; Guimarâes, C. R. W.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2005**, *1*, 817.

(63) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101.

(64) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.

(65) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A., Jr. *J. Chem. Phys.* **1996**, *104*, 2598.

(66) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532.

(67) Petersson, G. A.; Tensfeldt, T. G.; Montgomery, J. A., Jr. *J. Chem. Phys.* **1991** *94*, 6091.

(68) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999** *110*, 2822.

(69) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 1643.

(70) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244.

(71) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(72) (a) Jensen, F. *J. Chem. Phys.* **2001**, *115*, 9113. (b) Jensen, F.; Helgaker, T. *J. Chem. Phys.* **2004**, *121*, 3463.

(73) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(74) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 108.

(75) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407.

(76) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703.

(77) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.

(78) Li, T. H.; Chen, H. R.; Hu, W. P. *Chem. Phys. Lett.* **2005**, *412*, 430.

(79) Gordon, M. S.; Truhlar, D. G. *J. Am. Chem. Soc.* **1986**, *108*, 2.

(80) Fast, P. L.; Corchado, J.; Sanchez, M. L.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 3139.

(81) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 3898.

(82) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, 108, 4786.

(83) Perdew, J.; Schmidt, K. In *Density Functional Theory and its Applications to Materials*; Van Doren, V. E., Alsenoy, K., Geerlings, P., Eds.; American Institute of Physics: Melville, NY, 2001; p 577.

(84) (a) Tatewaki, H.; Huzinaga, S. *J. Comput. Chem.* **1980**, *1*, 205. (b) Thompson, J. D.; Winget, P.; Truhlar, D. G. *Phys. Chem. Comm.* **2001**, *4*, 4116. (c) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281. (d) Huzinaga, S.; Andzelm, J.; Klobukowski, M.; Radzio-Audzelm, E.; Sakai, Y.; Tatewaki, H. *Gaussian basis sets for molecular calculations*; Huzinaga, S., Ed.; Elsevier: Amsterdam, 1984. (e) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281.

(85) Lynch, B. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2004**, *111*, 335.

(86) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich,

S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian03 Revision D.01*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(87) Stewart, J. J. P.; Zheng, J.; Rossi, I.; Hu, W.-P.; Lynch, G. C.; Liu, Y.-P.; Chuang, Y.-Y.; Pu, J.; Li, J.; Cramer, C. J.; Fast, P. L.; Truhlar, D. G. *MOPAC-version 5.012mn*; University of Minnesota: Minneapolis, MN, 2006.

(88) Truhlar, D. G. http://comp.chem.umn.edu/mopac/ (accessed Oct 25, 2006).

(89) (a)Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947. (b) Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185.

(90) Werner, H.-J.; Knowles, P. J.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO, 2002.6*; University of Birmingham: Birmingham, 2002.

(91) Zhao, Y.; Truhlar, D. G. *MLGAUSS-version 2.0*; University of Minnesota: Minneapolis, MN, 2005.

(92) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843.

(93) Parthiban, S.; de Oliveira, G.; Martin, J. M. L. *J. Phys. Chem. A* **2001**, *105*, 895.

(94) Raghavachari, K.; Anderson, J. B. *J. Phys. Chem.* **1996**, *100*, 12960.

(95) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283.

(96) Dunlap, B. I. *J. Mol. Struct. (Theochem*) **2000**, *529*, 37.

(97) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 1009.

(98) Lee, T. J.; Taylor, P. R. *Int. J. Quantum Chem. Symp.* **1989**, *23*, 199.

CT600281G

JCTC Journal of Chemical Theory and Computation

The Use of Processor Groups in Molecular Dynamics Simulations to Sample Free-Energy States

Bruce Palmer,* Shawn Kathmann, Manojkumar Krishnan, Vinod Tipparaju, and Jarek Nieplocha

Pacific Northwest National Laboratory, Box 999, Richland, Washington 99352

Received August 9, 2006

Abstract: Molecular dynamics calculations composed of many independent simulations are frequently encountered in free-energy calculations, as well as many other simulation approaches. In principle, the availability of a large number of independent tasks should make possible the development of highly scalable parallel code that executes these tasks concurrently. This paper discusses the use of processor groups to write simulation codes of this type and describes results for a code that evaluates the volume dependence of the Helmholtz free energy for clusters of an immiscible fluid in a solvent. The results show that very high levels of scalability can be achieved using processor groups with corresponding reductions in the time to completion. The main limitation to scaling appears to be a load imbalance due to variations in the execution times of the individual tasks.

Introduction

Molecular dynamics (MD) simulations are becoming an increasingly important tool in understanding the properties and behavior of complex chemical and biochemical systems. Biochemists routinely use simulations to help understand problems such as protein−ligand interactions, the binding of drugs to active sites, and the role of conformation changes in enzymatic activity.[1] Simulations are also being used to understand many problems in materials science and chemistry, particularly with regard to reactions in aqueous phases.[2] The higher speeds and decreasing cost of processors have also contributed to the widespread use of molecular dynamics. Simulations that previously took days on a high-end supercomputer can now be performed on a relatively inexpensive workstation. Furthermore, the availability of clusters and other parallel architectures has continued to push the bounds on the size of the simulation that can be performed, to the point where the main limitation is, in many cases, not the size of the system that can be simulated but rather the time interval over which the simulation can be performed. For many problems, the goal is to simulate a fixed-size system for longer periods of time or to obtain better ensemble averages rather than increasing the size of the system.

The development of parallel molecular dynamics algorithms has been only partially successful in addressing the need for simulations that extend over longer time periods. Molecular dynamics algorithms are relatively communication-bound and thus scale poorly once the number of atoms per processor drops below a certain point.[3,4] Applying parallel programming to these algorithms has been much more successful in increasing the size of the system that can be simulated and relatively unsuccessful in decreasing the amount of time required to simulate a system for a given time interval. To address these difficulties, theoreticians have developed free energy techniques and other sampling protocols that are designed to provide indirect estimates of the population of events that are rare on the time scale of the simulation.[5−9] Instead of trying to simulate rare events directly, a collection of shorter simulations is used to estimate the relative free energy of these rare events and therefore estimate their relative population.

If the individual simulations are small enough to fit on a single processor, then one method of speeding up these calculations using parallel architectures is to farm out each simulation to an individual processor. This is commonly referred to as an embarrassingly parallel algorithm. However, if the calculations are too large to fit on a single processor, then each simulation or task must be distributed over several processors, and the only way to run the calculation is to run

* Corresponding author e-mail: bruce.palmer@pnl.gov.

10.1021/ct600260u CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/18/2007

many consecutive parallel tasks. This can be a major constraint if the individual tasks do not scale well above a certain number of processors. There is also no way to speed up the calculation if the number of processors available is more than the number of tasks, even if the individual tasks scale well beyond a single processor per task.

This paper will discuss the use of processor groups to achieve very high levels of scalability for algorithms that are based on multiple independent simulations. Processor groups are designed to allow the programmer to subdivide the domain of available processors (the "world group") into independent subdomains. Processor groups make it possible to run multiple concurrent independent calculations, each of which is running as a parallel, distributed task. This can be achieved with little additional effort beyond that required to write a conventional parallel code. These extensions have been incorporated into the Global Arrays Toolkit,[10] which has been used to obtain the results described below.

The remainder of this paper will give an overview of multiple task simulation strategies and then discuss the shared-memory programming model and the processor group extensions to it, the implementation of a spatial decomposition molecular dynamics algorithm using shared-memory programming, the development of a code that uses processor groups, and scaling results for the code.

Multiple Task Simulations

As discussed in the Introduction, multiple task simulations can be sped up in two ways using conventional parallel programming models. The first is to use an embarrassingly parallel approach to farm out tasks to individual processors; the second is to run multiple consecutive parallel tasks. For the embarrassingly parallel algorithm, the program starts with a list of independent tasks and assigns each processor a subset of tasks from the list. This is illustrated schematically in Figure 1a. A variant on this approach is the master−worker algorithm, where one processor (the master) assigns tasks to the remaining processors (the workers). As each worker processor finishes its task, it is assigned the next task on the list by the master. The embarrassingly parallel approach works very well if the individual tasks are relatively small calculations that can be run on single processors, but it breaks down for large simulations where each task must be distributed across more than one processor in order to be run at all. It also has the problem that there is no way to reduce the simulation time further once the number of processors exceeds the number of tasks.

Most general-purpose MD codes that perform free-energy calculations on parallel platforms use the approach of running consecutive parallel tasks. This is illustrated in Figure 1b. Each individual task is run in parallel, but only one task is run at a time. This approach works well if the individual tasks are too large to fit on a single processor but requires very good scaling of the parallel code to utilize large numbers of processors. This kind of scaling is difficult to achieve, however, and many molecular dynamics codes show very poor scaling once the number of atoms per processor drops to a few hundred.[3,4] Running consecutive parallel tasks is a natural extension of the way most parallel programs are written and can handle arbitrarily large systems, but it does not provide a good mechanism for utilizing large numbers of processors.

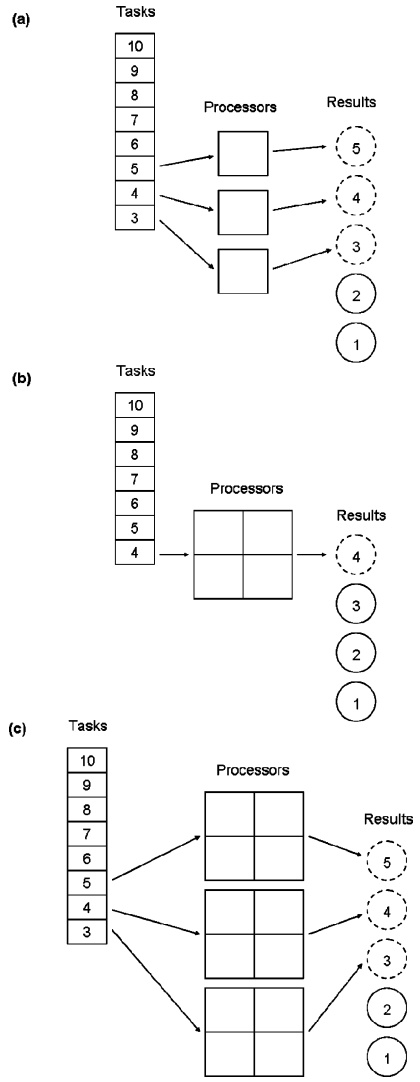


Figure 1. Schematic illustration of programming models discussed in the text: (a) embarrassingly parallel, (b) parallel sequential tasks, (c) concurrent parallel tasks using groups. Dotted circles indicate results that are currently being generated; solid circles represent completed tasks.

The approach discussed in this paper is to use processor groups to run multiple concurrent parallel tasks. This approach permits the use of very large numbers of processors to complete the calculation without taking the performance hit associated with poor scaling. The basic idea is to take the original collection of processors and subdivide them into smaller groups. Each group can then be used to run a parallel task that is independent of the other tasks. This is illustrated in Figure 1c. So far as the authors are aware, this approach has not been used in molecular dynamics, but it offers a way to greatly speed up calculations that consist of many independent tasks. The main disadvantage of this approach is that it is potentially much harder to program. However, by making use of the concept of a default processor group, discussed in more detail below, it turns out that a groups-based code is not significantly more complicated than a sequential parallel code.

Processor Groups and Shared Memory

The shared-memory programming model assumes memory can be divided into two categories, local and shared. Data structures that are created in local memory are visible only to the process that created them, while data structures in shared memory are visible to any process in the system. The message-passing programming model, on the other hand, assumes that all memory is local. Shared memory is a feature of some computer architectures in which multiple processors can directly access the same physical address space. Data objects created in this address space, such as arrays, can be written to and read from using a simple PUT/GET syntax that allows programmers to specify blocks of data in the shared array that need to be copied to or from local arrays. This style of programming tends to be much simpler than the distributed model used in message passing because shared memory preserves the global index space. Message passing, in contrast, requires the programmer to transform a data point from the global index space to a local index held on a particular processor. Communication in a shared programming model is accomplished by writing to and from arrays held in shared memory, while communication in a message-passing model is managed by sending data to a specific location on a specific processor. Although the shared-memory programming model originally reflected the physical layout of the memory on some parallel systems, it is also available on distributed memory platforms via the Global Arrays Toolkit.[10] This is a library of routines that allows programmers to create globally accessible data structures even if memory is completely distributed.

Most global operations in parallel programming are executed on the set of all available processors (the "world group"). These operations consist of synchronizations, global sums, and broadcasts. A synchronization forces all processes to wait until everyone has reached the same point in the execution before any process can proceed further, global sums add a value or values across all processors, and broadcasts send a value or values from one processor to all others. The global sum and broadcast operations implicitly synchronize the processors as well. Global operations are usually employed to guarantee data consistency across processors. Because they force processes to stop until everyone has reached the same point in the execution, global operations tend to degrade scalability in parallel codes, and their widespread use should be avoided. However, almost all parallel codes have at least a few global operations.

The presence of global operations is a major impediment to subdividing the world group into smaller groups that can run independent parallel tasks. Global operations in any task will force processes to wait until there is a corresponding global operation in other tasks. If the algorithm is such that other tasks will have different numbers or different sequences of global operations, the code will eventually hit a point where there is an unanswered global operation, and the program hangs. To avoid this, the concept of processor group has been introduced. Global operations executed on a processor group are restricted to only those processes in the group. It is also possible to restrict the scope of queries that return the number of processes and the local processor ID so that they act on subgroups instead of the world group.

For shared-memory programming models, processor groups also need to restrict the extent of shared arrays that are created in the context of a subgroup so that the address space represented by the shared array is accessible only from processes within the group. Implementing this functionality requires some modification of the memory allocation routines and additional layers of index translation, but this is transparent to the user. The main modification to the programming model is that programmers need to be able to specify that a given shared object is associated with a particular processor group, instead of an implicit assumption that it is being created on the world group. The Global Array Toolkit used in developing the cluster simulation application described here has been modified to include these capabilities.

Finally, an important concept associated with programming on processor groups is the default processor group. Support for this has been explicitly incorporated into the Global Arrays Toolkit, but the programming concept could also be used in codes based on message-passing interfaces (MPI), provided sufficient care is used. The default processor group is the group that all global operations and queries refer to unless another group is explicitly specified. For shared-memory programming, all shared data structures are created on the default group, again, unless explicitly specified otherwise. Conventional parallel programs have the world group as the default group. However, it is extremely useful to be able to specify some other group as the default. Parallel codes that have been written using the world group as the default can be repackaged as modules that run on a subgroup by wrapping them as a subroutine call. The calling program first creates the subgroup, sets the default group of all processes in the group to be the subgroup, and then calls the subroutine. The subroutine then runs exclusively on the subgroup. The advantage of this type of programming is that it allows modules to be developed as standard parallel codes before converting them to run on groups. This capability is explicitly supported in the Global Array Toolkit by subroutine calls that allow the programmer to set the default group, but the concept could also be executed using MPI by replacing explicit references to MPI_COMM_WORLD with a variable reference.

Parallel Molecular Dynamics Using Shared Memory-Style Programming

The parallel molecular dynamics code used in these studies is designed to simulate a Lennard-Jones fluid and fluid mixtures. There are no bonded interactions in this system, so there is no need for tables of bonded interactions and excluded atoms. The code is complicated by the need to incorporate a confining sphere that encloses one component of the Lennard-Jones mixture by forcing all atoms of the component to remain within a fixed distance of the center of mass of the cluster. This confining sphere uses a hard-sphere interaction to specularly reflect particles that cross the sphere boundary.

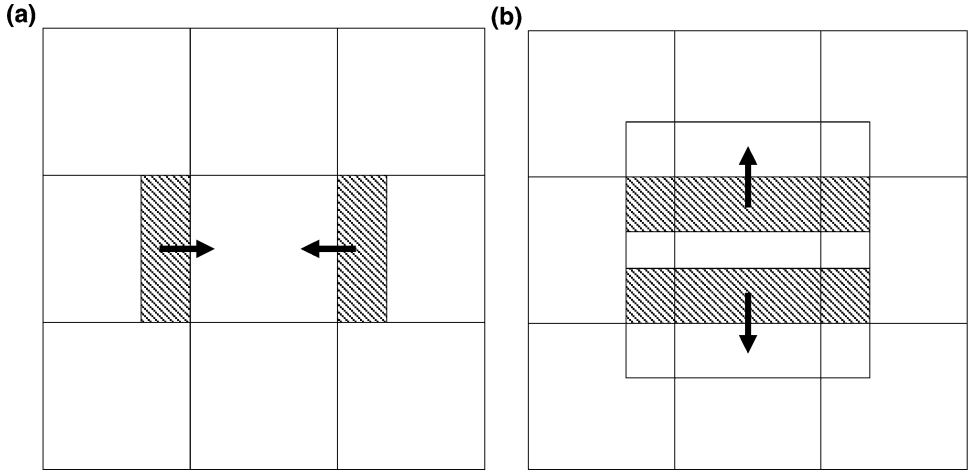


Figure 2. Schematic diagram of shift algorithm for a 2D system. (a) In the first stage, data that is within the boundary region of a neighboring east or west processor is sent to the neighboring processor. (b) In the second stage, data that is within the boundary region of a neighboring north or south processor is sent to the neighboring processor. This includes data that were transferred from east and west neighbors in the first stage.

The basic simulation kernel uses a spatial decomposition algorithm[11] to evaluate forces and distribute particle data. The original set of processors is decomposed into a three-dimensional block of processors (a set of eight processors would be divided into a cube of $2 \times 2 \times 2$ processors). The dimensions of the processor block grid are used to divide the simulation volume into a set of equal-sized cells. Particles in each cell of the simulation volume are then assigned to the corresponding processor. Forces in the Lennard-Jones fluid have a finite interaction length; that is, only particles that are within a finite cutoff distance R_c of each other have any interactions. To evaluate the forces at each time step, it is necessary to get the coordinates of all particles on surrounding processors that are within the cutoff distance. To reduce the number of data-transfer operations, it is usually desirable to pad the cutoff distance by an extra amount, δ. Thus, each processor must collect all particles that are within a distance $R_c + \delta$ of the spatial boundaries associated with the processor. This operation is accomplished using the shift algorithm.[11]

The shift operation consists of gathering particles that lie within a distance $R_c + \delta$ of the cell boundary along the x axis and sending their coordinates to the adjoining processors. This is illustrated in Figure 2a. The next step is to move all particles within a distance $R_c + \delta$ of the cell boundary along the y axis to adjoining processors. The particle locations from the adjoining cells along the x axis are already on the processor, so these can be sent along with particles located locally on the processor. This is illustrated in Figure 2b. After the transfer along the y axis is complete, the particles are transferred in a similar manner along the z axis. If all processors execute these transfers, then each processor will have a complete list of coordinates for all particles on neighboring processors within a distance of $R_c + \delta$ of the local cell boundaries. The fact that each of the updates is done sequentially along the axes reduces the total number of transfers to six, compared to the 26 transfers that would be required in three dimensions if data were collected from each of the neighboring processors independently. The shift algorithm generally needs to be performed twice during each time step. The first shift is before the force calculation when all particle coordinates must be exchanged. The second shift occurs after the force calculation, when all of the partial forces must be sent back to their home processors. This second shift can be eliminated by dropping the use of Newton's second law on all interactions between a locally held particle and a remotely held particle, but this can significantly increase the time spent in the force calculation.

The data transfer must be repeated at every time step. At periodic intervals, it is also necessary to redistribute the particles so that particles that have drifted outside the box associated with a given processor are assigned to new processors. Because the cutoff distance is padded by the amount δ, particles that have drifted a short distance outside the simulation cell do not cause any problems with the force calculation. Eventually, however, particles will move far enough that interactions will be missed if the particles are not reassigned to their appropriate processors. This reassignment is performed after a fixed number of steps. To summarize, the major communication steps in this algorithm are (1) the distribution of particles to processors on the basis of current coordinates, (2) the gathering of particle coordinates before evaluating forces, and (3) the scattering of partial forces back to particles after the force calculation. The bulk of the communication is from steps 2 and 3, which must be performed at every time step; the particle redistribution step occurs less frequently. Additional communication steps are associated with initialization and closeout of the calculation, but these are typically an insignificant fraction of the total execution time. There are also synchronizations and global summations of small quantities of data.

Individual shifts can be implemented using either message passing or using a shared-memory programming model. The message-passing model assumes point-to-point communication in which data are sent from a local array on one processor to a local array on another processor. Typically, this requires that both the sending and receiving processes post signals that they intend to send or receive a message,

and they must also specify the message length. Since the length is generally not known beforehand, this requires the exchange of an extra integer before the actual list of coordinates and so forth can be sent. Instead of exchanging individual messages, the shared-memory model creates a set of shared memory or otherwise globally accessible arrays that are used for communication. The shared memory arrays are either one-dimensional or quasi-one-dimensional (the second dimension is typically small). To communicate, two arrays are needed. The first array, referred to as the size array, is one-dimensional and only contains a number of elements equal to the number of processors. Each processor owns one data element that is used to store the number of particles that need to be moved. The second array, referred to as the communication array, is much larger. For concreteness, assume that the array will be used for moving coordinates around. A parameter is chosen that represents the maximum number of atoms that can be expected to reside in the communication array for any one processor at any point in the communication cycle. Call this parameter MAXAT. The size of the communication array for moving around a vector quantity such as the coordinates is then NPROCS $\times$ MAXAT $\times$ 3, where NPROCS is the total number of processors. Each processor then owns a MAXAT $\times$ 3 sized block of the total communication buffer.

A communication step takes place in two stages. In the first stage, all processors determine which particles need to have their coordinates sent to a neighboring processor. These coordinates are all gathered into a local buffer and then copied into the portion of the communication array owned by that processor, using a PUT operation. The number of particles that each processor puts in the communication array is also placed in the size array. After these operations are complete, a synchronization call is made to guarantee that all data is finished being copied to the communication array and the size array. The second stage begins by having each processor get the number of elements in the size array from the appropriate adjoining processor using a GET operation. Once this number has been obtained, the processor can get the block of data corresponding to the adjoining processor from the communication array. The entire sequence of operations is illustrated schematically in Figure 3 for a 2 $\times$ 2 array of processors. Each processor is receiving data from the processor to its right. Periodic boundary conditions are assumed, so each processor at the end of a row wraps around and gets data from the first processor in the row.

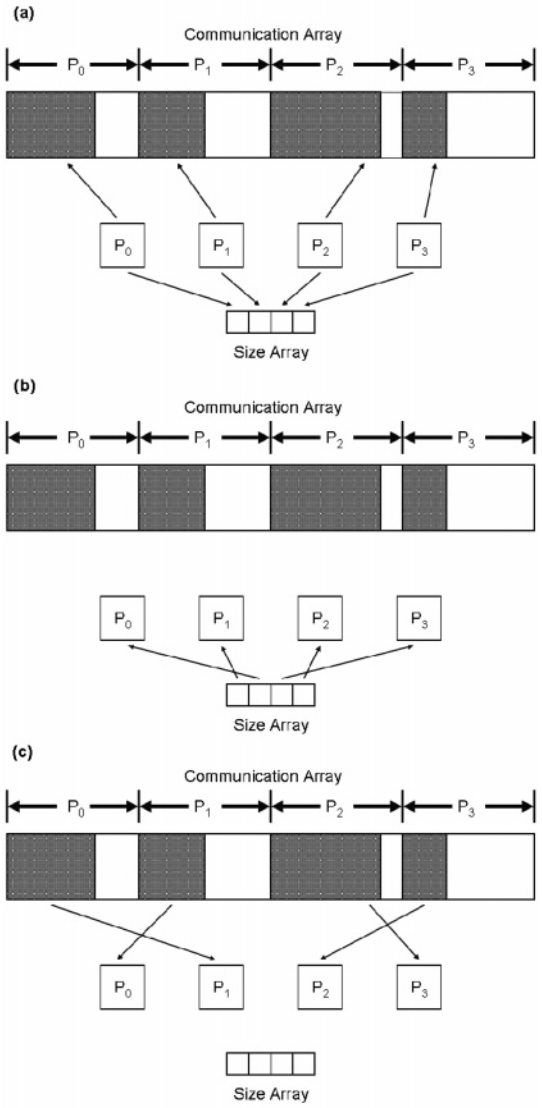


Figure 3. Schematic illustration of particle exchange for a four processor calculation which is decomposed into a 2 $\times$ 2 grid. (a) Each processor copies particle coordinates and so forth that need to be transferred to another processor to its locally held portion of a communication array. It also copies the number of particles that are in the communication array to a size array. (b) After a synchronization operation, each processor gets the amount of data from neighboring arrays from the size array. (c) Each processor gets the data for a neighboring process from the communication array.

Statistical Mechanics of Cluster Nucleation

The multiple task simulation targeted in this paper is the evaluation of atomic cluster size distribution functions for clusters of an immiscible fluid in a solvent. These calculations are not traditional free-energy calculations but share the feature of consisting of multiple independent simulations. The distributions that result from the individual simulations are converted into free-energy profiles for further analysis. The profiles lead directly to cluster dissolution rates. When combined with additional simulations of the relative free energies of clusters with different numbers of particles, they can also give the cluster nucleation rates. When such information is obtained for all cluster sizes up to the critical cluster (i.e., the cluster that defines the peak in the work of formation), a complete picture of the nucleation of phase separation can be developed.

A parallel single-task MD code was written that is designed to evaluate the cluster size probability distribution function

$$P(V) \propto \exp(-\beta p_{\text{ext}} V) \int \mathrm{d}\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N)] \prod_{i=1}^{M} \theta(r_i - r_{\text{cnf}})$$

where $\mathbf{r}^N$ is a collective coordinate representing the location of N particles, U is the potential energy of those particles, β

$= 1/k_BT$ is the inverse temperature, $V = 4\pi r_{cnf}^3/3$ is the volume of the confining sphere, r_{cnf} is the corresponding radius of the confining sphere, r_i is the distance of a particle to the cluster center of mass, $\theta(x)$ is the Heaviside step function, and the product runs over a subset containing M particles. The external pressure p_{ext} is an arbitrary applied pressure that guarantees that the probability distribution decays to zero at large volumes. This distribution can be directly related to the evaporation (or dissolution) rate of a cluster of size M.[12, 13] For evaporation, the system is chosen so that $M = N$, but for dissolution, the system would be configured so that $M < N$. Once evaluated, $P(V)$ can be converted to a free energy, $A(r_{cnf})$, using the expression[12]

$$A(r_{cut}) = -k_BT \ln[P(V)] - p_{ext}V = -k_BT \ln[P(r_{cut})/(4\pi r_{cnf}^2)] - p_{ext}V$$

The second expression is frequently more useful since the Monte Carlo sampling actually generates the distribution $P(r_{cnf}) = 4\pi r_{cnf}^2 P(V)$. Note that the explicit dependence of $A(r_{cnf})$ on p_{ext} drops out in this expression. The monomer dissolution rate of the cluster is directly proportional to the derivative of A with respect to r_{cnf}.

The problem of evaluating the distribution $P(V)$ can be reduced to simulating the behavior of a collection of particles inside a confining sphere of radius r_{cnf}. For simplicity, only the case in which all particles are contained in the confining sphere is discussed; the extension to the case where a subset of particles is restricted to the confining sphere is straightforward. The distribution $P(V)$ has been simulated previously using Monte Carlo techniques.[12,13] In this paper, the use of molecular dynamics simulations to evaluate $P(V)$ is explored. The confining sphere can be considered as an infinitely hard potential whose center corresponds to the center of mass of the confined particles. Whenever a particle hits the confining sphere, it is reflected specularly away from the sphere surface. The velocity of the remaining particles is also altered slightly because the momentum of the center of mass of the remaining particles is specularly reflected by the collision as well. A more detailed description of how these collisions are implemented will be provided elsewhere; for the purposes of this paper, it is important to note the following points:

(1) The presence of "hard-sphere"-type collisions means that time steps during which a collision occurs must be broken up into smaller intervals. Before and after the collision, the trajectories are smooth and can be integrated using a conventional second-order algorithm. The velocity−Verlet algorithm, recast as a second-order Gear predictor−corrector,[14] is especially convenient because it only needs the current coordinates and velocities to execute the next step. After determining that a collision with the confining sphere occurred during a time step, the time at which the collision occurred is calculated and the original step is broken up into pre- and postcollision intervals. The system is updated using a time step equal to the precollision interval, and then the velocities are modified using the rules for hard-sphere dynamics. Once the postcollision velocities are calculated, a second step is taken using a time step equal to the postcollision interval. If another collision is detected within the second interval, then the process is repeated until the original time step has been completed.

(2) For parallel simulations, hard-sphere dynamics associated with the confining sphere add several additional global operations. These are needed because each processor independently determines whether one of its particles has hit the confining sphere and, if so, when. This information must then be shared with all other processors. Additionally, processors also check for special cases, such as multiple collisions within a single time step and trajectories that loop into and out of the confining sphere within a time step, and communicate the results of these checks to other processors. Although these special cases occur infrequently, they cause significant problems if not handled correctly.

(3) The confining sphere radius is chosen using a conventional Monte Carlo procedure.[12] After a fixed number of steps, a new confining sphere volume is generated. If the new volume is smaller than the old volume, then a check is performed to see if any of the particles in the cluster have moved outside the confining sphere. If they have, the volume is rejected. If no particles are outside the sphere, then the sphere is accepted or rejected using a standard Monte Carlo acceptance test based on the change $p_{ext}\Delta V$. This test also adds an extra layer of synchronization.

Overall, the simulations proceed much like conventional molecular dynamics simulations. The main differences are that the coordinate update steps are more complicated and involve much higher levels of synchronization.

The system investigated in this study consists of a two-component Lennard-Jones fluid characterized by the potential function

$$U(\mathbf{r}^N) = \sum_{i<j} \varphi_{ij}(r_{ij})$$

where the Lennard-Jones interaction φ_{ij} has the form

$$\varphi_{ij}(r_{ij}) = \begin{cases} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] - \varphi_{ij}^0 & r_{ij} < R_{ij}^c \\ 0 & r_{ij} \geq R_{ij}^c \end{cases}$$

The constant φ_{ij}^0, representing the cutoff interaction energy, is

$$\varphi_{ij}^0 = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}^c}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}^c}\right)^{6}\right]$$

This choice guarantees that $\varphi_{ij}(r)$ goes continuously to zero as r approaches R_{ij}^c. For these simulations, the fluid contains two components denoted by A and B. For all interactions, σ_{ij} and ϵ_{ij} are set equal to 1. The particle masses are also set to 1. If both i and j belong to the same component, then R_{ij}^c is set equal to 2.5. If i and j belong to different components, then R_{ij}^c is set equal to $2^{1/6}$. These choices guarantee that interactions between particles in the same component are attractive at longer distances while interactions between particles in different components are purely repulsive. The pair interactions are illustrated in Figure 4. For these simulations, the component forming the cluster is B (referred to hereafter as the solute) while A is the majority component

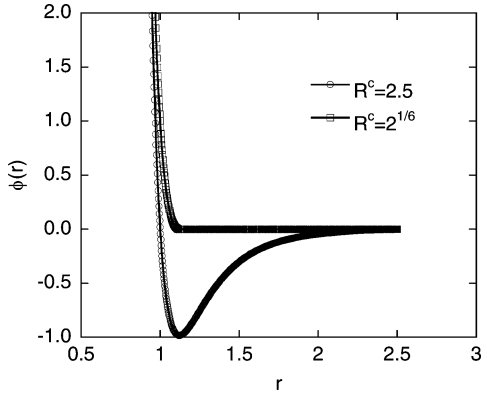


Figure 4. Plot of Lennard-Jones interaction potential using both long- and short-range cutoffs.

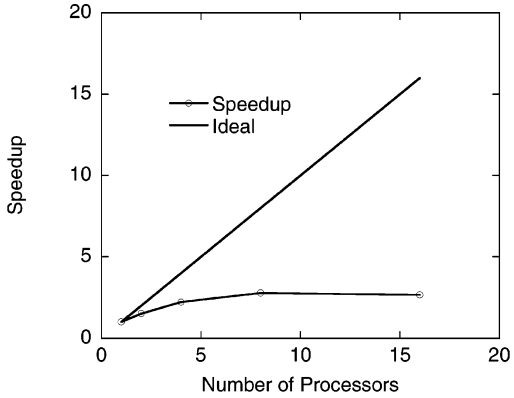


Figure 5. Single task scaling behavior of molecular dynamics simulation.

and acts as the solvent. The repulsive cross interaction means that B will tend to form clusters within A.

The output of the simulation is the distribution $P(r_{cnf})$ [which is readily converted to $P(V)$]. This is obtained by binning the corresponding values of the radius of the confining sphere that are selected at each Monte Carlo step. The program has also been modified to incorporate an equilibration protocol into each trajectory along with an additional setup routine that creates an initial condition by specifying the number of Lennard-Jones particles of each type that are used in the simulation. The initial conditions are chosen so that the minority B component is always located in a roughly spherical blob at the center of the simulation cell. To improve convergence, the range over which the confining sphere radius is sampled is restricted. The simulation is initially run for a short period during which the confining sphere radius is brought within the sampling range, it is then equilibrated further. After equilibration, data is collected and the distribution $P(r_{cnf})$ is evaluated. The data collection stage is much longer than the equilibration stage, primarily because the trajectory has to be run for a very long time to get good statistics for $P(r_{cnf})$.

Results

The scaling properties of the groups-based simulation code were investigated by first determining the scaling behavior of the isolated molecular dynamics module running as a standard parallel application. After determining the single-task scaling behavior, the scaling properties of multiple tasks running on groups were evaluated. The single-task scaling behavior was determined for a system consisting of 4000 solvent particles and two solute particles. The simulations were run for 50 000 steps using a time step of $\Delta\tau = 0.01$. These simulations are too short to provide valid scientific results, but they are long enough to characterize the scaling behavior of the code. Longer simulations that produced useful data are discussed below. The simulations were run using the isothermal/isobaric method of Nosé and Anderson[15,16] with a reduced temperature of 0.8 and a reduced pressure of 0.05. This locates the simulations in the liquid part of the Lennard-Jones phase diagram for the pure Lennard-Jones fluid.[17] The single-task scaling behavior of the molecular dynamics module on an HP dual Itanium2 1.5 GHz cluster with a Quadrics switch using the Elan communication libraries is shown in Figure 5. It can be seen that the scaling behavior of this calculation is fairly poor and drops off rapidly as the number of processors increases. The speedup from one to two processors is reasonable but starts falling off significantly in going from two to four processors. Beyond that, the speedup is relatively flat, and at 16 processors, the speedup has actually started to decrease. The poor scaling for this calculation is primarily due to the relatively small problem size and very poor load balancing associated with the initial condition. Simulations run for longer periods of time have better load balancing as the density equilibrates and density fluctuations present at the beginning disappear.

To perform multiple, concurrent parallel tasks, the molecular dynamics code was modified to divide the available processors in the world group into subgroups. The code then executes the tasks on the groups. When a task is completed on one group, the group gets the next task from a global counter. The global counter is implemented using the READ-INCREMENT functionality available in the Global Arrays Toolkit. A shared array, consisting of a single integer is created at the start of the calculation and initialized to zero. The READ-INCREMENT function reads the current value of an array variable and increments it by 1. As each processor group completes a task, the processor corresponding to process 0 in the group performs a read-increment on the array variable acting as the global counter and translates the current value of the variable into a set of task parameters.

Simulations of M tasks were run with each task representing a system of 4000 solvent particles and between two and $M + 1$ solute particles. For the timing studies described below, the number of tasks was set to 512. Most of the remaining parameters are the same as for the calculations described in the single-task scaling study, except that the number of time steps was reduced to 10 000 in order to reduce the overall simulation time. Because the tasks with larger numbers of solute particles were expected to take slightly longer, the tasks were executed in reverse order to improve load balancing. These simulations were run on between eight and 1024 processors on the HP cluster described above. On the basis of the single-task simulation results, the available processors were divided into groups of two processors. The results are shown in Figure 6.

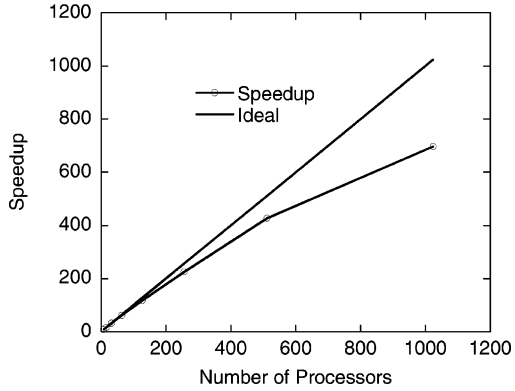


Figure 6. Speedup curve for 512 tasks as a function of the number of processors. The available processors were divided into groups containing two processors each.

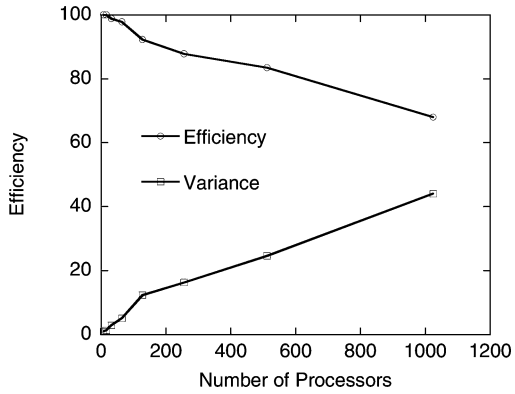


Figure 7. Parallel efficiency and maximum variance in execution times on each of the groups as a function of the number of processors for 512 tasks.

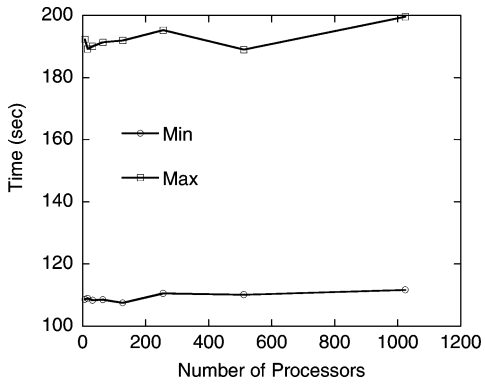


Figure 8. Plot of minimum and maximum task execution times for 512 tasks.

The results show good speedup all the way up to 1024 processors. The parallel efficiency, which is the actual speedup divided by the theoretical maximum speedup, is shown in Figure 7. The efficiency remains above 80% all the way up to 512 processors and is close to 70% even for 1024 processors. The main reason for the drop in efficiency appears to be the granularity of the execution times of the individual tasks. As the number of tasks per group drops, the variation in the amount of time required for each group to finish its tasks increases.

The variation in individual task execution times can be seen in Figure 8, where the minimum and maximum execution times for the individual tasks are plotted as a function of the number of processors. If the individual tasks are truly running independently, then each task should take about the same amount of time regardless of the total number of processors. This further implies that the minimum and maximum execution times for the collection of tasks should also be independent of the total number of processors, and this is seen in Figure 8, although there may be a small increase in execution times at very large processor counts. The figure also shows that there is a substantial variation in execution times. The maximum execution time is over 60% higher than the minimum execution time. This variation is much larger than the variation in the total number of particles in the different tasks (approximately 13%) and is due to the relatively short execution times and differences in short-time behavior caused by differences in the initial configuration.

The variance in execution time on individual groups was also calculated and is plotted along with the parallel efficiency in Figure 7. If T_i represents the total time spent on group i executing its tasks and $T_{\min}$ and $T_{\max}$ are the minimum and maximum values of T_i across the different groups, then the variance is $100(T_{\max} - T_{\min})/T_{\max}$. Note that $T_{\max}$ is almost identical to the overall execution time for the entire calculation. The variance is inversely correlated with parallel efficiency; as the variance increases, parallel efficiency decreases. This supports the conclusion that the main source of the decrease in parallel efficiency is the granularity of the individual tasks. This problem would probably decrease for longer simulations where the variation in initial conditions is not as important. For longer simulations, the execution times should track more closely with the total number of particles, and the execution times would be much closer together.

The short calculations described above were good only for evaluating scalability of the code. To produce useful scientific data, much longer calculations are needed. To demonstrate that such calculations are possible, a set of simulations consisting of 10^8 steps were run on systems containing two to five solute particles. In order to run simulations containing such a large number of steps in a reasonable amount of time, the number of solvent particles was reduced to 400. For the small solute clusters investigated in this study, 400 solvent particles should be sufficient. Simulations in this regime indicate that the correlation length of the radial distribution function for the pure Lennard-Jones fluid is on the order of $3-4\sigma$, while the typical dimension of the simulation cell is a little over 8σ. The first 2 000 000 steps were used to move the confining sphere to within the maximum value of 4.0σ and to equilibrate the system. The remainder of the simulation was used to collect data. The simulations were run on eight processors divided into four groups of two processors each. Some shorter runs on trajectories containing 1 000 000 steps indicate that the speedup in going from one to two processors is about 1.6. The total execution time for all four calculations was about 52 h, so the use of groups amounted to a substantial decrease in the overall execution time. There is no additional benefit to using groups containing more than two processors for a system this small since the speedup actually starts to decrease

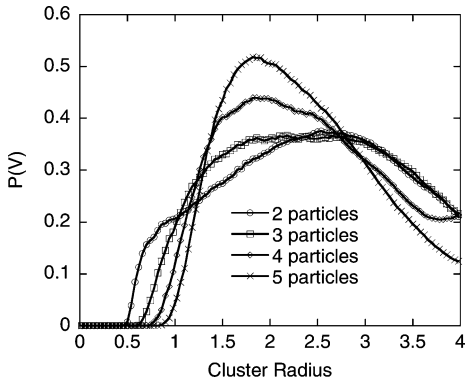


Figure 9. Confining sphere size distributions for simulations containing two to five solute particles and 400 solvent particles.

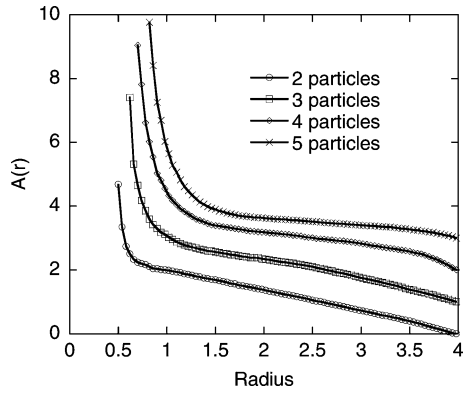


Figure 10. Cluster free energies as a function of size for simulations containing two to five solute particles and 400 solvent particles.

at four processors. For these longer simulations, the load imbalance issues are much smaller. The total variation in execution times for the different tasks is about 1.6%.

The probability distributions for the confining sphere radius are shown in Figure 9. A few things are worth noting. First, the leading edge of the distributions moves to larger radius values as the number of solute particles increases. This is expected because more particles imply a larger cluster. Second, even with 10^8 steps, the distributions still appear to contain noticeable amounts of noise. This is a common problem with simulations containing many degrees of freedom where the focus is to determine the behavior of a single coordinate. It is difficult to obtain accurate averages since each point in the distribution must be sampled many times but at the same time a substantial amount of work is required for each system update.

The actual quantity needed for evaluating dissolution rates is the free-energy curve $A(r_{cnf})$, which can be derived from the distributions $P(V)$. The curves for $A(r_{cnf})$ are shown in Figure 10. Because the $A(r_{cnf})$'s are proportional to the logarithm of the distribution functions, much of the noise in the $P(V)$ curves is suppressed. It is possible that, after applying some smoothing algorithms to these curves, reasonable values of both the derivative of $A(r_{cnf})$ and the location of the minimum value of the derivative could be obtained and used to evaluate the dissolution rate. A more extensive analysis of these results and results for larger clusters will be presented elsewhere.

As a final comment on the performance of multitask codes using groups, if only four processors had been available, the most efficient way to perform the simulations would be to run on groups containing only a single processor. This would be equivalent to the embarrassingly parallel algorithm. The reason for using only one processor per group is that efficiency is highest for single-processor calculations, and since the number of tasks exceeds the number of processors, the shortest time to solution will be obtained by running each task at the highest possible efficiency. This is generally true for any calculation with a limited number of processors. The shortest time to solution is obtained by running individual tasks on the smallest number of processors on which the calculation will fit. It is only when a large number of processors are available so that *all* tasks can be run concurrently that there is a time benefit to increasing the number of processors in the group.

Conclusions

The results presented here demonstrate that processor groups can be used to substantially reduce the amount of time required to perform multiple independent molecular dynamics simulations. The scenario of many independent calculations is often encountered in evaluations of free energies, which are an important class of calculations in biochemical simulations and other evaluations of chemical stability and reactivity. The results of these scaling studies indicate that processor groups can be used quite successfully to run multiple concurrent instances of a parallel calculation. The results also illustrate the utility of the default processor group concept, which allows developers to easily reconfigure parallel applications to run on groups with only minor adjustments to the code.

For free-energy calculations, processor groups can be used to either run multiple concurrent simulations when each individual simulation is too large to fit on a single processor or can be used to speed up simulations in which the number of available processors exceeds the number of individual simulations. Without processor groups, the only available alternatives are to run individual simulations on single processors following an embarrassingly parallel model or to submit multiple independent calculations to the computer. The first option can only be used if the individual simulations fit on a single processor, and the second option, besides being inconvenient, breaks down completely if more complicated simulation protocols are used in which initial results are used to bias the sampling and improve statistics.

Acknowledgment. This work was supported by Pacific Northwest National Laboratory (PNNL) and the U.S. Department of Energy's Office of Science through the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) and the Division of Chemical Sciences and was performed in part using the Molecular Science Computing Facility (MSCF) in EMSL. The MSCF and EMSL are funded by the Office of Biological and Environmental Research in the

U.S. Department of Energy. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC06-76RLO 1830.

References

(1) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzymatic Catalysis, Protein−Ligand, Protein−Protein, and Protein−Nucleic Acid Noncovalent Interactions. *Ann. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211.

(2) Garrett, B. C.; Schenter, G. K.; Morohita, A. Molecular Simulations of the Transport of Molecules across the Liquid/Vapor Interface of Water. *Chem. Rev.* **2006**, *106*, 1355.

(3) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781.

(4) Germain, R. S.; Zhestkov, Y.; Eleftheriou, M.; Rayshubskiy, A.; Suits, F.; Ward, T. J. C.; Fitch, B. G. Early Performance Data on the Blue Matter Molecular Simulation Framework. *IBM J. Res. Dev.* **2005**, *49*, 447.

(5) Straatsma, T. P.; McCammon, J. A. Multiconfiguration Thermodynamic Integration. *J. Chem. Phys.* **1991**, *95*, 1175.

(6) Tobias, D. J.; Brooks, C. L;, Fleischman, S. H. Conformational Flexibility in Free Energy Simulations. *Chem. Phys. Lett.* **1989**, *156*, 256.

(7) Tobias, D. J.; Brooks, C. L. Calculation of Free Energy Surfaces using the Methods of Thermodynamic Perturbation Theory. *Chem. Phys. Lett.* **1987**, *142*, 472.

(8) Widom, B. Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963**, *39*, 2808.

(9) Straatsma, T. P. Free Energy by Molecular Simulation. *Rev. Comput. Chem.* **1996**, *9*, 81.

(10) Nieplocha, J.; Palmer, B. J.; Tipparaju, V.; Krishnan, M.; Trease, H. E.; Apra, E. Advances, Applications and Performance of the Global Arrays Shared Memory Programming Toolkit. *Int. J. High Perf. Comput. Appl.* **2006**, *20*, 203.

(11) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J. Comput. Phys.* **1995**, *117*, 1.

(12) Kathmann, S. M.; Schenter, G. K.; Garrett, B. C. Dynamical Nucleation Theory: Calculation of Condensation Rate Constants for Small Water Clusters. *J. Chem. Phys.* **1999**, *111*, 4688.

(13) Schenter, G. K.; Kathmann, S. M.; Garrett, B. C. Variational Transition State Theory of Vapor Phase Nucleation. *J. Chem. Phys.* **1999**, *110*, 7951.

(14) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.

(15) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384.

(16) Nosé, S. A. Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511.

(17) Hansen, J.-P.; Verlet, L. Phase Transitions of the Lennard-Jones System. *Phys. Rev.* **1969**, *184*, 151.

CT600260U

JCTC Journal of Chemical Theory and Computation

Computational Requirements for Simulating the Structures and Proton Activity of Silicaceous Materials

Yuan Zhang, Zhen Hua Li, and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received September 18, 2006

Abstract: Structures of disiloxane and silanol and the energetics of their protonation, deprotonation, and proton exchange reactions have been studied with 14 density functionals in combination with eight basis sets. The geometries optimized by these 112 density functional methods are compared to those obtained by the second-order Møller−Plesset perturbation theory and the coupled cluster method, and the performance of all these methods on energetics is evaluated with benchmark Weizmann-1w results. The most accurate density functional for both geometries and energetics is M05-2X. Polarized augmented triple-ζ basis sets are found to be about a factor of 3−4 more accurate than polarized augmented double-ζ basis sets.

1. Introduction

Silicon is abundant in the earth's crust, where it occurs mainly as silica (silicon dioxide) and silicates. There have been many experimental and theoretical investigations of crystalline[1−12] and amorphous[13−22] silicates because of their importance in materials science, geophysics, and technology. In addition, fabricated silaceous mesoporous materials are receiving considerable attention for use and potential use as catalytic nanoreactors and nanotechnological compounds.[23−25] The fundamental structural building blocks of a variety of silica-containing materials are also found in small molecules, which may therefore serve as useful model compounds. In particular, disiloxane ($H_3Si-O-SiH_3$, Figure 1)[26−28] exhibits Si−O bond lengths and Si−O−Si bond angles that are virtually the same as those observed in bulk silica and silicates,[29] and the Si−OH hydroxyl group of silanol (H_3-Si−OH, Figure 2)[30−32] is a molecular mimic of the group found on the surfaces[33] and at the defect sites[34] of hydrous silica and zeolitic materials. These geometric features are very important in zeolites because they control the acidity of the lattice. Silicates and related aluminosilicate materials are major constituents of hydrothermal flows and play an important role in magmatic flow in the earth's crust; they have therefore been investigated extensively by many geologists.[35−37] The incorporation of H in silica is critical for its role in subducting oceanic crust.[38] Proton conduction in water-containing silica, the use of water as a defect passivant in semiconductor devices (and related applications),[39] and the presence of water in optic fibers[40] underline the technological importance of proton energetics in silicates.

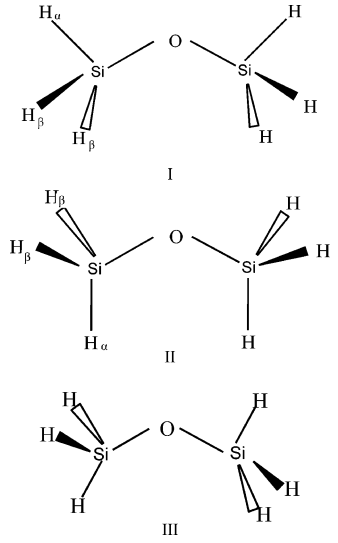


Figure 1. The three conformers of disiloxane.

Molecular force fields have been shown to be useful for reproducing the bulk properties of silica polymorphs, especially when the potential function was derived from extended-basis-set quantum mechanical electronic structure calcula-

* Corresponding author fax: 1-(612) 626-9390; e-mail: truhlar@umn.edu.

10.1021/ct6002884 CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/02/2007

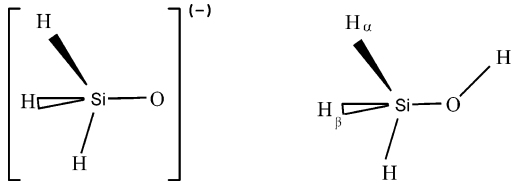


Figure 2. The structures of the silanol anion and silanol.

tions.[41] In fact, it is well-known that modeling silicon-containing compounds requires careful consideration of basis sets.[42] Advances in computational science allow quantum mechanical methods to be applied with large basis sets and high levels of theory to small systems and moderate basis sets and moderate levels of theory to larger and larger cluster models.[43−48] A validation study designed to show the computational requirements for modeling silicaceous material is a key starting point for examining the required levels of theory and sizes of basis sets. In particular, density functional theory (DFT) can be used to test the large number of molecular mechanics force fields[13,32,41,49−56] that have been proposed for silica, silicates, alkylsilanes, and disiloxane polymers. This is especially important because disiloxane has proven to be a highly useful benchmark for the accuracy of electronic structure methods for use in zeolite calculations.[27,57,58] Disiloxane serves not only as a model for silicone oxides and silicates but also as a model for polymeric organosiloxanes such as silicone oil[13,55,59,60] (polydimethylsiloxane polymers). In the present work, we use disiloxane and silanol as prototype molecules for testing and validating quantum mechanical computational methods.

We also study protonation and deprotonation of these two model systems, silanol and dimethylsiloxane. Knowledge of proton affinities is essential for studying proton transport in minerals, and the protonation of dimethylsiloxane may be considered to be a model for the protonation of an acid site in an all-silica zeolite or for diffusion of hydrogen in acidic silica or zeolites.[61−63]

For direct dynamics simulations, one wants to use the smallest possible basis set that yields the required accuracy, so it is essential to test both small and larger basis sets systematically. Many of the older studies of basis set extension were based on[42,64] wave function theory (WFT), but it is now widely recognized that DFT provides a more efficient and more accurate electronic structure approach for direct dynamics calculations. For example, both the B3LYP and M05-2X density functionals are more accurate than second-order-perturbation WFT.[65] Furthermore, it is now known that DFT has different[66−70] and sometimes smaller basis set requirements than WFT. It is also known that density functionals that are optimum for main-group chemistry may be very inaccurate for metals and vice versa,[71] with less systematic tests available for semiconductor elements like Si. These considerations motivate a systematic exploration of the accuracy of various density functionals and basis sets for prototype Si-containing molecules.

Both experimental and theoretical studies have shown that it is very difficult to determine with reasonable confidence the structures of disiloxane and silanol, in particular, the important Si−O−Si bond angle and the Si−O bond distance.

Table 1. Summary of the DFT Methods Assessed in This Study

	X[a]	year	type	ref(s)
HCTH	0	1998	local[b]	78−80
M06-L	0	2006	local	82
mPWPW	0	1998	local	83, 84
PBE	0	1996	local	85
VSXC	0	1998	local	89
MPWLYP1M	5	2005	HDFT[c]	70
TPSSh	10	2003	HDFT	87, 88
B3LYP	20	1994	HDFT	74, 75, 77
B97−1	21	1998	HDFT	78
B1LYP	25	1997	HDFT	74−76
M05	28	2005	HDFT	81
MPW1B95	31	2004	HDFT	53, 56, 78
PW6B95	46	2005	HDFT	86
M05-2X	56	2006	HDFT	71

[a] X denotes the percentage of Hartree−Fock exchange in the functional. [b] The term local denotes local density functional theory. [c] HDFT denotes hybrid density functional theory.

The geometries of these molecules obtained by theoretical methods are highly sensitive to the sophistication of the calculations. DFT is a powerful means by which zeolite structure, acidity, and reactivity can be elucidated.[72,73] DFT methods can provide very accurate results for geometries, energies, and other properties. However, the computational requirements of DFT are less demanding than those for wave function theory methods of comparable accuracy, allowing one to efficiently study the large systems needed for the realistic modeling of zeolitic materials. The accuracy of the DFT calculations depends on a variety of factors; the most important of these are the flexibility of the basis set and the form of the exchange-correlation functional. A larger, more flexible basis set gives a better description of the electron density but is computationally more demanding.

In the present study, the accuracy of 14 density functionals and eight basis sets was evaluated by calculating the geometries of disiloxane and silanol and the energies of their protonation/deprotonation and proton-transfer reactions. The 14 DFT functionals that we assess are B1LYP,[74−76] B3LYP,[74,75,77] B97-1,[78] HCTH,[78−80] M05,[81] M05-2X,[71] M06-L,[82] MPW1B95,[53,56,78] MPWLYP1M,[70] mPWPW,[83,84] PBE,[85] PW6B95,[86] TPSSh,[87,88] and VSXC.[89] All DFT methods in the present paper are listed in Table 1, ordered by the percentage X of Hartree−Fock (HF) exchange in the functional and, for a given X, are in alphabetical order.

The eight basis sets evaluated here include five polarized double-ζ basis sets, 6-31+B**,[90] 6-31+G**,[91] MIDI!,[92−94] MIDIX+,[95]and aug-pc1,[67,69] and three polarized triple-ζ basis sets, aug-cc-pV(T+d)Z,[96] aug-pc2,[67,69] and MG3S.[68] The aug-pc1 and aug-pc2 basis sets were created[67,69] expressly for use with DFT, so it is particularly interesting to test their accuracy and efficiency. Note that MIDI! is called MIDIX in the *Gaussian* program.

Two ab initio WFT methods, namely, Møller−Plesset second-order perturbation theory (MP2)[97] and coupled cluster theory with single and double excitations (CCSD),[98] were also used to optimize the geometries. The CCSD geometries are used as a benchmark in some cases (see below), but the

Table 2. Number of Contracted and Primitive Basis Functions Used in the Calculation Of Protonated Disiloxane ($H_6Si_2OH^+$)[a]

	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	av (PDZ)	MG3S	aug-pc2	aug-cc-pV(T+d)Z	av (PTZ)
contracted	64	76	97	100	140	95.4	205	307	317	276.3
primitive	108	120	193	193	274	177.6	293	508	536	445.7
ref(s)	92−94	95	90	91	66, 68		67	66, 68	96	

[a] PDZ denotes polarized double-ζ; PTZ denotes polarized triple-ζ; av denotes average.

MP2 calculations were carried out only for comparison purposes. High-level benchmark calculations were performed using the multicoefficient correlation method MCCM/3,[99] the balanced multicoefficient method BMC−CCSD,[90] and the very accurate multilevel Weizmann-1w (W1w)[100−102] method with fixed geometries calculated at the CCSD/MG3S and B97-1/aug-cc-pV(T+d)Z levels. B97-1 is the recommended method for geometry optimization of large molecules in the W2 protocol; see ref 101 for more information. Furthermore, aug-cc-pV(T+d)Z[96] is recommended by Martin[102] for geometry optimization in the W1w and W2w theories. To check that these geometries are sufficiently reliable for the W1w calculations, we also carried out W1w calculations at CCSD/MG3S geometries.

Parthiban and Martin[103] studied the accuracy of W1 and W2 theory for proton affinities and found that W1 "can basically be considered converged for this purpose"; in fact, they concluded that the uncertainty in W1 proton affinities is "considerably lower than that of the experimental values". Similar conclusions apply to W1w. Thus, our tests of the accuracy for proton affinities should be quite reliable.

2. Computational Details

All DFT, MP2, and CCSD energies and geometry optimizations were carried out using the *Gaussian 03*[104] and *MN-GFM*[105] programs, and the W1w calculations (using the Douglas−Kroll[106] relativistic correction as in ref 102) were carried out using the *MOLPRO*[107] program. H_6Si_2O (see Figure 1), H_3SiOH (see Figure 2), and the corresponding protonated and deprotonated molecules were fully optimized using the 14 DFT methods in combination with a series of basis sets. Vibrational frequency calculations were performed for all the stationary points, using each method and basis set, and these calculations verified that all structures are minima (no imaginary frequencies). We allowed the molecules complete flexibility during the analytical gradient geometry optimization except for reasonable symmetry constraints, for example, C_{2v} and C_s for disiloxane, C_s for silanol, and C_{3v} for the deprotonated silanol anion (H_3SiO^-). Because of the expected very shallow potential energy profile for fluctuations of the Si−O−Si bond angle of disiloxane, the geometry optimization criteria were set to stringent values (keyword ''opt=tight" in *Gaussian 03*). The integration grid used is a pruned (99, 590) grid ("ultrafine" as defined in *Gaussian 03*).

To illustrate the relative sizes of the various basis sets, Table 2 lists the total number of basis functions used in the calculation of protonated disiloxane.

3. Results and Discussion

Before considering our results, it is useful to note which levels of theory and basis sets have been used by other workers for some of the more accurate available studies of Si-containing compounds. Nachtigall et al.[108] employed the 6-311+G(3df,2pd) level and basis sets and found that the B3LYP functional is more accurate than local functionals for Si−Si bond cleavage and H_2 elimination from silanes (1996). Eichler et al.[109] used HF theory with a triple-ζ polarized basis set on O and a double-ζ polarized basis set on H and Si for an active site surrounded by a molecular mechanics environment (1997). Chatterjee et al.[34] used B3LYP/DNP for cluster calculations, where DNP is similar in quality to 6-31G(d,p) (1998). Demuth et al.[61] used the local PW91 density functional with a projector-augmented plane wave basis for periodic boundary condition calculations on mordenite (2000). Tossell and Sahaia[110] calculated $Si(OH)_3O^-\cdot 4H_2O$ at the MP2/6-311+G(2d,p)//HF/6-311+G-(2d,p) level. Walsh et al.[111] carried out cluster calculation where the active site was treated by B3LYP/6-31G(d) and the rest of the cluster by HF/STO-3G (2000). Tielens et al.[28] used density functional theory with aug-cc-pV*X*Z (*X* = D, T, Q) basis sets for calculation on silanol and disiloxane (2001). They found that B3LYP and B3PW91 performed the best, out of six functionals studied, and that they gave very similar results. Yuan et al.[112] carried out ZSM-5 zeolite calculations by ab initio HF or DFT methods (2002). Bussai et al.[113] carried out calculations on clusters at the HF/6-31G-(d) level (2002). Zwijnenburg et al.[62] carried out calculations on clusters at the B3LYP/6-31G(d) level (2002). Simperler et al.[48] carried out cluster calculations at the PW91/DNP level (2004). Tuma and Sauer[114] carried out cluster calculations with a PBE density functional and a plane wave basis set; they also used a combined MP2/DFT method (2004). Saengsawang et al.[33] carried out cluster calculations at the B3LYP/6-31++G(d,p) level (2005). Tossell[115] calculated the dimerization of silanol with B3LYP and MP2 using the 6-311+G(2d,p) basis set and with higher-level WFT calculations with smaller basis sets and low-level geometrics (2005). Ginhoven et al.[116] calculated water reactions in silica with the PW91 density functional and a plane wave basis set. Bakk et al.[117] modeled extended systems containing siloxane building blocks at the B3LYP/6-311+G(3df,3dp)//B3LYP/6-31+G(d) and MP2/6-311+G(3df,3pd)//MP2/6-31+G(d) levels (2006). Since most of the studies of silicaceous compounds used polarized double-ζ basis sets, it will be interesting to see if this is accurate enough.

3.1. Geometries of Disiloxane and Silanol. Disiloxane provides the simplest molecular representation of the key internal coordinates of the zeolite framework. This molecule has three possible conformers, two with C_{2v} symmetry, namely, the doubly staggered (I) and the doubly eclipsed (II), and the third with C_s symmetry (III), as shown in Figure 1. In the doubly staggered conformation, one of the

Table 3. Benchmark Values of Key Geometric Parameters and Dipole Moments for Silanol and Disiloxane (Distances in Å, Angles in Degrees, Dipole Moments in Debyes)

	silanol	disiloxane
	CCSD/MG3S	experiment
Si−O	1.645	1.634[29]
Si−O−H	118.6	
Si−O−Si		151.2[120]
μ	1.327	0.24[129]

hydrogens of both silyl groups is in a staggered position with respect to the opposite Si−O bond. In the doubly eclipsed conformation, both silyl groups are rotated by 180° relative to the doubly staggered position. Conformer III can be obtained by rotating one of the silyl groups in either conformer I or II by 180°. Earlier studies on conformers I and II indicate that conformer I is lower in energy.[29] This agrees with the experimental results on disiloxane crystals, which imply that disiloxane exhibits a distorted C_{2v} structure similar to that of conformer I.[30,118]

For disiloxane, both experimental and theoretical studies indicate that the energy change with respect to the Si−O−Si angle is very small, and the energy difference from the bent structure to the linear structure is only about 0.3 kcal/mol.[119,120] This extremely flat energy surface has resulted in large uncertainties in the experimental values of the Si−O−Si angle. An electron diffraction[29] study of gaseous disiloxane reported an angle of 144.1°; an X-ray crystal structure (solid phase)[121] of hexamethyl disiloxane gives 142.1°; a Raman/IR study[119] on crystalline disiloxane gave an angle of 149 ± 2°, and a more recent low-frequency Raman study[120] suggested a Si−O−Si angle of 151.2°. The key geometric parameters and dipole moments for silanol and disiloxane, both benchmark values and those predicted by the various levels of theory, are given in Tables 3−5.

In the present study, the results of most methods yield conformer I as stable, in agreement with early studies; however, for some methods, in particular, M05-2X/MG3S, M05-2X/aug-pc2, M05-2X/aug-cc-pV(T+d)Z, M06-L/MG3S, M06-L/aug-pc2, M06-L/aug-cc-pV(T+d)Z, and the VSXC functional with all basis sets except MIDIX+, conformer I is not stable; in all of these cases, conformer II is stable and was used for the calculation in the tables. At most levels, conformer III is a first-order transition state. We must emphasize that the potential energy surface of disiloxane is very flat with respect to the rotation of the silyl groups, and the absolute energy differences between the three conformers are very small, usually just a few hundredths of a kilocalorie per mole. As a result, the experimental results may correspond to an almost equal mixture of the different conformers.

In contrast to the considerable amount of data on disiloxane, there are no experimental data available on silanol, probably because it has a strong tendency to condense into disiloxane.[27] The benchmark values in Table 3 are from theory for silanol and from experimental results for disiloxane. The theoretical results for silanol are carried out with an augmented polarized triple-ζ basis set, in particular, by CCSD/MG3S. The use of geometries calculated at this level is motivated by previous experience in achieving a compromise of affordability and reliability for the geometry optimization of molecules with six or more atoms. It may also be justified by systematic work on smaller molecules. Consider, for example, the recent systematic studies of geometries for compounds containing second-row atoms by Coriani et al.[123] The found that bond lengths at the CCSD/cc-pVTZ level were not systematically improved by increasing the level to CCSD(T) or by increasing the basis to cc-pVQZ, and bond angles were not systematically improved by increasing the basis to cc-pVTZ. They found that the mean unsigned errors in their calculated CCSD/cc-pVTZ bond lengths and bond angles are 0.008 Å and 0.4°, respectively. Their average error for CCSD bond lengths would probably have been smaller if they had a third tighter *d* function for second-row atoms such as those in the MG3S basis, which does not suffer from this known deficiency of the cc-pVTZ and aug-cc-pVTZ basis sets. We note that their study did not include any molecules with more than five atoms (it involved one symmetric penta-atomic molecule, six tetra-atomics, 23 triatomics, and nine diatomics). Thus, conclusions about large molecules are subject to more uncertainty.

In Table 4, we present the values calculated for the equilibrium Si−O−Si and Si−O−H angles of disiloxane and silanol at various levels of theory. Results in the table indicate that the Si−O−Si angle is extremely sensitive to both the method and the basis set. In extreme cases, Si−O−Si is even predicted to be linear. For the polarized double-ζ 6-31+G**, MIDIX+, and 6-31+B** basis sets, 11, 7, and 2 out of 14 functionals, respectively, give a linear structure, while none of them give linear structures for the MIDI! and aug-pc1 basis sets. Grigoras and Lane have shown that the Si−O−Si angle is very sensitive to the values of the *d*-orbital exponent on Si,[122] which may explain the poor performance of some polarized double-ζ basis sets. MIDI! and aug-pc1 are the two best polarized double-ζ basis sets for the bond angle in disiloxane. Polarized triple-ζ basis sets generally give larger Si−O−Si angles than polarized double-ζ basis sets except for those polarized double-ζ basis sets giving linear structures. However, none of these polarized triple-ζ basis sets give a linear structure. Among the three polarized triple-ζ basis sets, except at the VSXC level, MG3S always gives the largest Si−O−Si angle, while aug-pc2 gives the smallest angle. It has been suggested by Nicholas et al. that, since the experimental values are not measured at 0 K, a thermal correction of about −5° should be applied to the experimental value.[30] After the correction, all of the theoretical values obtained using the polarized triple-ζ basis sets are further away from the experimental value. Since both of the available high-level ab initio methods, in particular, the present CCSD/MG3S calculation and the previous[124] CPF calculation, give a value greater than 152°, the uncorrected 151.2° of the latest measurement is perhaps the most accurate value. Examining the results given by the polarized triple-ζ basis sets indicates that all pure functionals give smaller Si−O−Si angles than the hybrid functionals. If one accepts that the recent experimental value of 151.2° is the most accurate

Table 4. Si−O−H and Si−O−Si Bond Angles of Silanol and Disiloxane (in Degrees)

	HCTH								M06-L							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
Si−O−H	112.3	115.7	117.3	117.4	116.0	117.8	117.9	117.7	113.7	115.9	118.0	117.7	116.0	118.5	117.7	118.1
Si−O−Si	142.1	143.7	140.7	146.6	138.4	149.5	148.1	148.7	134.2	136.6	143.6	180.0	139.1	153.8	148.6	153.5
MUE[a]	7.7	5.2	6.0	2.9	7.7	1.3	1.9	1.7	11.0	8.6	4.1	14.8	7.4	1.4	1.8	1.4
	mPWPW								PBE							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
Si−O−H	112.2	115.6	117.3	117.2	115.8	117.5	117.2	117.1	111.8	115.2	117.0	116.9	115.6	117.2	117.0	116.9
Si−O−Si	139.0	140.8	142.4	180.0	137.4	148.6	145.4	146.4	137.0	138.1	139.2	145.0	135.6	145.9	143.6	144.3
MUE[a]	9.3	6.7	5.1	15.1	8.3	1.9	3.6	3.1	10.5	8.3	6.8	4.0	9.3	3.4	4.6	4.3
	VSXC								MPWLYP1M							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
Si−O−H	114.1	117.7	120.1	120.1	118.1	119.7	119.7	119.4	112.9	116.5	118.2	118.2	117.1	118.7	118.5	118.4
Si−O−Si	134.3	180.0	134.1	135.0	131.8	135.8	136.3	136.2	144.4	146.4	147.3	180.0	141.8	153.8	150.3	151.6
MUE[a]	10.7	14.7	9.3	8.9	10.0	8.2	8.0	7.9	6.3	3.5	2.1	14.6	5.5	1.4	0.5	0.3
	TPSSh								B3LYP[b]							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
Si−O−H	113.5	116.9	118.6	118.4	116.6	118.5	118.3	118.1	113.6	117.1	119.0	119.0	117.7	119.5	119.2	119.1
Si−O−Si	145.0	180.0	149.9	180.0	140.6	155.6	151.0	152.0	145.3	180.0	180.0	180.0	143.8	158.9	152.2	156.6
MUE[a]	5.7	15.3	0.7	14.5	6.3	2.3	0.3	0.7	5.5	15.2	14.6	14.6	4.2	4.3	0.8	2.9
	B97-1								B1LYP							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
Si−O−H	112.8	116.3	118.3	118.3	116.7	118.5	118.3	118.0	113.7	117.3	119.2	119.2	117.8	119.7	119.4	119.3
Si−O−Si	139.5	143.2	143.5	180.0	138.8	151.3	148.6	149.9	146.2	180.0	180.0	180.0	144.5	160.5	155.7	158.1
MUE[a]	8.7	5.2	4.1	14.6	7.2	0.1	1.5	0.9	5.0	15.1	14.7	14.7	3.8	5.2	2.6	3.8
	M05								MPW1B95							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
Si−O−H	113.7	117.1	119.2	119.3	117.5	120.0	119.2	119.1	113.6	117.1	119.2	119.2	117.4	119.3	119.1	118.9
Si−O−Si	141.6	142.4	144.1	180.0	139.0	158.4	153.0	156.7	143.0	180.0	152.0	180.0	141.4	157.5	153.0	155.6
MUE[a]	7.3	5.2	3.9	14.7	6.7	4.3	1.2	3.0	6.7	15.2	0.7	14.7	5.5	3.5	1.1	2.3
	PW6B95								M05-2X							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
Si−O−H	113.5	117.1	119.2	119.2	117.4	119.3	119.1	119.0	113.8	117.8	106.1	120.4	118.3	120.4	120.1	120.2
Si−O−Si	143.8	180.0	152.3	180.0	142.0	157.7	153.2	155.7	142.3	180.0	146.5	180.0	140.1	156.7	152.2	156.2
MUE[a]	6.3	15.2	0.8	14.7	5.2	3.6	1.2	2.4	6.9	14.8	8.6	15.2	5.7	3.6	1.2	3.3

	MP2							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
Si−O−H	113.0	117.0	118.8	118.7	116.4	118.3	118.2	117.8
Si−O−Si	139.9	144.1	144.9	180.0	137.3	151.2	147.7	148.7
MUE[a]	8.5	4.4	3.2	14.4	8.0	0.2	2.0	1.7

[a] MUE denotes mean unsigned error. [b] For comparison, we note the following results of Tielens et al.:[28] 116.4, 135.6, and 8.9 with aug-cc-pVDZ; 118.5, 150.2, and 0.6 with aug-cc-pVTZ; and 119.1, 152.4, and 0.9 with aug-cc-pVQZ.

(and this value was selected as the benchmark in Table 3), then hybrid functionals perform better.

The Si−O−H angle was found to be the most sensitive geometrical parameter of silanol. MIDI! always gives the smallest angle, deviating from the high-level CCSD result of Table 3 by almost 5°, while other polarized double-ζ basis sets perform fairly well. Except for the M05 functional, the largest difference among the angles given by the three polarized triple-ζ basis sets never exceeds 0.4°.

The MUEs in Table 4 are the mean unsigned errors in these two bond angles. Not unexpectedly, most of the MUEs of the polarized triple-ζ basis sets are smaller than those of the polarized double-ζ basis sets. The performance of the three polarized triple-ζ basis sets depends on the functional.

Table 5. Mean Unsigned Error (in Å) of the Si−O Bond Distances in Silanol and Disiloxane

	PDZ[a]					PTZ[a]		
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-ccpV(T+d)Z
HCTH	0.028	0.044	0.040	0.032	0.040	0.007	0.012	0.011
M06-L	0.017	0.028	0.025	0.011	0.027	0.007	0.003	0.005
mPWPW	0.032	0.040	0.040	0.020	0.040	0.010	0.010	0.010
PBE	0.033	0.049	0.046	0.039	0.047	0.016	0.021	0.020
VSXC	0.020	0.027	0.032	0.026	0.033	0.007	0.010	0.009
MPWLYP1M	0.028	0.046	0.043	0.031	0.044	0.014	0.018	0.017
TPSSh	0.021	0.030	0.032	0.020	0.034	0.004	0.007	0.006
B3LYP[b]	0.019	0.029	0.025	0.019	0.032	0.005	0.010	0.004
B97−1	0.023	0.037	0.034	0.020	0.034	0.003	0.007	0.007
B1LYP	0.017	0.028	0.024	0.018	0.030	0.005	0.004	0.004
M05	0.019	0.036	0.029	0.016	0.030	0.006	0.004	0.004
MPW1B95	0.010	0.019	0.020	0.009	0.022	0.009	0.004	0.006
PW6B95	0.012	0.020	0.021	0.009	0.023	0.008	0.003	0.003
M05-2X	0.014	0.025	0.027	0.016	0.029	0.002	0.003	0.002
MP2	0.022	0.041	0.041	0.024	0.044	0.003	0.009	0.009

[a] PDZ denotes polarized double-ζ; PTZ denotes polarized triple-ζ. [b] For comparison, Tielens et al.[28] obtained 0.056 with aug-cc-pVDZ, 0.015 with aug-cc-pVTZ, and 0.007 with aug-cc-pVQZ.

For some functionals, one basis set gives smaller MUEs than others, while for some other functionals, another basis set gives smaller MUEs. In most cases, the MUEs of 6-31+B** are the smallest among the polarized double-ζ basis sets. Among all the possible combinations of functionals and basis sets, TPSSh/aug-pc2 and MPWLYP1M/aug-cc-pV(T+d)Z most closely reproduce the CCSD/MG3S and experimental results (MUE = 0.29 and 0.33°, respectively).

In Table S1 of the Supporting Information, we present the calculated equilibrium Si−O bond lengths of disiloxane and silanol at various levels of theory, and in Table 5, we present the mean unsigned error (compared to the experimental value for disiloxane and the theoretical CCSD/MG3S value for silanol), averaged over these two bond lengths. For the Si−O bond lengths of disiloxane, polarized triple-ζ basis sets always give shorter bond distances than polarized double-ζ for a given functional, closer to the experimental value of 1.634 Å and the high-level CCSD/MG3S result of 1.626 Å. For the Si−O bond lengths of silanol, among the five polarized double-ζ basis sets, MIDI! always gives the shortest bond distances, closer to the values given by the polarized triple-ζ basis sets. All density functionals perform very well with small MUEs. Among the five polarized double-ζ basis sets, MIDI! performs the best. The MUEs of the polarized triple-ζ basis sets are all smaller than those of the polarized double-ζ basis sets. The MG3S basis set performs better than the other two polarized triple-ζ basis sets. Taking only polarized triple-ζ basis sets into consideration, the M05-2X and B1LYP functionals have the best overall performances for the Si−O distances of dixiloxane and silanol.

3.2. Dipole Moments of Disiloxane and Silanol. Zwijnenburg et al.[62] studied the polarity of Si−O bonds in siliceous materials by calculating partial atomic changes. They found that two different methods of extracting such changes from DFT orbitals gave quite different results. We prefer, therefore, at least in the present study, to avoid such artificial quantities. We will instead use the dipole moment, an unambiguous physical observable, to gauge the accuracy of our charge distributions. The dipole moment depends strongly on the geometry, the basis set, and the density functional. The Si−O−Si bond angle plays an especially important role in determining the value of the dipole moment, which reaches a minimum (zero) for disiloxane when the Si−O−Si angle is linear.

In Table 6, we present the calculated dipole moments of disiloxane and silanol at various levels of theory. Examining the MUEs, MIDI! and aug-pc1 have the smallest MUEs among the five polarized double-ζ basis sets with all functionals. For the polarized triple-ζ basis sets, the MUEs of the MG3S basis set are always the smallest. Taking only polarized triple-ζ basis sets into consideration, the M05-2X functional has the best performance for dipole moments. Among all the combinations of functionals and basis sets, TPSSh/MG3S and M05-2X/MG3S are the best methods, with MUEs of just 0.014 and 0.008 D, respectively. Other hybrid functionals also perform very well.

3.3. Protonation/Deprotonation and Proton-Transfer Reactions. In this section, we discuss proton-transfer reactions. A proton-transfer reaction involves protonating the proton acceptor and deprotonating the proton donor. The proton affinities of both disiloxane and silanol are sensitive to the theoretical treatment. However, the accuracy with which we can calculate protonation energy is important for work with zeolites. The three possible proton-transfer reactions involving disiloxane and silanol are listed in Figure 3. The three reactions contain two protonation processes, $H^+ + (H_3Si)_2O \rightarrow (H_3Si)_2OH^+$ and $H_3SiOH + H^+ \rightarrow H_3SiOH_2^+$, and one deprotonation process, $H_3SiOH \rightarrow H^+ + H_3SiO^-$.

In Table 7, the W1w proton affinities of H_3SiO^-, H_3SiOH, and $(H_3Si)_2O$ are listed along with results from other accurate theoretical calculations and experiments where available. The G1 and G2 results are expected to be accurate to 2−3 kcal/mol, which is reasonably accurate but not as accurate as W1w. The W1w proton affinity of H_3SiO^- is 363.7 kcal/mol. The best available ab initio calculated value is 364.1 ± 1.1 kcal/mol.[125] With zero-point energy included, the proton affinity of the silanol anion is 355.7 kcal/mol. The

Table 6. Dipole Moments of Silanol and Disiloxane (in Debye)

	HCTH								M06-L							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
silanol	1.478	1.538	1.467	1.472	1.242	1.277	1.250	1.249	1.653	1.548	1.498	1.493	1.275	1.315	1.272	1.274
disiloxane	0.330	0.532	0.556	0.461	0.427	0.251	0.256	0.237	0.203	0.396	0.479	0.000	0.436	0.219	0.234	0.219
MUE[a]	0.120	0.252	0.228	0.183	0.136	0.030	0.047	0.041	0.181	0.188	0.205	0.203	0.124	0.017	0.031	0.037
	mPWPW								PBE							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
silanol	1.515	1.541	1.498	1.495	1.257	1.292	1.256	1.257	1.515	1.543	1.495	1.495	1.255	1.290	1.256	1.257
disiloxane	0.327	0.564	0.566	0.000	0.471	0.292	0.297	0.288	0.336	0.601	0.604	0.515	0.477	0.306	0.309	0.300
MUE[a]	0.137	0.269	0.248	0.204	0.151	0.043	0.064	0.059	0.142	0.289	0.266	0.221	0.155	0.052	0.070	0.065
	VSXC								MPWLYP1M							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
silanol	1.544	1.545	1.468	1.474	1.255	1.299	1.269	1.272	1.500	1.555	1.502	1.503	1.270	1.301	1.263	1.264
disiloxane	0.294	0.005	0.590	0.565	0.445	0.326	0.303	0.294	0.296	0.504	0.492	0.000	0.434	0.250	0.265	0.251
MUE[a]	0.136	0.226	0.245	0.236	0.138	0.057	0.061	0.055	0.114	0.246	0.213	0.208	0.126	0.018	0.044	0.037
	TPSSh								B3LYP[b]							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
silanol	1.524	1.552	1.504	1.501	1.274	1.308	1.275	1.277	1.509	1.557	1.504	1.505	1.284	1.313	1.280	1.281
disiloxane	0.292	0.000	0.456	0.000	0.440	0.231	0.254	0.244	0.292	0.001	0.000	0.000	0.412	0.200	0.251	0.209
MUE[a]	0.125	0.232	0.197	0.207	0.127	0.014	0.033	0.027	0.116	0.234	0.208	0.209	0.107	0.027	0.029	0.039
	B97-1								B1LYP							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
silanol	1.507	1.549	1.501	1.502	1.278	1.308	1.277	1.277	1.503	1.559	1.504	1.506	1.288	1.316	1.282	1.284
disiloxane	0.321	0.527	0.538	0.000	0.450	0.258	0.268	0.254	0.286	0.000	0.000	0.000	0.405	0.185	0.218	0.196
MUE[a]	0.130	0.254	0.236	0.207	0.130	0.019	0.039	0.032	0.111	0.236	0.209	0.209	0.102	0.033	0.033	0.044
	M05								MPW1B95							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
silanol	1.513	1.554	1.515	1.513	1.299	1.322	1.271	1.273	1.510	1.554	1.499	1.500	1.284	1.315	1.284	1.286
disiloxane	0.310	0.529	0.561	0.000	0.464	0.205	0.228	0.192	0.317	0.000	0.556	0.000	0.426	0.206	0.233	0.210
MUE[a]	0.128	0.258	0.254	0.213	0.126	0.020	0.034	0.051	0.130	0.234	0.244	0.206	0.115	0.023	0.025	0.036
	PW6B95								M05-2X							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
silanol	1.501	1.553	1.495	1.497	1.281	1.310	1.279	1.280	1.483	1.589	1.505	1.511	1.299	1.328	1.298	1.307
disiloxane	0.313	0.000	0.409	0.000	0.423	0.207	0.234	0.211	0.349	0.000	0.483	0.000	0.442	0.237	0.270	0.240
MUE[a]	0.129	0.233	0.178	0.213	0.106	0.019	0.031	0.042	0.146	0.234	0.207	0.206	0.123	0.008	0.037	0.021

	MP2							
	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	MG3S	aug-pc2	aug-cc-pV(T+d)Z
silanol	1.581	1.670	1.645	1.638	1.450	1.431	1.403	1.408
disiloxane	0.411	0.590	0.625	0.000	0.614	0.329	0.358	0.357
MUE[a]	0.213	0.347	0.351	0.275	0.248	0.096	0.097	0.099

[a] MUE denotes mean unsigned error. [b] For comparison, Tielens et al.[28] obtained 1.304, 0.554, and 0.169 with aug-cc-pVDZ; 1.278, 0.289, and 0.049 with aug-cc-pVTZ; and 1.276, 0.251, and 0.031 with aug-cc-pVQZ.

G2 result is 356.2 kcal/mol,[126] and the ab initio result is 356.2 ± 2 kcal/mol at 0 K,[125] in agreement with the W1w result. The ΔH°_{298} value calculated using W1w is 355.3 kcal/mol; this calculated value is in good agreement with the experimental value of 352 ± 4 kcal/mol.[127] The W1w proton affinity of silanol is 176.8 kcal/mol. The G2 value is 0.7 kcal/mol higher (177.5 kcal/mol); there is no experimental value for this quantity. The W1w proton affinity of disiloxane is 179.0 kcal/mol, which is the same as the G2 value; again, there is no experimental value for this quantity. Only low-level theoretical calculations have been reported,[126] and these give higher values for the proton affinity. The results in Table 7 indicate that the proton affinities of silanol and disiloxane are nearly the same.

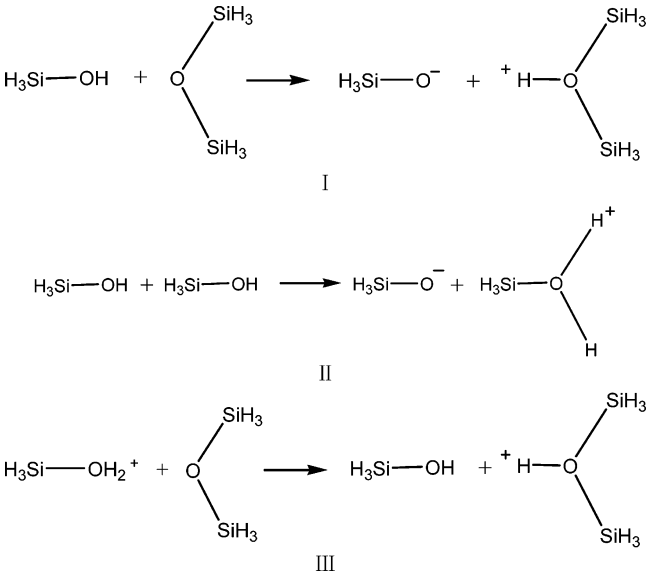


Figure 3. Proton-transfer reactions.

Table 7. Comparison of W1w Proton Affinities of Literatures Values (kcal/mol)

	present W1w	calculated G1 and G2[125]	Koput[124]
	Zero-Point Exclusive		
SiH_3O^-	363.7	363.3[a]	364.1 ± 1.1
SiH_3OH	184.2	184.3[a]	N/A
$(SiH_3)_2O$	186.3	185.9[a]	N/A
	ΔH_0		
SiH_3O^-	355.7	356.2[b]	356.2 ± 2
SiH_3OH	176.8	177.5[b]	N/A
$(SiH_3)_2O$	179.0	179.0[c]	N/A

[a] N/A denotes not available [b] G2 theory (ref 125) [c] G1 theory using MP2 for the 6-311+G(d,p) and 6-311G(d,p) basis set extensions (ref 125).

Table 8. Benchmark Values of Proton Affinities and Proton-Transfer Reaction Energies (Zero-Point Exclusive) at the W1w//B97-1/aug-cc-pV(T+d)Z Level

reaction	ΔE (in kcal/mol)
Proton Affinities	
$(H_3Si)_2OH^+ \rightarrow H^+ + (H_3Si)_2O$	186.3
$H_3SiOH_2^+ \rightarrow H^+ + H_3SiOH$	184.2
$H_3SiOH \rightarrow H^+ + H_3SiO^-$	363.7
Proton-Transfer Reaction Energies	
reaction I[a]	177.4
reaction II[a]	179.5
reaction III[a]	−2.1

[a] Reactions I, II, and III are illustrated in Figure 3.

The benchmark values for the protonation/deprotonation energies and the three proton-transfer reaction energies are give in Table 8. The MUEs of these proton affinities and proton-transfer reaction energies for high-level and multicoefficient methods are given in Table 9. All the errors in the tables are relative to the benchmark W1w results calculated using the geometries obtained at the B97-1/aug-cc-pV(T+d)Z level. The results obtained with DFT for the protonation/deprotonation energies and the overall proton-transfer reaction energies are given in Tables S4 and S5 (Supporting Information) and Table 10, respectively.

Table 9. Mean Unsigned Error of Three Proton Affinities and Three Proton Affinities and Three Proton-Transfer Reaction Energies by High-Level and Multicoefficient Methods (in kcal/mol)

	proton affinities	proton affinities and proton-transfer reaction energies
CCSD/MG3S	1.11	1.10
BMCCCSD//CCSD/MG3S	0.49	0.78
BMCCCSD//B97-1/aug-cc-pV(T+d)Z	0.49	0.57
MCG3/3 //CCSD/MG3S	1.12	0.86
MCG3/3//B97-1/aug-cc-pV(T+d)Z	1.07	0.83
W1w// CCSD/MG3S	0.05	0.05
W1w//B97-1/aug-cc-pV(T+d)Z	0	0

For proton affinities, the MIDIX+ basis set has the largest average MUE, in particular, 6.84 kcal/mol (Table S5, Supporting Information), and the MG3S basis set has the smallest average MUE, namely, 1.07 kcal/mol. The only functionals for which MG3S does not have the smallest error are the MPWLYP1M and M05-2X functionals. The functionals with a percentage of Hartree−Fock exchange less than or equal to 5% have MUEs systematically larger than those with 10% or more Hartree−Fock exchange. M05-2X with 56% HF exchange has the smallest average MUE values, in particular, 3.35 kcal/mol for polarized double-ζ basis sets and 0.23 kcal/mol using polarized triple-ζ basis sets. The same trends are observed in the MUEs of the three energies of the proton-transfer reactions in Figure 3. Tables S4 and S5 in the Supporting Information are full tables of the MUEs for the three proton affinities and the three energies of reaction, respectively, while Table 10 gives MUEs averaged over all six of these quantities. Table 10 shows that, for three proton affinities and three energies of reaction, MG3S has the smallest error in most DFT calculations, the exception being for the M05-2X functional. It is found that all of the average MUEs for polarized double-ζ basis sets are larger than 3 kcal/mol. The corresponding values for polarized triple-ζ basis sets are all 1.41 kcal/mol or less.

Among the polarized double-ζ basis sets, aug-pc1 is the largest and performs the best in many cases. MIDI! is the smallest basis, but it performs better than some bigger basis sets (6-31+B** and 6-31+G**). Among the polarized triple-ζ basis sets, MG3S is the best one for energies. MPW1B95 and M05-2X are the best functionals for energies with polarized double-ζ basis sets.

3.4. Composite Evaluation. In order to draw overall conclusions about the relative merits of the theoretical methods, we define a composite mean normalized percentage unsigned error as

$$\text{CMN\%UE} = [\text{MN\%UE(angles)} + \text{MN\%UE(distances)} + \text{MN\%UE(dipole moments)} + \text{MN\%UE(proton activity)}]/4 \quad (1)$$

Table 10. Mean Unsigned Error of Three Proton Affinities and Three Proton-Transfer Reaction Energies (in kcal/mol)

	X[a]	MIDI!	MIDIX+	6-31+B**	6-31+G**	aug-pc1	av (PDZ)[b]	MG3S	aug-pc2	aug-cc-pV(T+d)Z	av (PTZ)[b]
HCTH	0	3.56	4.74	5.24	5.01	3.96	4.50	2.04	2.50	2.48	2.34
M06-L	0	4.16	4.43	3.35	3.40	2.95	3.66	1.18	1.86	2.08	1.71
mPWPW	0	2.91	6.29	4.43	4.04	3.44	4.22	1.69	2.16	2.15	2.00
PBE	0	2.79	7.14	4.39	4.18	3.85	4.47	2.14	2.37	2.44	2.32
VSXC	0	3.54	5.58	3.86	3.73	3.00	3.94	1.72	1.93	1.98	1.88
MPWLYP1M	5	3.31	7.29	4.19	3.74	3.93	4.49	2.26	2.30	2.41	2.32
TPSSh	10	4.08	4.72	3.28	2.93	2.22	3.44	0.60	0.95	1.01	0.85
B3LYP	20	4.33	5.49	3.08	2.72	2.33	3.59	0.41	0.82	0.79	0.67
B97-1	21	4.21	4.77	3.82	3.62	2.44	3.77	0.88	1.38	1.26	1.17
B1LYP	25	4.59	5.30	3.01	2.75	2.14	3.56	0.22	0.60	0.56	0.46
M05	28	4.10	4.53	4.58	3.98	2.30	3.90	0.53	1.59	1.03	1.05
MPW1B95	31	4.47	5.04	2.78	2.48	1.61	3.28	0.31	0.55	0.36	0.41
PW6B95	46	4.48	4.96	3.10	2.76	1.85	3.43	0.34	0.77	0.58	0.56
M05-2X	56	4.95	4.79	2.34	2.38	1.60	3.21	0.46	0.13	0.25	0.28
MP2	100	4.08	6.61	3.68	3.50	2.82	4.14	1.58	0.89	1.47	1.31
av		3.97	5.44	3.68	3.41	2.70	3.84	1.09	1.39	1.39	1.29

[a] *X* denotes the percentage of Hartree−Fock exchange in the functional. [b] PDZ denotes polarized double-ζ; PTZ denotes polarized triple-ζ.

Table 11. Composite Mean Normalized Percentage Unsigned Error of 12 Geometrical Data, Dipole Moments, and Proton Activity Data on Basis of Calculations on Six Silicon-Containing Molecules

	X[a]	MIDI!	MIDIX+	6-31 B**	6-31+G**	aug-pc1	av (PDZ)[b]	MG3S	aug-pc2	aug-cc-pV(T+d)Z	av (PTZ)[b]
HCTH	0	6.00	10.20	9.52	7.39	6.73	7.97	1.53	2.28	2.02	1.94
M06-L	0	8.41	8.65	8.04	9.81	6.07	8.20	1.03	1.64	1.77	1.48
mPWPW	0	6.81	11.30	9.92	10.13	7.31	9.10	2.09	3.17	2.92	2.72
PBE	0	7.22	12.36	10.82	8.75	7.70	9.37	2.71	3.59	3.37	3.22
VSXC	0	6.90	11.05	10.40	9.94	7.01	9.06	3.61	3.76	3.54	3.63
MPWLYP1M	5	5.50	10.08	8.20	10.12	6.04	7.99	1.25	1.98	1.70	1.64
TPSSh	10	5.68	11.32	7.19	9.89	5.92	8.00	1.00	1.30	1.19	1.16
B3LYP	20	5.36	11.46	10.02	9.93	4.87	8.33	1.75	1.31	1.91	1.66
B97-1	21	6.46	10.25	9.19	10.03	6.20	8.42	0.78	1.79	1.45	1.34
B1LYP	25	5.09	11.47	10.03	9.97	4.59	8.23	2.12	1.67	2.21	2.00
M05	28	6.04	10.28	9.74	10.18	5.91	8.43	1.60	1.58	2.39	1.86
MPW1B95	31	5.89	11.26	8.49	9.72	5.15	8.10	1.54	1.15	1.70	1.47
PW6B95	46	5.63	11.23	6.17	9.70	5.13	7.57	1.63	1.24	1.77	1.55
M05-2X	56	6.10	11.80	8.96	10.09	5.29	8.45	0.79	1.24	0.98	1.00
MP2	100	9.02	13.32	12.76	12.19	10.29	11.52	3.36	3.68	3.76	3.60
av		6.41	11.07	9.30	9.86	6.28	8.58	1.79	2.09	2.18	2.02

[a] *X* denotes the percentage of Hartree−Fock exchange in the functional. [b] PDZ denotes polarized double-ζ; PTZ denotes polarized triple-ζ.

where the mean normalized percentage unsigned error of quantity Q is

$$\mathrm{MN\%UE}(Q) = \frac{\mathrm{MUE}(Q)}{\mathrm{MV}(Q)} \times 100\% \qquad (2)$$

and MV(Q) is the mean value of Q. For geometries, it can be calculated from Tables 4 and 5 that MV(angles) is 134.9°, and MV(distances) is 1.6395 Å. It can be calculated from Table 6 that MV(dipole moments) is 0.7835 D. The corresponding MUEs are in Tables 4−6. For proton activity, MV is the average of the absolute values of the six numbers in Table 8, which yields MV = 182.2 kcal/mol, and MUE is the value in Table 10.

Table 11 allows us to draw several important conclusions: (1) Polarized triple-ζ basis sets have errors about 4 times smaller than polarized double-ζ basis sets. Even the least accurate polarized triple-ζ basis set is about 3 times more accurate than the best double-ζ basis set. (2) MG3S is not only the smallest polarized triple-ζ basis set tested, it is also, on average, the most accurate. (3) MIDI!, although it was originally developed for Hartree−Fock calculations of geometries and charge distributions, and although it is the smallest polarized double-ζ basis set and has no diffuse functions, has a composite mean percentage error of only 6.35%, averaged over the functionals, which is better than three of the large polarized double-ζ basis sets and the same as the other, which has more than twice as many basis functions (see Table 2). (4) The most accurate density functional is M05-2X, which also has[71] excellent performance for a wide range of other quantities in main-group chemistry.

One reason for examining both local functionals (i.e., those with $X = 0$) and nonlocal functionals is that local functionals are more computationally efficient, especially when one uses density fitting,[128] which can only be used efficiently when $X = 0$. To illustrate this, we timed single-point energies on an oligo(dimethylsiloxane), namely, 3,5-tetramethyltetrasi-

loxane, with M05-2X (for which $X = 56$) and M06-L (for which $X = 0$), in both cases with the MG3S basis set, and in the latter case using density fitting. The ratio of computing times was a factor of 2.6. Furthermore, as the system size is increased and in the limit of large systems, the computing time for $X = 0$ scales as N^3, where N is the number of atoms, while that for $X \neq 0$ scales as N^4. Therefore, the reduction in cost when using local functionals can be quite significant for studying large silica systems.

In light of the favorable cost of local functionals, it is interesting to carry out a performance comparison focusing only on $X = 0$ functionals in Table 11. We see that the most accurate functionals at the polarized triple-ζ level are (in order) M06-L, HCTCH, mPWPW, and PBE. In fact, these four local functionals are all more accurate than MP2 (a WFT method), which scales as N^5.

4. Conclusions

A total of 14 DFT methods in combination with eight basis sets were assessed on a variety of properties, such as the structures, dipole moments, and proton affinities of disiloxane and silanol. For disiloxane conformers, the results obtained with the M05-2X functional in combination with polarized triple-ζ basis sets are in best agreement with the experimental and best estimate computational results. For silanol, the use of the MG3S basis set with DFT allows us to obtain values for several difficult properties that agree well with experimental results or high-level WFT. In particular, the M05-2X and M06-L calculations agree with the best estimates within, respectively, 1° and 4° for two bond angles, 0.002 and 0.003 Å for two bond distances, 0.01 and 0.02 D for two dipole moments, and 0.5 and 1.2 kcal/mol for six proton affinities and proton-transfer reaction energies.

On average, the most accurate density functional for both geometries and energetics is M05-2X, and M06-L is the most accurate local density functional. MIDI! is the smallest polarized double-ζ basis set tested but gives results in closer agreement, on average, with augmented polarized triple-ζ basis sets than do the more popular polarized double-ζ basis sets, and its performance is as good, on average, as the much larger aug-pc1 basis set. MG3S is not only the smallest polarized triple-ζ basis set tested, it is also, on average, the one that gives results closest to our best estimations of the correct values.

Acknowledgment. The authors are grateful to Dr. Mark Iron and Dr. Yan Zhao for assistance. This work was supported in part by the National Science Foundation under Grant No. ITR04-28774.

Supporting Information Available: Eight tables with key geometric parameters and error summaries for proton affinities and energies of proton-transfer reactions. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tsuneyuki, S.; Aoki, H.; Tsukada, M. *Phys. Rev. Lett.* **1990**, *64*, 776.

(2) Warren, M. C.; Ackland, G. J.; Karki, B. B.; Clark, S. J. *Mineral. Mag.* **1998**, *62*, 558.

(3) Dubrovinsky, L.; Saxena, S. K.; Ahuja, R.; Johansson, B. *Geophys. Res. Lett.* **1998**, *25*, 4253.

(4) Tjabane, M.; Lowther, J. E. *Physica B* **1999**, *270*, 164.

(5) Kihara, K. *Phys. Chem. Miner.* **2001**, *28*, 356.

(6) Gibbs, G. V.; Cox, D. F.; Ross, N. L. *Phys. Chem. Miner.* **2004**, *31*, 232.

(7) Tsuchiya, T.; Caracas, R.; Tsuchiya, J. *Geophys. Res. Lett.* **2004**, *31*, L11610.

(8) Li, L.; Weidner, D. J.; Brodholt, J.; Alfè, D.; Price, G. D.; Caracas, R.; Wentzcovitch, R. *Phys. Earth Planet. Inter.* **2006**, *155*, 249.

(9) Wentzcovitch, R. M.; Tsuchiya, T.; Tsuchiya, J. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 543.

(10) Umemoto, K.; Wentzcovitch, R. M.; Allen, P. B. *Science* **2006**, *311*, 983.

(11) Ma, Y.; Garofalini, S. H. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 174109.

(12) Sastre, G.; Corma, A. *J. Phys. Chem. B* **2006**, *110*, 17949.

(13) Susman, S.; Volin, K. J.; Price, L. D.; Grimsditch, M.; Rino, J. P; Kalia, R. K.; Vashishta, P.; Gwanmesia, G.; Wang, Y.; Liebermann, R. C. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1194.

(14) Tsuneyuki, S.; Matsui, Y. *Phys. Rev. Lett.* **1995**, *74*, 3197.

(15) Sarnthein, J.; Pasquarello, A.; Car, R. *Phys. Rev. Lett.* **1995**, *74*, 4682.

(16) Sarnthein, J.; Pasquarello, A.; Car, R. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1995**, *52*, 12690.

(17) Pasquarello, A.; Car, R. *Phys. Rev. Lett.* **1997**, *79*, 1766.

(18) Benoit, M.; Ispas, S.; Jund, P.; Jullien, R. *Eur. Phys. J. B* **2000**, *13*, 631.

(19) Valle, R. G. D.; Venuti, E. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 3809.

(20) Boureau, G.; Carniato, S.; Tetot, R.; Harding, J. H. *Mol. Simul.* **1997**, *20*, 27.

(21) Huang, L.; Duffrene, L.; Kieffer, J. *J. Non-Cryst. Solids* **2004**, *349*, 1.

(22) Hoang, V. V. *Diffus. Defect Data, Pt. A* **2005**, *242−244.*, 77.

(23) Zhao, D.; Feng, J. L.; Huo, Q.; Melosh, N.; Fredrickson, G. H.; Chmelka, B. F. *Science* **1998**, *279*, 548.

(24) Ying, J. Y.; Mehnert, C. P.; Wong, M. S. *Angew. Chem., Int. Ed.* **1999**, *38*, 56.

(25) Hatton, B.; Landskron, K.; Whitnall, W.; Perovic, D.; Ozin, G. A. *Acc. Chem. Res.* **2005**, *38*, 305.

(26) Earley, C. W. *J. Comput. Chem.* **1993**, *14*, 216.

(27) Nicholas, J. B.; Feyereisen, M. *J. Chem. Phys.* **1995**, *103*, 8032.

(28) Tielens, F.; Proft, F. D.; Geerlings, P. *THEOCHEM* **2001**, *542*, 227.

(29) Almenningen, A.; Bastiansen, O.; Ewing, V.; Hedberg, K.; Tretteberg, M. *Acta Chem. Scand.* **1963**, *17*, 2455.

(30) Nicholas, J. B.; Winans, R. E.; Harrison, R. J.; Iton, L. E.; Curtiss, L. A.; Hopfinger, A. J. *J. Phys. Chem.* **1992**, *96*, 7958.

(31) Stave, M. S.; Nicholas, J. B. *J. Phys. Chem.* **1993**, *97*, 9630.

(32) Pöhlmann, M.; Benoit, M.; Kob, W. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2004**, *70*, 184209.

(33) Saengsawang, O.; Remsungnen, T.; Fritzsche, S.; Haberlandt, R.; Hannongbua, S. *J. Phys. Chem. B* **2005**, *109*, 5684.

(34) Chatterjee, A.; Iwasaki, T.; Ebina, T.; Tsuruya, H.; Kanougi, T.; Oumi, Y.; Kubo, M.; Miyamoto, A. *Appl. Surf. Sci.* **1998**, *130*−*132*, 555.

(35) McMillan, P. F. *Rev. Mineral.* **1994**, *30*, 131.

(36) Schiano, P.; Clocchiatti, R.; Shimizu, N.; Maury, R. C.; Jochum, K. P.; Hofmann, A. W. *Nature* **1995**, *377*, 595.

(37) Kohn, S. C. *Miner. Mag.* **2000**, *64*, 389.

(38) Bromile, G.; Hilaret, N.; McCammon, C. *Geophys. Res. Lett.* **2004**, *31*, L04610.

(39) Helms, C. R.; Poindexter, E. H. *Rep. Prog. Phys.* **1994**, *57*, 791.

(40) Friebele, E. J.; Griscom, D. L. *Mater. Res. Soc. Symp. Proc.* **1996**, *61*, 319.

(41) Gibbs, G. V.; Hill, F. C.; Boisen, M. B.; Downs, R. T. Molecules as a Basis for Modeling the Force Field of Silica. In *Structure and Imperfections in Amorphous and Crystalline Silicon Dioxide*; Devine, R. A. B., Duraud, J. P., Dooryhee, E., Eds.; John Wiley and Sons Ltd.: London, 2000; p 153.

(42) Bar, M. R.; Sauer, J. *Chem. Phys. Lett.* **1994**, *226*, 405.

(43) Banavar, J. R.; Phillips, J. C. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1983**, *28*, 4716.

(44) Krishnamurty, S.; Pal, S.; Vetrivel, R.; Chandra, A. K.; Goursot, A.; Fajula, F. *J. Mol. Catal. A: Chem.* **1998**, *129*, 287.

(45) Ribeiro-Claro, P. J. A.; Amado, A. M. *THEOCHEM* **2000**, *528*, 19.

(46) Cheng, H.-P.; Barnett, R. N.; Landman, U. *J. Chem. Phys.* **2002**, *116*, 9300.

(47) Murashov, V. *J. Mol. Struct.* **2003**, *650*, 141.

(48) Simperler, A.; Bell, R. G.; Foster, M. D.; Gray, A. E.; Lewis, D. W.; Anderson, M. W. *J. Phys. Chem. B* **2004**, *108*, 7152.

(49) Abraham, R. J.; Grant, G. H. *J. Comput. Chem.* **1988**, *9*, 709.

(50) Feuston, B. P.; Garofalini, S. H. *J. Chem. Phys.* **1988**, *89*, 5818.

(51) Beest, B. W. H. v.; Kramer, G. J.; Santen, R. A. v. *Phys. Rev. Lett.* **1990**, *64*, 1955.

(52) Kramer, G. J.; Farragher, N. P.; Beest, B. W. H. v.; Santen, R. A. v. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 5068.

(53) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.

(54) Sun, H.; Rigby, D. *Spectrochim. Acta, Part A* **1997**, *53*, 1301.

(55) Smith, J. S.; Borodin, O.; Smith, D. G. *J. Phys. Chem. B* **2004**, *108*, 20340.

(56) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.

(57) Nicholas, J. B.; Winans, R. E.; Harrison, R. J.; Iton, L. E.; Curtiss, L. A.; Hopfinger, A. J. *J. Phys. Chem.* **1992**, *96*, 10247.

(58) Nicholas, J. B.; Hess, A. C. *J. Am. Chem. Soc.* **1994**, *116*, 5428.

(59) Litton, D. A.; Garofalini, S. H. *J. Appl. Phys.* **2001**, *89*, 6013.

(60) Smith, J. S.; Bedrov, D.; Smith, G. D. *Macromolecules* **2005**, *38*, 8101.

(61) Demuth, T.; Hafner, J.; Benco, L.; Toulhoat, H. *J. Phys. Chem. B* **2000**, *104*, 4593.

(62) Zwijnenburg, M. A.; Bromley, S. T.; Alsenoy, C. v.; Maschmeyer, T. *J. Phys. Chem. A* **2002**, *106*, 12376.

(63) Tuma, C.; Sauer, J. *Chem. Phys. Lett.* **2004**, *387*, 388.

(64) Ernst, C. A.; Allred, A. L.; Ratner, M. A.; Newton, M. D.; Gibbs, G. V.; Moskowitz, J. W.; Topiol, S. *Chem. Phys. Lett.* **1981**, *81*, 424.

(65) Fernandez-Ramos, A.; Miller, J. A.; Klippenstein, S. J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 4518.

(66) Martin, J. M. L.; El-Yazal, J.; François, J.-P. *Mol. Phys.* **1995**, *86*, 1437.

(67) Jensen, F. *J. Chem. Phys.* **2001**, *115*, 9113.

(68) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.

(69) Jensen, F.; Helgaker, T. *J. Chem. Phys.* **2004**, *121*, 3463.

(70) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107.

(71) Schultz, N.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 11127. Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364. Furche, F.; Perdew, J. P. J. *Chem. Phys.* **2006**, *124*, 44103. Harvey, J. N. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2006**, *102*, 203.

(72) Nicholas, J. B. *Top. Catal.* **1997**, *4*, 157.

(73) Sauer, J.; Ahlrichs, R. *J. Chem. Phys.* **1990**, *93*, 2575.

(74) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.

(75) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(76) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1997**, *274*, 242.

(77) Stephens, P.; Devlin, F.; Chabalowski, C.; Frisch, M. *J. Phys. Chem.* **1994**, *98*, 11623.

(78) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264.

(79) Boese, A. D.; Doltsinis, N. L.; Handy, N. C.; Sprik, M. *J. Chem. Phys.* **2000**, *112*, 1670.

(80) Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2001**, *114*, 5497.

(81) Zhao, Y.; Schultz, N. E.; Truhlara, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

(82) Yan, Z.; Truhlar, D. G. *J. Chem. Phys.* **2006**, submitted.

(83) The mPW of the later versions of Gaussian 98 and of Gaussian 03 differs slightly from the early version of mPWPW91 of Gaussian 98 owing to an error in Gaussian's coding of the original mPW exchange functional. See http://www.gaussian.com/g_tech/g03_rel.htm#updatedman (accessed Dec 9, 2006) or http://comp.chem.umn.edu/info/mpw1k.htm (accessed Dec 9, 2006). We always use the correct version.

(84) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.

(85) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(86) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656.

(87) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129.

(88) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(89) Voorhis, T. V.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.

(90) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 1643.

(91) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(92) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281.

(93) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192.

(94) Thompson, J. D.; Winget, P.; Truhlar, D. G. *Phys. Chem. Commun.* **2001**, *4*, 4116.

(95) Lynch, B. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2004**, *111*, 335.

(96) Dunning, T. H. J.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244.

(97) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(98) Purvis, G. D.; Bartlett. R. J. *J. Chem. Phys.* **1982**, *76*, 1910.

(99) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 3898.

(100) Martin, J. M. L.; Oliveira, G. d. *J. Chem. Phys.* **1999**, *111*, 1843.

(101) Oren, M.; Iron, M. A.; Burcat, A.; Martin, J. M. L. *J. Phys. Chem. A* **2004**, *108*, 7752.

(102) Martin, J. M. L. *THEOCHEM* **2006**, *771*, 19.

(103) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.

(104) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revisions B.02, B.05, and C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(105) Zhao, Y.; Truhlar, D. G. Minnesota-Gaussian Functional Module-version 2.0.1. http://comp.chem.umn.edu/mn-gfm/ (accessed Dec 9, 2006).

(106) Douglas, M.; Kroll, N. M. *Ann. Phys. (San Diego, CA, U.S.)* **1974**, *82*, 89. Hess, A. *Phys. Rev. A: At., Mol., Opt. Phys.* **1986**, *33*, 3742.

(107) Werner, H.-J. K. P. J.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P. C. D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Korona, T. L. R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E. N. A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Schumann, U.; Stoll, H. S. A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO*, version 2002.6; University of Birmingham: Birmingham, U. K., 2002.

(108) Nachtigall, P.; Jordan, K. D.; Smith, A.; Jonsson, H. *J. Chem. Phys.* **1996**, *104*, 148.

(109) Eichler, U.; Brandle, M.; Sauer, J. *J. Phys. Chem. B* **1997**, *101*, 10035.

(110) Tossell, J. A.; Sahaia, N. *Geochim. Cosmochim. Acta* **2000**, *64*, 4097.

(111) Walsh, T. R.; Wilson, M.; Sutton, A. P. *J. Chem. Phys.* **2000**, *113*, 9191.

(112) Yuan, S. P.; Wang. J. G.; Li, Y. W.; Peng, S. Y. *J. Mol. Catal. A: Chem.* **2002**, *178*, 267.

(113) Bussai, C.; Hannongbua, S.; Fritzsche, S.; Haberlandt, R. *Chem. Phys. Lett.* **2002**, *354*, 310.

(114) Tuma, C.; Sauer, J. *Chem. Phys. Lett.* **2004**, *387*, 388.

(115) Tossell, J. A. *Geochim. Cosmochim. Acta* **2005**, *69*, 283.

(116) Ginhoven, R. e. M. V.; Jonsson, H.; Park, B.; Corrales, L. R. *J. Phys. Chem. B* **2005**, *109*, 10936.

(117) Bakk, I.; Bóna, Á.; Nyulászi, L.; Szieberth, D. *THEOCHEM* **2006**, *770*, 111.

(118) Luke, B. T. *J. Phys. Chem.* **1993**, *97*, 7505.

(119) Durig, J. R.; Flanagan, M. J.; Kalasinsky, V. F. *J. Chem. Phys.* **1977**, *66*, 2775.

(120) Koput, J.; Wierzbicki, A. *J. Mol. Spectrosc.* **1983**, *99*, 116.

(121) Barrow, M. J.; Ebsworth, E. A. V.; Harding, M. M. *Acta. Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1979**, *35*, 2093.

(122) Grigoras, S.; Lane, T. H. *J. Comput. Chem.* **1987**, *8*, 84.

(123) Coriani, S.; Gauss, J.; Hättig, C.; Helgaker, T.; Jørgensen, P. *J. Chem. Phys.* **2005**, *123*, 184107.

(124) Koput, J. *Chem. Phys.* **1990**, *148*, 299.

(125) Koput, J. *Chem. Phys. Lett.* **2001**, *333*, 504.

(126) Curtiss, L. A.; Brand, H.; Nicholas, J. B.; Iton, L. E. *Chem. Phys. Lett.* **1991**, *184*, 215.

(127) Damrauer, R.; Simon, R.; Krempp, M. *J. Am. Chem. Soc.* **1991**, *113*, 4431.

(128) Dunlap, B. I. *THEOCHEM* **2000**, *529*, 37.

(129) Varma, R.; MacDiarmid, A. G.; Miller, J. G. *Inorg. Chem.* **1964**, *3*, 1754.

CT6002884

JCTC Journal of Chemical Theory and Computation

Theoretical Studies on the Color-Tuning Mechanism in Retinal Proteins

Kazuhiro Fujimoto,[†] Shigehiko Hayashi,[‡,§] Jun-ya Hasegawa,[†,||] and Hiroshi Nakatsuji*[,†,||]

Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Nishikyou-ku, Kyoto 615-8510, Japan, PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan, Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyou-ku, Kyoto 606-8520, Japan, and Quantum Chemistry Research Institute (QCRI), 58-8 Mikawa, Momoyama-cho, Fushimi-ku, Kyoto 612-8029, Japan

Received August 16, 2006

Abstract: The excited states of the three retinal proteins, bovine rhodopsin (Rh), bacteriorhodopsin (bR), and sensory rhodopsin II (sRII) were studied using the symmetry-adapted cluster-configuration interaction (SAC-CI) and combined quantum mechanical and molecular mechanical (QM/MM) methods. The computed absorption energies are in good agreement with the experimental ones for all three proteins. The spectral tuning mechanism was analyzed in terms of three contributions: molecular structures of the chromophore in the binding pockets, electrostatic (ES) interaction of the chromophore with the surrounding protein environment, and quantum-mechanical effect between the chromophore and the counterion group. This analysis provided an insight into the mechanism of the large blue-shifts in the absorption peak position of Rh and sRII from that of bR. Protein ES effect is primarily important both in Rh and in sRII, and the structure effect is secondary important in Rh. The quantum-mechanical interaction between the chromophore and the counterion is very important for quantitative reproduction of the excitation energy. These results indicate that the present approach is useful for studying the absorption spectra and the mechanism of the color tuning in the retinal proteins.

1. Introduction

The rhodopsin family of photoreceptors is among the best characterized membrane proteins. These proteins have a seven-transmembrane helical structure and function as photosensing and ion-pumps.[1] In a common visual photoreceptor, rhodopsin (Rh), the retinal chromophore shows the photoisomerization from 11-cis to all-trans forms. This reaction leads the protein to a signaling state, which is amplified biochemically through interaction with the G protein transducin.[2] In bacteriorhodopsin (bR) of *Halobacterium salinarum*, the photoisomerization of retinal from all-trans to 13-cis forms establishes an electrochemical gradient across the membrane and serves as a unidirectional proton transport.

These receptors consist of an apoprotein (opsin) and a retinal chromophore which is covalently bound to the apoprotein via a lysine residue by a protonated Schiff base (PSB) linkage. While the PSB form of retinal absorbs at about 440 nm in organic solvents, its maximal absorption (λ_{max}) drastically changes after binding to the apoprotein (opsin), which is known as "opsin shift".[3] The absorption maxima is regulated by opsin and widely spreads from 360 to 635 nm[4] to furnish the photoreceptors with color sensitivity, whereas the proteins include a common identical chromophore, retinal. In Rh environment, the chromophore has 11-cis form, and it is in an all-trans form in bR and sRII.

* Corresponding author e-mail: hiroshi@sbchem.kyoto-u.ac.jp.
[†] Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University.
[‡] PRESTO, Japan Science and Technology Agency.
[§] Department of Chemistry, Graduate School of Science, Kyoto University.
[||] Quantum Chemistry Research Institute (QCRI).

10.1021/ct6002687 CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/03/2007

The diversity of absorption maxima of rhodopsin has been investigated extensively,[5] and several explanations have been proposed.[6]

The spectral tuning mechanism can be analyzed in terms of the following three contributions. The first one is the chromophore-structural origin. Retinal would be distorted in the protein in order to accommodate the chromophore itself in the binding pocket. In fact, it has been shown by X-ray crystallographic studies that the polyene chain of the chromophore in Rh is strongly twisted, whereas those in bR and sRII are nearly planer.[7−9] Such torsion of the polyene chain is expected to sensitively alter the absorption energy, since the torsion disrupts conjugation of the π orbitals responsible for the excitation.

Second, ES (electrostatic) interaction between the chromophore and the surrounding protein environment could play a crucial role in the spectral tuning.[10−14] Mutagenesis experiments and theoretical analyses have shown that the absorption energies are strongly affected by the charged residues such as counter negative ion groups in the binding pocket.[15] The ES tuning mechanism exploits a characteristic in the excitation property of the PSB retinal molecule, i.e., significant change in the positive-charge distribution in the PSB along the polyene chain upon the excitation.[16] The positive charge redistribution upon the excitation creates a difference in ES interaction with the surrounding polar groups between the ground and excited states, which gives rise to the spectral shift. Especially, the counter negative ion groups contribute dominantly to the ES energy change upon the excitation so that mutual geometries of the PSB and the counterion groups are suggested to be one of the key factors for determining the absorption maximum.[10,15]

The third contribution includes higher-order interactions with the protein surroundings such as electronic polarization and charge-transfer interactions. The ES interaction described above induces the electronic polarization in both the chromophore and the protein surroundings. The electronic polarization enhances the ES interaction of the chromophore with the surroundings, leading to a shift of the absorption energy.[17−19] In addition to the electronic polarization, there is a strong electronic charge transfer interaction between the PSB and the counterion group,[20] which is suggested to significantly increase the absorption energy.[11,12,20]

In order to identify physical mechanism of the color tuning in the retinal proteins, several computational investigations have been performed by using modern quantum-chemistry methodologies.[10,11,20−23] Recent advances in computational technique have realized to predict the absorption energy of chromophore in protein whose X-ray crystallographic structures[7−9] have been solved. The computational studies mentioned above reported that the absorption maxima are reasonably reproduced with their own approaches.

However, the proposed mechanisms underlying the absorption spectra are still different in those studies, and they have not reached to consensus on the mechanism (see below). This difference arise from the fact that definitive elucidation of the underlying mechanism requires high methodological accuracy for *all* of the contributing factors, i.e., the electronic wave functions of the states involved, the protein structure including the chromophore, and the interaction of the chromophore with the surrounding environment. Errors in any of those contributing factors introduce ambiguity in the definitive determination of the molecular mechanism.

To accurately calculate the electronic energy for the ionic $\pi-\pi^*$ excited state of a polyene-like molecule, dynamic electron-correlation due to strong $\sigma-\pi$ polarization should be included appropriately.[24] In addition, the positive-charge migration on the Schiff base upon the excitation makes the electronic structures more complex. Various calculations have been performed so far by using modern methodologies in quantum chemistry, multireference (MR) perturbation theory (PT),[11,13,20,22,25,26] MR-configuration interaction (CI) method,[12] time-dependent density functional theory (TD-DFT),[23] and symmetry-adapted cluster-CI (SAC-CI) method.[14] Wanko et al. showed that TD-DFT gives qualitatively different excitation energy from the other methods when the C_6-C_7 bond is rotated.[12] This result indicates the rather complex nature of the excited-state wave function, which therefore requires careful and extensive treatment of the electron correlation.

Second, the absorption energy is highly sensitive to the structure of chromophore and protein. Wanko et al.[12] have shown that the absorption energy strongly depends on strength of bond alternation of the polyene chain of the chromophore, in addition to the torsional angle of the polyene chain as mentioned above. For example, the computed absorption energy at the geometry optimized by using the Hartree−Fock (HF) method significantly overestimates the experimental result,[10,20] which is due to too strong bond alternation in the HF geometry.[12] For proper description of the bond alternation of the chromophore, the dynamic electron-correlation should be taken into account. The second-order Moller−Plesset (MP2) perturbation theory or DFT was suggested for the geometry optimization.[12]

Geometry refinement of the protein surroundings is also crucial for the absorption energy calculation. As described above, the interaction of the chromophore with the surrounding protein is essential to the spectral tuning. Especially, it has been suggested that mutual distance between the PSB and its counterion group is one of the main factors for controlling the absorption maximum, and a small difference in the distance less than 1 Å could cause a significant change in the observed spectrum.[10,15]

However, spatial resolutions of X-ray crystallographic structures of the retinal proteins solved so far are more than 1.5 Å and hence are not sufficient to detect decisive structural differences completely. Thus it is necessary to refine the X-ray crystallographic protein structures before the absorption energies are computed. The geometry refinements for the entire protein structures were carried out by means of various computational approaches, such as semiempirical AM1[19,27] and tight binding DFT (DFTB) methods,[12,26,28,29] and hybrid ab initio quantum mechanical/molecular mechanical (QM/MM) methods (see below),[10,14,20] although in some studies the protein surroundings around the reaction center is kept fixed at the original coordinates of X-ray crystallographic models.[13,22]

Finally, the interaction of the chromophore with the surrounding has to be described properly. As described

above, the ES interaction plays a primary role in the color tuning. In order to describe the microscopic ES interaction, the QM/MM method is often employed, in which the ES field of the protein surrounding is represented with the effective point charges of the MM force field. However, the treatment using the point charges lacks the higher-order electronic effects of the protein surroundings such as the electronic polarization and the charge-transfer described above. Especially, the quantum-mechanical interaction of the chromophore with the counterion group contributes considerably to the absorption energies (0.2−0.5 eV)[11,12,20] and therefore has to be taken into account in the quantitative calculation.

In the present study, we have performed ab initio QM/MM and SAC-CI calculations to compute the absorption energies of three retinal proteins, bR, sRII, and Rh. The absorption maxima of sRII (497 nm)[30] and Rh (498 nm)[1] are largely blue-shifted by 70 nm compared with that of bR (568 nm).[31] It is interesting to note that the absorption maxima of sRII and Rh are similar to each other, even though the structures of the chromophores and the protein surroundings are distinctly different.[7,9] There must be different mechanisms for the blue-shift in sRII and Rh. We first carried out QM/MM geometry optimization for the entire protein, bR, sRII, and Rh. The QM/MM method divides the entire system into the QM segment and the MM segment. The QM segment is treated by the quantum-mechanical calculation to describe the electronic ground and excited states involved in the photoabsorption. The MM segment describes the steric and ES effects of the surrounding environment from the rest of the system by means of the molecular mechanics. With the QM/MM refined structures, we next calculated the absorption energies of the QM segment at the SAC-CI level of theory with the MM effective point charges representing the electrostatic field of the surrounding protein.

The SAC[32]/SAC-CI[33] method[34] is an accurate electronic-structure theory for the ground and excited states. This method has been established as a reliable theoretical method for calculating the ground and excited states of atoms and molecules.[34] The SAC method belongs to the cluster expansion method and treats the electron correlation in the ground state. Based on the correlated ground state described by the SAC method, the SAC-CI method was derived as the theory for the excited state.[33] The SAC-CI wave function satisfies the orthogonalities to the ground state SAC wave function. Therefore, the electron correlations in the ground and excited states are treated in a balanced way by the SAC/SAC-CI method. As indicated by the applications to more than 150 systems,[35] the SAC-CI method has been established as a powerful tool for studying the spectroscopy of the atoms and molecules. Owing to the perturbation selection technique of the excitation operators,[36] the computation program of the SAC-CI method is applicable to moderately large molecules. Recently, spectroscopy of biological systems[34] has become our target of applications.

Such advantages in the SAC-CI method realize the quantitative description of the electronic excited states of the retinal proteins. First, it is possible to compute the absorption energies for the whole chromophore molecule without truncating the β-ionone ring and methyl groups. We could further extend the systems to include the counterion and polar groups which have turned out to give important electronic-interaction effects as described above. In previous studies,[11,22,26] the truncated chromophore analogues were often used to reduce the high computational cost of the methods. However, such approximated models lead non-negligible error in the result of the computation.[13]

Moreover, all the valence orbitals were included in the active space of the coupled-cluster expansion in the SAC wave function, and thus the electron correlations among the π and σ orbitals were equivalently treated. This clearly contrasts with the previous calculations using CASSCF (complete active space SCF) or CAS-CI as the reference wave function.[11,13,22,26] Previous MRMP, CASPT2, and spectroscopy oriented CI (SORCI)[12,37] calculations included only π orbitals in the CAS-CI expansion. The multiconfigurational perturbation theories give excellent results, if the reference wave function is enough accurate. However, the first excited-state of retinal PSB involves large $\sigma-\pi$ polarization and large positive-charge migration as described above. In such a situation, it would be desirable to treat the electron correlation of all the valence electrons in equivalent way as in the SAC-CI method.

In our previous study,[10,14,20] the geometry of chromophore was optimized with the HF method, and the QM segment included only the retinal chromophore. In our present study, we improved these inaccurate treatments. Since the HF method emphasizes bond alternation in the retinal chromophore, we used DFT with B3LYP functional for optimizing the structure of retinals. Geometry obtained by DFT is expected to be similar to that obtained by MP2 calculation as shown in a previous report[38] and also in this study. We extended the QM segment to include the counterion residues around the retinal chromophore. The present calculations successfully reproduced the absorption energies of all three proteins, demonstrating accuracy and consistency of the method used. Based on these calculations, we further performed quantitative analysis on the molecular origins of the color tuning among the retinal proteins. The present result is also compared to the previous ones obtained by using the methodologies.[10−14,20,26] We found the source of the discrepancy in the previous studies: treatment of the electron correlation in the excited-state calculation, geometry of retinal and counter residues, and the quantum-mechanical interactions between the Schiff base and the protein residues, especially the counter-residue and the charged residues.

2. Computational Details

Figure 1(a,b) shows computational models of the PSB retinal chromophores in 11-cis and all-trans conformations, respectively. Schiff base (SB) retinal in an all-trans conformation was also examined. We note this structure is the C_6-anti form and is different from the C_6-syn structure used in the gas-phase experiment, the electrostatic ion storage ring in Aarhus (ELISA).[39,40] For the gas-phase calculations, geometry optimizations for the all-trans retinal were performed with the HF, DFT with B3LYP functional (B3LYP), and MP2 methods as well as the semiempirical AM1 method. The

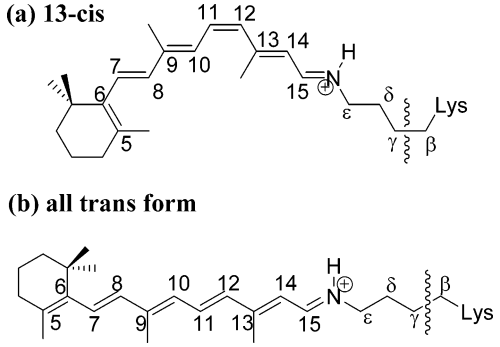


Figure 1. Retinal and all-trans PSB analog structure: (a) 11-cis form and (b) all-trans form.

absorption energies were then calculated at the SAC-CI and TD-DFT with B3LYP functional levels of theory for those optimized structures.

For the protein systems, the entire protein structures including the chromophore molecule were optimized by QM/MM calculation.[10,18,20,41,42] In the QM/MM method, the active site is treated by quantum-mechanical calculation, and the rest of protein is described at the MM level of theory. The QM/MM method used in the present study[20] takes into account ES interaction between the chromophore and the surrounding protein environment through restrained ES potential charge (RESP) operators, permitting us to efficiently determine the optimized structure of the entire protein.

The detail of the QM/MM method was reported elsewhere.[10,20] For visual receptor Rh, we constructed a starting structural model based on an X-ray crystallography structure recently reported by Okada and co-workers (PDB code: 1L9H).[7] For bR and sRII, the initial structures taken from PDB were 1C3W[8] and 1H68,[9] respectively. We carried out a geometry optimization for the whole protein using the QM/MM method.[10,20] The QM segment includes the whole retinal chromophore with the side chains of Lys296 and Glu113 (counterion group) and a proximal water molecule which has hydrogen-bonding to Glu113. The QM segment for bR and sRII was equivalent to that for Rh. The boundaries between the QM and MM segments were at the $C_\beta-C_\gamma$ of Lys296 (Lys216 in bR and Lys205 in sRII) and $C_\alpha-C_\beta$ bonds of Glu113 (Asp85 in bR and Asp75 in sRII). Hydrogen atoms were introduced for the link atom.

For the QM segment, the DFT with B3LYP functional was employed for the structure optimization. The basis functions used were Dunning's double-ζ plus polarization basis sets (D95(d))[43] for C atoms of the retinal π-system, N and H atoms of Schiff base, and O and C atoms of the carboxylate of the counter residue and the water molecule. For the other atoms, Dunning's double-ζ sets (D95)[43] were employed. In addition, single p-type anion functions ($\alpha=0.059$)[43] were augmented on the anionic O atoms of the counterion groups to properly describe the charge transfer interaction between the counterion groups and the chromophore. The coefficients of restraint terms in the RESP method are set to be 10^{-4} au for the atoms near the boundaries and 10^{-5} au for the other atoms.[10,20] The AMBER99 force field[44] was used in the MM calculation to describe the rest of the protein, and TIP3P[45] is used for water molecules. A residue-based 12 Å cutoff was used for van der Waals interactions. The QM/MM code[10,20,42] was incorporated in the QM program package GAMESS.[46]

The rms deviation between the X-ray and QM/MM optimized structures was measured for the residues within 6 Å from the retinal chromophores. The deviation is 0.6−0.9 Å for the three proteins, which is much smaller than the resolution of the X-ray structures. These deviations, however, depend on the origin common to the two structures. For the distance between the N(retinal) and O(glutamate), the deviation is about 0.3 Å.

To evaluate the absorption energies in the opsins, we performed the SAC-CI calculation. CI-singles (CIS) and TD-DFT calculations were also performed for comparison. Figure 2(a) shows the optimized structure of the QM segments: chromophore-Lys296 moiety, the counterion group Glu113, and a proximal molecule, Wat2b,[7] in Rh. In the single point SAC-CI calculations, the absorption energies were computed for the entire PSB retinal chromophore with the side chain of lysine shown in Figure 1(a,b). Furthermore, in order to examine the quantum-mechanical interaction with protein, we also include the counterion group and a proximal water molecule shown in Figure 2 (the retinal PSB active site; Glu113 and Wat2b for Rh,[7] Asp85 and Wat402 for bR,[8] and Asp75 and Wat402 for sRII[9]) which were carried out in the SAC-CI calculation. The ES effect of the other residues was included by the point charges of AMBER99.[44] The QM/MM system termed "AS" includes the chromophore, the counterion group, and the proximal water molecule in the QM segment. For comparison, we also computed the absorption energies for the QM/MM system including only the chromophore in the QM segment ("RET").

The detail of the SAC-CI method can be found elsewhere.[34] For the SAC-CI single-point calculations, we used the same basis functions as those used for the geometry optimizations. The active space of the SAC-CI calculations included all the valence orbitals. The 1s orbitals and the corresponding virtual orbitals were treated as the frozen orbitals. Total 451, 427, and 427 MOs were taken for the active orbitals in Rh, bR, and sRII, respectively. All the single excitation operators and the selected double excitation operators were included in the SAC/SAC-CI wave functions. Perturbation selection technique was used for selecting important double excitation.[36] In the perturbation selection, HF/CIS wave functions were used for the reference states to estimate second-order perturbation energy, and the energy thresholds of 5×10^{-6} and 5×10^{-7} au were used for the ground and excited states, respectively.

For SAC-CI, CIS, and TD-DFT calculations for the excited states and HF, B3LYP, MP2, and AM1 calculations for geometry optimization in the gas phase, we used a development version of Gaussian03 program system.[47] For the SAC-CI calculations, we have improved the computational algorithm for the perturbation selection of the two-electron operators, which realized the large-scale calculations of the PSB retinal active site.

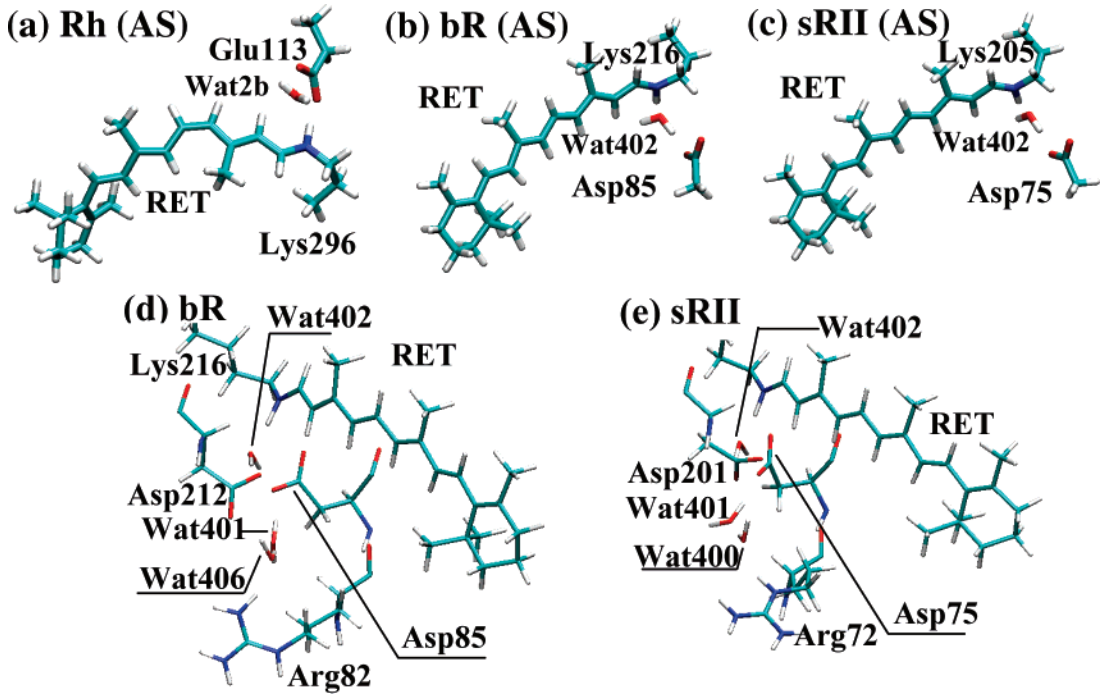


Figure 2. (a−c) QM/MM optimized structure of the active-site (AS) model (the retinal chromophore, counterion group, and a proximal water molecule) for (a) Rh, (b) bR, and (c) sRII. (d,e) The arrangement of the charged residudes close to the chromophore in (d) bR and (e) sRII.

3. Results

The first excited states which exhibit large oscillator strengths are assigned to the first absorption peaks for SAC-CI and CIS. The main character of the first excited states is $\pi-\pi^*$ excitation from the highest occupied molecular orbital (HOMO) to the lowest unoccupied MO (LUMO). In the TD-DFT calculations, it was found that the first excited states of the active site (AS) models do not possess the character of the HOMO−LUMO $\pi-\pi^*$ excitation, and the oscillator strength is rather small. We therefore assign the lowest $\pi-\pi^*$ excited states which exhibit large oscillator strengths to the first absorption peaks.

3.1. Assessment of Methodology for Geometry Optimization. We first carried out several test calculations to evaluate the methodologies for the geometry optimization. As reported by Wanko et al.,[12] the absorption energy is rather sensitive to the computational method used for the geometry optimization. Especially, bond alternation of the polyene chain in the chromophore tends to be overestimated in the HF and CASSCF optimized structure due to the lack of dynamic electron-correlation.[12] This structural deficiency causes overestimation of the excitation energy especially in the HOMO−LUMO transition, because the energy level of HOMO and LUMO are stabilized and destabilized, respectively. As shown in Figure 3 (a), the HOMO has bonding character at C5═C6, C7═C8, and C9═C10 double bonds, while the LUMO is antibonding at C11═C12, C13═C14, and C15═N double bonds.

In order to examine the bond-alternation effect on the SAC-CI absorption energies, we computed the absorption energies for all-trans retinal (Figure 1(b)) using several geometries optimized by HF, B3LYP, MP2, and AM1 methods. As shown in Table 1, HF and AM1 geometries give rather large absorption energies (1.44 and 1.37 eV, respectively) compared to the B3LYP and MP2 ones (1.18 and 1.19 eV, respectively). On the other hand, B3LYP and MP2 give structures very similar to each other.[12] We also confirmed that the SAC-CI absorption energy using the B3LYP geometry is in good agreement with that using an MP2 one, validating the use of B3LYP for the geometry optimization in the present study. These results are very similar to those obtained by Wanko and co-workers[12] and thus confirmed their previous study.

3.2. Absorption Energies of the Retinal Proteins. In Table 2, SAC-CI results are summarized and compared with experimental[1,30,31,48,49] and previous theoretical results.[10,13,20,22,26] The calculated excitation energy using the active site QM model ("AS") "in opsin" environment are 2.45 eV for Rh, 2.23 eV for bR, and 2.53 eV for sRII, which nicely agree with the experimentally observed absorption energies (2.49, 2.18, and 2.49 eV, respectively) for *all* the systems studied. The present results significantly improved the previous theoretical ones.[10,13,14,20,22,26] The root-mean-square error is around 0.04 eV, and the maximum error is 0.05 eV for bR.

It should be noted that the present result has been much improved over our previous one.[14] In the present study, the QM/MM geometry optimizations were carried out at the B3LYP/MM level, whereas the HF/MM method was employed for the optimizations in the previous study.[14] Moreover, in the previous calculations, we included only retinal PSB in the QM segment. The present model includes the counterion group and a water molecule close to the Schiff base in the QM segment in addition to the chromophore. The computed excitation energies are quite sensitive to these two factors. Although our previous results[14] were very close to the present results, this is actually due to a lucky

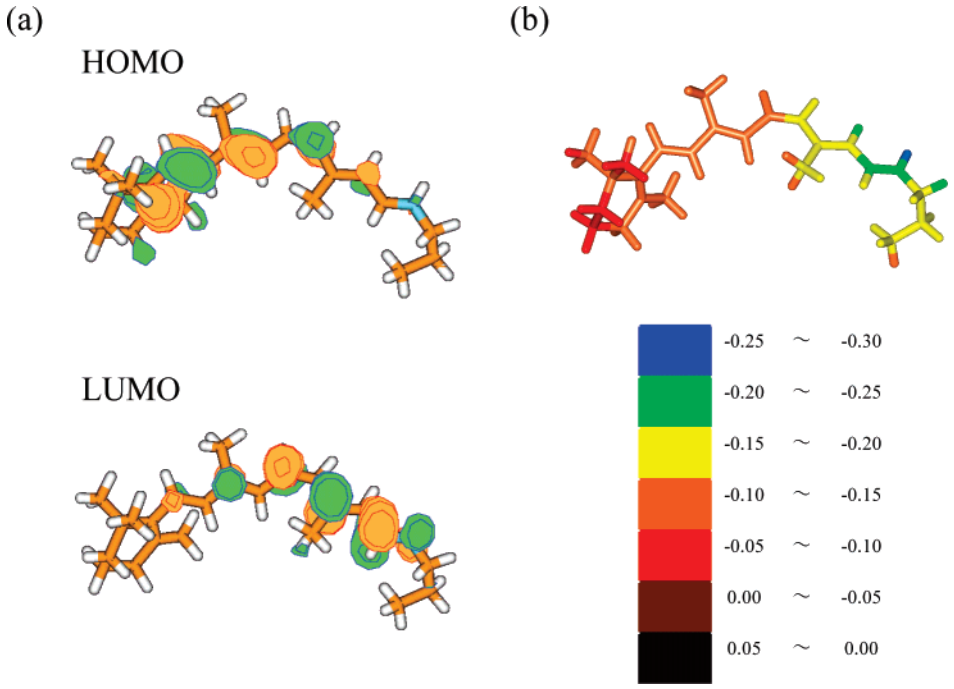


Figure 3. (a) HOMO and LUMO distribution for Rh and (b) ES potential by protein environment (opsin) in atomic unit.

Table 1. SAC-CI Absorption Energies and Oscillator Strength Calculated with the Gas-Phase Optimized Geometries

optimize	main config (|C|>0.3)	E_{ex}(eV)	f(au)
B3LYP	0.92(H → L)	1.18	1.01
MP2	0.92(H → L)	1.19	0.97
HF	0.92(H → L)	1.44	0.91
AM1	0.92(H → L)	1.37	0.81

cancellation between the two factors. The introduced error also affects the accuracy of the analysis on the color tuning mechanism. Detailed analysis is given in a later section.

The absorption energies of the chromophore are strongly affected by the interaction with the surrounding protein environments. In order to extract the effect of the protein environments, we computed the absorption energies of the chromophore in the absence of the protein environment (referred to as "bare" chromophore) and compared with those in the opsins. Note that the structures of the bare chromophore for the calculations are the same as those in the proteins. Thus, the differences in the absorption energies in the opsins from those of the bare chromophore represent the effects of the interaction with the protein environment but do not include contributions of changes in the chromophore structure upon the bindings to the opsins (see below). Table 2 lists the absorption energies of the bare chromophore. The absorption energies of the bare chromophore (1.36 eV for Rh, 1.30 eV for bR, and 1.31 eV for sRII) are considerably lower than those in the opsins by more than 1 eV, indicating that the interactions with the protein surroundings give rise to the large blue-shifts of the absorption energies.

The contributions of the chromophores' structural changes upon the bindings are also significant. We compared the results for "bare" chromophores with an all-trans chromophore (Figure 1(b)) whose structure was optimized in vacuo. The result shows that the absorption energies of the bare chromophore also exhibit blue-shifts by 0.18 eV for Rh, 0.12 eV for bR, and 0.13 eV for sRII from that for all-trans chromophore, 1.18 eV. Therefore, both the structural changes of the chromophore and the interaction with the protein surroundings cause blue-shift in the absorption energies under the opsin environment.

Table 3 lists oscillator strengths, f, calculated by the SAC-CI method for all the proteins studied. It was found that the oscillator strengths in the opsins are also significantly larger than those of the bare chromophore, indicating larger absorbances in the opsins. Note that the oscillator strength is proportional to the product of absorption energy, E_{ex}, and square of transition dipole moment, $|\mu_{eg}|$.[2] We therefore computed the transition dipole moments as well in order to identify the source of the increase of the oscillator strength. Table 3 also lists the computed transition dipole moments. In contrast to the oscillator strengths, the transition dipole moments of the chromophore in the opsins are smaller than those of the bare chromophore. Hence the larger oscillator strengths in the opsins are attributed to the large blue-shifts of the absorption energies in the opsins.

3.3. Spectral Blue-Shift Caused by the Protein Electrostatic Environment. The interaction of the chromophore with the surrounding protein environment contributes to the large blue-shift of the absorption energy as shown in Table 2. This is indicated by the significant decrease in the absorption energy of the bare chromophore. As suggested in the previous studies,[10,14] the main cause of the large blue-shift is difference in the ES interaction energy between the ground and excited states of the PSB retinal. The migration of the positive charge of PSB along the polyene chain upon the electronic excitation largely reduces the interaction with the protein environment, especially with the negatively

Table 2. Comparison of SAC-CI Excitation Energies with Other Results (eV)

				SAC-CI								
protein	model		environment	main config (|C|>0.3)	*f* (au)	$\Delta\mu$ (debye)	E_{ex} (eV)	exptl (eV)	CASPT2 E_{ex}(eV)	MRMP E_{ex}(eV)	SORCI E_{ex}(eV)	TD-B3LYP E_{ex}(eV)
Rh	WT[f]	AS	in opsin	0.94(H → L)	1.03		2.45	2.49[i]	2.86[a]			2.52
		RET		0.93(H → L)	0.88	−13.25	2.06		2.78,[b] 2.59[c]			2.44
		RET	bare	0.91(H → L)	0.63	−12.69	1.36	-	2.72,[b] 2.72[c]			2.53
bR	WT[f]	AS	in opsin	0.94(H → L)	1.29		2.23	2.18[j]				2.57
		RET		0.92(H → L)	1.15	−14.98	1.88			2.75[d]	2.34[e]	2.49
		RET	bare	0.92(H → L)	0.91	−13.29	1.30	-		2.05[d]	1.86[e]	2.31
	R82A[g]	AS	in opsin	0.94(H → L)	1.33		2.34	2.23[k]				
sRII	WT[f]	AS	in opsin	0.94(H → L)	1.42		2.53	2.49[l]				2.68
		RET		0.93(H → L)	1.27	−13.88	2.17					2.58
		RET	bare	0.92(H → L)	0.89	−14.00	1.31	-				2.30
	R72A[h]	AS	in opsin	0.94(H → L)	1.46		2.58	2.48[m]				

[a] CASPT2 result described in ref 26. [b] CASPT2 result described in ref 22. [c] CASPT2 result described in ref 13. [d] MRMP result described in ref 20. [e] SORCI result described in ref 12. [f] Shows "wild type". [g] Shows "R82A" mutant. [h] Shows "R72A" mutant. [i] Reference 1. [j] Reference 31. [k] Reference 48. [l] Reference 30. [m] Reference 49.

Table 3. Decomposition of Oscillator Strength[a]

protein	model	environment	*f* (au)	E_{ex} (eV)	$\|\mu_{eg}\|^2$
Rh	AS	in opsin	1.03 (1.63)	2.45 (1.80)	17.12 (0.92)
	RET		0.88 (1.40)	2.06 (1.51)	17.37 (0.93)
	RET	bare	0.63 (1.00)	1.36 (1.00)	18.65 (1.00)
bR	AS	in opsin	1.29 (1.42)	2.23 (1.72)	23.68 (0.85)
	RET		1.15 (1.26)	1.88 (1.45)	25.08 (0.90)
	RET	bare	0.91 (1.00)	1.30 (1.00)	28.01 (1.00)
sRII	AS	in opsin	1.42 (1.59)	2.53 (1.94)	22.95 (0.82)
	RET		1.27 (1.43)	2.17 (1.66)	23.91 (0.86)
	RET	bare	0.89 (1.00)	1.31 (1.00)	27.91 (1.00)

[a] The values in parentheses stand for the ratio, (in opsin)/(bare).

charged counterion groups in the vicinity of PSB[10] shown in Figure 2.

In order to estimate the contribution of the ES interaction energy, we computed the SAC-CI absorption energy of the QM/MM system where the QM segment includes only the chromophore (referred to as "RET"). The rest of the environmental effect was included as a point-charge model. Thus the interaction between the chromophore and the protein surroundings is only taken into account by the electrostatic potential. Table 2 lists the SAC-CI absorption energies of RET. The absorption energies were computed to be 2.06, 1.88, and 2.17 eV for Rh, bR, and sRII, respectively, and exhibit large blue-shifts from those of the bare chromophore by 0.58−0.86 eV, indicating the large contributions of the ES interaction.

Figure 3(a) depicts distributions of HOMO and LUMO of the chromophore in Rh, which are mainly responsible for the electronic excitation in the first excited state. As clearly seen, the distributions of the π orbitals are localized in the β-ionone ring and PSB halves of the polyene chain in HOMO and LUMO, respectively. Since the main configuration of the excited-state is an excitation from HOMO to LUMO, the excitation has charge-transfer character from the β-ionone ring side to the other side. In other words, the positive charge migration occurs from the PSB side to the other side, as mentioned above. The large changes in dipole moment of the chromophore upon the excitations, $\Delta\mu$ (−12.69 to −14.98 debye), listed in Table 2 clearly indicate the positive charge migrations of PSB. Figure 3(b) shows ES potential by the protein environment of Rh acting on atoms of the chromophore. Strong negative ES potential due to the counterion negative group Glu113 is observed in the PSB region. The ES potential gradually increases toward the β-ionone ring side, creating a gradient of the ES potential along the polyene chain. The large contribution of the ES interaction to the absorption energy is therefore clearly explained by the positive charge migration along the polyene chain against the gradient of the ES potential produced by the protein surroundings, especially the counterion group.

3.4. Impact of the Counterion Group on the Absorption Energy. It should be noted that the interactions of the chromophore with the protein surroundings are not fully recovered by the ES interactions described with the point-charge model. The ES treatment covers only 60−70% of the spectral blue-shift. As seen in Table 2, the absorption energies for the RET systems (1.88−2.17 eV) which only consider the ES interactions approximated by the effective point charges are still considerably lower than those of the experiments (2.18−2.49 eV) and for the AS systems (2.23−2.53 eV) where the quantum-mechanical interactions with the counterion and water molecules in the vicinity of PSB are taken into account by explicitly including them in the SAC-CI absorption energy calculations. The quantum-mechanical and ES interaction therefore provides large contributions to the blue-shifts, as suggested previously.[20] The large quantum-mechanical interaction due to the electronic polarization and the charge transfer in the ground state of bR was identified by Morokuma-Kitaura decomposition analysis in a previous study by Hayashi and Ohmine.[20] Thus the positive charge migration toward the β-ionone ring upon the excitation significantly diminishes the electronic interaction, leading to the blue-shift of the absorption energy.

The contributions of the quantum-mechanical interaction to the absorption energies, ΔE_{ex}^{ele}, are estimated with differences between the absorption energies of RET, E_{ex}(RET), and AS, E_{ex}(AS)

$$\Delta E_{ex}^{ele} = E_{ex}(\mathrm{AS}) - E_{ex}(\mathrm{RET}) \quad (1)$$

Table 4. Electronic Interaction Energies, ΔE_{ex}^{ele} (eV)

protein	ΔE_{ex}^{ele} SAC/SAC-CI	ΔE_{ex}^{ele} HF/CIS	ΔE_{ex}^{ele}(SAC/SAC-CI) − ΔE_{ex}^{ele}(HF/CIS)
Rh	0.386	0.057	0.329
bR	0.353	0.085	0.268
sRII	0.354	0.088	0.266

Table 4 lists the contributions estimated for the SAC/SAC-CI absorption energies. The contributions, 0.39, 0.35, and 0.35 eV for Rh, bR, and sRII, respectively, are considerably large and are more than one-third of the total blue-shifts by the interaction with the protein surroundings. Interestingly, dynamic electron-correlation is crucial for accurately describing the interaction. To see this, we also computed ΔE_{ex}^{ele} for the absorption energies at the HF/CIS level of theory. As seen in Table 4, the HF/CIS contributions of the electronic interaction were estimated to be 0.06−0.09 eV and are rather underestimated by 0.27−0.33 eV compared to the SAC/SAC-CI ones.

In addition, we checked the convergence of absorption energy with respect to the size of the QM segment. Since, SAC-CI calculation could not be performed due to the limitation in the computational resource, an ONIOM-like analysis was performed to examine the effect of the other charged residues close to the retinal chromophore.

$$E_{ex}^{Large,SAC-CI} \cong E_{ex}^{Small,SAC-CI} + (E_{ex}^{Large,CIS} - E_{ex}^{Small,CIS}) \quad (2)$$

$E_{ex}^{X,Y}$denotes the excitation energy for the X (X = Small or Large) QM region using the Y (Y = SAC-CI or CIS) method. The "Small" QM region is identical to the "AS" models as shown in Figure 2. The "Large" QM region additionally includes one arginine, one aspartate, and three water molecules in bR and sRII as shown in Figure 2(d,e). For the bR case, Arg82, Asp212, Wat401, Wat402, and Wat406 were included. In the sRII case, Arg72, Asp201, Wat400, Wat401, and Wat402 were added to the QM segment. With this estimation, we obtained the correction $E_{ex}^{Large,CIS} - E_{ex}^{Small,CIS}$ of 0.00 and 0.01 eV for bR and sRII, respectively. Therefore, the AS systems are suitable choices for the QM segment during the absorption energy calculation. This result indicates the quantum-mechanical effect from these secondary residues is negligible.

3.5. Effect of the Chromophore Structure on the Absorption Energy. As mentioned before, the change in the bond alternation significantly affects the absorption energy of the chromophore. With stronger bond alternation, more spectral blue-shift is expected as shown in Table 1. It is noteworthy that the bond alternation becomes strong in opsin due to the interaction with the protein surroundings.

Figure 4 illustrates the bond-alternation patterns of the all-trans PSB retinal chromophore in vacuo and in the bR opsin. The bond-alternation pattern of SB retinal is also shown in Figure 4 for comparison. It is clearly seen that the bond alternation in the bR opsin is more enhanced than that in vacuo. The result is consistent with a previous study by Sugihara et al.[28] using the DFTB method. The strongest bond alternation was observed in the SB retinal. The stronger bond

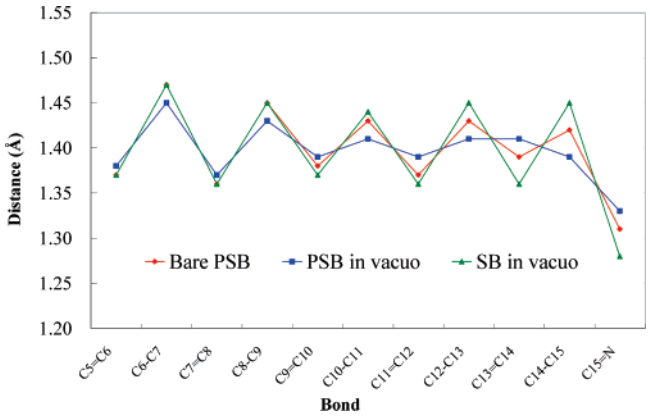


Figure 4. Bond length on the π-chain of PSB with B3LYP geometry (Å).

alternation of the PSB retinal in the opsin originates from the interaction of PSB with the counterion groups that suppresses resonance structures of the PSB retinal involving migrations of the positive charge along the polyene chain.[28] The pronounced bond alternation in the SB retinal is attributed to the absence of the resonance structures.

As shown above, the stronger bond alternation of the PSB retinal leads to the larger absorption energy. The main origin of the larger absorption energies of the bare chromophores than that of the stable chromophore in vacuo are therefore the stronger bond alternations of the bare chromophores induced by the interaction with the counterion groups in the opsins. Since the increases of the absorption energies of three proteins, 0.12−0.18 eV, are larger than the differences of the increases (maximally 0.06 eV), the stronger bond alternation is suggested to be a common mechanism giving rise to the main increases of the absorption energies for the three proteins. Overall, the interaction of the PSB retinal with the counterion groups contributes to the large blue-shift upon the binding of the chromophore into the opsin through not only the reduction of the interaction upon the electronic excitation discussed above but also the change in the bond alternation pattern.

The results clearly explain strong correlation between vibrational frequency of the ethylenic mode and the absorption energy revealed by Ebrey et al.[50] The resonance Raman experiments revealed that the retinal proteins with the larger absorption energies possess the higher ethylenic frequencies of the chromophore. The higher ethylenic frequency is attributed to the stronger bond alternation due to the stronger interaction of PSB with the counterion groups, which results in the larger absorption energy as discussed above.

There is a large difference in the bond alternation between the X-ray crystallographic structural model (1L9H) and the refined QM/MM one optimized at the DFT/B3LYP level of theory (B3LYP/MM). Figure 5 depicts the bond alternation patterns, showing much stronger bond alternation in the chromophore of 1L9H. We also carried out the SAC-CI absorption energy calculation for the chromophore of 1L9H, where positions of the missing hydrogen atoms in the X-ray crystallographic structures were determined with AMBER99 force field. Table 5 gives a comparison of the absorption energy between 1L9H and B3LYP/MM. The computed SAC-

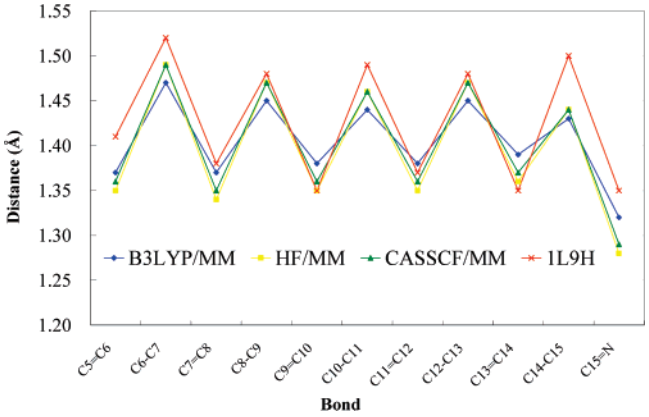


Figure 5. Bond length on π-chain of PSB in Rh (Å).

Table 5. SAC-CI Absorption Energies and Oscillator Strength for the Isolated Chromophores

optimize	main config(|C|>0.3)	E_{ex} (eV)	f (au)
B3LYP/MM	0.91 (H → L)	1.36	0.63
HF/MM[a]	0.89 (H → L)	1.72	0.53
CASSCF/MM[b]	0.89 (H → L)	1.70	0.52
1L9H[c]	0.72 (H → L)	1.72	0.42

[a] Reference 14. [b] Reference 13. [c] Reference 7.

CI absorption energy for 1L9H is 1.72 eV and significantly deviates from the B3LYP/MM one (1.36 eV), indicating again great impact of the bond alternation on the absorption energy. Note that the differences in the bond length responsible for the bond alternation between 1L9H and B3LYP/MM are less than 0.1 Å and much smaller than the resolution of the X-ray crystallographic model (2.6 Å).[7] Hence the X-ray crystallographic model is not "chemically precise" enough, and the structure refinement by the reliable methods is definitively requisite for the accurate prediction of the absorption energy. The bond alternation in the HF optimized structure is much stronger than that of the B3LYP one. The SAC-CI absorption energy of the bare chromophore using the HF/MM optimized structure was estimated to be 1.72 eV and is greatly overestimated compared to that for the B3LYP/MM optimized structure, as shown in Table 5.

Torsion of the polyene chain of the chromophore in the binding pocket also alters the absorption energy as mentioned above. The differences in the absorption energies of the bare chromophore mainly represent the contributions of the torsional distortion of the chromophore. The contributions of the torsion play an important role in the fine-tuning of the absorption energies among the retinal proteins as seen below.

Table 6 summarizes dihedral angles of the polyene chain of the chromophore in bR, sRII, and Rh. The chromophores in bR and sRII are both in all-trans conformation, and their structures are similar to each other. The polyene chains are almost planar except for the dihedral angles of C_{13}═C_{14}. Deviations of the C_{13}═C_{14} dihedral angles from planarity are by 21 and 15 degrees in bR and sRII, respectively, indicating crucial pretorsions for the selective photoisomerizations for those functions.[42] In Rh, the chromophore assumes a C_{11}-cis and C_6-syn conformation and is strongly twisted in the protein confinement. Again, torsion around the C_{11}═C_{12} bond by 14 degrees could play a role in the selective photoisomerization. In addition, a dihedral angle around the C_6−C_7 single bond largely deviates from planarity by −47 degrees. This value is larger than the one by Sugihara (−42 degrees)[26] and smaller than by Andruniów (−54 degrees).[13]

The single bond rotation around C_6−C_7 in Rh greatly contributes to the spectral shift. In order to illustrate this, we performed SAC-CI absorption energy calculations for the chromophore with torsions around the C_6−C_7 bond by 30 and 60 degrees. The absorption energies were computed for the chromophore in vacuo, and the geometries were optimized with constraints for the C_6−C_7 dihedral angles. Table 7 lists the dependence of the SAC-CI absorption energy on the bond rotation. The absorption energy increases as the torsion around C_6−C_7 is larger, consistent with calculations by Wanko et al. using other high level methods.[12] The blue-shift induced by the torsion around the single bond is due to shortened π conjugation of the polyene chain. The higher absorption energy of the bare chromophore of Rh is therefore suggested to be mainly attributed to the large torsion around C_6−C_7 by −47 degree.

4. Discussion

4.1. Mechanism of the Color Tuning. Based on the successful prediction of the absorption energies for those retinal proteins, we analyzed the spectral shifts in order to identify the molecular factors determining the color tuning. We consider the absorption energy of bR as a reference and analyzed a similar amount of the blue-shifts (0.31 eV) in the absorption energy of Rh and sRII from that of bR. As mentioned above, the apparently similar shifts are expected to involve different mechanisms of the color tuning, since the protein structures of Rh and sRII are quite different. The structure of sRII is found to be very similar to that of bR,[8,9] whereas the structure of Rh[7] is incompatible with those of bR and sRII.

The spectral shifts can be decomposed into three contributions. The first one is the structural distortion of the chromophore due to the protein confinement (structural effect). The second one is the ES interaction of the chromophore with the surrounding proteins (ES effect). The last one is the quantum effect of the counterion and a water molecule in the vicinity of PSB. This is a correction by the quantum-mechanical effect which cannot be described by the ES calculations (counterion QM correction). Those contributions can be deduced from the absorption energies listed in Table 2. The structural distortion effect is the difference of the absorption energies of the bare chromophores. The ES effect is the difference of the spectral shift due to the electrostatic environment modeled by the point-charges. The counterion QM correction is the difference in the absorption energy between the AS and RET systems.

As shown in Figure 6, the underlying mechanisms are completely different between Rh and sRII. In Rh, the ES contribution mainly contributes to the blue-shift (0.12 eV), which is more than the half of the total computed shift of 0.22 eV. This is because the interaction between PSB and the counterion group in Rh is stronger than that in bR. As

Table 6. Dihedral Angles (in deg) of the Retinal Chromophore in the QM/MM Refined Structures of Rh, bR, and sRII

	Rh				bR			sRII	
	B3LYP/MM[a]	HF/MM[b]	SCC-DFTB/MM[c]	CASSCF/MM[d]	B3LYP/MM[a]	HF/MM[e]	SCC-DFTB/MM[f]	B3LYP/MM[a]	HF/MM[e]
$N_\zeta-C_\epsilon$	93.12	94.48	100.76	109.36	−114.84	−112.87	−100.66	−95.56	−99.17
$C_{15}=N_\zeta$	179.36	−179.87	169.94	173.39	−166.73	−166.82	−168.91	−173.06	−169.23
$C_{14}-C_{15}$	172.78	172.91	178.59	177.74	−179.24	−178.15	174.36	177.85	179.12
$C_{13}=C_{14}$	−178.52	−178.94	173.29	178.79	−158.95	−161.34	−153.17	−164.99	−164.93
$C_{12}-C_{13}$	170.07	166.07	171.71	166.81	175.74	173.72	174.37	177.78	176.83
$C_{11}=C_{12}$	−13.93	−9.67	−18.25	−8.19	−172.60	−172.70	−171.71	−176.86	−175.56
$C_{10}-C_{11}$	175.00	168.85	175.34	166.24	−176.11	−174.98	−172.71	176.98	178.55
$C_9=C_{10}$	166.77	168.88	170.43	169.63	−179.23	−178.80	174.69	−174.28	−174.57
C_8-C_9	176.73	−176.46	171.30	−170.31	−173.57	−172.59	−164.50	177.36	178.89
$C_7=C_8$	175.81	174.96	−179.52	175.95	175.50	176.86	171.00	−173.90	−174.41
C_6-C_7	−46.73	−51.78	−42.09	−53.95	172.90	169.73	167.10	174.95	173.50

[a] Present geometry. [b] Reference 14. [c] Reference 26. [d] Reference 13. [e] Reference 10. [f] Reference 12.

Table 7. Dependence of Absorption Energies on C6−C7 Rotation of all-trans PSB

	TD-B3LYP			SAC-CI		
angle (deg)	main config |C|>0.3	*f* (au)	E_{ex}(eV)	main config |C|>0.3	*f* (au)	E_{ex}(eV)
≅0[a]	0.55(H → L)	2.03	2.41	0.92(H → L)	1.03	1.18
30.0[b]	0.56(H → L)	1.87	2.39	0.92(H → L)	0.98	1.21
60.0[b]	0.56(H → L)	0.81	2.17	0.92(H → L)	0.95	1.42

[a] Fully optimized geometry with B3LYP in the gas phase. [b] B3LYP geometry in the gas phase only fixed C6−C7 dihedral angle.

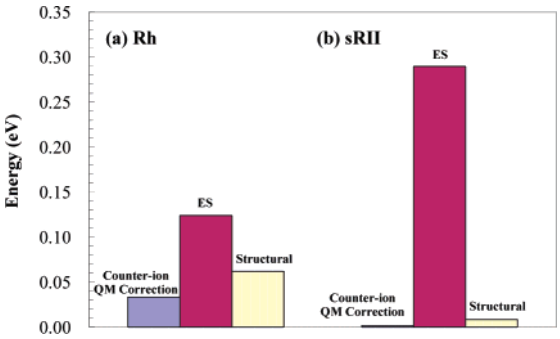


Figure 6. Decomposition of absorption energy difference from bR.

seen in Figure 2, the PSB in Rh forms a salt bridge directly to the counterion group, Glu113, whereas bR has a water molecule (Wat402) intervening between PSB and the counterion groups. The structural distortion and the QM correction contribute to the blue-shift by 0.06 and 0.04 eV, respectively. The structural contribution is mainly attributed to the torsion around the C_6-C_7 bond as described above.

In contrast to the case of Rh, the structural distortion is only a minor contribution of 0.01 eV to the overall blue-shift (0.30 eV). As seen above, the structures of the chromophore in bR and sRII are quite similar, resulting in the small contribution of the geometric structural distortion. Instead, the ES contribution (0.29 eV) dominates the overall blue-shift, as suggested previously.[10,14,20] In contrast to the shift of Rh, the counterions (0.00 eV) do not compensate the ES one.

On the large blue-shift of sRII from bR, a mechanism suggested in a previous study[51] is different in interaction of the chromophore with the positively charged guanidinium group of a proximal arginine between bR and sRII. The X-ray crystallographic studies revealed that orientation of the side chain of the proximal arginine, Arg72, in sRII is opposite to that of the corresponding arginine, Arg82, in bR.[8,9] The mechanism, however, has been questioned by other studies.[10,14] We therefore again analyzed the mechanism with the improved computational strategy used in the present study.

In order to examine the mechanism, we carried out theoretical mutation: the positive ES potentials produced by Arg82 of bR and Arg72 of sRII were replaced by that by alanine, a neutral residue, and the B3LYP/MM geometry optimizations were performed for the mutant proteins. As shown in Table 2, the SAC-CI absorption energies of the mutated bR and sRII are 2.34 and 2.58 eV, respectively, and undergo small changes from those of the native ones, 2.23 and 2.53 eV, respectively. This is consistent with results of the mutation experiments.[48,49] It is therefore concluded from the analysis that the ES interaction with the arginines plays a minor role in the spectral shift. This is because those arginines create the positive ES potential almost equally over the entire polyene chain of the chromophore, and the migration of the positive charge of PSB upon the excitation does not alter the ES interaction with the arginines.[10,14] It should be noted that the present analysis does not rule out non-negligible contribution of the electronic interaction with arginines through a hydrogen-bond network spanned between PSB and the arginines, as shown in Figure 2(d,e).

Recently, Hoffmann et al. proposed the two main and equally important factors about the color tuning mechanism between bR and sRII: the difference of neutral amino acids in the binding pocket and the difference in the extended hydrogen-bond network at the extracellular side of the proteins (counterion region).[52] We are also working on this subject, and the contribution of the electronic interaction has to be quantitatively determined in the near future.

The use of HF optimized structures overestimates the effect of the geometric distortion in the spectral blue-shift of Rh from bR. The contribution was estimated to be 0.21 eV for the HF/MM optimized structures, which is much larger than 0.06 eV for the B3LYP ones of the present study. The HF

Table 8. SAC-CI Absorption Energies, Oscillator Strength, and Difference Dipole Moment between the First Excited and Ground State in the Gas Phase (eV)

model		main config (|C|>0.3)	E_{ex}(eV)	f(au)	$\Delta\mu_{ex-gr}$ (debye)
PSB	bare[a]	0.92(H → L)	1.30	0.91	−13.29
	in vacuo[b]	0.92(H → L)	1.18	1.01	−9.36
SB	in vacuo[b]	0.94(H → L)	3.06	1.73	7.79

[a] B3LYP/MM geometry in bR. [b] B3LYP geometry in the gas phase.

method gives too strong bond alternation of the polyene structure, which overemphasizes torsions around the single and double bonds, respectively, although the overall distortion of the chromophore in the biding pocket is not greatly altered. Table 6 compares the torsions of the HF optimized structure with those of the B3LYP one, clearly indicating the errors of torsion. These errors lead to the overestimation of the contribution of geometric distortion. Note that the π bonding distributions of HOMO and LUMO locate mainly on the double and single bonds, respectively. Thus the overestimation of torsions around the single bonds lifts the energy level of LUMO. Similarly, the underestimation of torsions around the double bonds lowers that of HOMO. Both of the errors therefore give rise to the overestimated contribution of geometric distortion.

It is noteworthy that the absorption energy for the SB retinal in vacuo is computed to be 3.06 eV by the SAC-CI method as shown in Table 8. This result indicates that deprotonation at the Schiff base causes significant blue-shift in the absorption energy. In this sense, the present result supports the deprotonated form to be the UV pigments as suggested by Blatz et al. and Dukkipati et al.[53] We are now investigating the excited states for the UV pigments, and these will be reported in the future.

4.2. Comparison with Previous Theoretical Studies. The present results are compared with the previous studies using the modern electron-correlation methods.

In our previous SAC-CI study,[14] the QM region included only the chromophore. The counterion and water molecules in the vicinity of PSB were treated by the point-charge model. Furthermore, the HF method was employed for the QM/MM optimization.[14] The present study shows that these two inaccurate treatments introduce sizable errors in the absorption energy. As suggested previously[11,12,20] and confirmed in the present study, excluding the counterion groups from the QM segment significantly underestimates the absorption energy. As discussed in section 3.5, the HF optimized structure causes large overestimation of the absorption energy due to too strong bond alternation[12,25] and the torsion of the $\pi-$chain. The agreement in the previous study therefore is the result of cancellation between the underestimation due to the neglect of the electronic interaction and the overestimation due to the use of HF method for the geometry optimization.

Hayashi and Ohmine computed the absorption energy of bR at the MRMP level of theory.[20] With the HF/MM optimized structure and the QM segment including only the chromophore, they obtained the absorption energy of 2.75 eV, which rather overestimated the experimental data. The active space of the MRMP calculation was restricted to 12 electrons in 9 π orbitals instead of the full valence 12 electrons in 12 π orbitals. This would be another source of the overestimation in the absorption energy.

Recently, Andruniów et al. have revised their CASPT2 calculation by including the entire chromophore in the QM region.[13] The absorption energy was improved to be 2.59 eV, which nicely agrees with the experimental one (2.49 eV). However, in their mechanism proposed for the color tuning, the interaction of the chromophore with the protein surroundings provides a red shift in the absorption energy of the chromophore compared with the gas-phase result, which is significantly different from the present one: the absorption energy of the bare chromophore (2.72 eV) they obtained was larger than the absorption energy of the chromophore in the opsin (2.49 eV by the experiment[1] and 2.59 eV by CASPT2[13]). This gas-phase result is also much larger than those by Schreiber and Buss[11] using CASPT2 (1.88 eV) and Wanko et al.[12] using SORCI (1.93 eV). This discrepancy would originate from the CASSCF optimized structure used for the CASPT2 calculation. As shown in Figure 5, the CASSCF bond-alternation pattern of the chromophore[13] is very close to the HF one. We also note that the QM segment of their CASPT2 calculation did not include the counterion group and a water molecule proximal to PSB.

The QM/MM optimization procedure is also different. In the study of Andruniów et al.,[13] they used a X-ray crystallographic structure determined by Teller et al.[54] (PDB code: 1HZX), whereas the present study employed that determined by Okada et al. (PDB code: 1L9H).[7,29] The chromophore structure determined by Teller et al. is strongly distorted in the Schiff base region, showing remarkable deviation from the other three X-ray crystallographic structures by Palczewski et al.[55] and Okada et al.[7,29] In the QM/MM optimization, the protein atoms other than the Lys296 side chain and proximal water molecules were kept frozen, whereas the whole protein including the chromophore is fully and consistently relaxed in the present QM/MM optimization. Consequently, Andruniów et al. obtained the N_ζ(PSB)-O(Glu113) distance of 3.7 Å, which is much longer than that obtained in the present study (2.66 Å). The present structure is in accord with other QM/MM[26,29,56] and MD[57] studies. The large increase in the N_ζ-O distance breaks the salt bridge structure and can greatly lower the absorption energy as previously suggested.[10]

It should be noted, however, that the distance between PSB and the counterion group is still controversial in the experiment. NMR studies have suggested a longer distance and presence of a water molecule intervening between PSB and the counterion group.[58] In contrast, FTIR experiments have given the stronger hydrogen bonding between PSB and the counterion (Glu113) in Rh compared with that in bR.[59] In addition, low-temperature FTIR experiments of an internal water molecule in Rh has revealed a weaker hydrogen bonding of a water molecule (Wat2b) to the counterion (Glu113) compared with that of bR and sRII.[60] These data would indicate a shorter distance between PSB and the counterion group. In the latest X-ray crystallographic structure model by Okada et al.,[29] the distance is observed to be

~3.28 Å, although the resolution of the measurement (2.2 Å) does not seem to be high enough to detect the distance precisely. In this regard, further examination is required.

Wanko et al. clarified important methodological aspects in calculating the absorption energy of retinal proteins, such as the bond alternation and the distortion around C_6-C_7 bond of the chromophore.[12] They performed QM(SCC-DFTB[26,61])/MM calculations for the structural optimization and the SORCI calculations for the excited states. The computed absorption energy is 2.34 eV for bR, which is close to the experimental value of 2.18 eV. However, the calculation did not include the counterion groups in the QM segment. The computed absorption energy may become higher if the electronic interaction is taken into account.

Gascon and Batista computed the absorption energy of Rh with the TD-DFT method.[23] The absorption energy obtained (2.57 eV) is in good agreement with the experimental one (2.49 eV). However, the TD-DFT method is not appropriate for calculating the excited states of the PSB retinal chromophore. On twisting the C_6-C_7 bond, the TD-DFT gives even wrong trends as shown by Wanko et al.[12] and also in Table 7. In addition, TD-DFT failed to reproduce the absorption energy of Rh, bR, and sRII as shown in Table 2.

Schreiber et al. performed CASPT2 calculations in the gas phase for the PSB retinal chromophore and some counterion group (Glu113 and Wat2b).[11] They pointed out the remarkable influence of the counterion groups to the absorption energy. Sugihara et al. performed CASPT2//QM (SCC-DFTB)/MM calculations.[26] However, the chromophore was simplified, and the calculated excitation energy overestimated the experimental data.[11,26]

Recently, Andersen et al. measured the absorption spectrum of the C_6-syn PSB retinal chromophore in vacuo by using the ELISA, and the absorption maximum was observed at 2.03[39] and 2.00 eV.[40] As expected, this gas-phase absorption energy shows the red-shift of 0.79 eV compared with that in methanol (2.82 eV).[62] So far we have not performed calculations comparable to this experiment. With the geometry reported by Cembran et al.,[63] we obtained 1.87 eV by the SAC-CI method. However, this structure was optimized by the CASSCF method and is expected to be insufficient due to the overestimation in the bond alternation. The absorption energy also strongly depends on the torsional angle around the polyene chain. Therefore, it would be necessary for further investigation to simulate the ELISA experiment.

5. Conclusion

The excited states of Rh, bR, and sRII were studied by the SAC-CI calculations with the QM(B3LYP)/MM(AMBER99) optimized structures. This is the first study which successfully reproduced the major three retinal proteins, Rh, bR, and sRII. Based on the SAC-CI results, the mechanism of the color tuning in Rh, bR, and sRII was analyzed. The result shows that the spectral differences in these proteins are mainly due to the opsin ES interaction, and the chromophore conformation have only a minor effect on the color tuning in these proteins especially for bR and sRII. Our result also shows that the electronic interactions (quantum effect) between the chromophore and the counterion group are minor contributions to the color tuning. However the interactions are indispensable for predicting absolute absorption energies. There is so far no previous theoretical study reporting the electronic interaction due to the computational restriction, and thus current information would be much useful in a future study.

The effect of retinal bond alternation to the absorption energy was examined by comparison of several geometries both in vacuo and bare environment. The results clearly indicate that the retinal structure should be described precisely since the absorption spectrum is highly sensitive to retinal bond alternation. In addition, it is also shown that the dynamic electron-correlation being included in the wave function is necessary to expect the accurate absorption energy.

The present results indicate that the SAC-CI calculation with the QM/MM optimized structure is a promising approach to study the absorption spectrum of the retinal protein. This approach would be useful for studying the color tuning of the retinal proteins led by the theoretical prediction. Future studies of the color tuning are to account for the polarization effect in the protein. In the present study, though the polarization of the QM region (PSB, counteranion, water) by the protein environment is included, that of the protein environment by the chromophore is neglected. Houjou et al. suggested the large polarization effect on the absorption energies by the polarizable continuum calculations.[19,27] It is furthermore necessary to clarify the effect of protein polarization in an atomic detail. A study in this regard is ongoing.

Acknowledgment. This study was supported by a Grant-in-Aid for Creative Scientific Research and for Young Researcher from the Ministry of Education, Culture, Sports, Sciences, and Technology of Japan and also by Japan Science and Technology Agency. One of the authors (K. F.) gratefully acknowledges the Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science. A part of the computation was carried out in RCCS in Okazaki.

Supporting Information Available: The SAC-CI inputs on Gaussian03. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kandori, H.; Schichida, Y.; Yoshisawa, T. *Biochemistry (Moscow)* **2001**, *66*, 1197. Mathies, R. A.; Lugtenburg, J. In *Handbook of Biological Physics*; Stavenga, D. G., Grip, W. J. d., Pugh, E. N., Eds.; Elsevier Science B. V.: Amsterdam, 2000.

(2) Stavenga, D. G.; Grip, W. J. d.; Pugh, E. N. In *Molecular Mechanisms in Viral Transduction*; Elsevier Science: New York, 2000.

(3) Nakanishi, K.; Crouch, R. *Isr. J. Chem.* **1995**, *35*, 253.

(4) Kleinschmidt, J.; Harosi, F. I. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9181.

(5) Irving, C. S.; Byers, G. W.; Leermakers, P. A. *Biochemistry* **1970**, *9*, 858. Blatz, P. E.; Liebman, P. A. *Exp. Eye Res.* **1973**, *17*, 573. Honig, B.; Kinur, U.; Nakanishi, K.; Balogh-Nair, V.; Gawinowicz, M. A.; Arnaboldi, M.; Motto, M. G. *J. Am. Chem. Soc.* **1979**, *101*, 7084. Schulten, K.; Dinur, U.; Honig, B. *J. Chem. Phys.* **1980**, *73*, 3927. Béjá, O.; Spudich, E. N.; Spudich, J. L.; Leclerc, M.; DeLong, E. F. *Nature* **2001**, *411*, 786.

(6) Lin, S. W.; Kochendoerfer, G. G.; Carroll, K. S.; Wang, D.; Mathies, R.; Sakmar, T. P. *J. Biol. Chem.* **1998**, *273*, 24583.

(7) Okada, T.; Fujiyoshi, Y.; Silow, M.; Navarro, J.; Landau, E. M.; Shichida, Y. *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 5982.

(8) Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P.; Lanyi, J. K. *J. Mol. Biol.* **1999**, *291*, 899.

(9) Royant, A.; Nollert, P.; Edman, K.; Neutze, R.; Landau, E. M.; Pebay-Peyroula, E. Navarro, J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10131.

(10) Hayashi, S.; Tajkhorshid, E.; Pebay-Peyroula, E.; Royant, A.; Landau, E. M.; Navarro, J.; Schulten, K. *J. Phys. Chem. B* **2001**, *105*, 10124.

(11) Schreiber, M.; Buss, V.; Sugihara, M. *J. Chem. Phys.* **2003**, *119*, 12045.

(12) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606.

(13) Andruniów, T.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17908.

(14) Fujimoto, K.; Hasegawa, J.; Hayashi, S.; Kato, S.; Nakatsuji, H. *Chem. Phys. Lett.* **2005**, *414*, 239.

(15) Kandori, H. *Chem. Ind.* **1995**, *18*, 735.

(16) Mathies, R.; Stryer, L. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2169.

(17) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640. Sakurai, M.; Sakata, K.; Saito, S.; Nakajima, S.; Inoue, Y. *J. Am. Chem. Soc.* **2003**, *125*, 3108.

(18) Warshel, A.; Chu, Z. T. *J. Phys. Chem. B* **2001**, *105*, 9857.

(19) Houjou, H.; Inoue, Y.; Sakurai, M. *J. Phys. Chem. B* **2001**, *105*, 867.

(20) Hayashi, S.; Ohmine, I. *J. Phys. Chem. B* **2000**, *104*, 10678.

(21) Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125.

(22) Ferré, N.; Olivucci, M. *J. Am. Chem. Soc.* **2003**, *125*, 6868.

(23) Gascon, J. A.; Batista, V. S. *Biophys. J.* **2004**, *87*, 2931.

(24) Nakayama, K.; Nakano, H.; Hirao, K. *Int. J. Quantum Chem.* **1998**, *66*, 157.

(25) Hufen, J.; Sugihara, M.; Buss, V. *J. Phys. Chem. B* **2004**, *108*, 20419.

(26) Sugihara, M.; Hufen, J.; Buss, V. *Biochemistry* **2006**, *45*, 801.

(27) Houjou, H.; Koyama, K.; Wada, M.; Sameshima, K.; Inoue, Y. Sakurai, M. *Chem. Phys. Lett.* **1998**, *294*, 162.

(28) Sugihara, M.; Buss, V.; Entel, P.; Elstner, M.; Frauenheim, T. *Biochemistry* **2002**, *41*, 15259.

(29) Okada, T.; Sugihara, M.; Bondar, A.; Elstner, M.; Entel, P.; Buss, V. *J. Mol. Biol.* **2004**, *342*, 571.

(30) Chizhov, I.; Schmies, G.; Seidel, R.; Sydor, J. R.; Lüttenberg, B.; Engelhard, M. *Biophys. J.* **1998**, *75*, 999.

(31) Birge, R. R.; Zhang, C. F. *J. Chem. Phys.* **1990**, *92*, 7178.

(32) Nakatsuji, H.; Hirao, K. *J. Chem. Phys.* **1978**, *68*, 2053.

(33) Nakatsuji, H. *Chem. Phys. Lett.* **1978**, *59*, 362. Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 329. Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 334.

(34) Nakatsuji, H. In *Computational Chemistry - Reviews of Current Trends*; Leszcynski, J., Ed.; World Scientific: Singapore, 1997.

(35) Ehara, M.; Hasegawa, J.; Nakatsuji, H. In *Theory and Applications of Computational Chemistry*; Dykstra, C., Frenking, G., Kim, K., Scuseria, G., Eds.; Elsevier Science: New York, 2006.

(36) Nakatsuji, H. *Chem. Phys.* **1983**, *75*, 425.

(37) Neese, F. *J. Chem. Phys.* **2003**, *119*, 9428.

(38) Lee, H. M.; Kim, J.; Kim, C. J.; Kim, K. S. *J. Chem. Phys.* **2002**, *116*, 6549.

(39) Andersen, L. H.; Nielsen, I. B.; Kristensen, M. B.; Ghazaly, M. O. A. E.; Haacke, S.; Nielsen, M. B.; Petersen, M. Å. *J. Am. Chem. Soc.* **2005**, *127*, 12347.

(40) Nielsen, I. B.; Lammich, L.; Andersen, L. H. *Phys. Rev. Lett.* **2006**, *96*, 018304.

(41) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467. Gogonea, V.; Suárez, D.; van der Vaart, A.; Merz, K. W. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 217. Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 3093. Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700. Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227. Warshel, A.; Chu, Z. T.; Hwang, J.-K. *Chem. Phys.* **1991**, *158*, 303.

(42) Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2002**, *83*, 1281.

(43) Dunning, T. H., Jr.; Hey, P. J. In *Method of Electronic Structure Theory*; Shaefer, H. F., III, Ed.; Plenum Press: New York, 1977; p 1.

(44) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.

(45) Jorgensen, W. L.; Chandreskhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1982**, *79*, 926.

(46) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R. J. A.; Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.;

Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 2003.

(48) Balashov, S. P.; Govindjee, R.; Kono, M.; Imasheva, E.; Lukashev, E.; Ebrey, T. G.; Crouch, R. K.; Menick, D. R.; Feng, Y. *Biochemistry* **1993**, *32*, 10331.

(49) Ikeura, Y.; Shimono, K.; Iwamoto, M.; Sudo, Y.; Kamo, N. *Photochem. Photobiol.* **2003**, *77*, 96.

(50) Aton, B.; Doukas, A. G.; Callender, R. H.; Becher, B.; Ebrey, T. G. *Biochemistry* **1977**, *16*, 2995.

(51) Ren, L.; Martin, C. H.; Wise, K. J.; Gillespie, N. B.; Luecke, H.; Lanyi, J. K.; Spudich, J. L.; Birge, R. R. *Biochemistry* **2001**, *40*, 13906.

(52) Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Sci.* **2006**, *128*.

(53) Blatz, P. E.; Mohler, J. H.; Navangul, H. V. *Biochemistry* **1972**, *11*, 848. Dukkipati, A.; Kusnetzow, A.; Babu, K. R.; Ramos, L.; Singh, D.; Knox, B. E.; Birge, R. R. *Biochemistry* **2002**, *41*, 9842.

(54) Teller, D. C.; Okada, T.; Behnke, C. A.; Palczewski, K.; Stenkamp, R. E. *Biochemistry* **2001**, *40*, 7761.

(55) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Trong, I. L.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*.

(56) Röhrig, U. F.; Guidoni, L.; Laio, A.; Frank, I.; Rothlisberger, U. *J. Am. Chem. Soc.* **2004**, *126*, 15328.

(57) Röhrig, U. F.; Guidoni, L.; Rothlisberger, U. *Biochemistry* **2002**, *41*, 10799. Saam, J.; Tajkhorshid, E.; Hayashi, S.; Schulten, K. *Biophys. J.* **2002**, *83*, 3097.

(58) Eilers, M.; Reeves, P. J.; Ying, W.; Khorana, H. G.; Smith, S. O. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 487; Verhoeven, M. A.; Creemers, A. F. L.; Bovee-Geurts, P. H. M.; Grip, W. J. d.; Lugtenburg, J.; Groot, H. J. M. d. *Biochemistry* **2001**, *40*, 3283.

(59) Baasov, T.; Friedman, N.; Sheves, M. *Biochemistry* **1987**, *26*, 3210.

(60) Furutani, Y.; Shichida, Y.; Kandori, H. *Biochemistry* **2003**, *42*, 9619.

(61) Bondar, A.; Fischer, S.; Smith, J. C.; Elstner, M.; Suhai, S. *J. Am. Chem. Soc.* **2004**, *126*, 14668.

(62) Freedman, K. A.; Becker, R. S. *J. Am. Chem. Soc.* **1986**, *108*, 1245.

(63) Cembran, A.; González-Luque, R.; Altoè, P.; Merchán, M.; Bernardi, F.; Olivucci, M.; Garavelli, M. *J. Phys. Chem. A* **2005**, *109*, 6597.

CT6002687

JCTC Journal of Chemical Theory and Computation

The B(OH)−NH Analog Is a Surrogate for the Amide Bond (CO−NH) in Peptides: An ab Initio Study

Alpeshkumar K. Malde, Santosh A. Khedkar, and Evans C. Coutinho*

Department of Pharmaceutical Chemistry, Bombay College of Pharmacy, Kalina, Santacruz (E), Mumbai 400 098, India

Received August 7, 2006

Abstract: The conformational preferences of *N*-methyl-methylboronamide (NMB), a B(OH)−NH analog of the amide CO−NH in natural peptides, have been investigated at the Hartree−Fock; Becke's three-parameter exchange functional and the gradient-corrected functional of Lee, Yang, and Parr; and second-order Møller−Plesset levels of theory with the 6-31+G* basis set. The minima, saddle points, and rotation barriers on the potential energy surface of NMB have been located and the energy barriers estimated. Besides the global minimum, there are three local minima within 2.0 kcal mol^{-1} of the global minimum characterized by specific ω and τ torsion values. The energy barriers for rotation about the "ω angle" are 16.4−18.8 kcal mol^{-1} and are a consequence of the double-bond character of the B−N bond as revealed by natural bond orbitals calculations. The "ω angle" and the ω rotation barrier are nearly the same as those seen in natural peptides. The τ rotation barriers (B−O bond) are relatively low because of the single-bond character of the B−O bond. Ala-BON, the Ala-dipeptide derived from NMB, has been constructed as a model peptide to study the conformational preferences about the ϕ and ψ torsion angles. The study reveals a strong preference for α-helix, type-II β-turn, 2.27 ribbon, and antiparallel β-sheet conformations, and mirror images of both type-II β-turn and 2.27 ribbon motifs whose ϕ and ψ values fall in the "disfavored regions" of the Ramachandran map. Thus, the replacement of the carbonyl group by B−OH retains the geometry and barrier around the "ω angle" and induces a strong preference for regular secondary structure motifs and also structures with positive ϕ values. This makes the B(OH)−NH analog an important surrogate for the peptide bond, with the additional advantage of stability to proteolytic enzymes.

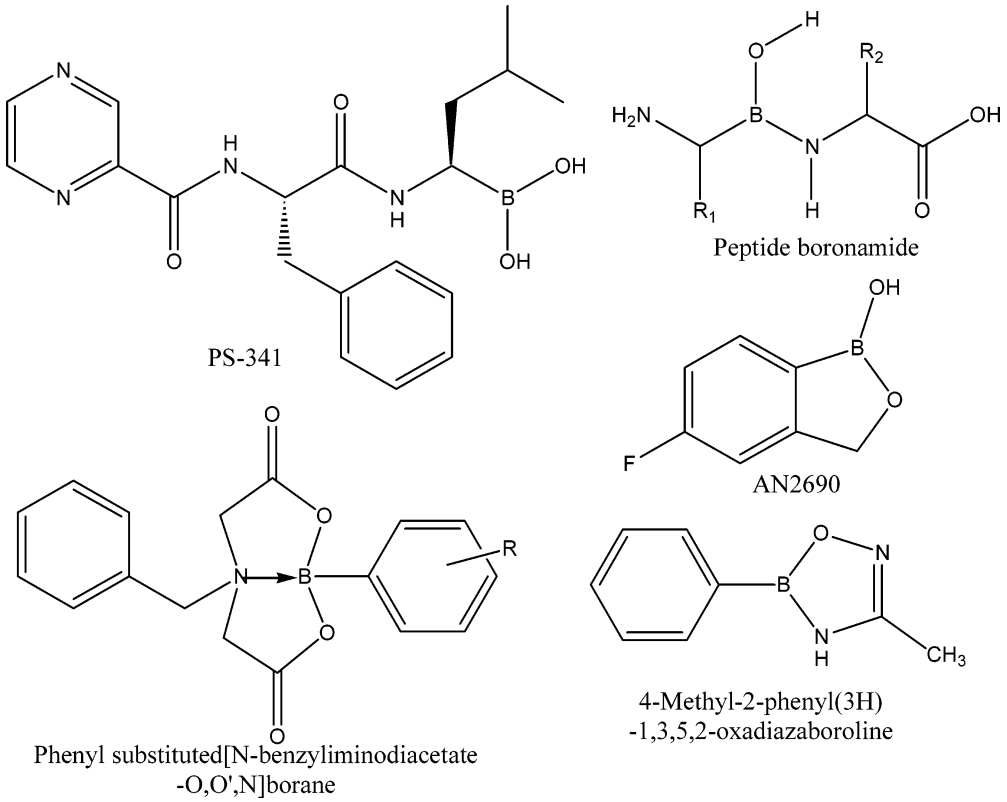
Introduction

The building blocks of most of the peptides and proteins are the 20 naturally occurring amino acids. Extensive experimental and theoretical studies have been carried out to understand the chemical and physical properties of these molecules. The use of physiologically relevant peptides in therapeutics is often hampered by their poor oral bioavailability.[1] Moreover, the pharmacological use of peptides or any drug demands high potency, good selectivity and specificity, metabolic stability, and low toxicity.[2] To achieve these demands, numerous modifications of the amino acid backbone and side chains have been suggested over the years. Side-chain modifications include the inversion of chirality of one or more key residues in the peptide,[3] introduction of a double bond between the Cα and Cβ atoms,[4] and moving the side chain from the Cα position to the N atom to form peptoids.[5] Some of the backbone modifications include elongation with β-, γ-, and δ-amino acids;[6] replacement of the amide bond[7] by sulfonamide, phosphonamidate, and carbamate; reduction of the amide bond;[8] *N*-methylation;[9] *N*-hydroxylation;[10] isosteric replacement of Cα by boron forming amino-carboxyboranes;[11] replacement of the amide NH by BH,[12,13] BOH,[14] or $BOCH_3$,[15] and replacement of the carbonyl carbon with boron, forming peptide boronic acid.[16]

* Corresponding author tel.: +91-22-26670871; fax: +91-22-26670816; e-mail: evans@bcpindia.org.

10.1021/ct600256s CCC: $37.00 © 2007 American Chemical Society
Published on Web 12/22/2006

Chart 1



In previous papers, we had reported the effects on the structure, electronic properties, and conformation of peptides when the amide nitrogen is replaced by boron.[12−15] The peptide boronic acids [$NH_2-AA_n-B(OH)_2$], with the carbonyl group replaced by B−OH in the C-terminal end, have been reported in the literature as one of the most potent serine protease inhibitors. The potential applications of such peptides include the inhibition of thrombin, proteasome, dipeptidyl peptidase IV, chymotrypsin, leucocyte elastase, pancreatic elastase, hepatitis C virus NS3 protease, and penicillin-binding protein.[16] The X-ray crystallography and NMR studies on peptide boronic acids reveal that peptide boronic acid inhibitors analogous to the substrate form a tetrahedral boron−serine adduct in the active site, while other boronic acids form boron−histidine adducts in the active site of serine protease.[17] The dipeptide boronic acid PS-341[16] (Chart 1) is a potent inhibitor of proteasome and exhibits anti-inflammatory activity, while AN2690[16] (Chart 1), an organic boronic acid derivative, is a potent antifungal agent and is under clinical trial for the treatment of onychomycosis topical treatment.

Peptide boronic acids reported in the literature have been designed by replacing carbon by boron at the C-terminal end of the peptide. Replacement of the carbonyl group of a residue other than the C-terminal, by boron, leads to peptide boronamides (Chart 1). Close analogs of these molecules are known; for example, phenyl-substituted [*N*-benzyliminodiacetate-O,O′,N]boranes (Chart 1) has been designed as a cytotoxic and anticancer drug with potential applications in boron neutron capture therapy for the treatment of tumors and melanomas. Besides this, 4-methyl-2-phenyl(3H-)1,3,5,2-oxadiazoboroline (Chart 1) containing the boronamide moiety has been reported in the literature.[17]

We have investigated the geometry, conformation, and electronic properties of peptide boronamides (Chart 1) by ab initio and density functional methods. *N*-methylacetamide (NMA, Figure 1) is well-established as a model for natural peptides and proteins.[18] *N*-methyl-methylboronamide (NMB, Figure 1) constructed by replacing CO in NMA with B−OH serves as a simple model to analyze the preferences for the "ω angle" in peptide boronamides. The alanine analog (Ala-BON, Figure 1), derived from Ala-dipeptide (Figure 1), has been adopted as a model to study the ϕ and ψ preferences of peptide boronamides. The hypersurface of NMB, with its associated ground and transition states, and the ground states of Ala-BON have been mapped by ab initio Hartree−Fock

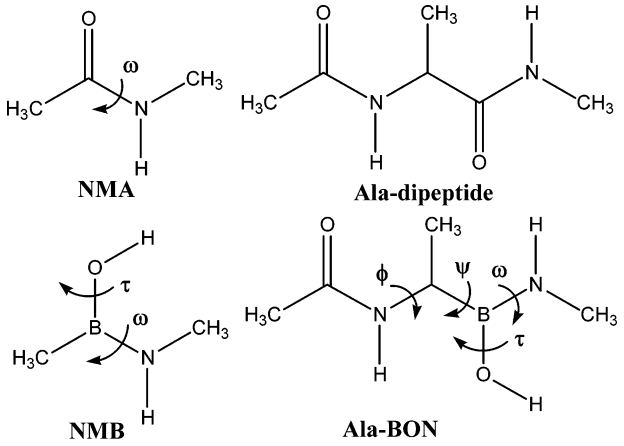


Figure 1. Structures of NMA, Ala-dipeptide, NMB, and Ala-BON.

Table 1. Energies (a.u.) and Relative Energies (kcal mol^{-1}) of Various Minima and Transition States on the PES of NMB at the HF, B3LYP, and MP2 Levels of Theory with the 6-31+G* Basis Set in the Aqueous Phase Using the PCM[a]

				HF/6-31+G*		B3LYP/6-31+G*		MP2(full)/6-31+G*			
		NIMAG	PG	a.u.[b]	rel.[c]	a.u.[b]	rel.[c]	a.u.[b]	rel.[c]	ω^d	τ^d
minima	GM	0	C_S	−234.3904417	0.0	−235.8715908	0.0	−235.109098	0.0	180	180
	LM1	0	C_S	−234.3898548	0.4	−235.8711296	0.3	−235.108948	0.1	0	180
	LM2	0	C_S	−234.3893319	0.7	−235.8706796	0.6	−235.107982	0.7	180	0
	LM3	0	C_S	−234.3879227	1.6	−235.8686787	1.8	−235.106071	1.9	0	0
ω rotation transition state (TS)	ωTS1	1	C_1	−234.3650603	15.9	−235.8455237	16.4	−235.082964	16.4	62/−62	−1
	ωTS2	1	C_1	−234.3621911	17.7	−235.8420428	18.5	−235.079209	18.8	113/−113	−1
	ωTS3	1	C_1	−234.3625781	17.5	−235.8427935	18.1	−235.079559	18.5	62/−62	179
	ωTS4	1	C_1	−234.3626095	17.5	−235.842479	18.3	−235.079838	18.4	118/−118	179
τ rotation TS	ωTS1	1	C_1	−234.3800096	6.5	−235.8611247	6.6	−235.097904	7.0	0	87/−87
	ωTS2	1	C_1	−234.3811566	5.8	−235.861951	6.0	−235.098847	6.4	183	88/−88

[a] NIMAG = number of imaginary frequency, PG = point group, GM = global minimum, LM = local minimum [b] Zero-point vibrational energy corrected values. [c] Relative energy in kcal mol^{-1}. [d] Torsion angle in degrees.

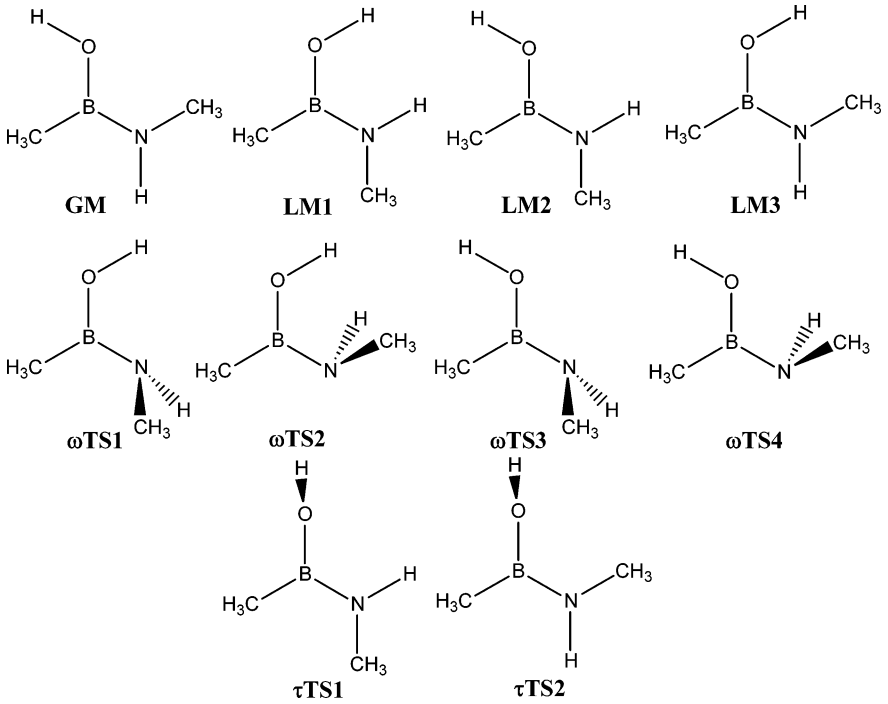


Figure 2. Structures of the minima and transition states on the PES of NMB.

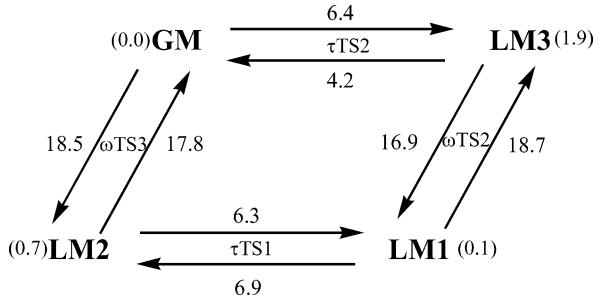


Figure 3. A schematic diagram showing interconversion between the various minima of NMB. The values on the arrows are energy barriers in kilocalories per mole, and the values in parentheses are relative energy values of minima in kilocalories per mole.

(HF), density functional, and post-HF methods. Second-order orbital interactions by the natural bond orbitals (NBO) method has also been carried out to understand the fundamental differences between the structures of peptide boronamides and their natural counterparts.

Computational Details

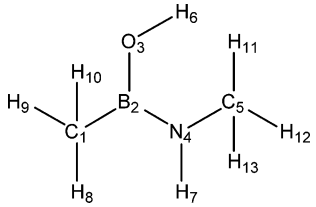
Ab initio molecular orbital[19] and density functional theory[20] calculations have been carried out with the Gaussian 03W (revision C.01)[21] package running on a 3.0 GHz Pentium IV processor with 1 GB of RAM. The stability of all wavefunctions was checked at the HF[22] and Becke's three-parameter exchange functional and the gradient-corrected functional of Lee, Yang, and Parr (B3LYP).[23] Second-order Møller−Plesset [MP2(full)][24] calculations were also carried out with the 6-31+G* basis set for all ground and transition state (TS) structures.

The two torsion angles, ω and τ, in *N*NMB are defined as shown in Figure 1. In NMB, the hydroxyl group can adopt two conformations around the B−O bond. In the first, the O−H bond is anti-periplanar with respect to the B−N bond, and the second arrangement is syn-periplanar. These initial

Table 2. Bond Length (Å) and Bond Angles (deg) of Minima and the TS of NMB Optimized at the MP2(full)/6-31+G* Level of Theory

parameter	GM	LM1	LM2	LM3	ω**TS1**	ω**TS2**	ω**TS3**	ω**TS4**	τ**TS1**	τ**TS2**
CB (1,2)	1.584	1.579	1.586	1.576	1.574	1.567	1.586	1.575	1.579	1.580
BO (2,3)	1.395	1.393	1.395	1.391	1.369	1.379	1.372	1.379	1.417	1.412
BN (2,4)	1.409	1.419	1.410	1.418	1.485	1.482	1.473	1.474	1.408	1.408
NC (4,5)	1.456	1.455	1.452	1.456	1.472	1.468	1.468	1.473	1.454	1.456
OH (3,6)	0.970	0.972	0.970	0.972	0.978	0.974	0.973	0.973	0.968	0.968
NH (4,7)	1.013	1.016	1.016	1.012	1.020	1.019	1.019	1.021	1.017	1.014
CBO (1,2,3)	121.4	116.4	122.4	116.8	118.3	117.4	121.4	122.4	118.3	119.5
CBN (1,2,4)	121.9	122.7	122.7	122.0	124.7	120.7	123.1	120.2	125.0	121.4
BOH (2,3,6)	112.3	114.2	112.6	114.2	108.1	112.2	112.2	112.8	116.5	116.9
BNC (2,4,5)	126.0	126.6	127.9	126.1	114.6	115.8	115.1	113.6	129.6	126.7
BNH (2,4,7)	118.8	120.0	117.0	118.6	111.6	114.3	112.5	111.0	115.9	118.4
CNH (5,4,7)	115.2	113.5	115.1	115.3	109.1	111.1	110.0	109.1	114.4	114.8
CBNC (ω) (1,2,4,5)	180.0	0.0	0.0	180.0	61.9/−61.9	114.0/−114.0	62.3/−62.3	118.8/−118.8	−0.3	−177.0
CBNH (1,2,4,7)	0.0	180.0	180.0	0.0	−62.6/62.6	−114.9/114.9	−64.9/64.9	−117.8/117.8	−175.7	0.9
NBOH (τ) (4,2,3,6)	180.0	0.0	180.0	0.0	−1.1	−1.3	179.2	178.8	87.0/−87.0	87.6/−87.6

Table 3. Partial Atomic Charges of NMB Calculated Using NPA and the "ESP Fit" as per the Merz−Singh−Kollman Scheme at the MP2(full)/6-31+G* Level and Using the PCM



atom	ESP fitted charges	NPA charges
C1	−0.745	−1.040
B2	0.849	1.179
O3	−0.850	−1.044
N4	−0.809	−1.030
C5	0.183	−0.409
H6	0.493	0.557
H7	0.384	0.448
H8	0.139	0.234
H9	0.160	0.238
H10	0.160	0.238
H11	0.031	0.225
H12	−0.002	0.203
H13	−0.002	0.203

two conformations around the B−O bond in NMB were chosen, and for each arrangement of τ, a scan in increments of 30° of the "ω torsion angle" was carried out at the HF, B3LYP, and MP2(full) levels of theory using the 6-31+G* basis set. Conformations with the ω values of 0° and 180° were found to be the lowest in energy. Now, for each conformation with a ω value of 0° or 180°, a τ scan in increments of 30° was run at the HF, B3LYP, and MP2-(full) levels of theory with the 6-31+G* basis set. The minima and saddle points for rotations about the ω and τ torsions were thus identified. All these conformations were further optimized, without constraints, at the B3LYP and MP2 levels of theory with the same basis set, and finally, the energies of the optimized structures were further calculated in the aqueous phase using the self-consistent reaction field (SCRF) polarizable continuum model (PCM). The stable conformations were confirmed by frequency calculations, which returned one imaginary frequency for each transition state and all positive frequencies for each ground state.

The NBO[25] analysis was carried out on the global minimum energy structure (GM) of NMB, optimized at the MP2(full)/6-31+G* level, to quantitatively estimate the second-order interactions as $E_{ij} = -2F_{ij}/\Delta E_{ij}$, where E_{ij} is the energy of the second-order interaction, $\Delta E_{ij} = E_i - E_j$ is the energy difference between the interacting molecular orbitals i and j, and F_{ij} is the Fock matrix element for the interaction between orbitals i and j. The "atomic partial charges" of the global minimum of NMB, optimized at the MP2(full)/6-31+G* level, were calculated using natural population analysis (NPA) as implemented in NBO and additionally by the "ESP fit" method formulated by Merz et al.[26]

For Ala-BON, the minima in the ϕ and ψ space was searched starting with the four different conformations for ω and τ identified (Table 1) previously for NMB. This corresponds to structures (i) $\omega = 180°$, $\tau = 180°$; (ii) $\omega = 0°$, $\tau = 180°$; (iii) $\omega = 180°$, $\tau = 0°$; and (iv) $\omega = 0°$, $\tau = 0°$ labeled as Ala-BON1, Ala-BON2, Ala-BON3, and Ala-BON4, respectively. For each (ω,τ) combination, 144 conformations were generated with 30° increments of the ϕ and ψ dihedrals. Each conformation was geometry-optimized first at the HF/3-21G level of theory with "constraints" on the initial ϕ and ψ angles. A Ramachandran map of the 144 conformations was constructed, and conformations within 10.0 kcal mol^{-1} of the global minimum were identified. These low-energy conformations were further optimized without constraints at the B3LYP/6-31+G* level of theory in the aqueous phase using the SCRF PCM.

Results and Discussion

All wavefunctions for molecules NMB and Ala-BON were found to be stable under the perturbations considered at the HF, B3LYP, and MP2 levels of theory.

Minima and Saddle Points of NMB. The conformations of the minima and various transition states for NMB are

Table 4. Torsion Angles (in degrees), Relative Energies (kcal mol^{-1}), and Secondary Structure Features of Minima of Ala-BON

	ω	τ	ϕ	ψ	rel. E (kcal mol^{-1})	structural feature (ideal values of the torsion angles)
Ala-BON1	180	180	61	129	0.0	type II′ β turn, second residue (60,120), positive ϕ
			−74	−122	0.9	type II β turn, second residue (−60,−120)
Ala-BON2	0	180	62	132	0.0	type II′ β turn, second residue (60,120), positive ϕ
Ala-BON3	180	0	−141	135	0.0	antiparallel β sheet (−139,135)
			68	−46	4.2	mirror image of 2.27 ribbon (78,−59), positive ϕ
Ala-BON4	0	0	−144	154	0.0	antiparallel β sheet (−139,135)
			−57	−46	1.7	α helix (−57,−47)

depicted in Figure 2, and their absolute and relative energies at the HF, B3LYP, and MP2(full) levels of theory with the 6-31+G* basis set and the PCM are given in Table 1. For NMB, besides the GM, there exist three local minima (LM1, LM2, and LM3). These local minima occur respectively at 0.1, 0.7, and 1.9 kcal mol^{-1} above the global minimum. Depending upon the pyramidal state of nitrogen (i.e., "pyramidal up" or "pyramidal down") and the relative orientation of the hydroxyl group (i.e., "τ angle"), there exist four transition states for rotation around the "ω angle" (ωTS1 to ωTS4). The natural peptide NMA exhibits two minima (global and local) and two transition states for the rotation around the "ω angle'.[27] The minima and transition states of NMB closely parallel those of NMA, with the exception that two additional minima and two additional transition states for rotation around the "ω angle" appear. This is a consequence of the hydroxyl group, which is absent in NMA. There are two transition states for rotation around the "τ angle", namely, τTS1 and τTS2. All minima of NMB exhibit C_S symmetry while the transition states have C_1 symmetry. The GM has $\omega = 180°$ and $\tau = 180°$, like the GM of NMA with $\omega = 180°$ and C_S symmetry.

Figure 3 depicts the interconversion between the four minima of NMB and the associated energy barriers. Although all three local minima lie within 2.0 kcal mol^{-1} of the global minimum, the barriers for interconversions are relatively high. The lowest barrier at 4.2 kcal mol^{-1} is associated with the conversion of LM3 to GM. The conversion of LM1 to LM3 has the highest barrier of 18.7 kcal mol^{-1}. The GM ↔ LM3 and LM1 ↔ LM2 interconversions exhibit relatively low energy barriers, while the GM ↔ LM2 and LM1 ↔ LM3 interconversions involve higher energy barriers. The interconversion between GM ↔ LM1 and LM2 ↔ LM3 follows a two-step process or involves a saddle point of index 2.

The geometric parameters of the minima and the transition states of NMB at the MP2(full)/6-31+G* level of theory are given in Table 2. There is a small increase of about 0.06−0.07 Å in the B−N bond length in the transition states for rotation about the "ω angle" compared to the ground states. Crystallographic data is available for a large number of "organic boronic acids" [R−B(OH)$_2$] and "peptide boronic acids" [AA−CONH−C(R)H−B(OH$_2$)], but no experimental data are yet available for boronamides. However, the crystal structure[28] of a cyclic boron derivative, 4-methyl-2-phenyl-(3H)-1,3,5,2-oxadiazaboroline (Chart 1), which is related to NMB has been reported. The B−O bond length in NMB is found to be 1.391−1.395 Å (Table 2) at the MP2(full)/6-31+G* level of theory. In comparison, the B−O bond length, reported in the literature for a range of organic and inorganic boron-containing molecules, varies from 1.34 to 1.42 Å (average 1.38 Å) for boron with trigonal planar geometry and 1.39 to 1.52 Å (average 1.48 Å) for boron with tetrahedral geometry.[29] In 4-methyl-2-phenyl-(3H)-1,3,5,2-oxadiazaboroline, the B−O bond length is 1.39 Å. Similarly, the B−N bond length of NMB is found to be 1.409−1.419 Å, while that in 4-methyl-2-phenyl-(3H)-1,3,5,2-oxadiazaboroline is 1.41 Å. The B−O and B−N bond lengths signify appreciable π character for these bonds. The C−B bond length of NMB is found to be 1.576−1.586 Å, while that in 4-methyl-2-phenyl-(3H)-1,3,5,2-oxadiazaboroline is 1.56 Å. In alkylboranes, the B−C (aliphatic carbon) bond length is about 1.590 Å, as is that for, for example, dimethylborane,[30] and it is 1.596 Å in dimesitylborane,[31] 1.570 Å in ditripylborane,[32] 1.571 Å in BMA (acetylmethylborane),[13] 1.564 Å in BMAOH (acetylmethylhydroxyborane),[14] and 1.612 Å in BMA−BOM (acetylmethylmethoxyborane).[15] The bond lengths for NMB calculated in this study are in the range of the experimental values for related compounds.

Rotation Barriers in NMB. The barrier to rotation about the "ω angle" in the natural peptide ranges from 16.0 to 25.0 kcal mol^{-1}. The "ω rotation" barrier in NMB at the MP2-(full)/6-31+G* level of theory is found to be 16.4−18.8 kcal mol^{-1}, which is comparable to the value in natural peptides. In sharp contrast, replacement of the nitrogen of the amide group by boron leads to a lowering (∼5.0 kcal mol^{-1}) of the "ω rotation" barrier in boron peptides.[13] The results of the NBO study reveal that the energy $E^{(2)}$ associated with $n_O \rightarrow \pi^*_{B-N}$ interaction (delocalization from lone pairs on oxygen into the π antibonding orbital of the B−N bond) is 51.4 kcal mol^{-1} and that for the $n_O \rightarrow \sigma^*_{B-N}$ interaction (delocalization from lone pairs on oxygen into the σ antibonding orbital of the B−N bond, i.e., negative hyperconjugation) is 3.6 kcal mol^{-1} in the GM of NMB at the MP2(full)/6-31+G* level of theory. In comparison, the energy associated with negative hyperconjugation, that is, $n_O \rightarrow \sigma^*_{C-N}$, is 32.8 kcal mol^{-1}, and that for $n_N \rightarrow \pi^*_{C-O}$ is 98.5 kcal mol^{-1} for the global energy minimum structure of NMA.[13] These two interactions are responsible for the high barrier to rotation about the C−N and the B−N bonds in NMA and NMB, respectively. The NBO analysis also yields a bond order of 2 for the B−N bond in NMB, which also explains the existence of the high ω rotation barrier in NMB. The barrier to τ rotation (B−O bond) in NMB is found to

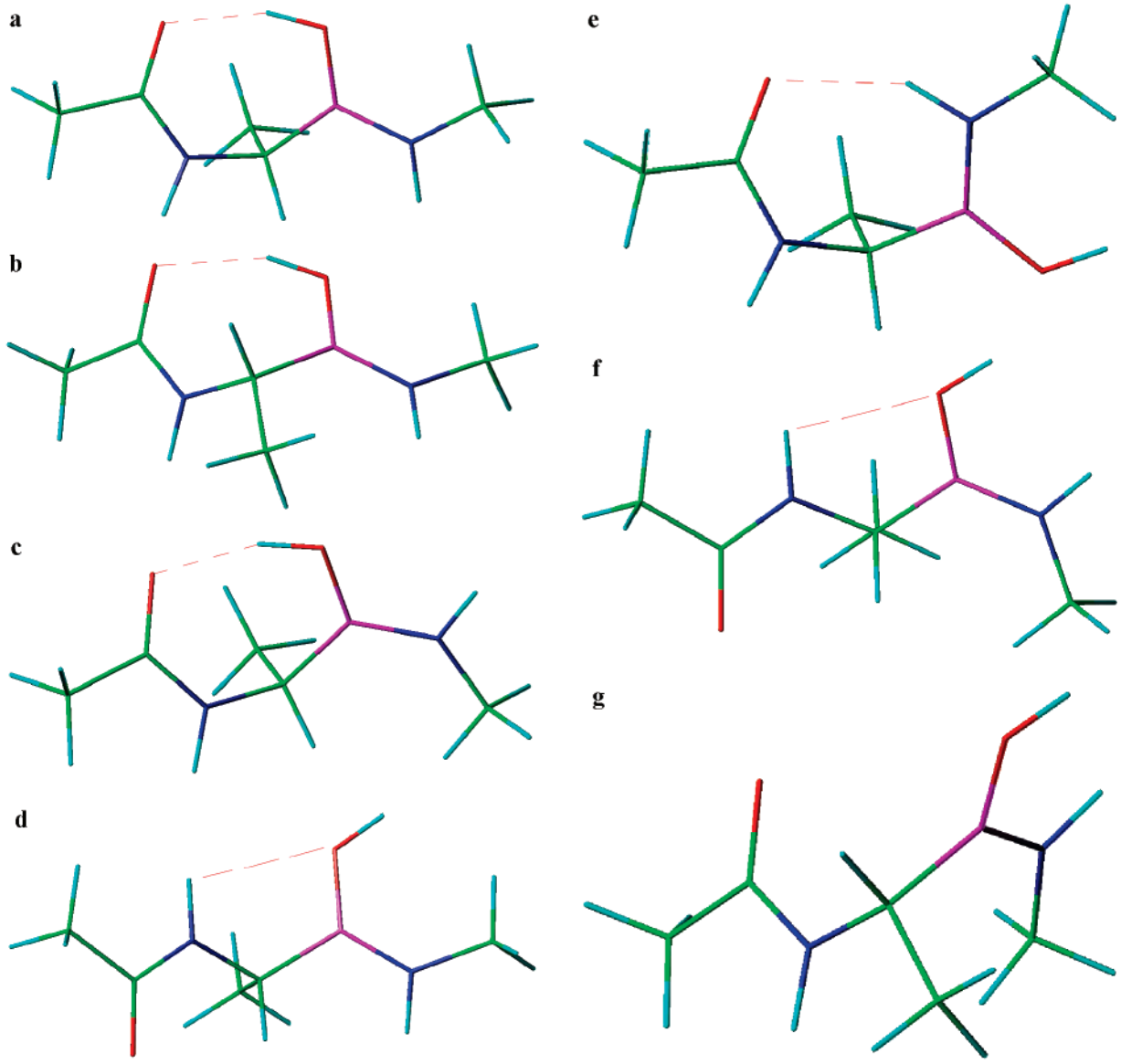


Figure 4. Preferred conformations of Ala-BON. Ala-BON1: (a) $\omega = 180°$, $\tau = 180°$, $\phi = 61°$, $\psi = 129°$ and (b) $\omega = 180°$, $\tau = 180°$, $\phi = -74°$, $\psi = -122°$. Ala-BON2: (c) $\omega = 0°$, $\tau = 180°$, $\phi = 62°$, $\psi = 132°$. Ala-BON3: (d) $\omega = 180°$, $\tau = 0°$, $\phi = -141°$, $\psi = 135°$ and (e) $\omega = 180°$, $\tau = 0°$, $\phi = 68°$, $\psi = -46°$. Ala-BON4: (f) $\omega = 0°$, $\tau = 0°$, $\phi = -144°$, $\psi = 154°$ and (g) $\omega = 0°$, $\tau = 0°$, $\phi = -57°$, $\psi = -46°$. Atom color codes: carbon, green; nitrogen, blue; oxygen, red; boron, magenta; hydrogen, cyan.

be between 6.4 and 7.0 kcal mol^{-1} at the MP2(full)/6-31+G* level of theory. The NBO results indicate a single bond for B−O, which is in line with the relatively low rotation barrier.

Partial Atomic Charges of NMB. The "natural charges" derived from NPA and ESP-fitted charges as per the Merz−Singh−Kollman scheme for the global minimum energy structure of NMB are listed in Table 3. The large positive charge on boron will make it the favored site for a nucleophilic attack leading ultimately to the formation of a stable tetrahedral adduct. The boronamides thus follow a chemistry similar to that of the peptide boronic acids, which are potent inhibitors of serine protease.[16] The peptide boronamides are therefore expected to act as strong inhibitors of serine protease.

Potential Energy Surface (PES) of Ala-BON. The four minima of NMB for the ω and τ torsions give rise to four different starting structures (Ala-BON1 to -4), whose minima in the ϕ and ψ space were scanned (Table 4). In Ala-BON1, with an ω value of 180° and a τ value of 180°, the global minimum corresponds to a structure with $\phi = 61°$ and $\psi = 129°$ (Figure 4a). These values are close to those for the second residue in a type-II′ β turn (ideal values $\phi = 60°$ and $\psi = 120°$). The local minimum energy structure with $\phi = -74°$ and $\psi = -122°$ (Figure 4b) corresponds to a β turn of type II (ideal values $\phi = -60°$ and $\psi = -120°$). In Ala-BON2, with an ω value of 0° and a τ value of 180°, there is only the global minimum energy structure with $\phi = 62°$ and $\psi = 132°$ (Figure 4c), which is very similar to the global minimum of Ala-BON1 with regard to τ, ϕ, and ψ angles but with a difference in the "ω angles". Both structures exhibit a positive ϕ value which is not favored by natural peptides. All minima of Ala-BON1 and Ala-BON2 have an intramolecular hydrogen bond (Figure 4a−c) between the

carbonyl O atom (acting as an acceptor) and the OH group attached to boron (acting as a donor).

Ala-BON3, with an ω value of 180° and a τ value of 0°, has one local minimum within 10.0 kcal mol^{-1} of the global minimum structure. The global minimum is a structure with $\phi = -141°$ and $\psi = 135°$ (Figure 4d) corresponding to an antiparallel β sheet (ideal values $\phi = -139°$ and $\psi = 135°$). The local minimum is a structure with $\phi = 68°$ and $\psi = -46°$ (Figure 4e) which is close to a 2.27 ribbon structure (ideal $\phi = -78°$ and $\psi = 59°$), and the ϕ value falls in the "disallowed region" of the Ramachandran map. The global minimum of Ala-BON3 exhibits a hydrogen bond (Figure 4d) between the amide N atom (acting as an acceptor) and the OH group attached to boron (acting as a donor). The local minimum structure has a hydrogen bond (Figure 4e) between the carbonyl O atom (acting as an acceptor) and the NH group attached to boron (acting as a donor).

The last structure, Ala-BON4, with an ω value of 180° and a τ value of 0°, also has one local minimum within 10.0 kcal mol^{-1} of the global minimum structure. The global minimum corresponds to a structure with $\phi = -144°$ and $\psi = -154°$ (Figure 4f) and resembles an antiparallel β sheet ($\phi = -139°$, $\psi = 135°$). The local minimum is a structure with $\phi = -57°$ and $\psi = -46°$ (Figure 4g), and these values are for an α helix ($\phi = -57°$, $\psi = -47°$). In Ala-BON4, only the global minimum displays a hydrogen bond (Figure 4f) between the amide N atom (acting as an acceptor) and the OH group attached to boron (acting as a donor).

The conformational preferences of natural peptides are governed by their ϕ and ψ values, which often lie within the "allowed regions" of the Ramachandran map. The replacement of CO by B−OH in natural peptides leads to structures with regular secondary motifs, in addition to few secondary structures with "disfavored" positive ϕ values, for, for example, mirror images of the type-II β-turn and the 2.27 ribbon motifs.

Conclusions

A novel boron analog of peptides has been designed by replacing the amide (CO−NH) moiety with the B(OH)−NH group. To understand the geometrical properties of these analogs around the ω angle, NMB has been established as a model peptide. The PES of NMB is characterized by four minima, four TSs for rotation around the "ω angle", and two TSs for rotation about the τ angle. The barriers for rotation about the "ω angle" are in the range of 16.4−18.8 kcal mol^{-1}, and those for rotation about the τ angle are 6.4−7.0 kcal mol^{-1}. The NBO study reveals that the B−N bond has appreciable double-bond character, and single-bond character for the B−O bond, which explains the relative differences in the rotation barriers of the ω and τ angles. The "ω angle" and its associated rotation barrier are remarkably similar to the corresponding values of natural peptides. The Ala-dipeptide derived from NMB favors α-helix, type-II β-turn, 2.27 ribbon, and antiparallel β-sheets conformations and structures which are mirror images of the type-II β-turn and the 2.27 ribbon motifs. These latter structures fall in the "disallowed regions" of the Ramachandran map. In conclusion, the replacement of CO by B−OH in peptides retains the backbone geometry in a trans configuration. This makes the B−OH moiety a good surrogate for the CO group in peptides, with the additional property that such analogs will be resistant to hydrolytic cleavage by the enzymes—amidases and peptidases.

Acknowledgment. This work is supported by the Department of Science and Technology (DST), New Delhi, through their FIST program (SR/FST/LS1−163/2003). A.K.M. and S.A.K. thanks the Council of Scientific and Industrial Research (CSIR), New Delhi, financial support.

References

(1) Gante, J. Peptidomimetics − Tailor-Made Enzyme Inhibitors. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1699−1720.

(2) Hruby, V. J. Designing Peptide Receptor Agonists and Antagonists. *Nat. Rev. Drug Discovery* **2002**, *1*, 847−858.

(3) Fischer, P. M. The Design, Synthesis and Application of Stereochemical and Directional Peptide Isomers: A Critical Review. *Curr. Protein Pept. Sci.* **2003**, *4*, 339−356.

(4) Thormann, M.; Hoffmann, H. J. Conformational Properties of Peptides Containing Dehydro Amino Acids. *THEOCHEM* **1998**, *431*, 79−96.

(5) Mohle, K.; Hoffmann, H. J. Secondary Structure Formation in N-Substituted Peptides. *J. Pept. Res.* **1998**, *51*, 19−28.

(6) (a) Cheng, R. P.; Gellman, S. H.; DeGrado, W. F. Beta-Peptides: From Structure to Function. *Chem. Rev.* **2001**, *101*, 3219−3232. (b) Hanessian, S.; Luo, X. H.; Schaum, R.; Michnick, S. Design of Secondary Structures in Unnatural Peptides − Stable Helical γ-Tetra-, Hexa- and Octapeptides and Consequences of α-Substitution. *J. Am. Chem. Soc.* **1998**, *120*, 8569−8570. (c) Grison, C.; Geneve, S.; Coutrot, P. Enantioselective Synthesis of α,β-Unsaturated γ- and δ-Lactams. *Tetrahedron Lett.* **2001**, *42*, 3831−3834.

(7) (a) Cho, C. Y.; Moran, E. J.; Cherry, S. R.; Stephans, J. C.; Fodor, S. P. A.; Adams, C. L.; Sundaram, A.; Jacobs, J. W.; Schultz, P. G. An Unnatural Biopolymer. *Science* **1993**, *261*, 1303−1305. (b) Rushing, S. D.; Hammer, R. P. Synthesis of Phosphonamide and Thiophosphonamide Dipeptides. *J. Am. Chem. Soc.* **2001**, 123, 4861−4862. (c) Baldauf, C.; Günther, R.; Hoffmann, H. J. Conformational Properties of Sulfonamido Peptides. *THEOCHEM* **2004**, *675*, 19−28.

(8) Cervini, L.; Theobald, P.; Corrigan, A.; Craig, A. G.; Rivier, C.; Vale, W.; Rivier, J. Corticotropin Releasing Factor (CRF) Agonists with Reduced Amide Bonds and Ser7 Substitutions. *J. Med. Chem.* **1999**, *42*, 761−768.

(9) Vogen, S. M.; Paczkowski, N. J.; Kirnarsky, L.; Short, A.; Whitmore, J. B.; Sherman, S. A.; Taylor, S. M.; Sanderson, S. D. Differential Activities of Decapeptide Agonists of Human C5a: The Conformational Effects of Backbone *N*-Methylation. *Int. Immunopharmacol.* **2001**, *12*, 2151−2162.

(10) Ye, Y.; Liu, M.; Kao, J. L.; Marshall, G. R. Peptide-Bond Modification for Metal Coordination: Peptides Containing Two Hydroxamate Groups. *Biopolymers* **2003**, *71*, 489−515.

(11) (a) Spielvogel, B. F.; Das, M. K.; McPhail, A. T.; Onam, K. D.; Hall, I. H. Boron Analogs of the α-Amino Acids. Synthesis, X-Ray Crystal Structure, and Biological Activity of Ammonia-Carboxyborane, the Boron Analog of Glycine.

J. Am. Chem. Soc. **1980**, *102*, *6343*−6344. (b) Miller, M. C.; Sood, A.; Spielvogel, B. F.; Hall, I. H. Synthesis and Antitumor Activity of Boronated Dipeptides Containing Aromatic Amino Acids. *Anticancer Res.* **1997**, *5A*, 3299−3306.

(12) Datar, P. A.; Coutinho, E. C. The ϕ, ψ Space of Boron Isosteres of Amino Acids: An *ab Initio* Study. *J. Theor. Comput. Chem.* **2004**, *3*, 189−202.

(13) Malde, A. K.; Khedkar, S. A.; Coutinho, E. C.; Saran, A. Geometry, Transition States, and Vibrational Spectra of Boron Isostere of *N*-Methylacetamide by ab Initio Calculations. *Int. J. Quantum Chem.* **2005**, *102*, 734−742.

(14) Malde, A. K.; Khedkar, S. A.; Coutinho, E. C. The ω, ϕ and ψ Space of *N*-Hydroxy-*N*-methylacetamide and *N*-Acetyl-*N*′-hydroxy-*N*′-methylamide of Alanine and Their Boron Isosteres. *J. Chem. Theory Comput.* **2006**, *2*, 312−321.

(15) Malde, A. K.; Khedkar, S. A.; Coutinho, E. C. The Stationary Points of the PES of N-Methoxy Peptides and Their Boron Isosteres: An *ab Initio* Study. *J. Chem. Theory Comput.* **2006**, *2*, 1664−1674.

(16) (a) Kettner, C.; Mersinger, L.; Knabb, R. The Selective Inhibition of Thrombin by Peptides of Boroarginine. *J. Biol. Chem.* **1990**, *265*, 18289−18297. (b) Archer, S. J.; Camac, D. M.; Wu, Z. J.; Farrow, N. A.; Domaille, P. J.; Wasserman, Z. R.; Bukhtiyarova, M.; Rizzo, C.; Jagannathan, S.; Mersinger, L. J.; Kettener, C. A. Hepatitis C Virus NS3 Protease Requires Its NS4A Cofactor Peptide for Optimal Binding of a Boronic Acid Inhibitor as Shown by NMR. *Chem. Biol.* **2002**, *9*, 79−92. (c) Pargellis, C. A.; Campbell, S. J.; Pav, S.; Graham, E. T.; Pitner, T. P. Inhibition of Dipeptidyl Petidase IV (CD26) by Peptide Boronic Acid Dipeptides. *J. Enzyme Inhib.* **1997**, *11*, 151−169. (d) Nicola, G.; Peddi, S.; Stefanova, M.; Nicholas, R. A.; Gutheil, W. G.; Davies, C. Crystal Structure of Escherichia coli PBP 5 Bound to a Tripeptide Boronic Acid Inhibitor: A Role for Ser110 in Deacylation. *Biochemistry* **2005**, *44*, 8207−8217. (e) Richardson, P. G.; Mitsiades, C.; Hideshima, T.; Anderson, K. C. Proteasome Inhibition in the Treatment of Cancer. *Cell Cycle* **2005**, *4*, 290−296. (f) Priestley, E. S.; De Lucca, I.; Ghavimi, B.; Erickson-Viitanen, S.; Decicco, C. P. P1 Phenethyl Peptide Boronic Acid Inhibitors of HCV NS3 Protease. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3199−3202. (g) Tapparelli, C.; Metternich, R.; Ehrhardt, C.; Zurini, M.; Claeson, G.; Scully, M. F.; Stone, S. R. In Vitro and in Vivo Characterization of a Neutral Boron-Containing Thrombin Inhibitor. *J. Biol. Chem.* **1993**, *268*, 4734−4741. (h) Soskel, N. T.; Watanabe, S.; Hardie, R.; Shenvi, A. B.; Punt, J. A.; Kettner, C. A New Peptide Boronic Acid Inhibitor of Elastase-Induced Lung Injury in Hamsters. *Am. Rev. Respir. Dis.* **1986**, *133*, 639−642. (i) Palombella, V. J.; Conner, E. M.; Fuseler, J. W.; Destree, A.; Davis, J. M.; Laroux, F. S.; Wolf, R. E.; Huang, J.; Brand, S.; Elliott, P. J.; Lazarus, D.; McCormack, T.; Parent, L.; Stein, R.; Adams, J.; Grisham, M. B. Role of the Proteasome and NF-kB in Streptococcal Cell Wall-Induced Polyarthritis. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 15671−15676. (j) Baker, S. J.; Zhang, Y. K.; Akama, T.; Lau, A.; Zhou, H.; Hernandez, V.; Mao, W.; Alley, M. R. K.; Sanders, V.; Plattner, J. J. Discovery of a New Boron-Containing Antifungal Agent, 5-Fluoro-1,3-dihydro-1-hydroxy-2,1-benzoxaborole (AN2690), for the Potential Treatment of Onychomycosis. *J. Med. Chem.* **2006**, *49*, 4447−4450.

(17) (a) Farr-Jones, S.; Smith, S. O.; Kettner, C. A.; Griffin, R. G.; Bachovchin, W. W. Crystal versus Solution Structure of Enzymes: NMR Spectroscopy of a Peptide Boronic Acid-Serine Protease Complex in the Crystalline State. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6922−6924. (b) Ivanov, D.; Bachovchin, W. W.; Redfield, A. G. Boron-11 Pure Quadrupole Resonance Investigation of Peptide Boronic Acid Inhibitors Bound to Alpha-Lytic Protease. *Biochemistry* **2002**, *41*, 1587−1590. (c) Davis, J. H.; Agard, D. A.; Handel, T. M.; Basus, V. J. Alterations in Chemical Shifts and Exchange Broadening upon Peptide Boronic Acid Inhibitor Binding to Alpha-Lytic Protease. *J. Biomol. NMR* **1997**, *10*, 21−27. (d) Tsilikounas, E.; Kettner, C. A.; Bachovchin, W. W. Boron-11 NMR Spectroscopy of Peptide Boronic Acid Inhibitor Complexes of α-Lytic Protease. Direct Evidence for Tetrahedral Boron in Both Boron-Histidine and Boron-Serine Adduct Complexes. *Biochemistry* **1993**, *32*, 12651−12655. (e) Mancilla, T.; Zamudio-Rivera, L. S.; Hiram; Beltran, I.; Santillan, R.; Farfan, N. Synthesis and Characterization of New (N→*B)* Phenyl Substituted [N-Benzyliminodiacetate-O,O′,N]boranes. *ARKIVOC (Gainesville, FL, U.S.)* **2005**, *6*, 366−376.

(18) Shipman, L. L.; Christoffersen, R. E. Ab Initio Calculations on Large Molecules Using Molecular Fragments. Model Peptide Studies. *J. Am. Chem. Soc.* **1973**, *95*, 1408−1416.

(19) Hehre, W. J.; Random, L.; Schlyer, P. V. R.; Pople, J. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1985.

(20) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; O.U.P.: New York, 1989.

(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.01; Gaussian, Inc.: Wallingford, CT, 2004.

(22) Roothan, C. C. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.

(23) (a) Becke, A. D. Density Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652. (b) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789. (c) Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13244−13249.

(24) (a) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618−622. (b) Martin-Head, G.; Pople, J. A.; Frisch, M. J. MP2 Energy Evaluation by Direct Methods. *Chem. Phys. Lett.* **1988**, *153*, 503−506.

(25) (a) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO*, version 3.1. (b) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, *735*−746. (c) Reed, A. E.; Weinhold, F.; Curtiss, L. A. Intermolecular Interactions from a Natural Bond Orbital, Donor-Acceptor Viewpoint. *Chem. Rev.* **1988**, *88*, 899−926.

(26) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.

(27) Villani, V.; Alagona, G.; Ghio, C. Ab Initio Studies on N-Methylacetamide. *Mol. Eng.* **1999**, *8*, 135−153.

(28) Raper, E. S. The Crystal and Molecular Structure of 4-Methyl-2-phenyl-(3H)-1,3,5,2-oxadiazaboroline. *Acta Crystallogr., Sect. B* **1978**, *34*, 3281−3284.

(29) (a) Filatov, S.; Shepelev, Y.; Bubnova, R.; Sennova, N.; Egorysheva, A. V.; Kargin, Y. F. The Study of $Bi_3B_5O_{12}$: Synthesis, Crystal Structure and Thermal Expansion of Oxoborate $Bi_3B_5O_{12}$. *J. Solid State Chem.* **2004**, *177*, 515−522. (b) Shishkov, I. F.; Khristenko, L. V.; Rudakov, F. M.; Vilkov, L. V.; Karlov, S. S.; Zaitseva, G. S.; Samdal, S. The Molecular Structure of Boratrane Determined by Gas Electron Diffraction and Quantum Mechanical Calculations. *J. Mol. Struct.* **2002**, *641*, 199−205.

(30) Vijay, A.; Sathyanarayana, D. N. Effects of Basis Set and Electron Correlation on the Structure and Vibrational Spectra of Diborane. *J. Mol. Struct.* **1995**, *351*, 215−229.

(31) Entwistle, C. D.; Marder, T. B.; Smith, P. S.; Howard, A. K.; Fox, M. A.; Mason, S. A. Dimesitylborane Monomer−Dimer Equilibrium in Solution, and the Solid-State Structure of the Dimer by Single Crystal Neutron and X-ray Diffraction. *J. Organomet. Chem.* **2003**, *680*, 165−172.

(32) Bartlett, R. A.; Rasikadis, H. V.; Olmstead, M. M.; Power, P. P.; Weese, K. J. Synthesis of the Monomeric $HBtrip_2$ (trip = 2,4,6-iso-$Pr_3C_6H_2$) and the X-ray Crystal Structures of $[HBMes_2]_2$ (Mes = 2,4,6,-$Me_3C_6H_2$) and $HBtrip_2$. *Organometallics* **1990**, *9*, 146−150.

CT600256S

JCTC Journal of Chemical Theory and Computation

Automated Parametrization of Biomolecular Force Fields from Quantum Mechanics/Molecular Mechanics (QM/MM) Simulations through Force Matching

Patrick Maurer, Alessandro Laio,† Håkan W. Hugosson,‡ Maria Carola Colombo, and Ursula Rothlisberger*

École Polytechnique Fédérale de Lausanne (EPFL), Institute of Chemical Sciences and Engineering, BCH-LCBC, CH-1015 Lausanne, Switzerland

Received September 14, 2006

Abstract: We introduce a novel procedure to parametrize biomolecular force fields. We perform finite-temperature quantum mechanics/molecular mechanics (QM/MM) molecular dynamics simulations, with the fragment or moiety that has to be parametrized being included in the QM region. By applying a force-matching algorithm, we derive a force field designed in order to reproduce the steric, electrostatic, and dynamic properties of the QM subsystem. The force field determined in this manner has an accuracy that is comparable to the one of the reference QM/MM calculation, but at a greatly reduced computational cost. This allows calculating quantities that would be prohibitive within a QM/MM approach, such as thermodynamic averages involving slow motions of a protein. The method is tested on three different systems in aqueous solution: dihydrogenphosphate, glycyl−alanine dipeptide, and a nitrosyl−dicarbonyl complex of technetium(I). Molecular dynamics simulations with the optimized force field show overall excellent performance in reproducing properties such as structures and dipole moments of the solutes as well as their solvation pattern.

1. Introduction

Molecular dynamics (MD) simulations using empirical force fields have become a standard tool to investigate the structure and dynamics of biological systems, such as proteins, nucleic acids, and membranes.[1] A variety of force fields tailored for biomolecular applications have been developed over recent decades, and they are still being continuously improved[2−5] (for a recent detailed overview, see refs 6 and 7).

These force fields provide parameters for simulating standard proteins, nucleic acids, lipids, some cofactors, and a number of solvents, so that a large variety of biological systems can routinely be studied. Often, however, one would like to simulate systems containing molecules or moieties for which force field parameters are not available. Examples include drug/receptor complexes or enzymes and nucleotide sequences that bind to a metal or contain special chemical modifications. The development of a reliable force field for these systems can be a cumbersome and time-consuming task. A common approach consists in deriving parameters from an electronic-structure calculation of the molecule or of a model compound in the gas phase. Atomic point charges are obtained by fitting to the quantum electrostatic potential (ESP) estimated on a grid surrounding the molecule.[8] The calculation of vibrational properties provides force constants for bonded interactions. Since these calculations are usually performed in the gas phase, polarization effects in the condensed phase are accounted for in a somewhat arbitrary—but nevertheless successful—fashion by using a basis set (6-31G*) that is known to overestimate polarization.[8] The resulting parameters need to be tested against experimental data and, if necessary, readjusted.

* Corresponding author phone: ++41 (0)21 693 0325; fax: ++41 (0)21 693 0320; e-mail: ursula.roethlisberger@epfl.ch.

† Current address: International School for Advanced Studies (SISSA), I-34014 Trieste, Italy.

‡ Current address: Laboratory of Theoretical Chemistry, Royal Institute of Technology, S-10691 Stockholm, Sweden.

10.1021/ct600284f CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/20/2007

Alternatively, parametrization can be completely circumvented by performing hybrid quantum mechanical/molecular mechanical (QM/MM) simulations.[9−13] In QM/MM simulations, the system is partitioned into a core region (the QM region), which is treated with a quantum chemical method, usually density functional theory (DFT) or semiempirical approaches, and an environment (the MM region) described with a classical force field. The QM region can be chosen in such a way that it includes all components of the system for which no parametrization is available. The QM/MM approach allows for an accurate description of the QM region under the steric and electrostatic influence of the environment. However, the high degree of accuracy is achieved at the price of a significant computational cost.

Ideally, one would like to perform simulations with the accuracy of a QM/MM treatment at the computational cost of classical MD. With this goal in mind, we propose here to exploit QM/MM simulations for deriving a force field designed in such a way as to reproduce the steric, electrostatic, and dynamic properties of the QM subsystem. In the QM/MM implementation used in this work,[13] the charge density of the quantum region is explicitly polarized by the electric field generated by the atomic point charges of the classical environment. In addition, effects such as temperature and pressure are automatically taken into account.

The force field determined in this manner can be used to efficiently sample the configurational space without the severe limitations of the simulation time of first-principles-based QM/MM schemes. This allows the calculation of, at a low computational cost, quantities that would be prohibitive within a QM/MM approach, such as, for the ligand docking problem, thermodynamic averages involving slow motions of the target protein. The potential constructed by this method is tailored for the long-time propagation of a specific system. As a consequence, the transferability of this tailored force field is expected to be limited. For example, the parameters determined with this method for the same solute solvated in water or acetone might be different and not necessarily transferable between the two cases. This approach resembles the optimal potential method[14] and the "learn-on-the-fly'' approach,[15] introduced a few years ago for solid-state physics applications. A more transferable potential can be obtained by choosing several reference systems, with, for example, different solvents or different protein environments.

As one of the reviewers of this article pointed out, using conformations from QM/MM simulations to generate ESP-derived charges has an additional advantage over the traditional use of gas-phase optimized structures. Especially in the case of highly polar species, gas-phase optimized structures are usually "closed-up", for example, as a result of the formation of intramolecular hydrogen bonds, which leads to nontransferable charges. The explicit hydration of such species allows for breaking intramolecular hydrogen bonds, resulting in more "open" conformations and more transferable charges.

In the work presented here, the parameters are derived exploiting the so-called force-matching technique. This approach dates back to the work of Ercolessi and Adams,[16] who developed an empirical potential for aluminum by fitting the parameters so as to reproduce a set of forces obtained from ab initio calculations. Force matching has since been applied with great success to parametrize empirical potentials for several systems of increasing complexity. Examples include Fe,[14] Si/SiO_2,[15,17,18] tantalum,[19] alkaline earth oxides,[20,21] and H_2O.[22]

Here, we apply force matching to parametrize a standard, nonpolarizable, biomolecular force field, without any modifications of the functional form. We optimize all of the charges and torsional, bonding, and bending parameters. The Lennard-Jones parameters are not optimized but are kept fixed to the standard force field value. This choice is consistent with the QM/MM interaction Hamiltonian we use that retains the Lennard-Jones parameters from the classical force field.[13] Leaving the functional form unchanged allows for a seamless integration of the fitted parameters for a subregion of the system into an accurate and well-tested existing set of parameters for the surrounding protein. The choice of a nonpolarizable force field has been motivated by the following reasons. Although there is consensus in the field that the explicit inclusion of polarization plays an important role in the development of more accurate force fields,[6,7] and considerable success has already been achieved,[23−32] such force fields cannot yet be considered standard methods. Using a nonpolarizable force field provides an accuracy benchmark against which more sophisticated models can be compared. The force-matching procedure can thus be used to identify situations where a higher-level force field leads to an important improvement, as opposed to situations where a minor improvement would not justify the additional computational cost.

In this work, the fit is performed using finite-temperature QM/MM trajectories of a few picoseconds. During this relatively short simulation time, it is unlikely to observe a dihedral transition in the QM subsystem. Therefore, the procedure described in this work provides a parameter set that can be safely used only for a specific conformer of the system. It is possible to generate a force field that also reproduces torsional barriers if the QM/MM dynamics are performed under the action of a bias potential that induces transitions in the available computational time.[33−36]

We will show that, after reparametrization, a standard, nonpolarizable, biomolecular force field performs in fact remarkably well in reproducing QM/MM results even for an electronically complex compound such as a nitrosyl−dicarbonyl complex of technetium(I).

2. Methods

The parametrization of a force field for a subregion of a system is carried out in three steps: First, a QM/MM simulation at finite temperature is performed, with the fragment to be parametrized being included in the QM subsystem. During this simulation, the forces on all the atoms of the QM subsystem as well as the electrostatic potential and field on the nearby MM atoms are stored. Second, a set of atomic point charges $\{q_i\}$ that reproduce the electrostatic potential and forces that the QM system exerts on the surrounding classical atoms is derived. Third, the nonbonded contributions, computed with the charges obtained in the

second step and given Lennard-Jones parameters, are subtracted from the total forces on the QM atoms. The remaining forces are assumed to be derived from bonded interactions. The parameters for bonded interactions (torsions, bending, and bonds) are thus adjusted in order to reproduce the remaining forces. The entire procedure is described in detail in the following sections.

2.1. Reference Forces. The force field for a fragment is optimized, requiring that the classical forces reproduce a set of forces computed in configurations generated by a finite-temperature QM/MM molecular dynamics run. The MM and QM subsystems are explicitly coupled by a potential that describes their steric and electrostatic interaction.[13] The QM subsystem is treated at the DFT level, with classical atoms within a distance of 8 Å being explicitly coupled to the quantum charge density. Additional coupling schemes for more distant MM atoms, for example, via classical point charges or multipole expansion of the density,[13] were not used in order to facilitate the derivation of atomic point charges (see below in section 2.2.1).

The method is tested by generating force fields for three different molecules in aqueous solution. The three test systems and the corresponding computational setups are detailed in section 2.3. In all three cases, the QM subsystem consists of the solute molecule to be parametrized, while the surrounding solvent is treated at the classical level. Thus, all interactions between the QM and MM subsystems are of a nonbonded nature. For each system, a second QM/MM trajectory is generated for validating the newly generated force field and assessing its accuracy.

2.2. Force Matching. We perform the fitting procedure with the GROMOS96 functional form[2] for the classical force field. The potential energy E_p is given by the sum of the nonbonded potential E_p^{nb} and the bonded (covalent) potential E_p^{b}, where E_p^{nb} is given by (in atomic units)

$$E_p^{nb} = \sum_i^N \sum_j^{N_i^{nb}} \left(\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \quad (1)$$

The index i runs over all N atoms in the system, while the index j runs over all N_i^{nb} atoms within a cutoff distance r_c of atom i, excluding first and second neighbors, that is, atoms that are connected to i by one or two covalent bonds. q_i and q_j are fractional atomic point charges. A_{ij} and B_{ij} are the coefficients of the Lennard-Jones potential between atoms i and j. The contribution E_p^{b} is given by

$$E_p^{b} = \sum_{n=1}^{N^{bon}} \frac{1}{4} k_{b_n} (b_n^2 - b_{0_n}^2)^2 + \sum_{n=1}^{N^{ang}} \frac{1}{2} k_{\theta n} (\cos\theta_n - \cos\theta_{0n})^2 + \sum_{n=1}^{N^{imp}} \frac{1}{2} k_{\xi n} (\xi_n - \xi_{0n})^2 + \sum_{n=1}^{N^{tor}} k_{\varphi n} [1 + \cos(\delta_n)\cos(m_n \varphi_n)] \quad (2)$$

The first term assigns a bond-stretching potential with force constant k_b to all atom pairs connected by a covalent bond and keeps the bond length b close to its equilibrium value b_0. The second term keeps bond angles θ close to an equilibrium value θ_0. The third term corresponds to a harmonic (or so-called improper) dihedral angle potential and is used to maintain the planarity of groups such as sp^2-hybridized carbon atoms. The fourth term represents a periodic torsional potential with a barrier height of $2k_\varphi$ and a phase-shift δ (0 or π). m is the multiplicity, that is, the number of minima in the interval $[-\pi, \pi]$.

The force matching is performed in two separate steps: first, the nonbonded parameters (i.e., atomic charges) are optimized, and in a subsequent step, the parameters for bonded interactions are optimized.

2.2.1. Fit of Atomic Point Charges. The method to derive atomic point charges is closely related to the method of dynamically generated electrostatic-potential-derived charges (D-RESP) introduced by Laio et al.[37] In the D-RESP method, the grid used for charge fitting is defined by the positions of all classical atoms that are explicitly coupled to the QM charge density (the so-called NN atoms).[13,37] Since QM/MM coupling requires the computation of the electrostatic potential on the NN atoms at each step of a QM/MM simulation,[13] the charges can be derived "on-the-fly'' with minimal computational overhead, in a single computational step that consists in solving a system of linear equations.[37] This step also introduces a restraint of the atomic charges to their Hirshfeld values.[38] Such a restraint is necessary as unrestrained fitting results in an overdetermined problem for which many nearly equivalent solutions exist, leading to charge sets that are purely "best-quality-of-fit'', with values that are often chemically not meaningful and show a strong dependence on the conformation of the molecule.[8,37]

Here, we have modified the original D-RESP scheme in a few important aspects. In order to reproduce in the best possible manner the solvation structure around the QM subsystem, we require the charge set to also reproduce the electric field on the NN atoms originating from the QM charge density, in addition to the potential. Chemically equivalent atoms, such as the hydrogen atoms of a methyl group, are required to carry identical charges. While D-RESP charges are derived at each MD step, we here derive from the QM/MM trajectory a single set of charges that performs best on average over all reference conformations used for the fit. It was shown by Reynolds et al. that including multiple conformations results in more transferable charges compared to those derived from a single conformation.[39] As a welcome side effect, this allows for a weaker restraint to the Hirshfeld charges than is necessary when deriving charges for a single conformation. The set of atomic point charges $\{q_i\}$ is obtained by minimizing a target penalty χ^2:

$$\chi^2 = \sum_{l=1}^{L} \Big[\sum_{j \in NN_l} w^V (V_{jl}^{MM} - V_{jl}^{\rho})^2 + \sum_{j \in NN_l} w^E ||\mathbf{E}_{jl}^{MM} - \mathbf{E}_{jl}^{\rho}||^2 \Big) + \sum_{i \in QM} w^H (q_i - q_{il}^{H})^2 \Big] + w^Q (Q^{tot} - \sum_{i \in QM} q_i)^2 \quad (3)$$

The index l runs over all L conformations; j runs over all NN_l classical atoms that are explicitly coupled to the quantum system in configuration l, and i refers to atoms of the quantum system. w^V, w^E, w^H, and w^Q are arbitrarily chosen weighting factors whose values will be specified in section 3.1.1. V_{jl}^{ρ} and $\mathbf{E}_{jl}^{\rho}$ are the electrostatic potential and field, respectively, on the classical atom j in configuration l due

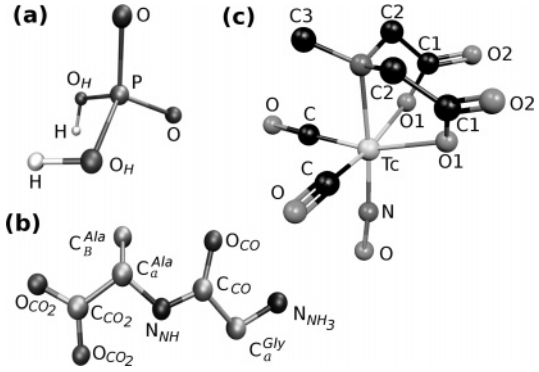


Figure 1. Graphical representation of test compounds and the naming convention used for atom types: (a) $H_2PO_4^-$, (b) Gly-Ala dipeptide, and (c) TNDM. In b and c, hydrogen atoms have been omitted for clarity.

to the presence of the QM system, while V_{jl}^{MM} and $\mathbf{E}_{jl}^{\mathrm{MM}}$ are the potential and field resulting from the classical point charges $\{q_i\}$. The third term in eq 3 restrains the charges q_i to the Hirshfeld charges q_{il}^{H}, and the last term finally restrains the total charge of the subsystem to the correct value Q^{tot}. The minimization of χ^2 corresponds to solving an overdetermined[40] system of linear equations in $\{q_i\}$ in the least-squares sense, a task that can conveniently be solved with an algorithm such as *QR* factorization.[41] The original D-RESP scheme is recovered by choosing values $L = 1$, $w^V = 1$, $w^E = 0$, and $w^H = 0.1$.[37]

2.2.2. Bonded Interactions. In a second step, the set of charges $\{q_i\}$ is used to calculate the total nonbonded forces on all of the QM atoms, taking into account exclusions and a scaling of 1−4 interactions[2] according to the GROMOS96 functional form. We denote these forces by $\mathbf{F}_{li}^{\mathrm{MM_{nb}}}$ (for the QM atom i and the configuration l). The set of parameters $\{x_n\} = \{b_{0n}, \theta_{0n}, k_{bn}, k_{\theta n}, k_{\xi n}, k_{\varphi n}\}$, describing bonded interactions, is then determined by minimizing the function σ^2, given by

$$\sigma^2 = \sum_{l=1}^{L} \sum_{i \in \mathrm{QM}} ||\mathbf{F}_{li}^{\mathrm{MM_b}} - (\mathbf{F}_{li}^{\mathrm{QM}} - \mathbf{F}_{li}^{\mathrm{MM_{nb}}})||^2 \qquad (4)$$

The minimization of σ^2 corresponds to a nonlinear least-squares optimization. Since the derivatives $\partial\sigma/\partial x_n$ are trivial to calculate analytically, a gradients-based optimization algorithm can be used. Specifically, we use the Levenberg−Marquardt algorithm as implemented in the MINPACK collection of routines.[42,43]

2.3. Test Systems. The method is validated through simulations of aqueous solutions of three different molecules. Their structures and the naming scheme used for atom types are shown in Figure 1. In all three test cases, the QM subsystem consists of the molecule to be parametrized, while the surrounding solvent is treated at the classical level. The three test cases are as follows:

(i) The Anion Dihydrogenphosphate ($H_2PO_4^-$). $H_2PO_4^-$ is used as a prototype for the functional group R−PO_4−R′ that is ubiquitous in biomolecules. Since it is an anion, the solvent is expected to significantly polarize the system.

(ii) The Zwitterionic Form of the Dipeptide Glycyl-Alanine (Gly-Ala). In previous work,[44] we have compared the solvation structure of this dipeptide at different levels of theory, from classical MD to QM/MM to a full QM treatment of the dipeptide and its first and second solvation shells. It was found that classical simulations with the Amber/parm94 and GROMOS96 force fields exhibit significant differences in the solvation structure, compared to QM/MM and full QM simulations, especially for the charged termini. This system thus provides a suitable test to show that reparametrization can improve the solvation properties with respect to a (higher-level) reference simulation.

(iii) The Transition Metal Complex [Tc(NO)(CO)$_2$(MIDA)] (MIDA = N-Methyl-iminodiacetic Acid) (TNDM). Technetium compounds are used in nuclear medicine for the diagnosis and localization of tumor cells, while their rhenium analogues have potential use in therapeutic intervention.[45] The $[Tc(NO)(CO)_2]^{2+}$ core is thought to be highly promising for the labeling of biomolecules, because of its small size and low molecular weight.[46] The generation of force field parameters for this compound enables the investigation of interactions between labeled biomolecules and their biological targets by means of well-established techniques within the framework of classical MD simulations. We consider this compound an excellent test case for our approach, as it presents a number of challenging features, such as metal−ligand bonds and subtle differences between the isoelectronic ligands CO and NO^+. Moreover, the complex is highly inert, so that a description of metal−ligand bonds by simple harmonic potentials appears justified.

2.3.1. Computational Details. The QM/MM implementation used in this work[13] combines the packages CPMD[47,48] for the quantum system and GROMOS96[49] for the classical part. The QM subsystem is treated at the DFT level, using norm-conserving pseudopotentials of the Martins−Trouiller type.[50] The plane-wave basis was expanded up to a cutoff of 70 Ry. Classical atoms within a distance of 8 Å were explicitly coupled to the quantum charge density.[13] All simulations were performed at 300 K under constant volume conditions.

System sizes and simulation times were as follows. (i) $H_2PO_4^-$: cubic quantum box with an edge of 9 Å; BLYP exchange-correlation functional[51,52] for the QM subsystem; MM subsystem consisting of 763 water molecules (SPC model[53]) in a cubic box with an edge of 14 Å; a total QM/MM simulation time of 20 ps, of which 7 ps were used for parametrization and 13 for validation. (ii) Gly-Ala dipeptide: quantum box of size 15 × 12 × 12 Å^3; BLYP functional; MM subsystem consisting of 1718 water molecules (SPC) in a cubic box with length 37 Å; 3 ps of QM/MM simulation used for parametrization and 7 for validation. (iii) TNDM: cubic quantum box with length 14 Å; BP86 exchange-correlation functional,[51,54] which for Tc compounds provide a better agreement with experimental structures;[55] MM subsystem consisting of 1663 water molecules (TIP3P model[56]) in a cubic box with length 37 Å; 3 ps of QM/MM simulation for parametrization and 4 for validation. The standard GROMOS force field does not provide Lennard-Jones parameters for Tc^+ and the nitrosyl ligand. Since our goal was to validate the force-matching procedure, and Lennard-Jones interactions are treated identically in the QM/

MM and classical calculations, we decided to avoid large computational efforts for an accurate determination of these parameters. Instead, an ad hoc approach was chosen to determine the missing parameters: The nitrogen and oxygen atoms of the nitrosyl ligand were assigned the parameters of an amine nitrogen and a carbonyl oxygen, respectively. For technetium, we used the parameters describing a Cu^+ ion, because the charges are identical and, among those metal ions for which the standard force field provides parameters,[2] Cu^+ is most similar in size to Tc^+. For simulations of TNDM for purposes other than validation of the force-matching approach, the missing Lennard-Jones parameters clearly need to be determined in a more careful fashion.

In order to probe the influence of explicit solvation on the charges, an additional set of charges was derived from the electrostatic potential calculated in the gas phase. These charges were derived according to the RESP scheme,[8] but using solute structures from the QM/MM simulations instead of gas-phase optimized ones. Gaussian 03[57] was used to perform single-point HF/6-31+G* calculations, and the program resp from the Amber suite[58] was used to compute the RESP charges.

For each set of fitted parameters, a classical MD run was performed to generate a trajectory to be compared to the QM/MM reference. Initial coordinates and velocities, the time step, and total simulation time were all chosen to be identical to those of the corresponding QM/MM trajectory used for validation.

3. Results and Discussion

3.1. Quality of Fit. 3.1.1. Electrostatics. To measure the quality of a charge set $\{q_i\}$, we compute the relative standard deviation (SD) of the electrostatic potential (SD_V) and field (SD_E) with respect to the QM reference values over all L configurations, with SD_V and SD_E defined as

$$SD_V = \sqrt{\frac{\sum_l^L \sum_{j \in NN_l} (V_{jl}^{MM} - V_{jl}^{\rho})^2}{\sum_l^L \sum_{j \in NN_l} (V_{jl}^{\rho})^2}} \quad (5)$$

$$SD_E = \sqrt{\frac{\sum_l^L \sum_{j \in NN_l} ||\mathbf{E}_{jl}^{MM} - \mathbf{E}_{jl}^{\rho}||^2}{\sum_l^L \sum_{j \in NN_l} ||\mathbf{E}_{jl}^{\rho}||^2}} \quad (6)$$

In order to probe the influence of the weighting parameters in eq 3, a series of charge sets were derived for different values of w^V, w^E, w^H, and w^Q. The results for $H_2PO_4^-$ are shown as a representative example in Figure 2a.

As expected, increasing the weight w^E improves the quality of the electric field, that is, of the forces on the MM atoms. However, this is achieved at the price of a poorer quality of the electrostatic potential. Moreover, high values of w^E in combination with a weak restraint to Hirshfeld charges (i.e., small w^H) lead to considerable deviations of the total charge from its correct value, an effect that was found to be especially strong in the case of $H_2PO_4^-$. This behavior is remedied by restraining the overall charge to the correct value through a high value of w^Q. The effect does not occur when fitting the potential only, so that the above constraint is not needed when $w^E = 0$ (as in the original D-RESP procedure).

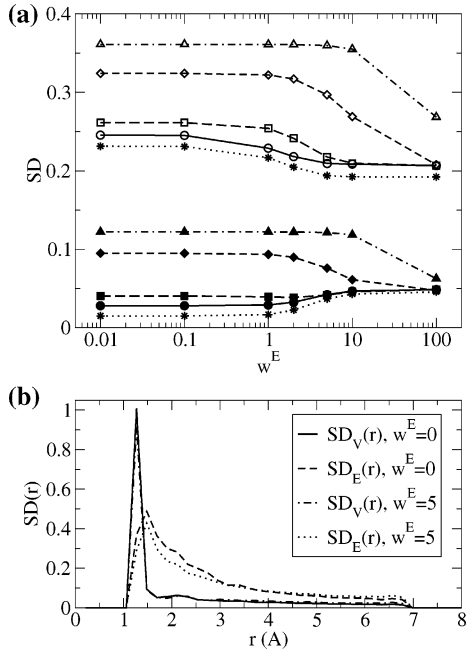


Figure 2. (a) SD_V (filled symbols) and SD_E (empty symbols) for $H_2PO_4^-$ as a function of w^E for different values of w^H: 0 (circles), 0.01 (squares), 0.1 (diamonds), and 1.0 (triangles). $w^V = 1$ and $w^Q = 1000$ in all cases. Stars indicate the best possible quality of fit mentioned in the text. (b) SD_V and SD_E as a function of distance to the nearest QM atom. $w^H = 0.01$.

Remarkably, the fit can only be marginally improved by allowing for fluctuating charges. Star symbols in Figure 2a indicate the average SD_V and SD_E of charges that were derived for single snapshots, without imposing any equivalencies between atoms. Thus, these charges represent the best quality of fit that can be achieved, for given values of w^V, w^E, w^H, and w^Q, with a polarizable model that uses exclusively atom-centered point charges, such as the fluctuating charge model.[23,59] This result suggests that the accuracy limiting factor is not so much the use of fixed charges instead of fluctuating ones but that it is rather the model of exclusively atom-centered point charges itself. This view is in agreement with work by Masia et al.,[60] who showed that polarizable models with additional interaction sites are superior to the fluctuating-charge model in their ability to reproduce the polarization of a quantum system by classical point charges.

Figure 2b shows SD as a function of the distance to the nearest QM atom. As can be seen, the deviation is largest at short distances, and this effect is more pronounced for the field than for the potential. Most likely, these deviations result from the inability of a simple atomic point charge model to describe the highly inhomogeneous field in the immediate vicinity of the QM system. The deviations are very large only for small distances (<1.7 Å) that correspond to the short-range tail of the radial distribution function of hydrogen

Table 1. Atomic Charges of $H_2PO_4^-$

atom[a]	MDR	MDRF	D-RESP[b]	RESP[c]
P	+0.011	+0.513	−0.038	+1.547
O	−0.596	−0.658	−0.599	−0.902
O_H	−0.326	−0.580	−0.214	−0.770
H	+0.416	+0.482	+0.336	+0.398

[a] Atom types (see Figure 1a). [b] Time-averaged values over the QM/MM reference trajectory. See also ref 62. [c] Gas-phase HF/6−31+G*.

Table 2. Atomic Charges of Gly-Ala

atom[a]	MDR	MDRF	D-RESP[b]	RESP[c]	G96[d]	Amber[e]
H_{NH_3}	+0.379	+0.444	+0.227	+0.334	+0.248	+0.164
N_{NH_3}	−0.395	−0.658	+0.145	−0.395	+0.129	+0.294
C_α^{Gly}	+0.045	+0.150	+0.039	−0.041	+0.127	−0.010
H_α^{Gly}	+0.081	+0.051	+0.080	+0.132	0.000	+0.090
C_{CO}	+0.544	+0.493	+0.069	+0.591	+0.380	+0.616
O_{CO}	−0.681	−0.593	−0.457	−0.587	−0.380	−0.572
N_{NH}	−0.571	−0.707	+0.083	−0.579	−0.280	−0.382
H_{NH}	+0.417	+0.471	+0.236	+0.313	+0.280	+0.268
C_α^{Ala}	+0.265	+0.472	+0.086	+0.168	0.000	−0.175
H_α^{Ala}	+0.099	+0.013	+0.123	+0.082	0.000	+0.107
C_β^{Ala}	−0.162	−0.164	−0.040	−0.293	0.000	−0.209
H_β^{Ala}	+0.054	+0.035	+0.038	+0.078	0.000	+0.076
C_{CO_2}	+0.710	+0.456	−0.064	+0.861	+0.270	+0.773
O_{CO_2}	−0.865	−0.736	−0.574	−0.810	−0.635	−0.806

[a] Atom types (see Figure 1b). [b] Time-averaged values over the QM/MM reference trajectory. See also ref 62. [c] Gas-phase HF/6−31+G*. [d] GROMOS 43A1 force field.[2] [e] Parm94 force field.[3]

bonds.[61] Therefore, as we will show in the following, this error in the fit has only a marginal influence on the solvation structure.

Two different sets of parameters were derived via the force-matching procedure: a first one, in which charges were derived by fitting only to the electrostatic potential, that is, $w^E = 0$, and a second one with charges derived from both the potential and the field, with $w^E = 5$. We will refer to the former as MDR (multiconformation D-RESP charges) and to the latter as MDRF (F for field). The remaining weighting factors were in both cases set to $w^V = 1$, $w^H = 0.01$, and $w^Q = 1000$. The charge sets derived from gas-phase HF/6-31+G* calculations (see section 2.3.1) are denoted as RESP. Tables 1−3 compare the MDR and MDRF charges of the three test compounds to RESP and time-averaged D-RESP charges and, in the case of Gly-Ala, also to the values used in the Amber/parm94 and GROMOS96 force fields.

In general, MDR(F) charges are similar to those used in the Amber force field, which is not surprising since both are based on a quantum mechanical origin. The most striking difference is observed for the terminal ammonium group of Gly-Ala, which is strongly polarized in the MDR(F) charge sets. The positive charge of the ammonium group is localized on the hydrogen atoms, while the nitrogen atom carries a negative partial charge. In the Amber and GROMOS force fields, in contrast, the positive charge is more evenly distributed. This difference in charge distribution most likely arises from the fact that MDR(F) charges are derived for an explicitly hydrated species in a condensed system.

The optimized atomic charges for TNDM are shown in Table 3. Compared to the formal charges of Tc^+, NO^+, and CO, the optimized charges appear to indicate a considerable amount of charge transfer from the metal center to the nitrosyl and carbonyl ligands. This result can be interpreted as a manifestation of the known strong π back donation from metal d orbitals into antibonding π^* orbitals of these ligands, resulting in metal−ligand bonds of highly covalent character.

Table 3. Atomic Charges of TNDM

atom[a]	MDR	MDRF	D-RESP[b]
Tc	+1.791	+1.664	+1.794
N	−0.353	−0.331	−0.341
O_{NO}	+0.102	+0.090	+0.102
C	−0.231	−0.068	−0.284
O_{CO}	+0.059	−0.056	+0.098
C1	+0.111	+0.479	−0.026
O1	−0.494	−0.661	−0.472
O2	−0.520	−0.580	−0.414
C2	+0.024	−0.069	+0.015
H_{C2}	+0.129	+0.096	+0.131
N1	−0.069	−0.056	−0.051
C3	−0.031	−0.045	−0.017
H_{C3}	+0.048	+0.068	+0.052

[a] The naming scheme for atom types is shown in Figure 1c. [b] Average over the QM/MM reference trajectory. See also ref 62.

Table 4. Relative Error of the Total Force SD_T

	MDR	MDRF	G96
$H_2PO_4^-$	0.44	0.42	1.36[a]
Gly-Ala	0.28	0.28	1.01[b]
TNDM	0.34	0.33	N/A

[a] See ref 63. [b] GROMOS96 43A1 force field.[2]

3.1.2. Bonded and Total Forces. In analogy to SD_E in eq 6, we define standard deviations SD_B to measure the quality of the bonded forces and SD_T to measure the quality of the total force. Bonded forces are reproduced with values of SD_B between 0.24 (Gly-Ala) and 0.28 (TNDM). The relative errors in the total force SD_T are shown in Table 4. Both charge sets yield similar SD_T values for a given molecule, with MDRF parameters outperforming MDR by a few percent. The error is largest in the case of $H_2PO_4^-$ with $SD_T = 0.44$, and the best performance is obtained in the case of Gly-Ala with $SD_T = 0.28$. Interestingly, when looking at the SD_T per atom in Gly-Ala, the largest error is made on the oxygen atoms with $SD_T = 0.4$, similar in size to that in $H_2PO_4^-$. These findings are a clear indication that the main source of error lies in the description of electrostatics with a simple, atom-centered, point-charge model. This simplification will have the strongest impact when the charge distribution around an atom is anisotropic, as is the case with lone pairs located on oxygen atoms.

3.2. MD with Optimized Parameter Sets. 3.2.1. Solvation Structure. The ability to reproduce the solvation structure of the reference QM/MM simulation provides a sensitive test for the quality of the optimized charge sets. The solvation structure is analyzed by calculating radial distribution functions $g(r)$ between solute and solvent atoms. We denote water oxygen and hydrogen atoms as O_W and H_W, respectively.

Figure 3 compares radial distribution functions $g(r)$ of the phosphate ion in water. The corresponding coordination

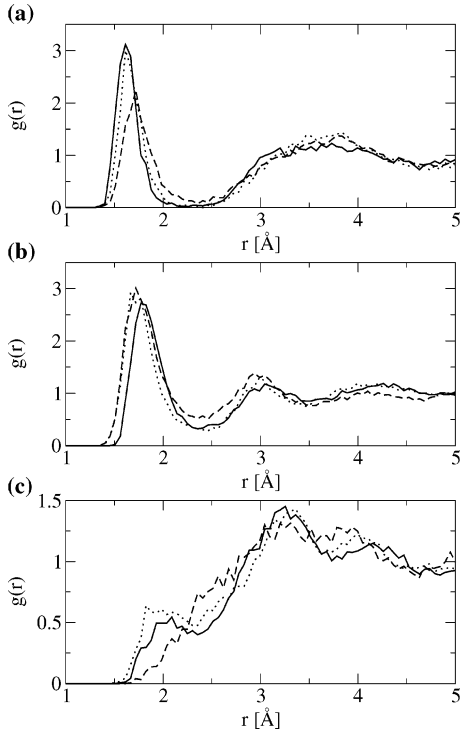


Figure 3. Radial distribution functions $g(r)$ of $H_2PO_4^-$ in water. Solid lines indicate the QM/MM reference simulation; dashed lines indicate the simulation with the charge set MDR, and dotted lines indicate the one with the charge set MDRF. Shown are the radial distribution functions (a) $H-O_W$, (b) $O-H_W$, and (c) O_H-H_W. The names of the atom types are shown in Figure 1a.

Table 5. Hydrogen-Bond Coordination Numbers N_C (per Atom) for the First Solvation Shell of $H_2PO_4^-$ (Top) and Gly-Ala (Bottom)

	QM/MM	MDR	MDRF	RESP	G96
H	1.0	1.0	1.0	1.0	N/A
O_H	1.4	1.5	1.7	1.8	N/A
O	3.5	4.3	3.6	5.6	N/A
H_{NH_3}	1.0	1.1	1.0	1.0	1.2
H_{NH}	1.0	1.0	1.0	1.0	0.9
O_{CO}	2.4	2.4	2.0	1.9	1.0
O_{CO_2}	3.3	3.7	3.4	3.1	3.6

numbers for hydrogen bonding in the first solvation shell[64] are given in Table 5. The naming scheme for atom types is shown in Figure 1a. Overall, the charge set MDRF performs clearly better than the set MDR. For the $O-H_W$ radial distribution function (Figure 3b), both charge sets yield a first peak that is shifted to a shorter distance by ≈0.06 Å compared to the QM/MM reference. Thus, both charge sets somewhat overestimate the basicity of the O atoms. However, MDRF charges yield a first peak that is merely shifted, with the shape of the peak and the hydrogen-bond coordination number N_C (Table 5) being in good agreement with QM/MM. In contrast, MDR charges lead to a broadening of the peak and severe overestimation of N_C. The shape and depth of the first minimum around 2.4 Å indicates that the orientation of water molecules in the first solvation shell is more rigid in the QM/MM and MDRF simulations compared to that in the simulation with MDR charges.

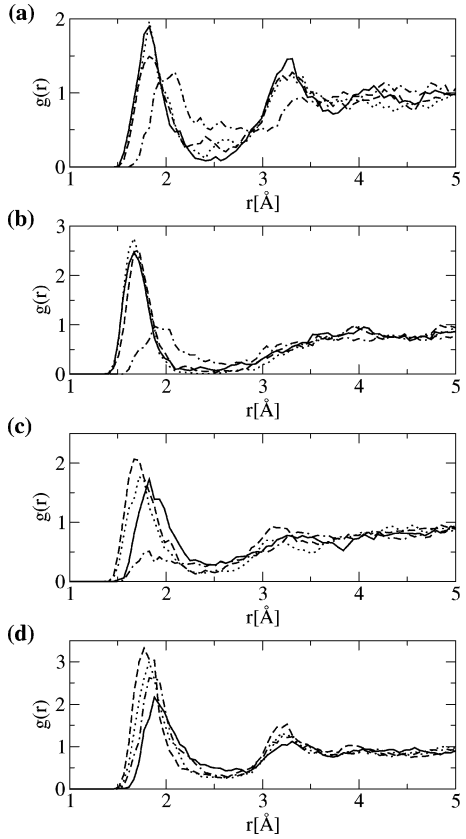


Figure 4. Radial distribution functions $g(r)$ of polar groups of the Gly-Ala dipeptide. The line convention is the same as in Figure 3. Dashed−dotted lines refer to a simulation using original GROMOS parameters. Shown are the radial distribution functions (a) $H_{NH_3}-O_W$, (b) $H_{NH}-O_W$, (c) $O_{CO}-H_W$, and (d) $O_{CO_2}-H_W$.

Significant differences are also found in the solvation of the hydroxyl group. The $H-O_W$ radial distribution function in Figure 3a shows excellent agreement between MDRF and the QM/MM reference, while MDR charges severely underestimate the acidity of the protons. In the case of the hydroxyl oxygen atoms O_H, MDR charges qualitatively fail to reproduce the small first peak (Figure 3c). In contrast, MDRF charges slightly overestimate hydrogen bonding to the solvent, but the solvation structure is in overall good agreement with the QM/MM reference. Since there is no well-defined first solvation shell in the case of the O_H atoms, a comparison of coordination numbers does not appear to be meaningful.

Solute−solvent radial distribution functions $g(r)$ of polar groups of the Gly-Ala dipeptide are shown in Figure 4; the corresponding coordination numbers for hydrogen bonding[64] are given in Table 5. As in the case of $H_2PO_4^-$, the solvation of hydrogen-bond donors is reproduced better than that of acceptors. MDRF charges in fact reproduce the solvation of the terminal NH_3^+ group almost perfectly. MDR charges on the other hand underestimate the height of the first peak of the $H_{NH_3}-O_W$ radial distribution function but correctly reproduce the position of the first maximum. Both charge sets reproduce well the position and width of the second peak

at 3.3 Å. For the solvation of the hydrogen H_{NH} in the amide bond, the agreement with the QM/MM reference is excellent for both of the fitted charge sets.

The hydrogen-bond acceptor strengths of the carbonyl and carboxyl oxygens are clearly overestimated by both charge sets, with MDRF performing slightly better than MDR. In the case of the terminal carboxylate group, the best agreement with the QM/MM reference of all classical simulations is obtained with the original GROMOS parameters. For the other polar groups, however, we observe a clear improvement compared to GROMOS charges by using the fitted charge sets.

Since Lennard-Jones interactions are identical in the QM/MM and classical simulations, the shifts in $g(r)$ involving oxygen atoms must be attributed to an inability of the classical model to reproduce the charge distribution found in QM/MM. Most likely, the major flaw of the classical model consists in the use of atom-centered point charges that do not allow to account for anisotropies of the charge distribution around an atom. In addition, it appears that polarization of the QM/MM charge density also leads to an overestimation of hydrogen-bond acceptor capabilities. This view is supported by the following observation: gas-phase derived RESP charges still lead to an overestimation of the acceptor capability of the Gly-Ala oxygen atoms, but to a lesser extent than MDR(F) charges (Figure S-2, Supporting Information). In fact, we expect the charge distribution to be less anisotropic in the gas phase compared to that in the hydrated species, where the lone pairs will be strongly oriented toward the protons of the first hydration shell. On the other hand, RESP charges significantly underestimate the acidity of the proton of the amide group of Gly-Ala (Figure S-2, Supporting Information). In the case of $H_2PO_4^-$, charges derived in the gas phase underestimate the acidity of the protons and overestimate the basicity of the oxygen atoms. This result can be attributed to a charge distribution that is rather diffuse in the gas phase, while explicit hydration leads to a stronger localization of the charge. Consequently, charges derived from a polarized charge density give a better description of the solvation structure.

In view of the above observations, it appears crucial for an accurate description of the solvation structure of polar groups that the charges be derived from a QM/MM wavefunction that is polarized by the solvent. For some acceptor groups, however, polarization leads to an overestimation of hydrogen bonding, as in the case of the oxygen atoms of Gly-Ala. The inclusion of off-atomic interaction sites, such as lone pairs, should lead to a better description of hydrogen-bond acceptors.[28,60]

In the case of TNDM, the only hydrophilic groups are the carboxyl ligands, the solvation structure of which is reproduced well (Supporting Information, Figure S-1). In contrast to $H_2PO_4^-$ and Gly-Ala, hydrogen bonding of the oxygen atoms is slightly underestimated. The remaining groups are of hydrophobic character, which is well-reproduced by the optimized parameters in the case of the aliphatic groups. The optimized charge sets also reproduce the hydrophobic character of the nitrosyl and carbonyl ligands. In fact, they appear slightly more hydrophobic in the classical simulations than in the QM/MM reference, where we find a more pronounced peak at around 3 Å in the $O-O_W$ radial distribution function (Figure S-1, Supporting Information).

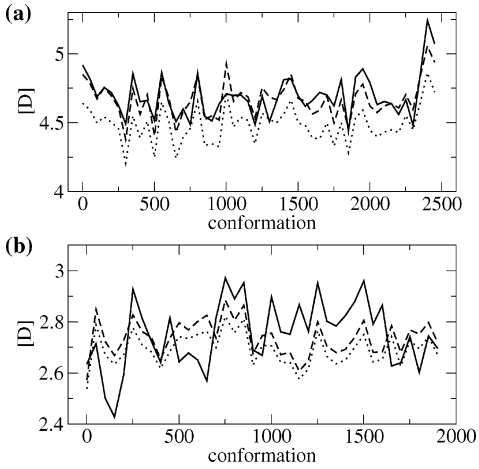


Figure 5. Dipole moments (in Debye) for conformations along the QM/MM reference trajectory. Shown are the dipole moment in the QM/MM simulation (solid lines) and dipole moments calculated with MDR (dashed lines) and MDRF charge sets (dotted lines). The corresponding average values and standard deviations are given in Table 6. (a) Gly-Ala dipeptide. (b) TNDM.

Table 6. Average and Standard Deviation of the Dipole Moment along the QM/MM Reference Trajectory (in Debye)

	Gly-Ala	TNDM
QM/MM	4.69 ± 0.15	2.72 ± 0.15
MDR	4.66 ± 0.13	2.71 ± 0.08
MDRF	4.48 ± 0.12	2.67 ± 0.08

3.2.2. Dipole Moments. An additional test to assess the quality of the fitted charge sets consists in comparing the total dipole moment of the solute to that of the QM/MM reference. Figure 5 compares the quantum mechanical dipole moments in conformations taken from the QM/MM reference simulation to the equivalent quantity as described with MDR and MDRF charge sets. Averages and standard deviations are reported in Table 6.

In general, the dipole moment is reproduced well by both charge sets, with absolute deviations below 0.4 D (0.3 D for TNDM), corresponding to relative deviations below 10%. The average dipole of MDR charges agrees within 0.03 D (i.e., better than 1%) with the QM/MM reference. MDRF charges, on the other hand, systematically underestimate the dipole, especially in the case of Gly-Ala. The smaller fluctuations in the classical simulations compared to the QM/MM reference must be attributed to the fixed charge model, as changes of the dipole moment arise from changes of the molecular geometry only. In the QM/MM simulations, on the other hand, there is an additional contribution from instantaneous charge fluctuations.

It becomes apparent from Figure 5 that MDR charges perform clearly better than MDRF charges, not only in reproducing the average dipole but also in reproducing the

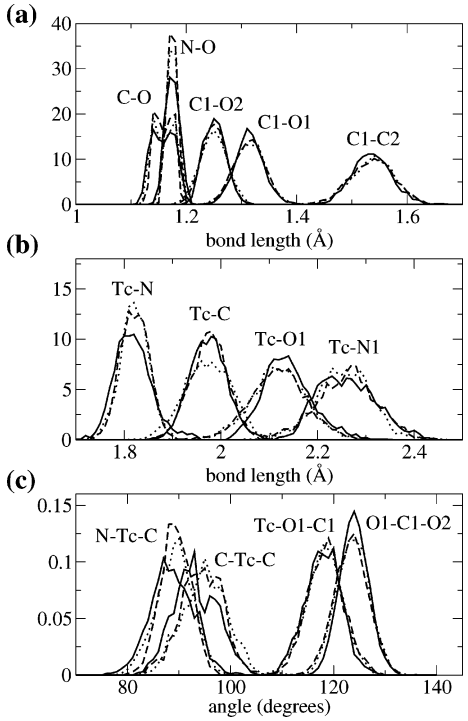


Figure 6. Normalized distribution of (a, b) selected bond lengths and (c) selected bond angles in TNDM. The line convention is identical to that in Figure 3. The atom naming convention for atom types is shown in Figure 1c.

dipole fluctuations along the trajectory. This can be interpreted as an effect of the different charge-fitting procedures. Including the electric field in the fitting leads to a better description at short range, because of the $1/r^2$ dependence of the field. Evidently, this is achieved at the expense of a poorer quality at longer range, compared to fitting of the potential only (Figure 2). Including the field in the fitting procedure apparently leads to a different weighting of the fitting, in the sense that the optimized charges represent well the electronic properties of functional groups but to a lesser extent those of the entire molecule. Thus, MDRF charges perform better in reproducing more "local" properties, such as the structure of the first solvation shell around a functional group, while MDR charges perform better for more "global" properties like the total dipole moment. A possible way to overcome these limitations of MDRF charges could consist in explicitly including the dipole moment (and possibly higher multipoles) in the fitting procedure.

3.2.3. Molecular Structures. We will limit the discussion of structural properties to TNDM, as it presents the most challenging features for a force-matching approach to reproduce. However, the results obtained for TNDM are representative also for the other two test compounds. The naming convention for the atom types of TNDM is shown in Figure 1c. Figure 6 shows normalized distributions of selected bond lengths and bond angles in the QM/MM reference simulation and in the simulations with optimized parameters. The overall agreement can be considered excellent, both for the position of the maximum and the width of the distribution. The small differences in length and strength of the bond in the CO and NO^+ ligands are reproduced faithfully by both parameter sets. The same observation holds

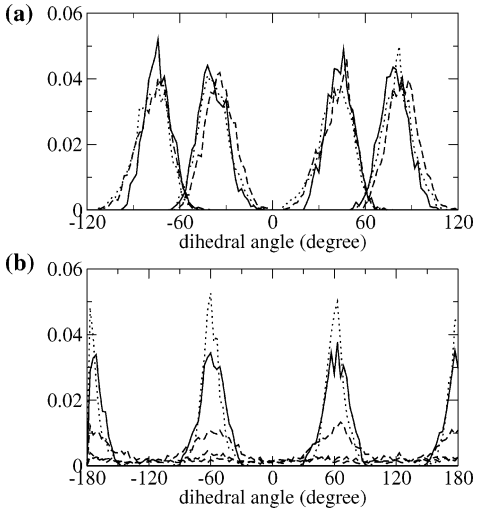


Figure 7. Normalized distribution of selected dihedral angles in TNDM. (a) C3−N1−C2−H_{C2}. (b) Tc−N1−C3−H_{C3}. The line convention is identical to that in Figure 3.

for the different metal−ligand bonds. The method also captures the different C−O bond lengths in the carboxyl ligand. Both parameter sets yield almost identical distributions, with the exception of the metal−carbonyl bond, the strength of which is underestimated by the MDRF parameter set. For the MDR parameters, on the other hand, we find a distribution for this bond that is practically identical to the QM/MM reference. The difference most likely results from the different intramolecular electrostatic interaction between the carbonyl and carboxyl ligands, for which the two charge-fitting schemes yield rather different charges (Table 3).

Figure 7a shows normalized distributions of the four dihedral angles C3−N1−C2−H_{C2}. These dihedrals are part of a rigid ring system resulting from chelation of the metal center by the ligand. All four distributions are well-reproduced by both fitted parameter sets. The picture looks different for the N1−C3 bond, which is the only bond in TNDM around which rotation can occur. Figure 7b shows distributions of the three dihedral angles Tc−N1−C3−H_{C3}, for which the minima of the torsional potential are reproduced correctly. However, while there are no dihedral transitions in the QM/MM reference simulation, they do occur in the simulation with the parameter set MDR, clearly indicating that the height of the rotational barrier is underestimated in the fitting process. In contrast, no dihedral transitions occur with MDRF parameters and the distribution is reproduced remarkably well. This indicates a value for the torsional barrier that is closer to the "true" value in the QM/MM description. In fact, the narrower distribution indicates that the barrier height is slightly overestimated with MDRF. For a quantitative assessment, one would need to compare the frequency at which dihedral transitions occur, which in turn requires simulation times of sufficient length to observe a significant number of these events. Unless enhanced sampling techniques are used, the necessary simulation time is beyond what can currently be achieved with reasonable computational effort within a DFT/MM approach.

4. Conclusions

We introduce a force-matching approach to derive parameters for nonpolarizable biomolecular force fields from QM/MM simulations. As a consequence of using QM/MM calculations for obtaining the reference forces, the parameters are derived for a species that is explicitly polarized by the environment, and finite temperature and pressure are automatically taken into account. We have optimized all interaction parameters except those for van der Waals interactions, which are retained from the classical force field. This choice is made in order to remain consistent with the QM/MM Hamiltonian[13] but is also dictated by the failure of DFT to account for dispersion interactions.

MD simulations with optimized parameters perform remarkably well in reproducing properties from the QM/MM simulation. The agreement can certainly be improved further by using a more sophisticated force field, for example, by including lone pairs as additional interaction sites or by abandoning the fixed point-charge description for an explicitly polarizable model. In fact, our procedure can point out the necessity for using a more sophisticated model and allows quantification of the improvement brought about by such a model. For example, the results presented here suggest that a fluctuating charge model with atomic point charges as the sole interaction sites is insufficient for significantly improving the description of electrostatics, at least for the systems considered here and if the ability to reproduce the electrostatic field is taken as a measure.

The charge scheme MDRF is slightly, sometimes clearly, superior to its counterpart MDR in reproducing solvation structures. For O_H in $H_2PO_4^-$, MDR fails to describe the correct (QM/MM) solvation structure, while MDRF gives excellent agreement. For larger molecules, a charge-fitting procedure aimed at reproducing the forces on the MM atoms that are close to the QM subsystem might become inappropriate, as charges of QM atoms that are far from the QM/MM boundary will become ill-defined. In these cases, it might be necessary to combine the derivation of bonded and nonbonded parameters into a single, necessarily nonlinear, minimization scheme. As a benefit, such an approach could also lead to an improved balance between electrostatic and covalent intramolecular interaction parameters.

An obvious drawback of using DFT for the reference calculations lies in its failure to describe dispersion interactions within the quantum system, so that the parametrization of Lennard-Jones potentials is not possible within a standard QM/MM scheme based on a purely local approximation to DFT. The recent development of dispersion-corrected pseudopotentials[65] might open a way to overcome this limitation and develop a "DFT-consistent" force field including Lennard-Jones parameters. A further limitation concerns the derivation of torsional barriers, since the time scale that is accessible to the QM/MM reference simulation will in general be insufficient for the sampling of complete torsional profiles. This limitation could be overcome by exploiting an enhanced sampling technique.[33−36]

Since the parameters are derived for a specific environment, the method could be very useful in the context of free energy perturbation methods for the accurate determination of receptor/ligand binding free energies. For these applications, the ligand (and the receptor binding site) could be parametrized in several states along the reaction coordinate, providing an accurate estimate of the binding free energy that can account for long-time fluctuations of the protein environment. The method appears to be ideally suited to parametrize compounds whose stability is crucially influenced by the environment, such as reaction intermediates in enzymatic cycles. The method can further serve as a tool to analyze the influence of the environment on charge distributions or bond strengths.

Currently, the method is being applied to study the binding of ruthenium and platinum anticancer drugs to DNA.[66,67] Results of these studies will be published in separate articles.[68,69]

Acknowledgment. We thank the Swiss Center for Scientific Computing (CSCS) in Manno, Switzerland, for a generous allocation of CPU time. This work was supported by the Swiss National Science Foundation (Grant No. 21.57250.99).

Supporting Information Available: The solvation structure of $H_2PO_4^-$ and Gly-Ala using HF/6-31+G*-derived charges (Figures S-1 and S-2), the solvation structure of TNDM (Figure S-3), and CPU requirements for QM/MM and classical simulations (Table S-1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. F. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190−196.

(2) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide;* vdf Hochschulverlag AG: Zürich, 1996.

(3) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(4) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(5) MacKerell, A. D., Jr.; Bashford, D.; Bellott, R. L.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B.* **1998**, *102*, 3586−3616.

(6) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27−85.

(7) MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584−1604.

(8) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(9) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(10) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.

(11) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(12) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452−10467.

(13) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941−6947.

(14) Laio, A.; Bernard, S.; Chiarotti, G. L.; Scandolo, S.; Tosatti, E. *Science* **2000**, *287*, 1027−1030.

(15) Csányi, G.; Albaret, T.; Payne, M. C.; De Vita, A. *Phys. Rev. Lett.* **2004**, *93*, 175503.

(16) Ercolessi, F.; Adams, J. B. *Europhys. Lett.* **1994**, *26*, 583−588.

(17) Lenosky, T. J.; Sadigh, B.; Alonso, E.; Bulatov, V. V.; Diaz de la Rubia, T.; Kim, J.; Voter, A. F.; Kress, J. D. *Modell. Simul. Mater. Sci. Eng.* **2000**, *8*, 825−841.

(18) Umeno, Y.; Kitamura, T.; Date, K.; Hayashi, M.; Iwasaki, T. *Comput. Mater. Sci.* **2002**, *25*, 447−456.

(19) Li, Y.; Siegel, D. J.; Adams, J. B.; Liu, X.-Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2003**, *67*, 125101.

(20) Aguado, A.; Madden, P. A. *J. Chem. Phys.* **2003**, *118*, 5718−5728.

(21) Aguado, A.; Madden, P. A. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2004**, *70*, 245103.

(22) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896−10913.

(23) Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1996**, *118*, 672−679.

(24) Liu, Y.-P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. *J. Chem. Phys.* **1998**, *108*, 4739−4755.

(25) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741−754.

(26) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(27) Ferenczy, G. G.; Reynolds, C. A. *J. Phys. Chem. A* **2001**, *105*, 11470−11479.

(28) Cieplak, P.; Caldwell, J.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1048−1057.

(29) Tabacchi, G.; Mundy, C. J.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **2002**, *117*, 1416−1433.

(30) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515−1531.

(31) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621−627.

(32) Patel, S.; MacKerell, A. D., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1504−1514.

(33) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472−477.

(34) VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2000**, *113*, 4863−4868.

(35) VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 203−208.

(36) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.

(37) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Phys. Chem. B* **2002**, *106*, 3700−3707.

(38) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129−138.

(39) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 9075−9079.

(40) With a cutoff of 8 Å, the number of NN atoms in a single conformation is typically on the order of a few hundred.

(41) Lawson, C. L.; Hanson, R. J. *Solving Least Squares Problems*; Prentice Hall: Englewood Cliffs, NJ, 1974.

(42) Moré, J. J. The Levenberg−Marquardt Algorithm: Implementation and Theory. In *Numerical Analysis*; Watson, G. A., Ed.; Springer-Verlag: New York, 1977; pp 105−116.

(43) Moré, J. J.; Garbow, B. S.; Hillstrom, K. E. *User Guide for MINPACK-1*; Report ANL-80-74; Argonne National Laboratory: Argonne, IL, 1980.

(44) Hugosson, H. W.; Laio, A.; Maurer, P.; Rothlisberger, U. *J. Comput. Chem.* **2006**, *27*, 672−684.

(45) Dilworth, J. R.; Parrott, S. J. *Chem. Soc. Rev.* **1998**, *27*, 43−55.

(46) Schibli, R.; Schubiger, P. A. *Eur. J. Nucl. Med.* **2002**, *29*, 1529−1542.

(47) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474.

(48) *CPMD*, v. 3.8; IBM Corp: Armonk, NY, 1990−2004. Copyright MPI fuer Festkoerperforschung Stuttgart 1997−2001.

(49) Scott, W. P. R.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.

(50) Trouiller, N.; Martins, J. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993−2006.

(51) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(52) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(53) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.

(54) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822−8824.

(55) Maurer, P.; Magistrato, A.; Rothlisberger, U. *J. Phys. Chem. A* **2004**, *108*, 11494−11499.

(56) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(57) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.;

Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.03; Gaussian Inc.: Pittsburgh, PA, 2003.

(58) Case, D. A.; Darden, T. A.; Cheathem, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.

(59) Rappé, A. K.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358−3363.

(60) Masia, M.; Probst, M.; Rey, R. *J. Chem. Phys.* **2004**, *121*, 7362−7378.

(61) The very large deviation of the potential below 1.7 Å results from a small fraction of atoms (less than 10 out of several thousand). The small number makes them essentially irrelevant for the fitting.

(62) D-RESP charges correspond to MDR charges but are derived for a single conformation and with a stronger restraint to the Hirshfeld values ($w^H = 0.1$). The stronger restraint represents a compromise between avoiding large charge fluctuations between conformations and achieving a good quality of the fit. See ref 37 for details.

(63) Starting from an existing parametrization of H_3PO_4 (Spieser, S. A. H.; Leeflang, B. R.; Kroon-Batenburg, L. M. J.; Kroon, J. *J. Chem. Phys. A.* **2000**, *104*, 7333−7338), charges were adjusted manually in order to reproduce the solvation structure of a QM/MM reference. Colombo, M. C.; Vande-Vondele, J.; Rothlisberger, U. To be published.

(64) The first solvation shell is assumed to extend up to 2.5 Å.

(65) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.

(66) Dorcier, A.; Dyson, P. J.; Gossens, C.; Rothlisberger, U.; Scopelliti, R.; Tavernelli, I. *Organometallics* **2005**, *24*, 2114−2123.

(67) Spiegel, K.; Rothlisberger, U.; Carloni, P. *J. Phys. Chem. B* **2004**, *108*, 7963−7968.

(68) Gossens, C.; Tavernelli, I.; Maurer, P.; Rothlisberger, U. To be published.

(69) Spiegel, K.; Magistrato, A.; Maurer, P.; Ruggerone, P.; Rothlisberger, U.; Carloni, P.; Reedijk, J.; Klein, M. Submitted for publication.

CT600284F

JCTC Journal of Chemical Theory and Computation

Multiscale Optimization of a Truncated Newton Minimization Algorithm and Application to Proteins and Protein−Ligand Complexes

Kai Zhu,[†] Michael R. Shirts,[†] Richard A. Friesner,[†] and Matthew P. Jacobson*[,‡]

Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, New York 10027, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158-2517

Received April 7, 2006

Abstract: We optimize a truncated Newton (TN) minimization algorithm and computer package, TNPACK, developed for macromolecular minimizations by applying multiscale methods, analogous to those used in molecular dynamics (e.g., r-RESPA). The molecular mechanics forces are divided into short- and long-range components, with the long-range forces updated only intermittently in the iterative evaluations. This algorithm, which we refer to as MSTN, is implemented as a modification to the TNPACK package and is tested on energy minimizations of protein loops, entire proteins, and protein−ligand complexes and compared with the unmodified truncated Newton algorithm, a quasi-Newton algorithm (LBFGS), and a conjugate gradient algorithm (CG+). In vacuum minimizations, the speedup of MSTN relative to the unmodified TN algorithm (TNPACK) depends on system size and the distance cutoffs used for defining the short- and long-range interactions and the long-range force updating frequency, but it is 4 to 5 times greater in the work reported here. This algorithm works best for the minimization of small portions of a protein and shows some degradation (speedup factor of 2−3) for the minimization of entire proteins. The MSTN algorithm is faster than the quasi-Newton and conjugate gradient algorithms by approximately 1 order of magnitude. We also present a modification of the algorithm which permits minimizations with a generalized Born implicit solvent model, using a self-consistent procedure that increases the computational expense, relative to a vacuum, by only a small factor (∼3−4).

Introduction

Minimization is a core functionality in protein molecular mechanics programs. Minimization of a protein structure taken from the Protein Data Bank (PDB) is a standard step in nearly every modeling effort, and an intrinsic component of many conformational search algorithms (e.g., Monte Carlo plus minimization[1−3]). Consequently, an algorithm that can substantially reduce the computational effort for minimization can have a major impact on studies of protein structure and protein−ligand interactions.

Efforts to optimize minimization algorithms have focused primarily on the algorithm that determines the size and direction of the geometry steps. These algorithms differ in the details of how they obtain a search step, and also what kind of information they require. For example, gradient-based methods only require the first derivatives, while Newton-type methods require the second derivatives. A number of different approaches exist, such as conjugate gradient,[4] Broyden−Fletcher−Goldfarb−Shanno (BFGS),[5] and truncated Newton (TN),[6−12] all of which have been shown to have advantages and disadvantages depending upon the context (for example, the size of the molecule to be optimized and how far one is from the minimum). However, less work

* Corresponding author e-mail: matt@cgl.ucsf.edu.
[†] Columbia University.
[‡] University of California.

10.1021/ct600129f CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/14/2007

has been done in attempting to reduce the cost of the energy/gradient evaluations required at each step. This is in contrast to molecular dynamics simulations where extensive effort has gone into reducing the computational effort per step, via such techniques as the multiple time scale method[13,14] and fast multipole method.[15]

In the case of minimization, the issue of reducing computational effort per step is ultimately tied to which algorithm should be used to generate the geometry steps. Newton−Raphson-type algorithms that utilize second derivatives usually have a faster convergence rate while requiring more storage space. A key issue is whether second-derivative information can be made inexpensive enough to use profitably for large systems. Brute force calculation of all second derivatives, followed by inversion of the Hessian to rigorously solve the Newton−Raphson equations, is grossly inefficient when dealing with the thousands or tens of thousands of coordinates present in a typical protein minimization problem. However, more sophisticated methods such as truncated Newton, in conjunction with preconditioned iterative algorithms for solving the approximate Newton−Raphson equations defined by the truncation, are much more promising. A novel implementation of the truncated Newton method, TNPACK of Schlick and co-workers,[8−12] has shown significant speed advantages for the minimization of macromolecular systems. Its protocol for advancing the geometry involves both an inner loop (iterative solution of the truncated Newton−Raphson equations) and an outer loop (each consisting of one truncated Newton−Raphson step, at which the energy and gradient have to be calculated). The use of even approximate second derivatives can dramatically reduce the required number of outer iterations as compared to competitors such as quasi-Newton methods or conjugate gradient methods that do not utilize such information.[10,16] Solving the Newton−Raphson equations inexactly also reduces the computational expense while retaining enough accuracy and thus rapid convergence. Thus, the critical issue is the cost of energy and gradient evaluations; if these can be done with sufficiently inexpensive approximations, very large gains as compared to alternative approaches can be realized.

In this paper, we present an integrated approach to the optimization of the truncated Newton approximation as applied to protein minimization. We examine both local minimization (in which a region of the protein is kept fixed—this situation arises frequently, for example, in modeling active sites and structure prediction of the side chains and loops) and global minimization of the entire protein. The basic idea is to utilize ideas developed in the context of molecular dynamics simulations, principally, multiple time- and length-scale approximations (which we refer to, compactly, as a "multiscale" methodology), to accelerate truncated Newton minimizations. Each aspect of the truncated Newton technology is examined in detail, and an optimized set of approximations is designed. Performance is evaluated for a substantial number of test cases by comparison with alternative approaches, as well as with that of the original TNPACK employed without the enhancements described herein.

In addition to presenting a methodology for optimization of a molecular mechanics energy function in the gas phase, we also adapt our approach to the minimization of a system in a continuum solvent, specifically, the surface generalized Born (SGB) model[17,18] that we have described previously in a number of publications.[19−24] The approximations required to handle the continuum solvation calculation efficiently, while employing the same general principles as in the gas phase, are notably different in details.

Methods

Multiscale Truncated Newton Minimization Algorithm. All Newton-type methods are based on approximating the objective function locally by assuming a quadratic model and minimizing or approximately minimizing that model. We denote by x_k the current approximation to the solution vector $\mathbf{x^*}$, and by g_k and H_k the gradient and Hessian evaluated at x_k, respectively. The new estimate for x_k is then obtained from the Taylor series expansion, up to quadratic terms, along a search direction p_k:

$$E(x_k + p_k) \approx E(x_k) + g_k{}^T p_k + \frac{1}{2} p_k{}^T H_k p_k$$

To find the search direction p_k, we find the minimum of the quadratic function by solving the equation

$$H_k p_k = -g_k$$

Subsequently, a line search or trust region method is applied to find the appropriate step length λ_k and generate the next iterate according the following equation:

$$x_{k+1} = x_k + \lambda_k p_k$$

Thus, given a starting point x_0, a series of iterates $\{x_0,x_1,...,x_k,x_{k+1},...\}$ will be generated until some convergence criterion is satisfied.

The truncated Newton method distinguishes itself from other Newton-type methods by solving the Newton equation inexactly. Formally, a truncation criterion is applied, as in TNPACK:[10]

$$||H_k p_k + g_k|| \leq \Phi_k ||g_k||$$

This truncation criterion is easily satisfied at regions distant

$$\Phi_k = \min\left\{\frac{1}{k}, \left\|g_k\right\|\right\}$$

from the local minimum; as a local minimum is approached, the condition becomes more stringent, and this leads to increasingly accurate solutions for the search direction. The truncation is justified by the approximate nature of the Newton method at each iteration.

Figure 1 schematically depicts the TN algorithm. Each iteration of the outer loop serves to choose a search direction and make a move along it. The inner loop solves for the search direction using a preconditioned conjugate gradient (PCG) algorithm, which is also an iterative process. Once the search direction is chosen, a one-dimensional line search iteratively determines the step size along the search direction.

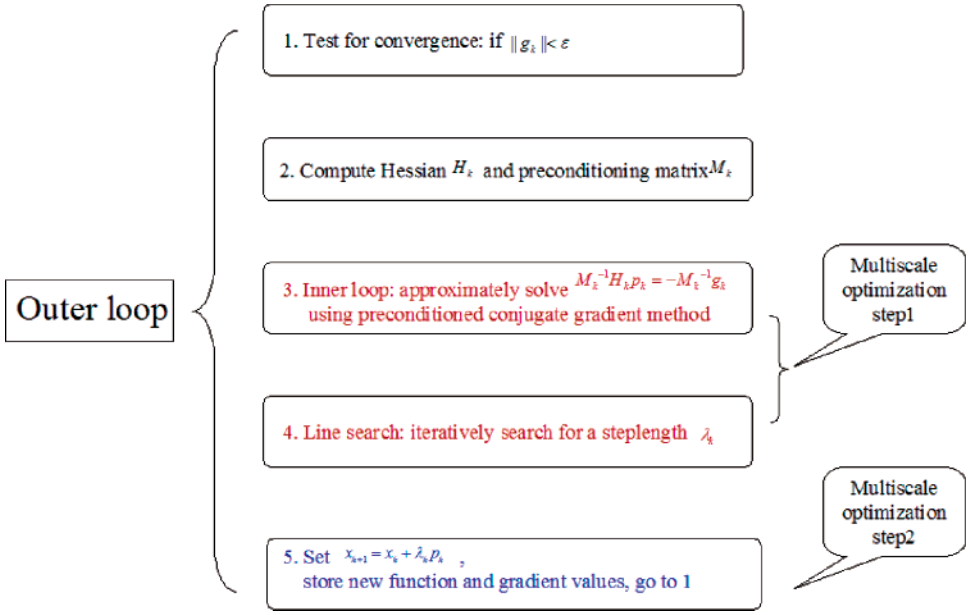


Figure 1. Schematic of the multiscale optimization of the truncated Newton implementation. Multiscale optimization step 1 reduces the energy and gradient evaluation costs of the PCG inner loop and line search, while step 2 optimizes those of the outer loop.

The key innovation in our multiscale optimization is the division of the molecular mechanics interactions into short- and long-range components, in analogy to multiscale molecular dynamics methods such as RESPA.[13,14] That is,

$$E = E_s + E_l$$

$$g = g_s + g_l$$

$$H = H_s + H_l$$

(Subscripts 1 and s refer to the long and short components, respectively.) The short-range forces contain all covalent interactions and nonbonded interactions between atoms separated by small distances, and the long-range interactions include all other nonbonded interactions. The detailed partitioning of the interactions is discussed below. This division can significantly improve efficiency because (1) the slowly varying long-range components can be updated less frequently than short-range components and (2) the long-range components can require significantly greater computational expense than the short-range components, depending on the distance cutoffs that are used. Specifically, we invoke the following approximations:

- $H_l = 0$. The TN algorithm uses only a sparse approximation to the Hessian to aid convergence in the PCG inner loop. We use only the covalent terms and 1−4 interactions from the molecular mechanics force field in the Hessian preconditioner.
- The long-range component of the gradient, g_l, is updated infrequently, as described in detail below.
- The long-range component of the energy, when it is not computed directly, is updated from the long-range component of the gradient according to $E_l = g_l'(x - x') + E_l'$, where x', g_l', and E_l' represent the coordinates, gradient, and long-range energy, respectively, from the most recent update of the long-range interactions.

We apply the multiscale optimization to all components of the TN algorithm that require iterative evaluations of the energy and gradient. In the PCG inner loop and the line search, we never update the long-range gradient. The PCG inner loop only approximately solves for the Newton−Raphson search direction, regardless of any approximations to the energies and gradient. Likewise, the line search only approximately minimizes the energy, using truncation criteria to terminate the one-dimensional search. Thus, the additional approximations invoked by assuming that the long-range gradient is constant can in principle affect the number of outer steps required for convergence but do not affect accuracy. It should also be noted that the step size in the line search becomes small as the minimization converges on a local minimum, which means that approximating the long-range gradient as constant in the inner loop and line search becomes increasingly accurate. In the outer loop, we update the long-range energy and gradient periodically; the number of outer steps between updating the long-range forces is an adjustable parameter, discussed further below.

It is noted that TNPACK uses the local energy terms to approximate its Hessian preconditioner in the PCG inner loop. In some sense, this is similar to our approximate treatment of gradient and energy in the PCG inner loop. However, there is a significant distinction between these two approximations. The preconditioner is only a tool for accelerating convergence and does not alter the function and gradient values being provided to define the minimization task in the linear system, while the multiscale approximation does. Thus, standard convergence criteria that guarantee convergence for the original algorithm cannot be directly applied, and behavior can only be assessed in practice in comparison to the unmodified TNPACK, as well as other

minimization algorithms, as we discuss below. In principle, it would be possible to use the multiscale approximation in early parts of the minimization but revert to unmodified TNPACK as a minimum is approached. However, in practice, we have found this to be unnecessary, at least with a convergence criterion of 0.001 kcal/mol/Å root-mean-squared gradient (RMSG).

Clearly, the performance of this optimization strategy depends on the definition of long- and short-range forces and the updating frequency of the long-range energy and gradient in the outer loop. The overall efficiency is a balance between minimizing the computational expense associated with the energy evaluations, by treating as many nonbonded interactions as possible as "long-range" and updating these infrequently, and minimizing the number of outer loop iterations, which will increase if the approximations employed in the energy evaluations are too severe. We have addressed this tradeoff by empirically optimizing the updating frequency for the long-range forces and the division of molecular mechanics forces into short- and long-range. In this work, we use residue-based cutoffs for distinguishing between short- and long-range nonbonded interactions. Short-range forces include all bond, angle, torsion, and 1−4 nonbonded interactions. The nonbonded interactions are partitioned into short- and long-range using distance cutoffs which depend on the amino acid types involved in the interaction. We employ an absolute cutoff of 30 Å for charged−charged residue pairs, 20 Å for charged−neutral, and 15 Å for neutral−neutral, with no smoothing. That is, all interactions beyond these distance cutoffs are ignored, because they are small. The distance cutoffs for partitioning the three types of nonbonded interactions into short- and long-range are adjustable, as discussed below. The all-atom optimized potential for liquid simulations (OPLS) force field[25,26] is used for all tests performed here. Parameters for the ligands were obtained using atom-typing capabilities provided in IMPACT.[27]

Our baseline implementation of the minimization algorithm TNPACK (Algorithm 702 in the ACM Digital Library)[7−12] uses default values for most parameters unless otherwise noted. Specifically, we utilize a residual-based truncation criterion of 0.25 for the inner PCG loop and the unconventional modified Cholesky factorization. Hessian-vector products are obtained by finite difference. We assess convergence of the minimization exclusively using the root-mean-squared gradient, calculated over all degrees of freedom included in the minimization. All minimizations are performed in Cartesian coordinates in double precision. The multiscale truncated Newton (MSTN) implementation made minimal modifications to the TNPACK Fortran source code to enable the multiscale optimization but otherwise used unmodified TNPACK with the same parameters.

Other Minimization Algorithms Used. In addition to assessing the impact of our multiscale modifications to truncated Newton, we also make compararisons to a quasi-Newton minimization algorithm (LBFGS)[5] and a conjugate gradient algorithm (CG+).[4] For these two software packages, we use the default parameters and do not attempt further optimization. The quasi-Newton algorithm is used, retaining information from the prior seven steps, and without preconditioning. The conjugate gradient algorithm is used with a positive Pollak−Ribiere update.

Minimization with Generalized Born Solvent. The generalized Born (GB) implicit solvation model is well-suited for performing rapid minimizations because the solvent-induced screening between a pair of charges is treated using an analytical formula:

$$\Delta G_{ij} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2\, e^{-D_{ij}}}}$$

where q_k is the partial charge on atom k, ϵ is the dielectric constant of the solvent, α_k is the Born α radius for atom k, r_{ij} is the distance between a pair of atoms i and j, $\alpha_{ij} = \sqrt{\alpha_i \alpha_j}$, and $D_{ij} = r_{ij}^2/(2\alpha_{ij})^2$. The Born α's can be computed in a variety of ways, including analytical expressions and surface or volume integration, depending on the specific implementation.

The GB pair term is of course differentiable with respect to atomic coordinates, but the resultant expression for the gradient involves derivatives of the Born α's with respect to the atomic coordinates, which must be determined numerically for GB models that calculate the Born α's by a grid-based integration. Because these derivatives are expensive to compute, we employ a self-consistent minimization, in which the Born α's are held fixed during the course of the minimization, then updated prior to another minimization, and so on until the energy ceases to decrease by more than 1 kcal/mol. This threshold is precise enough for most protein modeling purposes. In practice, self-consistency rarely requires more than two to three cycles of TN minimization, and the second and subsequent minimizations are generally very fast.

The GB implementation used here is based on the surface generalized Born implementation of Ghosh et al.[18] The updating of Born α's is quite expensive and can be the bottleneck of the simulation if there are a large number of conformations that need to be energy-minimized. In cases where only a portion of the protein or protein−ligand complex is minimized, our implementation updates only those portions of the surface and surface integrals that change.

The GB pair terms are partitioned into the short- and long-range interactions using the same criteria described above for the nonbonded interactions. The GB pair terms corresponding to atoms separated by three bonds or less are also included in the sparse Hessian preconditioner.

Results and Discussion

Test Set. One of the simplest macromolecular applications of a minimization algorithm is to simply energy-minimize an entire protein, for example, in preparation for molecular dynamics. We have chosen a set of 20 proteins for such a test, with 1162−5929 atoms. The PDB codes for these proteins are 1HOE, 2HSP, 1ECI, 1CCD, 1J8B, 3CRD, 1AB2, 1HPW, 1BUO, 1HT9, 1BZS, 1BHD, 1EJF, 1HYT, 1NJ4, 1CVM, 1DIM, 1EM2, 1AXN, and 2IF1. However, in the context of homology modeling and other modeling applica-

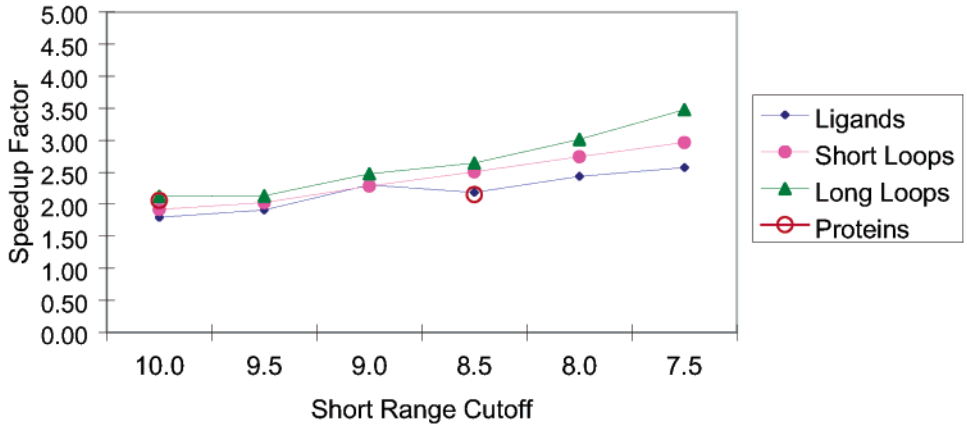


Figure 2. Average speedup factors of MSTN, with the multiscale optimization applied only to the inner loop, relative to unmodified TN on the minimization of ligands, short loops, long loops, and full proteins. The decrease in computational expense due to inner loop optimization is related primarily to the ratio of computational expense for determining the short- and long-range forces, which in turn is determined by the distance cutoff for partitioning the nonbonded interactions into short- and long-range. In these tests, the minimizations are considered converged when the RMS gradient decreases below 0.001 kcal/mol/Å. Details of the test cases used can be found in Tables 2 and 3 and in the Supporting Information.

tions, it is also common to energy-minimize relatively small portions of a protein, ranging from single side chains to loops or secondary structure elements, that is, in the context of side chain,[19,20] loop,[21,22] and helix[24] optimization. For this reason, we have also chosen a set of 25 proteins loops with 13-residue lengths from our previous work.[22] To study the possible effects of loop size, we also cut these loops in half and composed another 25 protein loops with six-residue lengths. Finally, in the context of small molecule docking and the estimation of protein−ligand relative binding affinities,[23,36−38] it is frequently desirable to energy minimize a ligand in a rigid protein receptor, and we have performed such a test on 20 ligands in a single protein receptor. Altogether, the test cases range from 51 to 17 787 degrees of freedom. All computations are performed using single PIII 1.4 GHz processors.

Inner Loop versus Outer Loop Optimization. As discussed in the Methods section, the multiscale optimization is applied in both the inner and outer loops of the truncated Newton algorithm. To assess the effects of our optimization, we perform it in two steps. In the first step, we apply the multiscale optimization only to the PCG inner loop and line search, where the long-range interactions are never updated, as discussed in the Methods section. In the second step, we also apply the multiscale optimization to the outer loop by only updating the long-range forces periodically. This is the full optimization, denoted by MSTN. We evaluate the decrease in computational expense to reach a local minimum relative to the unmodified truncated Newton minimization, which is denoted by TN. In this work, we use exclusively the RMSG as the convergence criterion for the minimization (0.01, 0.001, and 0.0001 kcal/mol/Å in various tests).

The application of the multiscale optimization to the inner loop reduces the cost per Newton−Raphson step. As shown in Figure 2, the decrease in computational expense due to inner loop optimization is related primarily to the ratio of computational expense for determining the short- and long-range forces, which in turn is determined by the distance cutoff for partitioning the nonbonded interactions into short- and long-range. For the ligand, short loop, and long loop test sets, the decrease in computational expense improves as this distance cutoff decreases, down to 7.5 Å. Below 7.5 Å, many minimizations have serious convergence problems. That is, the approximation that the search direction and line search stepsize can be determined solely by the short-range interactions begins to break down. For the full protein minimizations, the convergence problems become serious when the shor/long-range cutoff is decreased below 8.5 Å, as discussed further below. In all of the following results, the short/long-range cutoff is fixed to 7.5 Å for the ligand, short loop, and long loop test sets and 10 Å for the full protein minimizations.

The application of the multiscale optimization to the inner loop accounts for most of the reduction in the computational expense of MSTN relative to TN. As shown in Figure 3, some additional decrease in computational expense can be obtained by also only periodically updating the long-range forces in the outer loop. However, when the long-range gradient is updated less frequently than once every five outer iterations, some of the ligand and loop minimizations start to converge more slowly. Thus, we use an outer-loop updating frequency of every five steps in the remaining MSTN results reported here. The full protein minimizations require more frequent updating to remain stable, and we use an updating frequency of two in the results below.

Computational Expense of MSTN Relative to TN and Other Minimization Algorithms. The results using the optimized set of parameters are summarized in Table 1. Details of the test cases are reported in Tables 2 and 3 and in the Supporting Information. The effect of the multiscale optimization is to decrease the computational expense of the truncated Newton minimization by a factor of approximately 4 for the ligand, short loop, and long loop minimizations. The acceleration is somewhat less for full proteins, as discussed in more detail below.

It is known that Newton-type minimization algorithms have a quadratic convergence rate; that is, they converge much more quickly as they approach a local minimum.

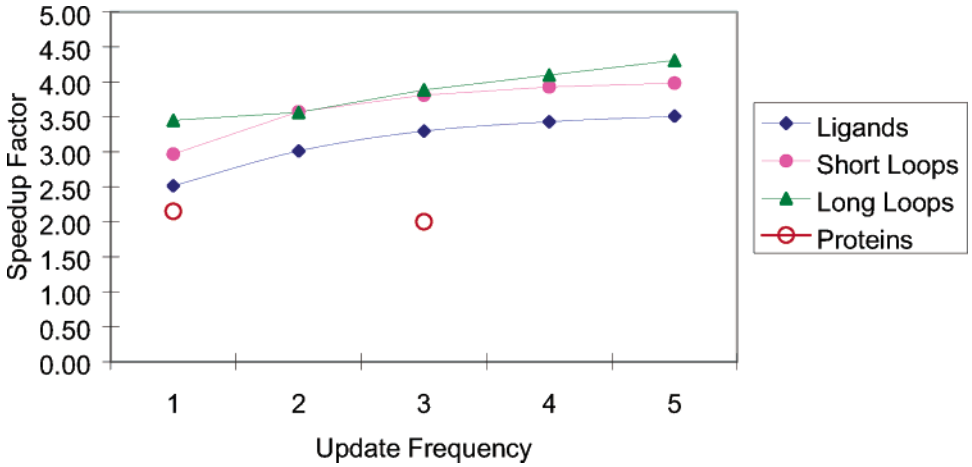


Figure 3. Speedup factors of MSTN, with the multiscale optimization applied to both the inner and outer loops, relative to unmodified TN on the minimization of ligands, short loops, long loops, and full proteins. The decrease in computational expense when applying the multiscale optimization to the outer loop is determined primarily by the number of outer-step iterations between updates to the long-range gradient ("update frequency" on the *x* axis). In these tests, the minimizations are considered converged when the RMS gradient decreases below 0.001 kcal/mol/Å, and the distance cutoff for partitioning the nonbonded interactions into short- and long-range is kept fixed at 7.5 Å (8.5 Å for proteins, see text for discussions). Details of the test cases used can be found in Tables 2 and 3 and in the Supporting Information.

Table 1. Speedup Factors of MSTN Relative to Unmodified TN, LBFGS, and CG+ with Different Convergence Criteria (Defined by the Root-Mean-Squared Gradient, in Units of kcal/mol/Å)

	different convergence criteria of RMSG (kcal/mol/Å)								
	TN/MSTN			QN/MSTN			CG/MSTN		
	0.01	0.001	0.0001	0.01	0.001	0.0001	0.01	0.001	0.0001
ligands	3.8	3.5	3.6	14.3	14.9	15.0	14.5	17.2	19.5
short loops	4.0	4.0	4.0	20.0	21.2	22.9	17.2	22.0	27.9
long loops	3.8	4.3	4.3	12.6	13.8	16.2	14.1	17.5	25.1
proteins	2.6	2.0	N/A	2.5	2.5	N/A	3.6	5.1	N/A

Truncated Newton methods, although they solve the Newton−Raphson equations approximately, largely retain this advantage. Although our optimization introduces additional approximations that alter the function and gradient, and hence standard convergence results cannot be applied, Table 1 shows that the speedup factors of MSTN do not vary significantly when three different convergence criteria, from loose to tight, are used. We can see from the table that MSTN has rather stable speedup factors and does not degrade even when the tightest convergence criterion of 0.0001 kcal/mol/Å root-mean-squared gradient is used.

The acceleration of MSTN relative to CG+ (hereafter CG) and LBFGS (hereafter QN) minimization, using the same energy function, is at least 1 order of magnitude for the ligand and loop test sets and is a somewhat more modest factor of 3−5 for the full proteins, as shown in Table 1. The QN minimization is generally somewhat faster than CG, and this advantage is more prominent when the convergence criterion is tighter, that is, 0.0001 kcal/mol/ Å.

Effects of Multiscale Optimization Other Than Decreasing Computational Expense. Up to this point, we have focused exclusively on the total CPU time required to converge to a local minimum. The number of iterations (outer loops) of the TN minimization can give a sense of whether the approximations employed in MSTN affect the rate of convergence (although the overall computational expense is of course less). As shown in Table 2 and the Supporting Information, for the ligand and loop minimizations, the numbers of iterations to reach convergence for TN and MSTN are nearly identical, indicating that not updating the long-range forces in the inner loop does not strongly affect convergence. A comparison of the corresponding minimization trajectories also shows that they are nearly identical. Full protein minimizations are discussed below.

Another question is whether MSTN finds different local minima than TN, when starting from the same initial conformation. It is important to emphasize that identifying the same local minima is not a requirement for the success of MSTN. It is well-known that different algorithms often converge to different local minima if the energy surface is very complex, even when starting from the same point. For example, the truncated Newton method can identify different local minima than those found by the full Newton method for a complex energy function. Indeed, in our tests, minimizations starting from identical initial states can diverge when running on different processors or with executables built with different compilers (this is true for several different minimization algorithms including TN).

For the ligand minimizations, the optimized MSTN and original TN algorithms identify minima with almost exactly the same final energy and structures. Usually, the minimized structures differ from each other only in the second or third

Table 2. Comparison of MSTN with TN on Short Loops in Terms of Number of Iterations, Final Energy, Final Gradient (kcal/mol/Å), ΔE (kcal/mol), and RMSD (Å)[a]

			starting grams	time (s)		iterations		final gradient		ΔE (kcal/mol)	RMSD (Å)
				TN	MSTN	TN	MSTN	TN	MSTN		
1ojq	A:167	A:172	8.38	20.44	5.04	36	37	0.00001	0.00001	0.00	0.01
1dpg	A:352	A:357	7.10	6.26	1.52	16	16	0.00001	0.00056	0.01	0.02
1xyz	A:645	A:650	8.82	6.85	2.67	18	27	0.00051	0.00008	0.01	0.01
1eok	A:147	A:152	9.71	1.98	0.51	9	9	0.00019	0.00007	0.00	0.01
1p1m	A:327	A:332	7.93	9.26	2.66	19	20	0.00025	0.00014	0.00	0.00
1ock	A:43	A:58	7.06	30.17	7.09	25	27	0.00022	0.00002	0.00	0.00
1hnj	A:191	A:196	7.65	6.35	1.44	27	27	0.00015	0.00011	0.14	0.06
1o6l	A:386	A:391	7.30	19.15	3.64	22	19	0.00002	0.00064	0.00	0.01
1bkp	A:51	A:56	15.47	5.16	1.33	15	16	0.00033	0.00004	0.00	0.01
1f46	A:64	A:69	7.91	1.88	0.43	10	9	0.00004	0.00015	0.00	0.00
1jp4	A:153	A:158	13.42	4.77	1.37	15	16	0.00019	0.00004	0.03	0.03
1nln	A:26	A:31	15.60	8.28	2.21	18	16	0.00010	0.00063	0.00	0.00
1kbl	A:793	A:798	9.72	9.83	2.11	19	18	0.00002	0.00053	0.00	0.00
1l8a	A:691	A:696	8.28	22.80	5.55	31	32	0.00015	0.00007	0.07	0.03
1cnv	_:110	_:115	8.56	23.25	5.47	42	41	0.00002	0.00015	0.00	0.00
1mo9	A:107	A:112	9.75	7.17	2.13	18	21	0.00045	0.00005	0.01	0.01
1gpi	A:308	A:313	13.30	3.50	0.84	21	17	0.00041	0.00082	0.00	0.00
1lki	_:62	_:67	7.96	4.47	1.03	17	17	0.00011	0.00037	0.00	0.01
1qqp	2:161	2:166	10.11	2.81	0.70	10	10	0.00015	0.00019	0.00	0.00
1qs1	A:389	A:394	7.06	12.56	2.91	16	15	0.00012	0.00002	0.00	0.01
1d0c	A:280	A:285	11.47	4.67	1.28	21	22	0.00046	0.00009	0.00	0.00
1krh	A:131	A:136	7.92	4.01	1.00	16	15	0.00006	0.00030	0.02	0.02
2hlc	A:91	A:96	9.20	2.42	0.62	11	11	0.00009	0.00006	0.00	0.00
1ako	_:203	_:208	7.36	4.25	1.21	23	23	0.00006	0.00060	0.03	0.00
1ed8	A:67	A:72	15.17	2.63	0.71	10	10	0.00006	0.00007	0.00	0.00

[a] ΔE is defined as the energy of the MSTN structure minus the energy of the corresponding TN structure. The RMSD is calculated using all heavy atoms of the loop region.

decimal places of the Cartesian coordinates. For the loops, the final energies are usually slightly different and, occasionally, several kilocalories per mole for long loops. Table 2 shows a summary for the short loop test set. We calculate the root-mean-square deviation (RMSD) between the local minima identified by MSTN and TN and find that these RMSDs are mostly negligible. A few structures in the long loop set have significant RMSDs up to 0.5∼1 Å, and these minimizations also have quite different final energies.

To quantify the divergence of MSTN and TN, we also compare the final energies and structures minimized by LBFGS and CG+ with TN, as shown in the Supporting Information. Similar to the comparison of MSTN and TN, LBFGS and CG+ frequently converge to the same local minima as TN. The number of cases where significant divergence is seen (i.e., different local minima as indicated by different final energies and RMSDs) between LBFGS/CG+ and TN is about the same as that in the comparisons of MSTN and TN, but the magnitude of the divergence is often greater.

Minimization of Full Proteins. The minimizations of MSTN on full proteins are quite different compared with the ligand and loop minimizations. MSTN and TN minimizations seldom converge to the same energy minima; they can take very different numbers of iterations, and the corresponding minima have significant RMSDs. This is due to the fact that the energy surface of an entire protein is so complex enough that there are many available local minima. On average, the MSTN minimization reduces the computational expense by a nontrivial factor (roughly 2), but for a small number of cases, the MSTN minimization actually requires greater CPU time to converge than unmodified TN. Some minimizations fail at the line search because no acceptable step size can be found. This situation also happens with unmodified TN, but less frequently. Simply restarting the minimization solves this problem. To obtain a meaningful comparison, we add the two times together.

Altogether, it is clear that minimizations of entire macromolecules cannot be subjected to the same level of multiscale optimization as minimizations of relatively small portions of the macromolecule, or small molecules interacting with macromolecules. We note that the advantage of unmodified TN relative to less sophisticated minimization algorithms such as CG and LBFGS is also less for full protein minimizations than for minimizations with a smaller number of degrees of freedom. As Table 1 shows, unmodified TNPACK is 3−5 times faster than LBFGS and CG+ for the test sets of ligands and loops, but it is only slightly faster than LBFGS and 1−2 times faster than CG+ for entire proteins. We speculate that this behavior may relate to the way the convergence properties of the PCG inner loop vary with the number of degrees of freedom or the complexity of the function being minimized.

Crystal Environment. We have applied MSTN in a series of applications, including side-chain optimization,[19,20] loop prediction,[21,22] helix prediction,[39] and estimation of protein−

Table 3. Minimization of Ligands in Generalized Born Solvent[a]

ligand (KEGG ID)	degrees of freedom	iterations for ddd	time for ddd(s)	iterations for SGB	pure min (s)	α update (s)	SGB total/ddd
C02717	126	12	0.85	15	1.80	1.11	3.40
C09321	87	13	0.54	16	0.96	0.86	3.35
C04303	156	29	1.34	24	3.65	1.17	3.60
C02006	102	16	0.65	19	1.33	1.00	3.61
C00355	75	9	0.26	24	1.16	0.87	7.70
C01302	108	51	2.93	22	2.39	1.04	1.17
C05364	51	7	0.17	10	0.39	0.77	6.97
C05662	63	10	0.28	13	0.60	0.81	4.95
C00852	126	23	1.33	37	4.01	1.38	4.07
C04771	72	8	0.20	10	0.42	0.87	6.35
C01850	123	10	0.62	15	1.35	1.09	3.92
C03300	138	33	2.44	18	1.79	1.14	1.20
C05983	132	18	1.01	21	1.86	1.11	2.94
C05401	105	11	0.48	27	2.12	0.94	6.41
C06019	81	18	0.76	30	2.17	0.84	3.96
C09126	153	46	2.91	62	6.55	1.49	2.77
C04711	93	10	0.34	14	0.86	0.97	5.48
C04194	222	31	5.57	20	4.89	1.23	1.10
C06585	78	14	0.37	16	0.55	0.65	3.22
C04498	123	8	0.43	10	0.86	0.92	4.12
Average							4.02

[a] The ligands are docked into the binding site of methylaspartate ammonia lyase (PDB ID: 1kkr) using Glide[40−42] and then minimized to a 0.001 kcal/mol/Å RMS gradient. The ligands are taken from the Kyoto Encyclopedia of Genes and Genomics (KEGG) database, which contains metabolite ligands, toxins, inhibitors, and pollutants.[43] The times are divided into the computational expense of the minimization steps (pure min) and the expense of updating the Born α's between each minimization in the self-consistent procedure (α update). The right column is the ratio of the total computational expense of minimization in a GB solvent versus in a vacuum.

ligand relative binding affinities.[23,36−38] In some of these applications, we have explicitly modeled the crystal environment in order to make a more direct and realistic comparison to the structure determined by X-ray crystallography. In these applications, we create and update the symmetric copies according to the space group information of the PDB file for every putative candidate during the minimization. The inclusion of crystal packing effects in this way leads to a larger computational cost, because the nonbonded energy evaluations extend over multiple copies of the asymmetric unit. Therefore, it is interesting to see the performance of MSTN compared with TN when the crystal environment is included. Using a convergence criterion of 0.001 kcal/mol/Å, the average speedup factors of MSTN2 on short and long loops are 5.0 and 5.4, respectively. Without the crystal environment, these two factors are 4.0 and 4.3, respectively. The increased speedup factors clearly occur because inclusion of the crystal packing environment increases the number of long-range interactions more greatly than the number of short-range interactions.

Minimizations in Generalized Born Implicit Solvent. The computational costs of the minimizations in implicit solvent are divided into the cost of the minimization steps and the cost of updating the Born α's between iterations of the self-consistent procedure. The latter, which requires updating and integrating over a large portion of the protein surface, can be nontrivial compared with the energy evaluations. For the four test sets, the total computational costs of minimization in implicit solvent are, on average, 4.0, 3.2, 2.7, and 2.4 times greater than in a vacuum. Table 3 shows the details for the ligand minimizations in a SGB solvent; the results for the other tests sets are shown in the Supporting Information. It is important to keep in mind that the minimizations in a vacuum and implicit solvent do not converge to the same local minimum; that is, the energy surface is different. The computational cost of each force evaluation is approximately 50% greater when using GB, because of the additional cost of calculating the pair term. The overall computational expense of the minimization steps in GB solvent is greater than this largely because of the cost of updating the Born α's.

Conclusions

We have demonstrated a simple method for accelerating macromolecular minimizations on the basis of the partitioning of forces into short- and long-range components. As with similar multiscale methods applied to molecular dynamics, the long-range forces are updated less frequently than the short-range forces. In principle, this simple idea can be applied to several different classes of gradient-based minimization algorithms, but we have focused on implementing it for the powerful truncated Newton method as implemented in the TNPACK package. The acceleration of our multiscale implementation depends on the exact partitioning of the forces into short- and long-range components and updating strategy. In the work reported here, the speedup relative to truncated Newton without the multiscale implementation is about 4−5 for a number of systems ranging from ligands to 13-residue-long loops. For entire proteins, this algorithm works less well but still shows a speedup factor of roughly 2. We also compare to commonly used conjugate gradient and quasi-Newton methods, using the same energy function

but no multiscale implementation; the MSTN method is faster by 1 order of magnitude.

Finally, we have implemented a self-consistent procedure for minimizations in a generalized Born implicit solvent, which increases the computational expense, relative to a vacuum, by only a factor of ~3. This self-consistent procedure can be understood as another multiscale approximation, which decomposes the solvent "forces" in the generalized Born model into rapidly varying (short-range pair term) and slowly varying (long-range pair term and self-term) components.

We have already deployed the MSTN technology in multiple applications, including side-chain optimization,[19,20] loop prediction,[21,22] helix prediction,[39] and estimation of protein−ligand relative binding affinities.[23,36−38] In each of these applications, we use the same energy function as in the work reported here and enumerate hundreds or thousands of local minima on the energy surface. Minimization is a rate-limiting step in each application, and the efficiency of the MSTN algorithm has been critical to the success of these works. Further optimization of the MSTN method may be possible through more rigorous theoretical and empirical studies of the convergence properties than we have attempted here.

Acknowledgment. This work was supported by NSF Grant 0346399 to M.P.J. and by NIH Grants GM52018 to R.A.F. and GM071790 to M.P.J. M.P.J is a consultant to Schrödinger, Inc.

Supporting Information Available: Detailed computational data for each of the 40 test cases, including comparisons of MSTN with TN, LBFGS, and CG+ in terms of computational time, number of iterations, final structures, final energies, and final gradients; detailed comparisons of MSTN in a vacuum and in a generalized Born implicit solvent for each test case. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611−6615.

(2) Pillardy, J.; Arnautova, Y. A.; Czaplewski, C.; Gibson, K. D.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12351−12356.

(3) Trosset, J. Y.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8011−8015.

(4) Gilbert, J. C.; Nocedal, J. *SIAM J. Optimiz.* **1992**, *2*, 21.

(5) Liu, D. C.; Nocedal, J. *Math. Prog.* **1989**, *45*, 503−528.

(6) Dennis, J. E. J.; Schnabel, R. B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*; Prentice Hall: Englewood Cliffs, NJ, 1983; pp 239−255.

(7) Schlick, T. Optimization Methods in Computational Chemistry. *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publisher: New York, 1992; Vol. III, pp 1−71.

(8) Schlick, T. *SIAM J. Sci. Stat. Comput.* **1993**, *14*, 424−445.

(9) Schlick, T.; Fogelson, A. *ACM T. Math. Software* **1992**, *18*, 141.

(10) Schlick, T.; Overton, M. *J. Comput. Chem.* **1987**, *8*, 1025−1039.

(11) Xie, D.; Schlick, T. *ACM T. Math. Software* **1999**, *25*, 108−122.

(12) Xie, D.; Schlick, T. *SIAM J. Optimiz.* **1999**, *10*, 132−154.

(13) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1991**, *94*, 6811−6815.

(14) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.* **1992**, *97*, 1990−2001.

(15) Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1997**, *135*, 280−292.

(16) Derreumaux, P.; Zhang, G.; Brooks, B.; Schlick, T. *J. Comput. Chem.* **1994**, *15*, 532−552.

(17) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2001**, *23*, 517−529.

(18) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.

(19) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. *J. Mol. Biol.* **2002**, *320*, 597−608.

(20) Jacobson, M. P.; Kaminski, G. A.; Friesner, R. A.; Rapp, C. S. *J. Phys. Chem. B* **2002**, *106*, 11673−11680.

(21) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins* **2004**, *55*, 351−367.

(22) Zhu, K.; Pincus, D. L.; Zhao, S.; Friesner, R. A. *Proteins* **2006**, *65*, 438−452.

(23) Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. *Biochemistry* **2005**, *44*, 2059−2071.

(24) Li, X.; Jacobson, M. P.; Friesner, R. A. *Proteins* **2004**, *55*, 368−382.

(25) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(26) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(27) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2005**, *26*, 1752−1780.

(28) Sherman, W.; Day, T. J. F.; Jacobson, M. P.; Friesner, R. A.; Farid, R. *J. Med. Chem.* **2006**, *49*, 534−553.

(29) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. *J. Chem. Inf. Model.* **2006**, *46*, 243−253.

(30) Bernacki, K.; Kalyanaraman, C.; Jacobson, M. P. *J. Biomol. Screen.* **2005**, *10*, 675−681.

(31) *Glide*; Schrodinger Inc: New York, 2004.

(32) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(33) Goto, S.; Okuno, Y.; Hattori, M.; Nishioka, T.; Kanehisa, M. *Nucleic Acids Res.* **2002**, *30*, 402−404.

CT600129F

JCTC Journal of Chemical Theory and Computation

Theoretical Study of the Photochemical Isomerization of Colchicine

Ivo Cacelli,*,† Maurizio D'Auria,‡ and Vincenzo Villani‡

Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, 56126 Pisa, Italy, and Dipartimento di Chimica, Università della Basilicata, Via N. Sauro 85, 85100 Potenza, Italy

Received October 17, 2006

Abstract: The photochemical reaction of colchicine to β- and γ-lumicolchicine, through a mechanism involving a disrotatory cyclization, is studied by theoretical methods. The energetics of the reaction, including one or two methanol solvent molecules, are studied at the DFT-B3LYP and multireference perturbation levels of theory using the 6-31G(d) basis set. The results show that, in agreement with experimental results, the first excited state of colchicine at ∼3.6 eV can lead to both β- and γ-lumicolchicine, whose energy is about 15 kcal mol^{-1} above the colchicine energy. Owing to the high steric tension of the condensed four- and five-atom rings arising from cyclization, the two *trans*-lumicolchicines are higher in energy (>60 kcal mol^{-1}), and their formation appears much less probable. A partial inclusion of the solvent effects through the addition of two solvent molecules does not alter the general conclusions based on the free energy in the gas phase. The photochemical reaction path is studied by choosing the distance between the two carbon atoms which form the new σ bond as the leading coordinate of the minimum-energy path of both the ground and the first singlet excited states. The energies are computed by Multi Configurational self-consistent-field calculations on a model molecule, retaining those atoms that presumably play an active role in the reaction. A reasonable mechanism starting from colchicine in the first singlet excited state and leading to γ-lumicolchicine is proposed. On the contrary, a high-energy transition state is found for *trans*-lumicolchicines, whose formation, although not strictly forbidden for energetic reasons, appears to be rather improbable.

Introduction

Colchicine (CC; **1**) is the main alkaloid of the poisonous plant meadow saffron (*Colchicum autumnale* L.),[1] a common plant of European and North African origin that flowers in autumn on a leafless stalk. Besides their use as a poison, the active ingredients of *Colchicum* species belong to the oldest known drugs and have been used for more than 2000 years in the treatment of acute gout. Colchicine (**1**) was first isolated in 1820 by Pelletier and Caventou[2] and is an important bioactive compound used in the treatment of a broad variety of diseases and still remains the sole drug for the therapy of acute gout and familial Mediterranean fever. Moreover, colchicine acts an antimitotic agent by binding to tubulin: it distorts the tubulin/microtubule equilibrium in such a manner that mitosis is arrested in metaphase. Therefore, this compound can be used to selectively damage rapidly proliferating cancer cells.

The irradiation of colchicine leads to the formation of β- and γ-lumicolchicine (β- and γ-LCC; **3** and **4**; Chart 1). Prolonged irradiation times lead to the formation of α-lumicolchicine (**2**) (Chart 1).[2−16]

In a previous study, we reported a pump and probe spectroscopic study on colchicine photoisomerization.[17] After the first excited state was populated by means of irradiation at 360 nm, femtosecond transient spectroscopy showed an

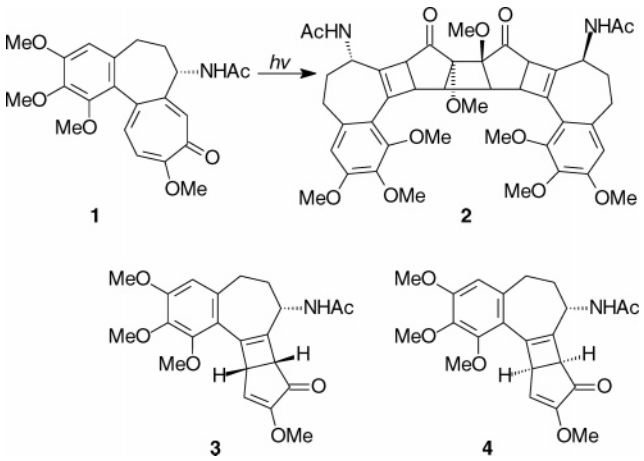
* Corresponding author tel.: +39 050 2219-213; fax: +39 050 2219-260; e-mail: ivo@dcci.unipi.it.

† Università di Pisa.

‡ Università della Basilicata.

10.1021/ct600306t CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/02/2007

Chart 1. The Photoisomerization of Colchicine



instantaneous strong band with a maximum at 510 nm which disappeared within a few hundred femtoseconds, leaving a broad, structureless band with a maximum around 470 nm. A second band was observed around 410 nm. The analysis in time showed that the 510 nm component appeared instantaneously and decayed following a biexponential law with time constants of 300 ± 100 fs and 40 ps. The kinetics studied by a probe wavelength of 420 nm showed a measurable rise time of 300 ± 150 fs. Transient spectroscopy, as well as theoretical calculations, was in agreement with a mechanism involving a disrotatory cyclization of colchicine in its first excited singlet state to give β- and/or γ-lumicolchicine. The triplet states were considered to play no active role in the process.

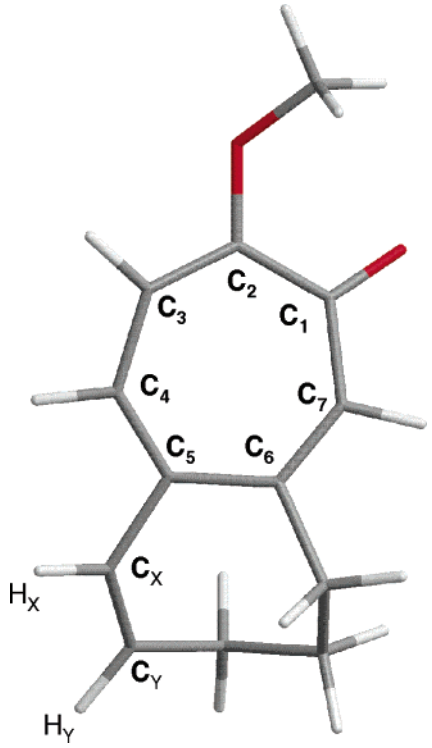
In this paper, we report the results of extensive theoretical calculations on colchicine and on the four possible isomers of lumicolchicine. The first study is concerned with the free-energy calculations of the involved molecules; the second step focuses on the decay mechanism from the first excited singlet state of colchicine to the photochemical products.

Method of Calculation

The geometry of the ground state of colchicine and lumicolchicines was optimized without symmetry constraints, both by the Hartree−Fock (HF) method and by density functional theory (DFT) in the B3LYP implementation[18,19] using the 6-31G* basis, which includes a *d* polarization function on the C, N, and O atoms. In order to include some solvent effects, these calculations were repeated including two methanol molecules. They were initially put on the side of the seven-atom ring at a large distance, in order that they be driven by the energy gradient towards the more favorable energetic positions. The solvation energy in methanol was also estimated by the polarizable continuum model[20] (PCM) in which the solvent is treated as a continuum system surrounding the molecule outside a cavity modeled on the molecular shape. HF and DFT calculations were performed with the Gaussian 03 package.[21]

Multireference perturbation theory (MR-PT) calculations were then carried out at the geometries previously optimized at the DFT level, in order to confirm the relative stability of reactants and products. We adopted the CIPSI algorithm[22,23] in which the configurational space is gradually enlarged by

Chart 2. Model System of Colchicine Adopted for the MC-SCF Calculations of the MEP for the S_0 and S_1 States[a]



[a] The atom numbering in the reactive seven-atom ring is shown.

exploiting the first-order perturbative correction of the single and double excitations to the ground state. The sequence is repeated until a reasonable compromise between the level of accuracy and the computational effort is reached. In the present work, the final configurational space includes about 6000 detours. The final energy is obtained by second-order diagrammatic perturbation theory in the Möller−Plesset partition scheme, considering the contribution of single and double excitations of this final reference space.

The photochemical reaction under study includes the geometry rearrangement from the first singlet excited state S_1 of colchicine to the ground state S_0 of the four lumicolchicine isomers. Therefore, at least two electronic states have to be considered. In view of the large number of internal coordinates involved in the present reaction, the possibility of building a high-dimensional potential energy surface (PES) for each electronic state is a very complicated task, well beyond the scope of the present paper. A simpler possibility is provided by the minimum energy path (MEP) concept, in which a reaction path is built up by selecting the leading coordinate and optimizing all the remaining internal coordinates at fixed values of the former. An approximate reaction path was thus studied by considering the C_4-C_7 (Chart 2) distance as the leading reaction coordinate, that is, the one which undergoes the most relevant changes along the photochemical reaction. A MEP has to be determined separately both for S_0 and for S_1, in order to account for any possible population of the two states. This requires the nontrivial problem of optimizing the geometry for an excited state without the possibility of exploiting any symmetry restriction.

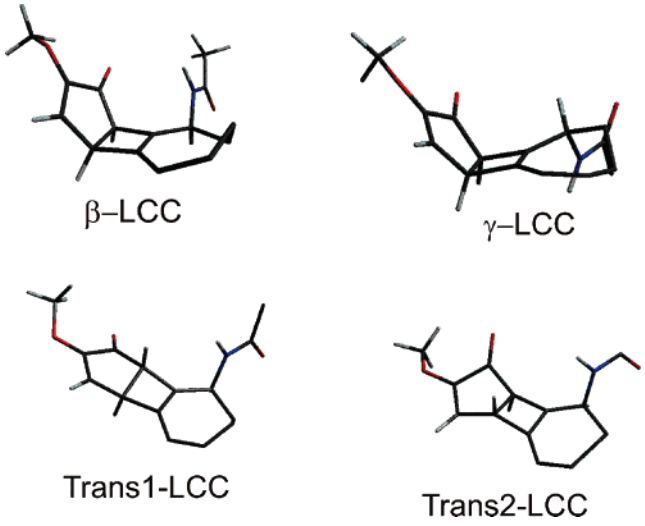
To save computational time and, at the same time, to avoid possible complications due to mixings of near degenerate excited states localized on the reactive part of the molecule and on the inert six-member ring, the calculation of the reaction path was computed on the model system reported in Chart 2. This was obtained by conserving the reactive molecular region and the aliphatic seven-atom rings adjacent to it. In order to preserve the original steric features, the aromatic ring was substituted by a double CC bond with the length fixed at the typical aromatic value of 1.38 Å. Moreover, the bonded hydrogen atoms H_X and H_Y were constrained to form a HCCH dihedral of 0° ($H_XC_XC_YH_Y$ in Chart 2) in order to mimic the constraints arising from the original aromatic ring. The amide group was substituted by a hydrogen atom. The effectiveness of this stratagem was tested by optimizing the energy of the complete molecule and of the model system. Since the bond lengths of the two calculations do not differ by more than 0.02 Å, the model system was considered adequate for the present study.

The MEPs of both the S_0 and S_1 states were computed by Multi Configurational self-consistent-field (MC-SCF) calculations in the restricted active space formulation, including the five higher occupied molecular orbitals (MOs) and the five lowest empty MOs in the active space. All the single and double excitations in the active space were included in the configurational space. The active space (10 MOs and 10 electrons) involves all the π and π^* orbitals in CC, including even the C_X-C_Y π bond (Chart 2), which is not involved in conjugation. In LCC, there are only four π orbitals, and the active space includes the lone-pair orbital on the methoxy group with the exclusion of the newly formed C_4-C_7 bond. Therefore, the active space appears to be slightly unbalanced for the reagent and the products. On the other hand, a better description of the formation of the C_4-C_7 bond would require the inclusion of all the valence orbitals, but this is not possible for computational reasons. The two states were computed using state-averaged two-states calculations with equal weights, in order to avoid root flipping problems. Geometry optimization was thus performed at several C_4-C_7 distances considering both S_0 and S_1 as the reference state. The MOLPRO program[24] was used for MC-SCF calculations.

Results and Discussion

Energy and Geometrical Structure of the Reagent and Products. Chart 3 displays a schematic picture of the four isomers β, γ, trans-1, and trans-2 of lumicolchicine. The β and γ isomers differ from each other in the position of the amide group with respect to the five-membered ring containing the carbonyl and methoxy groups. In the β isomer, the amide and carbonyl groups are on the same side with respect to the $C_4-C_5-C_6-C_7$ plane and may form an intramolecular hydrogen bond. Conversely, in the γ isomer, these groups are much more distant and cannot interact appreciably. In both the trans conformers, the two condensed rings lie roughly in the same plane and no direct interaction with the amide group can occur. Their geometrical difference can be again referred to the amide group: in the trans-1 conformer, the C_7-H bond is on the same side as the amide group, whereas in the trans-2 conformer, the C_7-H bond is on the opposite side.

Chart 3. The Reactive Molecular Energy Region of the Four Isomers of Lumicolchicine



The relative energies of CC and the four isomers of LCC at the optimized geometries obtained by several methods are reported in Table 1. The table includes also the low-level HF/6-31G results for comparison and the zero-point energy (ZPE) correction as well as an estimate of the solvation energy with the explicit inclusion of two methanol molecules and with the PCM method. Some structural parameters referred to the B3LYP-optimized geometry are reported in Table 2.

The HF and B3LYP calculations give similar results, and also the correlated MR-PT treatment shows that the trans forms of LCC are much higher in energy than the β and γ conformers. This is due to the steric tension of the reactive seven-atom ring in both the *trans*-LCCs where the trans position of the hydrogen atoms bonded to the C_4 and C_7 atoms generates a large distortion from planarity in both of the two condensed rings [see the root-mean-square (RMS) column of Table 2]. The small difference in stability between the β and γ conformers can probably be ascribed to an intramolecular hydrogen bond between the carbonyl group and the H atom of the amide group, which occurs for β-LCC. This is clearly indicated by the O−N distance reported in the last column of Table 2: 2.1 Å for β-LCC versus 4.5 Å for γ-LCC.

Since we are interested in the photochemical reaction of CC in solution, this apparent stability of β-LCC versus γ-LCC obtained in gas-phase calculations (about 4 kcal mol^{-1}) could have a weak meaning. For this reason, we have performed a further calculation in which two methanol solvent molecules have been added to the systems under scrutiny. As explained in the previous section, the geometry optimization starts with the two solvent molecules placed at a large distance from the reactive molecular region, in order not to show bias between the different possibilities of forming solute−solvent hydrogen bonds. In the optimized geometry of β-LCC, one methanol molecule interposes between the carbonyl group and the amidic hydrogen, destroying the intramolecular hydrogen bond observed in the isolated

Table 1. Relative Energy of Colchicine (CC) and Lumicolchicines (LCC) at Their Equilibrium Geometries (kcal mol^{-1})[a]

method	basis set	CC	β-LCC	γ-LCC	*trans*-1-LCC	*trans*-2-LCC
HF/6-31G	6-31G	0.00[b]	14.03	18.38	74.47	71.06
B3LYP	6-31G*	0.00[c]	12.54	16.83	62.09	61.73
MR-PT[d]	6-31G*	0.00[e]	16.39	17.13	61.94	61.88
B3LYP + 2S[f]	6-31G*	0.00	10.73	11.15	64.92	64.80
B3LYP+PCM	6-31G*	0.00	16.53	16.89	61.78	62.53
ZPE (B3LYP)	6-31G*	277.18	276.99	276.74	275.30	276.24
B3LYP+ZPE	6-31G*	0.00	12.35	16.39	61.21	60.79

[a] Apart from HF, all other results are obtained with the 6-31G* basis set. [b] Absolute energy HF/6-31G is −1350.628764 H at HF optimized geometry. [c] Absolute energy B3LYP/6-31G* is −1359.417458 H. [d] Calculations at the geometry optimized at the B3LYP level. [e] Absolute energy −1340.416360 H. [f] Including two methanol molecules.

Table 2. Value of Some Relevant Internal Coordinates at Equilibrium Geometry (Å or Degrees) by B3LYP/6-31G* Calculations

	C_4-C_7	C_4-C_5	C_5-C_6	C_6-C_7	θ (RMS)[a]	C_1-C_2	O−HN
CC	3.06	1.38	1.44	1.37	3.3 (0.002, 0.028)	1.48	4.70
β-LCC	1.56	1.54	1.36	1.53	62.0 (0.003,0.009)	1.49	2.13
γ-LCC	1.56	1.55	1.36	1.52	61.5 (0.000,0.001)	1.49	4.51
trans-1-LCC	1.55	1.55	1.37	1.54	13.7 (0.047,0.202)	1.55	4.80
trans-2-LCC	1.55	1.55	1.37	1.54	14.2 (0.041,0.208)	1.55	3.86

[a] Angle between the two least-squares planes given by the $C_4-C_5-C_5-C_7$ and $C_1-C_2-C_3-C_4-C_7$ atoms. The root-mean-square (RMS) value (in Å) of the two least-squares planes is reported in parentheses.

β-LCC molecule. Apart from this particular behavior observed only for the β-LCC conformer, in all the other cases, the methanol molecules form hydrogen bonds with the ether, carbonyl, or amide groups of both CC and LCCs. The relative energies from these new calculations show, however, that the fraction of solvation energy accounted for two solvent molecules does not upset the relative stability of the conformers in the gas-phase calculations. The slightly smaller solvation energy of β-LCC with respect to γ-LCC can probably be ascribed to the destruction of the already discussed intramolecular hydrogen bond between the carbonyl and the amide groups. The geometries optimized by accounting for the solvent effects through the PCM method show no trace of an intramolecular hydrogen bond, and the PCM solvation energy is similar for CC and for the four LCC conformers. All of these results allow us to assess that the solvent effect does not affect significantly the relative stability of the four LCC isomers.

Even the inclusion of the ZPE in the total molecular energy has small effects on the energetic gaps between the several LCCs (last row of Table 1). The free energy obtained by the inclusion of the entropic contribution of the vibrational degrees of freedom changes the energy differences by no more than 0.5 kcal mol^{-1}. Therefore, the computed relative energies of the LCC isomers appear to be rather stable both versus solvation and versus vibrational contributions.

The relative stabilities of CC and the LCCs are now considered in the light of the photochemical reaction in which colchicine undergoes a transition to the first singlet excited state and decays to γ-LCC and β-LCC. Since the absorption band of CC is centered at 3.65 eV (84 kcal mol^{-1}), from the energy criterion, all four of the LCCs could in principle be obtained, even if the energy of the *trans*-LCCs lie only about 20 kcal mol^{-1} below the maximum energy allowed for the products. However, the higher stability of both the cis conformers with respect to the *trans*-LCCs makes the former much more probable, and the present theoretical data are in full agreement with experimental results. Moreover, the high energy of all the LCCs with respect to the thermal energy precludes the possibility of thermal interconversion from CC to LCC, in agreement with observations.

The present data confirm the theoretical investigations reported in previous papers.[17,25] In the hypothesis that the first excited state of CC at 3.65 eV corresponds essentially to a singlet highest occupied molecular orbital → lowest unoccupied molecular orbital (HOMO → LUMO) transition, the spatial distribution of the LUMO of CC was carefully analyzed,[4] and it was verified that a disrotatory process could lead to a C_4-C_7 bonding, rather than antibonding, σ orbital. This feature was considered consistent both with the presence of a low-energy transition state and with the formation of *cis*-LCC isomers. The present data confirm the previous analysis and add the valuable information that the trans conformers are less stable.

From the C−C bond lengths reported in Table 2, it is apparent that the C_4-C_7 distance does not change along the LCC series. Even the other C−C bond lengths in the two condensed rings show the typical values of single and double C−C bonds for all LCC isomers: for instance, the C_4-C_5 and C_6-C_7 bond lengths are consistent with a double and, respectively, single C−C bond. From the C−C bond lengths of colchicine, it seems that the seven-atom ring has a reduced π aromaticity, since the C_4-C_5 and C_6-C_7 bonds show the typical double-bond length (1.34 Å). On the contrary, the C_5-C_6 bond length in CC is shorter than the typical C−C single bond distance (>1.5 Å). The latter feature may probably be ascribed to the fact that the C_5-C_6 bond is included in the nonplanar seven-atom ring (lower part of Chart 2) which, due the presence of both sp^2 and sp^3 carbon atoms, is affected by some degree of steric tension.

Whereas in the *cis*-LCCs (both β and γ) the two condensed rings are nearly planar and form an angle of about 60° (sixth

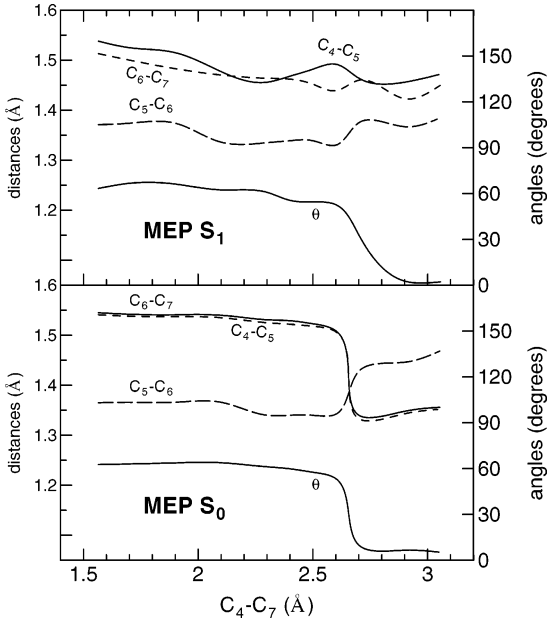


Figure 1. Values of some internal coordinates (Å or deg) versus the C_4-C_7 distance along the MEP of the S_0 and S_1 states connecting the γ-lumicolchicine with the colchicine. All the remaining internal coordinates are optimized for the model system of Chart 2 at the MC-SCF/6-31G* level of theory. θ is the angle between the two least-squares planes given by the $C_4-C_5-C_5-C_7$ and $C_1-C_2-C_3-C_4-C_7$ atoms (Chart 2).

column of Table 2), in the trans conformers, the same rings are nearly coplanar but their atoms strongly deviate from coplanarity, as indicated by the high RMS values of the least-squares planes. This supports a high steric tension in both rings, which is the basic reason for the instability of the trans conformers. Finally, the distance between the carbonyl oxygen and the amidic hydrogen, reported in the last column of Table 2, indicates an intramolecular hydrogen bond in β-LCC, as above-mentioned.

Geometrical Rearrangements along the S_0 and S_1 MEPs. In Figure 1, we report the values of some internal coordinates which undergo large changes along the transformation from CC to LCC, versus the leading internal coordinate C_4-C_7 distance. The reported values refer to geometries separately optimized for S_0 and S_1 at the MC-SCF level for the model system of Chart 2. In order to avoid the complications arising from the intramolecular hydrogen bond involving the amide group in β-LCC, we have considered the MEP connecting CC to the γ-LCC isomer. The corresponding energy curves of the MEPs are reported in Figure 2 [$E_0(0)$ and $E_1(1)$ respectively for S_0 and S_1] and will be discussed later on.

Since the S_1 state roughly corresponds to a $\pi \rightarrow \pi^*$ (HOMO $\rightarrow$ LUMO) excitation[4] localized on the reactive region both for CC and for LCCs, a general decrease of the bond orders and an increase of the bond lengths are expected in the seven-atom rings, on going from S_0 to S_1. From Figure 1, it is apparent that this effect is more pronounced in colchicine, indicating that the electronic excitation causes a destabilization on the seven-atom ring. The main geometrical changes are concerned with the bond length of the $C_4-C_5-C_6-C_7$ sequence, which roughly corresponds to $C_4=C_5-C_6=C_7$ and $C_4-C_5=C_6-C_7$ for S_0 and S_1, respectively. This is a very interesting feature since the bond lengths undergo moderate changes along the S_1 MEP and are similar to those of γ-lumicolchicine. Therefore, the geometrical relaxation of S_1 in the reagent (and presumably in the early stage of the reaction) gives rise to a structure consistent with the observed photochemical reaction.

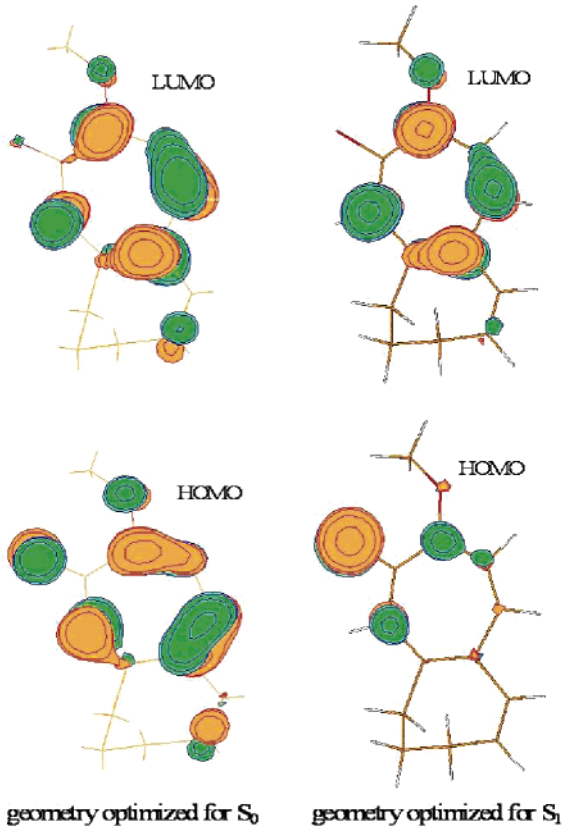


Figure 2. Map of the HOMO and LUMO of the colchicine model system of Chart 2. Left: the geometry is that optimized for the ground state. Right: relaxed geometry for the S_1 state. The MOs are the natural orbitals of MC-SCF calculations.

The profile of the MEP curves of Figure 1 indicates that the most relevant geometrical changes occur in a small range of C_4-C_7 distances, namely, in the 2.6−2.7 Å range, where the seven-membered ring bends towards an angle of about 60°. This geometrical change is more evident for S_0 where the θ change is accompanied by sharp changes in the bond lengths, corresponding to a shift from a single to double bond or vice versa. For S_1, the change of the interplane angle θ is less rapid, and it is accompanied by smaller changes of the C−C bond lengths with no net increment/decrement of one unit of bond order. As discussed above, the flatness of the curves in the S_1 MEP probably derives from the fact that the S_1 geometrical relaxation in colchicine (without changing the C_4-C_7 distance) gives bond lengths similar to those observed in the four-atom ring of the products.

Analysis of the Molecular Orbitals. A partial rationale of this behavior can be found by analyzing the HOMO and LUMO of the S_0 and S_1 MEPs at a value of the C_4-C_7 distance of 3.05 Å, corresponding to CC. The MOs are displayed in Figure 2. The HOMO and LUMO at the S_0 MEP are typical of conjugate systems: both are π orbitals, and

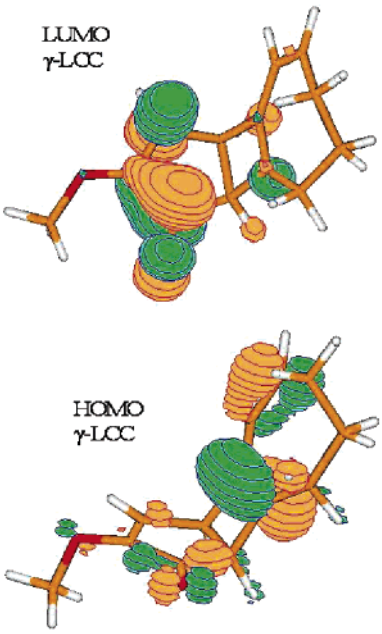


Figure 3. Map of the HOMO and LUMO of the γ-lumicolchicine model system of Chart 2. The MOs are the natural orbitals of MC-SCF calculations.

the LUMO has one more node than the HOMO. The HOMO seems to give a noticeable contribution to the double bonds of the seven-atom ring with some contribution on the carbonylic oxygen. In the relaxed geometry (i.e., optimized for S_1 at $C_4-C_7 = 3.05$ Å), the LUMO is similar to the previous one but the HOMO is almost completely localized on the carbonylic oxygen (nonbonding), and it is very different from the HOMO at the S_0 MEP. This finding is in accord with the data of Figure 1 for CC, where on going from the S_0 MEP to the S_1 MEP a remarkable increase of the C_4-C_5 and C_6-C_7 distances was observed, together with a decrease of the C_5-C_6 bond length. Indeed, since the HOMO at the S_0 MEP is bonding for the C_4-C_5 and C_6-C_7 bonds and antibonding for the C_5-C_6 bond, its migration towards the carbonyl group reinforces the first two bonds and weakens the latter.

The two LUMOs of Figure 2 remain similar along their own MEPs, and both move towards the C_1-C_2 bond (see Figure 3). The HOMO of the S_0 MEP localizes in the C_4-C_7 region along the path, gives a relevant contribution to the formation of the new σ bond, and finally becomes a π bonding orbital on the C_5-C_6 bond in the four-atom ring of LCC (see Figure 3). The HOMO of the S_1 MEP (also called HSOMO) is localized on the carbonylic oxygen in CC, shifts partially to the C_7 atom, and becomes similar to the HOMO of the S_0 MEP in γ-LCC. Thus, it appears that for γ-lumicolchicine the HOMO and the LUMO are localized on different rings and we may expect a weak transition to the first excited state.

Energetics of the Reaction Path. The MC-SCF energies of the ground state S_0 and of the first singlet excited state S_1 along the MEP connecting colchicine and γ-lumicolchicine are displayed in Figure 4. The four curves (the *trans*-1-LCC curve will be considered later on) refer to the energy of the S_0 and S_1 states computed along the S_0 MEP and S_1 MEP

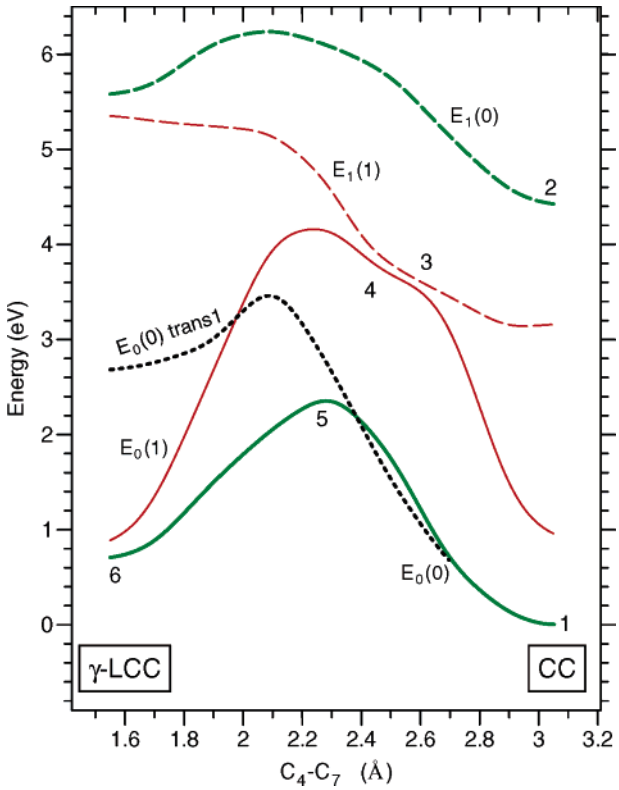


Figure 4. Energy of the S_0 (E_0, full lines) and of the S_1 (E_1, dashed lines) states along their own MEP leading to γ-lumicolchicine. The 0 in parentheses indicates the S_0-MEP (green heavy lines); the 1 indicates the S_1-MEP (red light lines). The MEP of S_0 leading to trans-1-lumicolchicine is also reported. The numbers 1−6 refer to the discussion in the text.

(label in parentheses). For instance, the curve labeled $E_1(0)$ in Figure 4 refers to the S_1 energy computed at the geometries optimized for S_0. Thus, the $E_1(0) - E_0(0)$ energy difference represents the vertical excitation energy at the geometry optimized for S_0 for a given C_4-C_7 distance. $E_1(0) - E_1(1)$ is the relaxation energy of the excited state due to the change of the internal coordinates other than the fixed C_4-C_7 distance. The CI expansion coefficients of the MC-SCF states indicate that at all geometries S_0 and S_1 always have a closed shell and a single excited configuration, respectively.

The $E_0(0)$ energy curve represents the minimum energy path of a thermal photoisomerization. The barrier for this process is about 2.4 eV (~52 kcal mol^{-1}), which prevents this reaction from spontaneously proceeding, in agreement with the experimental evidence.

A clear difference between the two shown MEPs is that the $S_0 \rightarrow S_1$ excitation energy calculated along the S_0 MEP $[E_1(0) - E_0(0)]$ is always higher than 4 eV, whereas the same quantity calculated along the S_1 MEP $[E_1(1) - E_0(1)]$ is high in proximity to the reagents and products but is small in the 2.2−2.8 Å range. In particular, S_0 and S_1 are near degenerate along the S_1 MEP, in the 2.4−2.6 Å range, and this could enhance the probability of surface hopping. It is noteworthy that, in this range, the inter-ring angle is close to 60°; that is, the seven-atom ring is bent (see Figure 1). Therefore, on going from CC to γ-LCC, the bending of the seven-atom ring atom occurs when the C_4-C_7 distance is well higher than the typical C−C single bond length.

On the basis of the data reported in Figure 4, we may attempt to sketch a possible reaction mechanism with the help of the numbers 1−6, as reported in the figure. Colchicine (point 1) absorbs one photon, and the S_1 state at the same geometry (point 2) is populated through a sudden vertical photoexcitation. The S_1 state undergoes a geometry relaxation, and the representative point moves from point 2 to point 3 towards smaller C_4-C_7 values and smaller energies. From Figure 1, it is apparent that in the 2 → 3 step the θ angle increases and the bond distances involving the C_4, C_5, C_6, and C_7 atoms move towards the values they take in LCC. Notice that in this step the energy of S_0 increases by more than 2 eV [$E_0(1)$ curve] and approaches the $E_1(1)$ curve. The next step towards point 4 involves a possible nonadiabatic transition to the S_0 PES. The 4 → 5 step is concerned with the geometrical relaxation of S_0, and finally, the system can flow on the PES of S_0 towards point 6, co-incident with the γ-LCC product.

It is worth noticing that all of the hypothesized steps proceed with a decrease of the potential energy and that no barrier is observed for the proposed mechanism. The rate-determining step is likely to be the nonadiabatic transition from S_1 to S_0 around 2.5 Å.

The dotted curve reported in Figure 4 is the energy of the ground state along the reaction path of S_0 connecting colchicine with *trans*-1-lumicolchicine, computed as previously done for the γ conformer. For C_4-C_7 values greater than 2.6 Å, the reaction paths for *trans*- and *cis*-LCCs are identical, since the seven-atom ring is nearly planar. For lower values of C_4-C_7, the *trans*-1-LCC curve shows an energy maximum of about 3.4 eV at 2.1 Å. Since both the energies of the final state and of the transition state are lower than the energy of point 2, the formation of a low quantity of *trans*-1-LCC cannot be completely excluded by the present data.

Finally, a pictorial view of the electronic rearrangement along the S_0 MEP is reported in Figure 5. The contour lines refer to the projection of the electron density of S_0 onto the plane containing both the C_4 and C_7 atoms and bisecting the angle formed by the planes of the four- and five-atom rings. The disrotatory cyclization process and the C_4-C_7 σ-bond formation is apparent for C_4-C_7 distances shorter than 2.60 Å, where the two condensed rings bend towards an interplane angle of about 60°. The geometrical changes of the hydrogen atoms bonded to C_4 and C_7 can be estimated by observing the deformation of the density from the circular form. These hydrogen atoms are seen to deviate from a nearly collinear $H-C_4-C_7-H$ sequence in colchicine to C−C−H angles close to the typical tetrahedral value in γ-lumicolchicine. The corresponding contour map of S_1 is very similar with a small increase of density for C_4-C_7 in the 2.40−2.40 Å range and does not allow quantitative conclusions about the different electronic structures of the two states.

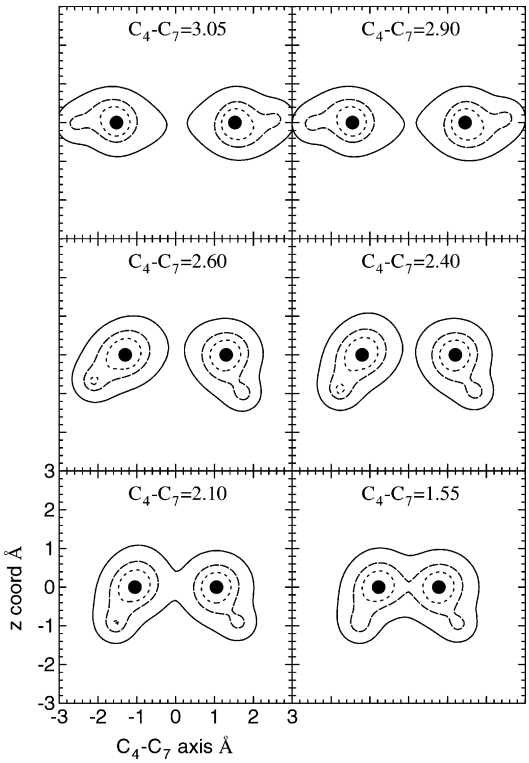


Figure 5. Contour map of S_0 electron density along the S_0-MEP. The density is projected onto the plane containing both the C_4 and C_7 atoms and bisecting the angle formed by the $C_4-C_5-C_5-C_7$ and $C_1-C_2-C_3-C_4-C_7$ least square. The abscissa axis is the line connecting C_4 and C_7; the ordinate (*z* coordinate) is the orthogonal line on the plane. The C_4 and C_7 atoms are indicated as full circles. The contour lines correspond to density values of 0.05, 0.15, and 0.25.

Conclusions and Final Remarks

From the computed energies of colchicine and the four possible isomers of lumicolchicine, it is argued that the β and γ conformers are strongly favored as possible products of the intramolecular photocyclization reaction. The energy profiles of the ground and the first singlet excited states allow tracing a possible reaction path, using the distance of the two carbon atoms forming the new σ bond as the leading coordinate. The geometrical rearrangement of the excited state is consistent with the geometry of the products, and all of the results are consistent with the experimentally observed formation of β and γ conformers.

In the present paper, the theoretical study has been driven by a hypothesized reaction path in accordance with experimental evidence. However, this paper pretends neither to explain completely the reaction under study nor to exclude other possible paths leading to different products. For instance, for similar systems like tropolone derivatives, it was found that an analogous photochemical reaction leads to the formation of a new σ bond between the C_1 and C_5 atoms.[26,27] The study of the MEP provides useful information but does not account for the dependence on the initial dynamic conditions of the reactant nor for possible deviations of the reactive trajectories from the MEPs. However, it has been stated that reactions in solution have more of a chance to proceed along MEPs since the excess energy is more easily dissipated through the solvent interaction. The experimentally observed rapid decay of S_1 with time constants of 300 fs could indicate an instantaneous geometrical rearrangement, in agreement with the present results which show that the relaxation energy of S_1 is about 1 eV. The second slow decay pattern with a time constant of 40 ps is consistent with the

possibility that the isomerization reaction may only occur through a well-defined path, which, presumably, involves small deviations from the MEP.

The hypothesized mechanism does not include possible intersystem crossing with triplet states. In order to check for this possibility, which however does not seem in agreement with experimental evidence, the energy curve of the lowest triplet state T_1 and the spin−orbit coupling between S_1 and T_1 were computed along the MEP of S_1 for C_4-C_7 values greater than 2.4 Å. It was found that T_1 always has a lower energy than S_1 with no crossing and that the spin−orbit coupling between these two states does not exceed 0.5 cm^{-1}. From these results, it appears that the role of T_1 in the studied photochemical isomerization is irrelevant. This is in accord with an experimental study in which no quenching effect was observed in the reactivity, in the presence of several triplet acceptors.[28] The opposite is expected for the same reaction in tiocolchicine, where no photoisomerization was observed and S_1 dissipates its excess energy in a different way.[25] This study will be the object of a future work.

Acknowledgment. I.C. acknowledges the MURST (Ministero della Ricerca Scientfica e Tecnologica: cofinanziamento 2004) for financial support. V.V. acknowledges CASPUR (University La Sapienza, Rome, Italy) for the use of computational resources.

References

(1) Capraro, H. G.; Brossi, A. In *The Alkaloids*; Brossi, A., Ed.; Academic Press: New York, 1984; Vol. 23, pp 1−70.

(2) Pelletier, P. J.; Caventou, J. B. *Ann. Chim. Phys*. **1820**, *14*, 69.

(3) Canonica, L.; Danieli, B.; Manitto, P.; Russo, G. *Tetrahedron Lett.* **1969**, 607−608.

(4) Chapman, O. L.; Smith, H. G.; King, R. W. *J. Am. Chem. Soc*. **1963**, *85*, 803−812.

(5) Forbes, E. J. *J. Chem. Soc*. **1955**, *9*, 3864−3870.

(6) Gardner, P. D.; Brandon, R. L.; Haynes, G. R. *J. Am. Chem. Soc*. **1957**, *79*, 6334−6337.

(7) Gasic, O.; Popovic, M. *Herba Hung.* **1981**, *41*, 51−54.

(8) Husek, A.; Sutlupinar, N.; Sedmera, P.; Voegelein, F.; Valka, I.; Simanek, V. *Phytochemistry* **1990**, *29*, 3058−3060.

(9) Malichova, V.; Potesilova, H.; Preininger, V.; Santavy, F. *Planta Med.* **1979**, *36*, 119−127.

(10) Merchant, J. R.; Joshi, V. *Indian J. Chem., Sect. B: Org. Chem. Incl. Med. Chem.* **1976**, *14*, 908−911.

(11) Pijsewska, L.; Kaul, J. L.; Joshi, R. K.; Santavy, F. *Collect. Czech. Chem. Commun.* **1967**, *32*, 158−170.

(12) Potesilova, H.; Acaraz, C.; Santavy, F. *Collect. Czech. Chem. Commun.* **1969**, *34*, 2128−2133.

(13) Potesilova, H.; Wiedermannova, J.; Santavy, F. *Collect. Czech. Chem. Commun.* **1969**, *34*, 3642−3645.

(14) Santavy, F. S. P.; Snatzke, G.; Reichstein, J. *Helv. Chim. Acta* **1971**, *54*, 1084−1095.

(15) Sutlupinar, N.; Husek, A.; Potesilova, H.; Dvorackova, S.; Hanus, V.; Sedmera, P.; Simanek, V. *Planta Med.* **1988**, *54*, 243−245.

(16) Takur, R. S.; Potesilova, H.; Santavy, F. *Planta Med.* **1975**, *28*, 201−209.

(17) Bussotti, L.; Cacelli, I.; D'Auria, M.; Foggi, P.; Lesma, G.; Silvani, A.; Villani, V. *J. Phys. Chem. A.* **2003**, *107*, 9079−9085.

(18) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(19) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(20) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.04; Gaussian, Inc.: Pittsburgh, PA, 2003.

(22) Cimiraglia, R. *J. Chem. Phys.* **1985**, *83*, 1746−1749.

(23) Huron, B.; Malrieu, J. P.; Rancurel, P. *J. Chem. Phys.* **1973**, *58*, 5745−5759.

(24) Werner, H.-J.; Knowles, P. J. *MOLPRO*, version 2002.1; University College Cardiff Consultants Limited: Cardiff, Wales, U. K., 2002.

(25) Bussotti, L.; D'Auria, M.; Foggi, P. G. L.; Righini, R.; Silvani, A. *Phothochem. Photobiol.* **2000**, *71*, 29−34.

(26) Calucci, L.; Cavazza, M.; Veracini, C. A.; Zandomeneghi, M. *J. Photochem. Photobiol., A* **1998**, *117*, 43.

(27) Joy, A.; Kammumalle, L. S.; Ramamurthy, V. *Org. Biomol. Chem.* **2005**, *3*, 3045.

(28) Croteau, R.; Leblanc, R. *Photochem. Photobiol.* **1978**, *28*, 33−38.

CT600306T

JCTC Journal of Chemical Theory and Computation

Structure and Stability of Small ZnC_n Clusters

Carmen Barrientos,* Pilar Redondo, and Antonio Largo

Departamento de Química Física, Facultad de Ciencias, Universidad de Valladolid, 47005 Valladolid, Spain

Received May 23, 2006

Abstract: A theoretical study of the small ZnC_n ($n = 1-8$) clusters has been carried out at the B3LYP/6-311+G(d) level. Different molecular properties for open-chain and cyclic species have been calculated. The computed properties include electronic energies, vibrational frequencies, dipole moments, and rotational constants for individual species. In addition, the relative stability of the different clusters is also discussed in terms of the incremental binding energy. In the case of open-chain clusters, the lowest-lying states correspond to triplet states with the exception of ZnC_2, whereas the electronic ground state is found to be a singlet state for cyclic clusters. The incremental binding energy graph shows a smooth even−odd parity effect in the incremental binding energy, *n*-even species being, in general, more stable than the adjacent odd-numbered ones. It was also found that the first members in the series, excepting ZnC_2, prefer open-chain structures, whereas when $n \geq 6$ cyclic ground states are favored.

1. Introduction

Carbon clusters have been the subject of study from both experimental and theoretical points of view for many decades.[1,2] These studies have involved small clusters (2−25 atoms) as well as larger clusters (up to 100 atoms), and also charged clusters have been considered. The interest in carbon clusters is mainly due to their applications in several fields. For instance, different molecules of this kind have been detected in astrophysical sources. In particular, carbon chains have been proposed as contributors to the formation of the long-chain cyanopolyynes, carbon dust, polycyclic aromatic hydrocarbons, and as possible carriers of diffuse interstellar bands.[3,4] Furthermore, carbon clusters are known to be present in the nucleation of carbon particles and formation of soot in hydrocarbon flames. Carbon clusters are also of interest in gas-phase chemistry, since they could act as intermediates in the chemical vapor depositions process for the production of thin diamond and silicon carbide films.[5,6] In addition to such practical implications, carbon clusters present interest in themselves because of the complexities in the properties and spectroscopy of these species.

The study of heteroatom-doped carbon clusters has also engaged great interest in recent years. In particular, the interaction between a transition metal and carbon is important in understanding several cluster materials including endohedrofullerenes, the catalytic growth of carbon nanotubes, and metallocarbohedrenes. Experimentally, anionic clusters can be produced from laser ionization and laser ablation.[7−11] An important conclusion obtained from observed records of time-of-flight mass spectra is that some clusters exhibit a dramatic even/odd alternation in signal intensity.

In order to explore these interesting experimental observations, several theoretical investigations have been conducted on XC_n clusters in which the heteroatom is a first- or second-row element in the periodic table. On the other hand, the information about carbon clusters doped with transition metals is scarcer.

Recently, we have studied the $MgC_n/MgC_n^+/MgC_n^-$ [12,13] and $CaC_n/CaC_n^+/CaC_n^-$ [14,15] isovalent clusters. Furthermore, we have also carried out studies on the linear[16] and cyclic[17] carbon clusters doped with the first transition metal, scandium. In a recent paper,[18] the structure of VC_n clusters has also been addressed. Linear and cyclic CrC_n clusters have been studied by Zhai et al.[19] It would be interesting to provide information about carbon clusters doped with Zn, the isovalent element of Mg and Ca "bracketing" the first

* Corresponding author fax: 34-983-423482; e-mail: cbb@qf.uva.es.

10.1021/ct6001797 CCC: $37.00 © 2007 American Chemical Society
Published on Web 01/26/2007

transition metal row, since, to the best of our knowledge, no data have ever been published for ZnC_n clusters.

In the present work, we have carried out a theoretical study of the ZnC_n ($n = 1-8$) open-chain and cyclic isomers, reporting their geometry structures and some spectroscopic data. It must be noted that the presence of 3d electrons increases the number of low-lying electronic states. Thus, information on these states can serve as helpful guidelines for the synthesis of related materials as well as for future theoretical studies of heteroatom-doped carbon clusters. In addition, we have also investigated the isomers' stabilities, incremental binding energy, and the competition between open-chain and cyclic isomers.

2. Computational Details

In our study, we have employed density functional theory (DFT). In particular, the popular hybrid B3LYP[20,21] exchange-correlation functional has been selected. This includes the Lee−Yang−Parr[22] correlation functional in conjunction with a hybrid exchange functional first proposed by Becke.[23] The latter is a linear combination of local density approximation, Becke's gradient correction,[24] and the Hartree−Fock exchange energy based on Kohn−Sham orbitals.[25]

As in previous works, we have employed the triple split-valence d-polarized 6-311G+(d)[26] basis set including an extra set of diffuse functions.

This hybrid formalism has been previously proved in accord with more rigorous ab initio methods in the study of carbon clusters and medium-size heteroatom-doped carbon clusters.[27]

For each optimized structure, harmonic vibrational frequencies were computed using analytic gradient techniques. These frequencies were employed to estimate the zero-point vibrational energy correction and enable us to check the nature of stationary points. Other properties, such as dipole moments or rotational constants, were also computed.

All calculations were carried out with the Gaussian 98[28] program package.

3. Results and Discussion

We have searched for different possible ZnC_n isomers, but we will only report the results for those which might be competitive in stability. For each ZnC_n species, we have considered different multiplicities: singlet, triplet, and quintet states. In particular, we include in the present study two different types of structures: open-chain structures with the zinc atom located at one end of the carbon chain and cyclic structures where zinc is bonded essentially to the two terminal carbon atoms of the C_n unit. Other possible open-chain structures with the zinc atom in a different position were also considered but are not shown here since they lie much higher in energy, according to the fact that C−C bonds are much stronger than Zn−C ones. For the same reason, other types of cyclic isomers, such as those in which a carbon atom or a zinc atom is in a bridge position outside a ring, were found to lie higher in energy, and they will not be included here. It should be pointed out that we have not found ZnC_n fan structures, that is, structures where the zinc atom is side-bonded to the entire C_n unit; all of our attempts to obtain these structures led to cyclic conformations.

Table 1. Electronic Energies, $<S^2>$ Values, Dipole Moments, and Relative Energies (Including the ZPE Correction) for Open-Chain ZnC_n Clusters at the B3LYP/6-311+G(d) Level

isomer	state	$-E$ (a.u)	$<S^2>$	μ (D)	ΔE (kcal mol^{-1})
ZnC	$^1\Delta$	1817.193 612		3.27	33.96
	$^3\Sigma^-$	1817.247 618	2.1202	2.98	0.00
	$^5\Sigma^-$	1817.198 368	6.0045	1.87	31.09
ZnC_2	$^1\Sigma^+$	1855.363 605		7.48	0.00
	$^3\Pi$	1855.346 568	2.0052	4.14	11.53
	$^5A'$	1855.212 773	6.0224	2.54	93.06
ZnC_3	$^1A'$	1893.445 079		3.69	0.00
	$^3\Pi$	1893.437 249	2.0207	5.24	5.13
	$^5\Sigma^-$	1893.386 328	6.0688	1.50	32.34
ZnC_4	$^1A'$	1931.532 646		9.13	4.15
	$^3\Pi$	1931.539 267	2.0198	6.53	0.00
	$^5\Pi$	1931.442 319	6.0336	5.67	59.61
ZnC_5	$^1A'$	1969.626 375		5.15	0.00
	$^3\Pi$	1969.626 207	2.0339	7.16	0.36
	$^5\Sigma^-$	1969.575 183	6.1202	1.97	31.52
ZnC_6	$^1A'$	2007.712 418		10.77	6.28
	$^3\Pi$	2007.723 288	2.0307	8.28	0.00
	$^5\Pi$	2007.648 520	6.0552	7.64	45.28
ZnC_7	$^1A'$	2045.805 034		6.62	2.58
	$^3\Pi$	2045.810 183	2.0468	8.84	0.00
	$^5\Sigma^-$	2045.760 506	6.1670	2.46	30.79
ZnC_8	$^1A'$	2083.891 754		12.36	7.54
	$^3\Pi$	2083.905 250	2.0407	9.89	0.00
	$^5\Pi$	2083.844 736	6.0749	9.43	36.19

In order to analyze systematic trends, in the different properties considered, with the size of the clusters, we will present the results for each type of structure separately. Thus, we first present the results for the open-chain structures; second, cyclic isomers will be considered, and finally, we will discuss the competition between open-chain and cyclic isomers. On the other hand, in the next subsections, we will give some properties such as the electronic energies, dipole moments, and so forth, for the lowest-lying ZnC_n species on various potential surfaces; however, other molecular properties which might be helpful in an experimental search for these species, such as harmonic vibrational frequencies and rotational constants, are provided as Supporting Information (Tables S1 and S2).

3.1. ZnC_n Open-Chain Isomers. In Table 1, we show absolute electronic energies, $<S^2>$ expectation values, dipole moments, and relative energies at the B3LYP/6-311+G(d) level of theory, for the lowest-lying open-chain ZnC_n species on the singlet, triplet, and quintet potential surfaces. In addition, in Figure 1, the optimized geometries for the most stable open-chain structures at the B3LYP/6-311+G(d) level of theory are shown.

It can be readily seen from Table 1 that, except for the first member of the series ZnC, the $<S^2>$ values are uniform and deviate slightly from the pure spin value; consequently, it can be considered that the wave functions employed are

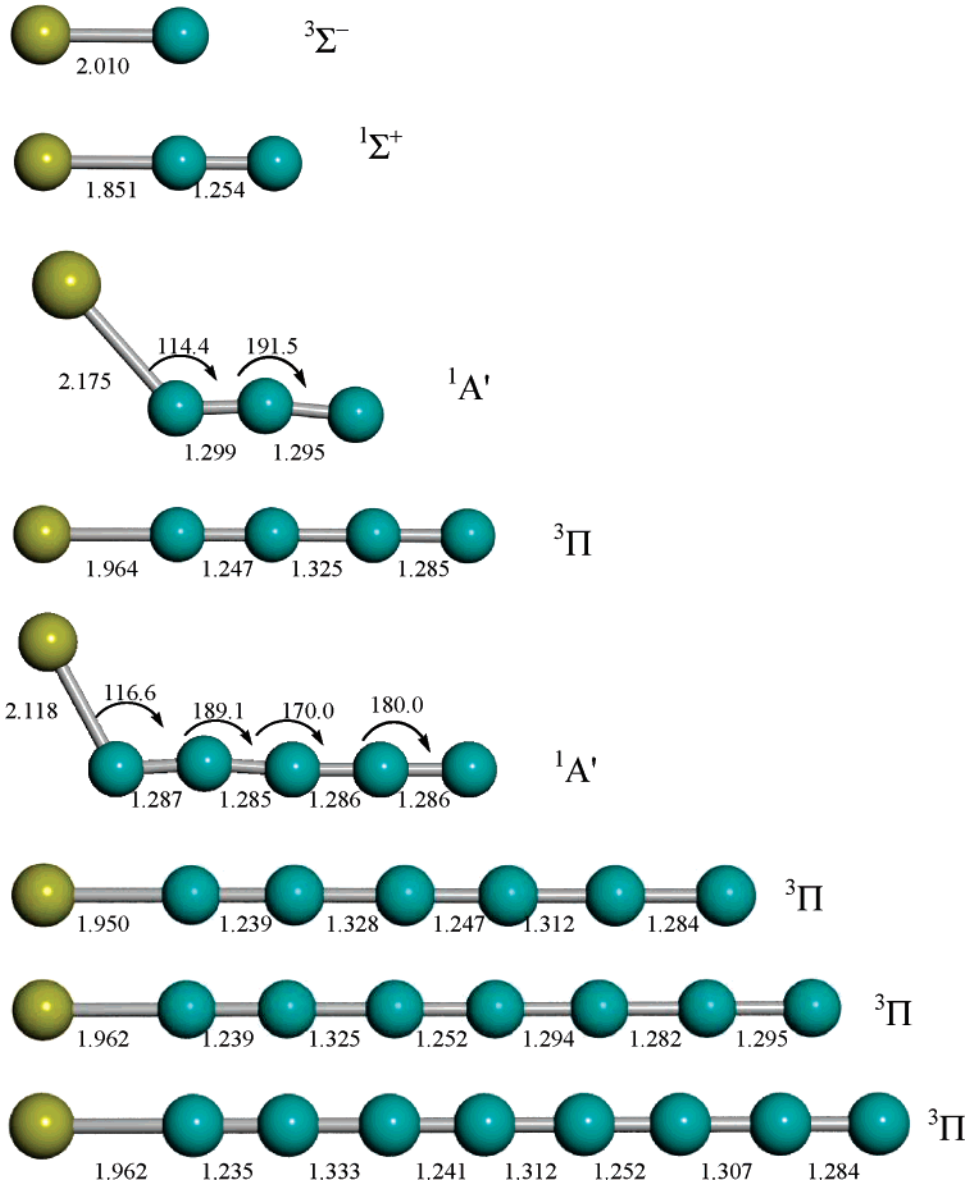


Figure 1. Equilibrium geometries of ZnC_n open-chain clusters at the B3LYP/6-311+G(d) level of theory. Distances are given in angstroms and angles in degrees.

nearly spin-pure, and in these species, spin contamination should not be a problem.

The reported lowest-lying open-chain structures all have real frequencies (Table S1), thus confirming that they are true minima on their respective potential surfaces.

As can be seen in Table 1, in most cases, the lowest-lying structure for ZnC_n open-chain clusters was found to correspond to a triplet state, with the exception of ZnC_2, which presents a singlet state ($^1\Sigma^+$). In addition, in the cases of ZnC_3 and ZnC_5, singlet nonlinear structures are found to lie slightly below the triplet species. Nevertheless, the general trend, especially for larger clusters, is to favor triplet ground states.

The triplet lowest-energy states can be explained in terms of their valence electronic configuration. Linear zinc-doped carbon clusters, ZnC_n, present $4n + 12$ (n being the number of carbon atoms) valence electrons, which are distributed over the valence orbitals as follows:

$$\{\text{core}\}\ 1\sigma^2...1\pi^4 1\delta^4...(n+3)\sigma^1\left(\frac{n+3}{2}\right)\pi^1$$

for n-odd members

$$\{\text{core}\}\ 1\sigma^2...1\pi^4 1\delta^4...(n+3)\sigma^1\left(\frac{n+2}{2}\right)\pi^3$$

for n-even members

except for

$$\text{ZnC: } \{\text{core}\}\ 1\sigma^2 1\pi^4 1\delta^4 2\sigma^2 3\sigma^2 2\pi^2(^3\Sigma^-)$$

$$\text{ZnC}_2\text{: } \{\text{core}\}\ 1\sigma^2 2\sigma^2 1\pi^4 1\delta^4 3\sigma^2 4\sigma^2 2\pi^4(^1\Sigma^+)$$

ZnC_n clusters contain four valence δ electrons, $2n + 5$ valence σ electrons, and $2n + 3$ valence π electrons. A total

of $2n + 4$ valence σ electrons of the $2n + 5$ total ones fully occupy $n + 2$ σ orbitals, and the one valence σ electron remaining leads to a half-filled σ orbital, σ^1.

In *n*-even clusters, $2n$ is a four-multiple and the $2n + 3$ valence π electrons are distributed over π orbitals, leading to final π^3 distributions. On the other hand, in the *n*-odd ZnC_n species, $2n + 2$ is a four-multiple and the $2n + 3$ π electrons lead to π^1 distributions. Both $\sigma^1\pi^1$ and $\sigma^1\pi^3$ configurations give $^3\Pi$ electronic states.

Before discussing our results, we are going to compare them with the previous theoretical results that have been reported for the ZnC system. The ground state of Zn is of ^{1}S symmetry with a $4s^2$ configuration; its first excited state is of ^{3}P symmetry with a $4s^14p^1$ configuration, and the experimental $^3P \leftarrow S$ energy gap is about 4.054 eV.[29] On the other hand, C presents a ground state of ^{3}P symmetry with a $2s^22p^2$ configuration, and the corresponding energy gaps are (^{5}S, ^{1}S, $^1D \leftarrow {}^3P$) 1.260, 2.084, and 4.179 eV, respectively. Therefore, the lowest ZnC dissociation channels in ascending energy order are (Zn + C) = ($^1S + {}^3P$, $^1S + {}^1D$, $^1S + {}^1S$...).

Recently, a controversy has taken place about the ground state of diatomic carbides MC.[30−37] In these carbides, the $^3\Sigma^-$ state correlates to M(s^2;^{1}S) + C(^{3}P) and competes with a $^5\Sigma^-$ state correlating to M(s^1p^1;^{3}P) + C(^{3}P). The interplay between the atomic energy separation $M(^3P) \leftarrow M(^1S)$ and the binding energy of the $^3\Sigma^-$ and $^5\Sigma^-$ states dictates ultimately the ground state of MC. However, in zinc carbide, this antagonism would probably be hindered by the $^3P \leftarrow {}^1S$ energy separation (4.054 eV); thus, in principle, a $^3\Sigma^-$ ground state could be expected.

While there is an absence of experimental results on ZnC, some theoretical studies can be found in the literature involving zinc carbide. For instance, Boldyrev and Simons,[38] from a quadratic configuration interaction method, predicted a $^3\Sigma^-$ to be the ground state of ZnC with a $^5\Sigma^-$ state lying 23.9 kcal/mol higher in energy [QCISD(T)/6-311++G(2df,-2f) level of theory]. Kerkines et al.[34] tested a state-specific multireference Brillouin−Wigner coupled cluster (MRBWCCSD) method against the single-reference CCSD method in the competing $^3\Sigma^-$ and $^5\Sigma^-$ states of zinc carbide. From their MRBWCCSD treatment, the $^5\Sigma^- \leftarrow {}^3\Sigma^-$ gap was predicted to lie around 22−24 kcal/mol. Recently, Tsouloucha et al.[39] carried out a study of the ground and excited states of zinc carbide using multireference methods in conjunction with quantitative basis sets. They predicted the $^5\Sigma^-$ state to lie 26.3 or 23.8 kcal/mol [with or without relativistic effects MRCI-(d10)+Q] above the ground $^3\Sigma^-$ state. Finally, Gutsev et al.[40] have performed a theoretical study of 3d-metal monocarbides using DFT and hybrid DFT methods.

We have studied the lowest-lying states in the singlet, triplet, and quintet ZnC potential surfaces, and our B3LYP calculations predicted a $^3\Sigma^-$ ground state with $^5\Sigma^-$ and $^1\Delta$ lying about 31 kcal/mol and 34 kcal/mol, respectively, higher in energy.

The valence electronic configuration for ZnC is

$$\{\text{core}\}\ 1\sigma^2 1\pi^4 1\delta^4 2\sigma^2 3\sigma^2 2\pi^2 (^3\Sigma^-)$$

which could be represented by the following valence bond structure:

$$\text{Zn:} \rightarrow \overset{\bullet}{\underset{\bullet}{\text{C}}}\text{:}$$

where Zn and C are held by a σ dative bond and the two π electrons remain practically located on C; see also, ref 39.

In ZnC, our computed Zn−C distance is 2.010 Å, in good agreement with previous data.[34−39]

In the case of ZnC_2, our calculations predicted a $^1\Sigma^+$ ground state with a $^3\Pi$ state lying 11.53 kcal/mol higher in energy. We found a relatively short Zn−C bond length, 1.851 Å, suggesting a moderately strong double bond, and a C−C distance of 1.254 Å. These geometrical parameters are compatible with a bonding picture resulting from a dative σ bond between Zn and the C_2 unit [resulting from donation of the 4s electrons to the C_2 lowest unoccupied molecular orbital (LUMO) $3\sigma_g$ orbital] enhanced by partial π bonding from d(Zn) → next-LUMO ($1\pi_g$) donation:

$$\text{Zn}{=}\text{C}{=}\text{C:}$$

The $^3\Pi$ state is associated mainly with the following valence-bond scheme (supported by the spin densities):

$$\bullet\text{Zn}{-}\overset{\bullet}{\text{C}}{=}\text{C:}$$

From Figure 1, it can be seen that the $^3\Pi$ states of ZnC_n clusters ($n \geq 3$) present similar Zn−C distances ∼ 1.96 Å. The C−C bond lengths are all in the range 1.253−1.334 Å and can be assimilated to moderately strong double bonds. In addition, spin densities show an unpaired electron located at zinc and the second one distributed along the carbon chain (mainly at C_1). Consequently, the most important valence structure for these clusters is a cumulene type:

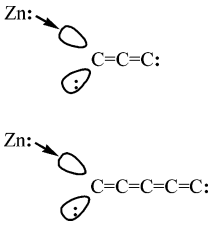
$$\bullet\text{Zn}{-}\overset{\bullet}{\text{C}}(=\text{C})_{n-2}{=}\text{C:}$$

Nevertheless, an alternation in the C−C bond distances can be noted, $C_{odd}-C_{even}$ being shorter than $C_{even}-C_{odd}$. This suggests some degree of polyacetylenic character with alternating triple C−C bonds:

$$\bullet\text{Zn}(-\text{C}{\equiv}\text{C})_{(n-1)/2}{-}\overset{\bullet}{\text{C}}\text{:} \qquad n \text{ odd}$$

$$\bullet\text{Zn}(-\text{C}{\equiv}\text{C})_{(n-2)/2}{-}\text{C}{\equiv}\text{C}\bullet \quad \left(\bullet\text{Zn}(-\text{C}{\equiv}\text{C})_{n-2/2}{-}\overset{\bullet}{\text{C}}{=}\text{C:}\right) \qquad n \text{ even}$$

It should be noted that ZnC_3 and ZnC_5 present $^1A'$ states slightly lower in energy than the $^3\Pi$ ones (5.13 kcal/mol lower in energy for ZnC_3 and 0.36 kcal/mol for ZnC_5). In these cases, we observe longer Zn−C distances (2.175 Å for ZnC_3 and 2.118 Å for ZnC_5) than in the other clusters, and similar C−C bond lengths. In these species, deviation from linearity is important, since the ∠ZnCC angle takes values of 114° and 116° respectively for ZnC_3 and ZnC_5. The geometrical parameters as well as the molecular orbitals suggest a bonding picture resulting from a dative σ bond between Zn and an empty orbital of the terminal carbon of the C_n unit



The Mulliken population analysis indicates that the zinc atom bears a positive charge (between +0.45 and +0.78); thus, the Zn−C bond shows a certain degree of ionic character. As a consequence of the ionic character of the Zn−C bond, relatively high dipole moments are found.

The dipole moments, shown in Table 1, regularly increase with the number of carbon atoms running from 2.98 for ZnC to 9.89 D for ZnC_8. It can be remarked that the high magnitude of the dipole moments might help in the possible experimental detection of these species.

By comparing zinc-doped carbon clusters to their magnesium and calcium isovalent systems, we observed that ZnC_n compounds exhibit a rather similar behavior. MgC_n[12] clusters present $^3\Pi$ ground states (except MgC, $^3\Sigma^-$, and MgC_2,$^1\Sigma^+$), whereas CaC_n[14] *n*-even clusters have singlet ground states (even though the singlet−triplet difference is rather small).

3.2. ZnC_n Cyclic Isomers. The electronic energies, $<S^2>$ expectation values, dipole moments, and relative energies at the B3LYP/6-311+G(d) level of theory, for the lowest-lying cyclic ZnC_n species on the singlet, triplet, and quintet potential surfaces, are given in Table 2. In Figure 2, the optimized geometries for the predicted most stable cyclic species are shown.

As we have mentioned before, no fan structures have been found after an exhaustive search on the respective potential surfaces. Our attempts to obtain fan species collapsed into the corresponding cyclic structures. The only species which resembles a fan isomer is that corresponding to ZnC_4 in Figure 2. This is the most stable species on the ZnC_4 potential surface, and the Zn−C distances suggest the possibility of interaction of the metal with the entire carbon chain (typical of a fan structure). A topological analysis of the charge density[41] shows that this is a truly monocyclic species with formal peripheral Zn−C bonds and no transannular Zn−C bonding. Nevertheless, the geometrical parameters (Zn−C bonds and ∠CCC angles around 140°) might suggest that the metal interacts with the whole quasi-linear carbon unit. All other species are clearly monocyclic structures, where the metal essentially interacts with the terminal carbon atoms of the C_n unit.

All cyclic ZnC_n structures have singlet lowest-lying states. Quintet states are clearly much less stable. On the other hand, triplet states are not so high in energy, especially for *n*-odd species. In fact, in the case of ZnC_3, singlet and triplet states are very close in energy, the triplet one lying slightly higher in energy. Therefore, in such a case, no definitive conclusion about the lowest-lying cyclic state can be drawn, and higher-level calculations should be required.

The Zn−C distances are larger for the cyclic structures than for their open-chain counterparts for the first members of the series ($n = 2-5$), whereas they are very similar to the linear isomers for the larger species ($n = 6-8$). In all cases, the C−C distances show clear alternations, with $C_{odd}-C_{even}$ distances shorter than $C_{even}-C_{odd}$ ones. As expected, dipole moments are always lower than those found in the open-chain isomers. The general trend is to decrease as the number of carbon atoms increases, although there are some oscillations in this behavior.

Table 2. Electronic Energies, $<S^2>$ Values, Dipole Moments, and Relative Energies (including the ZPE Correction) for Cyclic ZnC_n Clusters at the B3LYP/6-311+G(d) Level

isomer	state	$-E$ (a.u)	$<S^2>$	μ (D)	ΔE (kcal mol^{-1})
ZnC_2	$^1A_1$	1855.372 935		6.35	0.00
	$^3A_2$	1855.357 356	2.0125	1.91	13.04
	$^5A_1$	1855.212 791	6.0097	0.76	99.42
ZnC_3	$^1A_1$	1893.428 832		1.75	0.00
	$^3A_2$	1893.428 559	2.0197	2.20	0.07
	$^5A_1$	1893.339 773	6.0540	2.97	55.51
ZnC_4	$^1A_1$	1931.509 083		3.75	0.00
	$^3A_2$	1931.494 300	2.0093	2.43	8.11
	$^5A_1$	1931.404 511	6.0292	1.85	63.56
ZnC_5	$^1A_1$	1969.604 686		2.24	0.00
	$^3B_1$	1969.588 413	2.1125	2.10	10.17
	$^5A_2$	1969.528 952	6.0613	1.19	47.38
ZnC_6	$^1A_1$	2007.733 103		1.13	0.00
	$^3A''$	2007.681 770	2.0410	2.77	30.95
	$^5A_2$	2007.620 343	6.0781	0.01	69.12
ZnC_7	$^1A_1$	2045.812 691		1.70	0.00
	$^3B_1$	2045.800 458	2.1338	1.09	7.11
	$^5B_1$	2045.711 463	6.1253	3.18	62.13
ZnC_8	$^1A_1$	2083.914 092		0.76	0.00
	$^3A''$	2083.868 988	2.0379	1.79	26.89
	$^5A_1$	2083.834 998	6.0941	1.82	48.08

3.3. Relative Stabilities of ZnC_n Clusters. In order to compare the relative stability of the clusters with different sizes, we will use the concept of incremental binding energies.[42,43] The incremental binding energy can be defined as the change in energy accompanying the following process:

$$ZnC_n \rightarrow ZnC_{n-1} + C$$

and it can be computed as the consecutive binding energy differences between adjacent ZnC_n and ZnC_{n-1} clusters.

The results for the incremental binding energy, as a function of the number of carbon atoms, for the different open-chain and cyclic ZnC_n clusters are shown in Figure 3. It can be observed that there is an even−odd alternation for the open-chain isomers, with *n*-even species being comparatively more stable than *n*-odd ones. However, this alternation is not very pronounced, and in fact, the incremental binding energy varies smoothly along the series. This parity effect can be attributed to the number of available $\pi-$ valence electrons, mainly the electron number in π-type highest occupied molecular orbital (HOMO).[44] As we pointed out before, open-chain ZnC_n clusters ($^3\Pi$) present $\sigma^1\pi^1$ configurations (*n*-odd clusters) and $\sigma^1\pi^3$ configurations (*n*-even clusters). In this sense, *n*-even species are comparatively more stable because of the presence of three electrons in

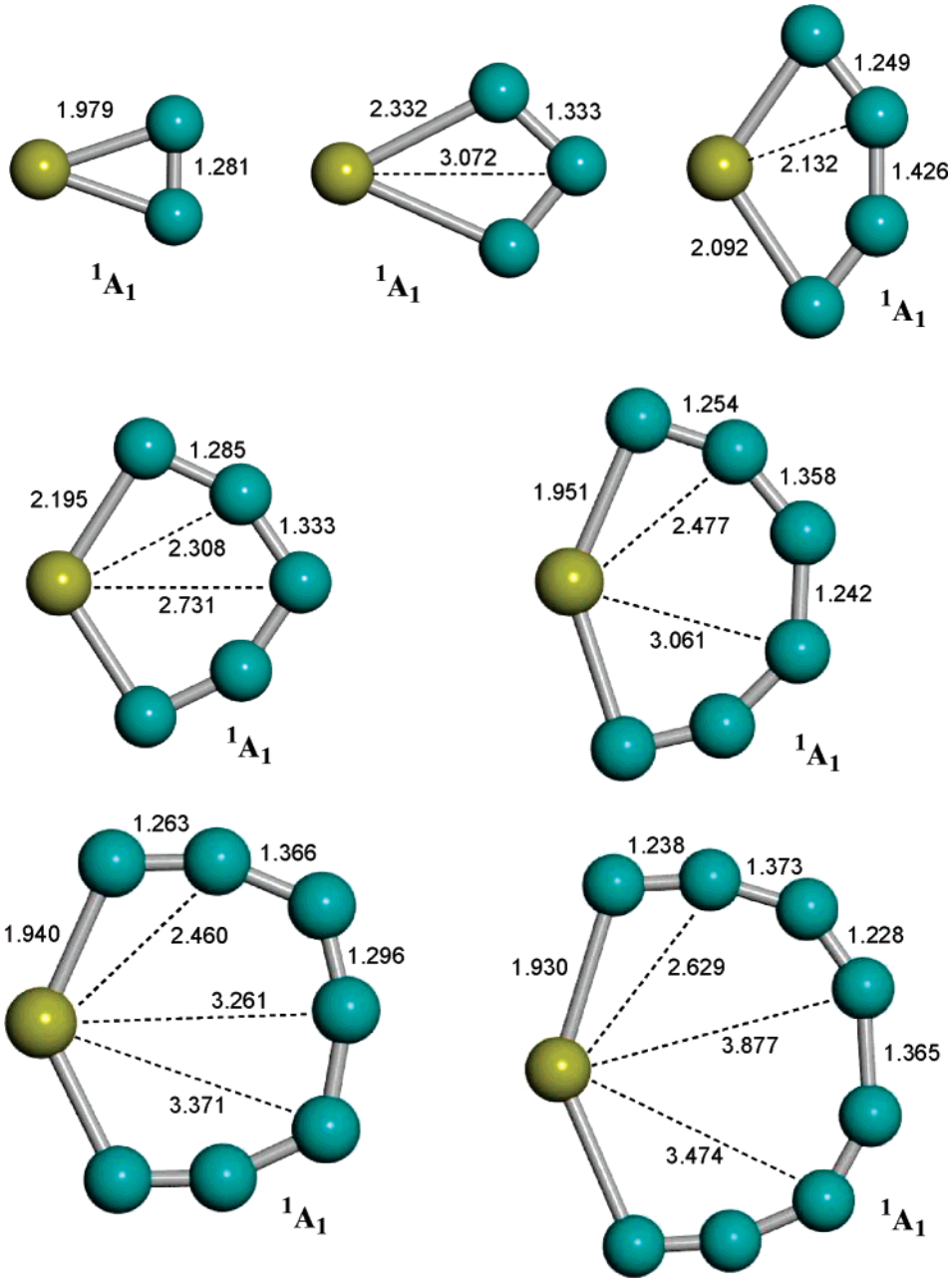


Figure 2. Equilibrium geometries of ZnC_n cyclic clusters at the B3LYP/6-311+G(d) level of theory. Distances are given in angstroms and angles in degrees.

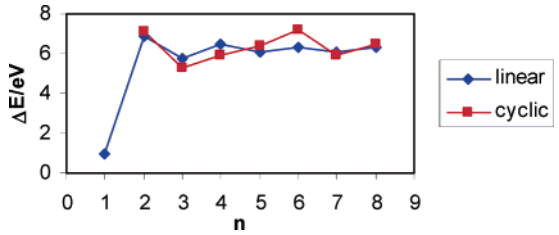


Figure 3. Incremental binding energies (eV) for ZnC_n open-chain and cyclic clusters at the B3LYP/6-311+G(d) level of theory vs the number of carbon atoms.

the HOMO, instead of just one electron as in the case of *n*-odd clusters. However, in both cases, the π-type HOMO is not fully filled and a similar stability for the different open-chain species can be expected. Consequently, the incremental binding energy varies smoothly for these systems.

Incremental binding energies for cyclic isomers do not exhibit a strict parity alternation effect. There seems to be a preference for *n*-even clusters, as in the case of open-chain isomers, but this trend breaks for ZnC_4.

The energy differences between cyclic and open-chain species are shown in Figure 4 as a function of the number of carbon atoms. The general trend is that open-chain isomers are preferred for small clusters, up to $n = 5$, with the exception of the first member of the series ZnC_2. On the other hand, cyclic isomers are preferred for larger clusters, although in these cases, the energy difference between both structures is rather small.

We may compare these results with those obtained for other carbon clusters doped with transition metals.[16−18] It seems that early transition metals, such as Sc or V, clearly favor cyclic (or fan) isomers. In those cases, the open-chain

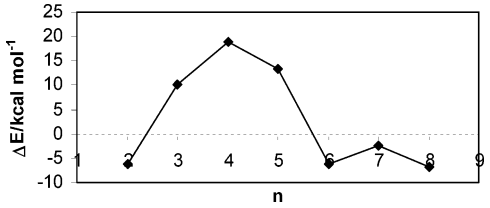


Figure 4. Relative energies, including the ZPE correction (kcal/mol), of the cyclic ZnC_n isomers with respect to the corresponding open-chain structures vs the number of carbon atoms.

isomers are relatively unstable, with energy differences between open-chain and cyclic isomers within the range 15−35 kcal mol^{-1} for ScC_n[17] clusters and 10−22 kcal mol^{-1} for VC_n[18] species. On the other hand, for late transition metals (Zn being a prototypical case), the open-chain isomers are comparatively more stable, and in fact, the smaller clusters have open-chain ground states. This behavior is probably related to two concomitant effects. In first place, charge donation from the metal to the carbon unit (which is expected to play an important role in the stabilization of transition metal carbides[45]) is more favorable along C_{2v} symmetry than for linear arrangements and should favor cyclic isomers. However, the energy of the metal 4s (and 3d) orbital lowers along the transition series, and consequently, this interaction is less favorable for late transition metals. In second place, back-donation from the carbon unit (which is expected to contribute to some extent to the stabilization of these species and is also in principle more favorable in C_{2v} than in $C_{\infty v}$ symmetry) should be hindered to a large extent for late transition metals, because of the large occupancy of the metal 3d orbitals (which are fully occupied for Zn).

4. Conclusions

A theoretical study on the open-chain and cyclic ZnC_n clusters ($n = 1-8$) has been carried out by using hybrid density theory with the 6-311G+(d) basis set.

According to our calculations, open-chain ZnC_n clusters usually present triplet ground states. ZnC_2 has a singlet ground state, whereas for ZnC_3 and ZnC_5, the nonlinear singlet lies slightly below the linear triplet structure. All cyclic species studied in this study have singlet ground states.

The relative stability of the different clusters has been estimated employing the concept of incremental binding energies. For open-chain ZnC_n clusters, a slight alternation stability effect is observed with *n*-even clusters being more stable than *n*-odd ones. These odd−even alternation trends have been interpreted according to the different electronic configuration. In the case of cyclic isomers, even though there is no strict parity effect, it seems that *n*-even species are also favored.

Our study suggests that, for late transition metals, such as Zn, open-chain isomers are comparatively more stable than in the case of early transition metals, which clearly favor cyclic ground states. For small ZnC_n, the open-chain isomer is the most stable one.

Predictions for some molecular properties such as rotational constants, dipole moments, and vibrational frequencies, which could be useful for an eventual experimental characterization, have been also provided (see the Supporting Information).

Acknowledgment. The authors acknowledge the Ministerio de Ciencia y Tecnología of Spain (Grant CTQ2004-07405-C02-01) and the Junta de Castilla y León (Grant VA 085/03) for financial support.

Supporting Information Available: Rotational constants and vibrational frequencies for open-chain and cyclic species (Tables S1 and S2). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Weltner, W., Jr.; Van Zee, R. J. *Chem. Rev.* **1989**, *89*, 1713.

(2) Van Orden, A.; Saykally, R. J. *Chem. Rev.* **1998**, *98*, 2313.

(3) Fulara, J.; Lessen, D.; Freivogel, P.; Maier, J. P. *Nature* **1993**, *366*, 439.

(4) Freivogel, P.; Fulara, J.; Maier, J. P. *Astrophys. J.* **1994**, *431*, L151.

(5) Allendorf, M. D. J. *Electrochem. Soc.* **1993**, *140*, 747.

(6) Koinuma, H.; Horiuchi, T.; Inomata, K.; Ha, H.-K.; Nakajina, K.; Chaudhary, K. A. *Pure Appl. Chem.* **1996**, *68*, 1151.

(7) Consalvo, D.; Mele, A.; Stranges, D.; Giardini-Guidoni, A.; Teghil, R. *Int. J. Mass. Spectrom. Ion Processes* **1989**, *91*, 319.

(8) Leleyter, M.; Joyes, P. *Surf. Sci.* **1985**, *156*, 800.

(9) Orth, R. G.; Jonkmann, H. T.; Michl, J. *Int. J. Mass Spectrom. Ion Processes* **1982**, *43*, 41.

(10) Becker, S.; Dietze, H. J. *Int. J. Mass Spectrom. Ion Processes* **1988**, *82*, 287.

(11) Huang, R. B.; Wang, C. R.; Liu, Z. Y.; Zheng, L. S.; Qi, F.; Sheng, L. S.; Yu, S. Q.; Zhang, Y. W. *Z. Phys. D: At., Mol. Clusters* **1995**, *33*, 49.

(12) Redondo, P.; Barrientos, C.; Cimas, A.; Largo, A. *J. Phys. Chem. A* **2003**, *107*, 4676.

(13) Redondo, P.; Barrientos, C.; Cimas, A.; Largo, A. *J. Phys. Chem. A* **2003**, *107*, 6317.

(14) Largo, A.; Redondo, P.; Barrientos, C. *J. Phys. Chem. A* **2004**, *108*, 6421.

(15) Redondo, P.; Barrientos; Largo, A. *J. Phys. Chem. A* **2004**, *108*, 11132.

(16) Redondo, P.; Barrientos, C.; Largo, A. *J. Phys. Chem. A* **2005**, *109*, 8594.

(17) Redondo, P.; Barrientos, C.; Largo, A. *J. Phys. Chem. A* **2006**, *110*, 4057.

(18) Redondo, P.; Barrientos, C.; Largo, A. *J. Chem. Theory Comput.* **2006**, *2*, 885.

(19) Zhai, H. J.; Wang, L. S.; Jena, P.; Gutsev, G. L.; Bauschlicher, C. W., Jr. *J. Chem. Phys.* **2004**, *120*, 8996.

(20) Becke, A. D. *J. Chem. Phys.* **1986**, *84*, 4524.

(21) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547.

(22) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(23) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053.

(24) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(25) Kohn, W.; Sham, L. J. *Phys. Rev. A: At., Mol., Opt. Phys.* **1965**, *140*, 1133.

(26) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(27) Martin, J. M. L.; Taylor, P. R. *J. Phys. Chem.* **1996**, *100*, 6047.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millan, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelly, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian Inc: Pittsburgh, PA, 1998.

(29) Moore, C. E. *Atomic Energy Levels*; NSRDS-NBS Circular No. 35; U.S. GPO: Washington, DC, 1971.

(30) Castro, M. A.; Canuto, S.; Müller-Plathe, F. *Phys. Rev. A: At., Mol., Opt. Phys.* **1992**, *46*, 4415.

(31) Da Silva, C. O.; Teixeira, F. E. C.; Azevedo, J. A. T.; Da Silva, E. C.; Nascimento, M. A. C. *Int J. Quantum Chem.* **1996**, *60*, 433.

(32) Serrano, A.; Canuto, S. *Chem. Phys. Lett.* **1997**, *269*, 193.

(33) Takada, H. H.; Pelegrini, M.; Roberto-Neto, O.; Machado, F. B. C. *Chem. Phys. Lett.* **2002**, *363*, 283.

(34) Kerkines, I. S. K.; Pittner, J.; Čarsky, P.; Mavridis, A.; Hubač, I. *J. Chem. Phys.* **2002**, *117*, 9733.

(35) Papakondylis, A.; Mavridis, A. *J. Phys. Chem. A* **2003**, *107*, 7650.

(36) Halfen, D. T.; Apponi, A. J.; Ziurys, L. M. *Astrophys. J.* **2002**, *577*, L67.

(37) Pelegrini, M.; Roberto-Neto, O.; Ornellas, F. R.; Machado, F. B. C. *Chem. Phys. Lett.* **2004**, *383*, 143.

(38) Boldyrev, A. I.; Simons, J. *Mol. Phys*. **1997**, 92, 365.

(39) Tsouloucha, A.; Kerkines, I. S. K.; Mavridis, A. *J. Phys. Chem. A* **2003**, *107*, 6062.

(40) Gutsev, G. L.; Andrews, L.; Bauschlicher, C. W. *Theor. Chem. Acc.* **2003**, *109*, 298.

(41) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Clarendon Press: Oxford, United Kingdom, 1990.

(42) Pascoli, G.; Lavendy, H. *J. Phys. Chem. A* **1999**, *103*, 3518.

(43) Raghavachari, K.; Binkley, J. S. *J. Chem. Phys.* **1987**, *87*, 2191.

(44) Boldyrev, A. I.; Simons, J. *J. Phys. Chem. A* **1997**, 101, 2215.

(45) Rayón, V. M.; Redondo, P.; Barrientos, C.; Largo, A. *Chem.—Eur. J.* **2006**, *12*, 6963.

CT6001797

JCTC Journal of Chemical Theory and Computation

Geometry Optimization of a Ru(IV) Allyl Dicationic Complex: A DFT Failure?

Maria José Calhorda,[†] Paul S. Pregosin,[‡] and Luis F. Veiros*[,§]

Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal, Laboratory of Inorganic Chemistry, ETHZ, Hönggerberg CH-8093 Zürich, Switzerland, and Centro de Química Estrutural, Complexo I, Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal

Received November 22, 2006

Abstract: Five pure and four hybrid DFT functionals associated with VDZP, VTZP, and VQZP basis sets are tested (Gaussian 03) for their performance on the geometry optimization of [Ru(η^5-C_5H_5)(η^3-$CH_2CHCHC_6H_5$)(CH_3CN)$_2$]$^{2+}$. When the calculated geometries were compared with the X-ray structure determination for the analogous complex with permethylated cyclopentadienyl, it was found that in all cases the coordination mode of the η^3-allyl was very poorly described, despite the functional used. The Ru−C bond distance corresponding to the substituted allyl carbon was overestimated by 0.23−0.50 Å, depending on the functional and the basis set used. These results were reproduced by further testing carried out with the ADF program and larger basis sets. MP2 leads to an acceptable value for the same Ru−C distance, with an underestimation of 0.07 Å, suggesting that, at least in the case of the functionals tested, DFT does not provide an accurate description of a weak Ru−C interaction.

1. Introduction

Any survey covering the recent chemistry literature shows that DFT methods[1] have become a widely used tool in computational studies applied to all fields of chemistry and, in particular, to organotransition-metal chemistry.[2] A comparatively low computational cost, when compared to wave-function based methods, and the quality of the results obtained justify the number of published studies ranging from structural and thermochemical studies[3] to reaction modeling and mechanistic investigations.[2,4] Often the geometries calculated in the computational studies cannot be tested, owing to the absence of experimental results related to the transient nature of the species under study. On the other hand, it is well documented that when dealing with ground-state structures, an almost perfect agreement can be reached between calculated and experimental parameters.[2,5−8] The accuracy of DFT in transition-metal chemistry has been recently reviewed by Harvey.[9]

Among all functionals currently favored by chemists, the B3LYP emerges as one of the most popular, if not the most used functional, providing, in general, good geometries and reliable energies.[10] B3LYP is a hybrid functional including a 20% mixture of Hartree−Fock[11] exchange with DFT exchange-correlation, given by Becke's three-parameter functional[12] with the Lee, Yang, and Parr correlation functional, which includes both local and nonlocal terms.[13,14]

The nature of the functionals for exchange and correlation currently used, however, remains approximated, and constant effort is being made to develop new and more accurate functionals. As recently stated by Perdew et al.,[15] DFT users can play an important role in this task, by reporting relevant failures or "pathologic cases", thus helping the developers to understand weak points of the available functionals in order to improve them.

During our recent computational investigations on the role of the cyclopentadienyl Ru(IV) complex, [Ru(η^5-Cp)(η^3-CH_2-$CHCHC_6H_5$)(CH_3CN)$_2$]$^{2+}$ (Cp = C_5H_5), as a possible catalyst of organic reactions such as allylic alkylation[16] and Friedel−Crafts reaction,[17] we came across one of such "pathologic cases". In this work, we report the failure of B3LYP as well as other hybrid and pure DFT functionals, associated with currently used VDZP and VTZP basis sets, to describe

* Corresponding author phone: +351-218 419 283; fax: +351-218 464 457; e-mail: veiros@ist.utl.pt.
[†] Universidade de Lisboa.
[‡] ETHZ.
[§] Instituto Superior Técnico.

10.1021/ct600341r CCC: $37.00 © 2007 American Chemical Society
Published on Web 02/24/2007

correctly the coordination of the allyl ligand, $CH_2CHCHC_6H_5$, to ruthenium on the former complex. The calculated geometries are compared with the experimental X-ray structure of the analogous Cp* complex, $[Ru(\eta^5\text{-}C_5Me_5)(\eta^3\text{-}CH_2\text{-}CHCHC_6H_5)(CH_3CN)_2]^{2+}$ ($Cp^* = C_5Me_5$; $Me = CH_3$).[17]

2. Computational Details

The calculations were performed using the Gaussian 03 software package[18] or ADF (2005.01)[19−21] without symmetry constraints. Both pure and hybrid DFT functionals were tested in the geometry optimization of the complex studied. Along the text, the usual acronyms identify the combinations of exchange, pure or hybrid, with correlation functionals. The exchange functionals used included Becke's 1988 functional (B),[22] the exchange component of Perdew and Wang's 1991 functional (PW91),[23−27] and the 1996 functional of Perdew, Burke, and Ernzerhof (PBE).[28,29] These were combined with the following correlation functionals: Perdew 1981 local correlation functional with gradient corrections (P86),[30] the correlation functional of Lee, Yang, and Parr which includes both local and nonlocal terms (LYP),[13,14] Perdew and Wang's 1991 gradient-corrected correlation functional (PW91),[23−27] and the 1996 gradient-corrected correlation functional of Perdew, Burke, and Ernzerhof (PBE).[28,29] Three hybrid functionals, including a mixture of Hartree−Fock exchange with DFT exchange, were also tested, Becke's three-parameter functional (B3),[12] modified Perdew−Wang exchange (mPW1), as implemented by Adamo and Barone,[31] and the hybrid functional of Perdew, Burke, and Ernzerhof (PBE1).[29] Two wavefunction based methods were also used in the geometry optimizations: Hartree−Fock (HF)[11] and second-order Møller-Plesset correlation energy correction (MP2).[32−37] Three basis sets were used for Ru. Of these, two included effective core potentials (ECP), Los Alamos ECP with valence double-ζ (LanL2DZ),[38−41] here named I_M, and Stuttgart/Dresden ECP with valence triple-ζ (SDD),[42−44] labeled II_M. The third metal basis set tested is an all-electron basis, a standard 3-21G[45−50] (labeled III_M). In all three cases, I_M, II_M, and III_M, a f-polarization function was added for Ru.[51] The main group elements belonging to the ligands, C, H, and N, were described by Pople's split-valence basis, 4-31G(d) (basis I_L),[52−55] the same set with diffuse functions added only to the three allylic carbon atoms, 4-31(+)G(d) (basis II_L), standard 6-311G(d,p) (basis III_L),[56−62] Ahlrichs' TZVP (basis IV_L),[63,64] and Dunning's correlation consistent cc-pVTZ (basis V_L)[65] and cc-pVQZ (basis VI_L).[66] Thus, for example, B3LYP/(II_MI_L) stands for a B3LYP calculation using basis II_M for the metal (SDD plus f-polarization function) and I_L basis for C, H, and N, that is, 4-31G(d). Similarly, BP86/(III_MII_L) means a BP86 calculation using III_M basis set for the metal (3-21G plus f-polarization) and II_L basis for C, H, and N, i.e., 4-31G(d) plus diffuse functions on the three allylic C atoms.

The calculations performed with the ADF program consisted of gradient corrected geometry optimizations,[67,68] without symmetry constraints, using the Local Density Approximation of the correlation energy (Vosko−Wilk−Nusair)[69] and the Generalized Gradient Approximation (Perdew−Wang[25] and Becke and Perdew exchange and correlation corrections).[22,30] Relativistic effects were treated with the ZORA approximation.[70] TZ2P basis sets (triple ζ Slater-type orbitals, STO) were used to describe the valence shells of Ru (4p, 4d, 5s), C and N (2s and 2p), and H (1s), with a set of two polarization functions[71,72] added, and frozen core orbitals for Ru ([1−4]s, [2−3]p, 3d), C (1s), and N (1s).[73] The QZ4P is an all-electron basis set with quadruple ζ Slater-type orbitals, with four added polarization functions for Ru (one p and three f functions), C, N (two d and two f functions), and H (one s, two d, and two f functions).

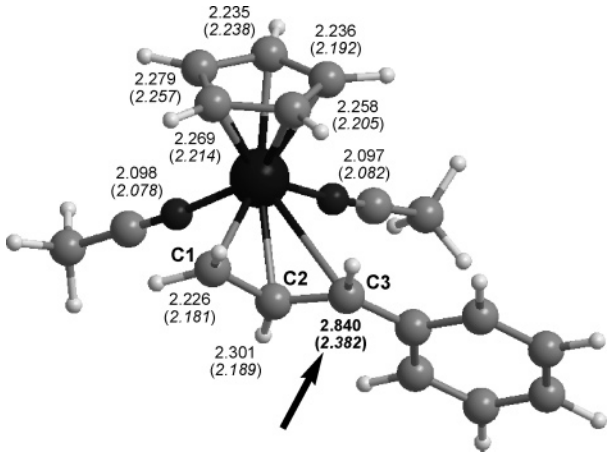


Figure 1. Optimized geometry of $[Ru(\eta^5\text{-}C_5H_5)(\eta^3\text{-}CH_2\text{-}CHCHC_6H_5)(CH_3CN)_2]^{2+}$ obtained at the B3LYP/(I_MI_L) level, showing the calculated bond distances (Å) between the metal and all coordinating atoms, and experimental values taken from the X-ray structure of the analogous Cp* complex (italics).

3. Results and Discussion

During the course of our recent computational studies on the mechanism of the ruthenium catalyzed regioselective allylic alkylation reaction, several ruthenium cyclopentadienyl complexes, such as $[Ru(\eta^5\text{-}C_5H_5)(\eta^3\text{-}CH_2CHCHC_6H_5)(CH_3CN)_2]^{2+}$ and $[Ru(\eta^5\text{-}C_5H_5)(\eta^3\text{-}CH_2CHCHC_6H_5)(OCO_2\text{-}CH_3)]^{+}$, were checked as possible active species. While reasonable optimized geometries were obtained for the carbonate complex,[16] the geometry calculated for the bis-(acetonitrile) species at the B3LYP/(I_MI_L) level (see Computational Details for basis set labels used throughout the text) led to an unexpected and very long Ru−C bond to one of the allylic carbon atoms. It is represented in Figure 1 as well as the relevant bond lengths around the metal (Ru−X) and the experimental values taken from the X-ray structure of the related Cp* complex, $[Ru(\eta^5\text{-}C_5Me_5)(\eta^3\text{-}CH_2\text{-}CHCHC_6H_5)(CH_3CN)_2]^{2+}$.[17]

The molecule in Figure 1 represents a typical complex with pseudopiano stool geometry. The cyclopentadienyl ligand, Cp, is coordinated through its entire π-system, that is, in a η^5 mode, with the five carbon atoms bonded to the metal. The two acetonitrile ligands are coordinated end-on by the nitrogen atoms, while the allyl moiety completes the coordination sphere of the metal with its three carbon atoms binding the metal, in a formal η^3-allyl ligand.

The coordination of cyclopentadienyl and the two acetonitrile ligands is reasonably well described in the optimized geometry, with absolute deviations between the calculated and the experimental values for the Ru−X bond distances

ranging between 0.003 and 0.055 Å, and this holds for all the computational approaches tested. However, this is not the case for the allyl coordination and especially not for the bond distance between the metal and the carbon adjacent to the phenyl ring (C3, marked with an arrow in Figure 1). The calculated Ru−C3 bond length is unacceptably overestimated by 0.458 Å, when compared with the experimental distance. In other words, the allyl coordination asymmetry is greatly enhanced in the calculated structure. This asymmetry is reflected by the difference between the Ru−C1 and Ru−C3 distances (terminal carbon atoms), which is 0.20 Å in the X-ray structure and rises to 0.61 Å in the calculated geometry. This result is difficult to understand since the method employed, B3LYP/($I_M I_L$), performed well in the optimization of analogous and isoelectronic Ru(IV) complexes, [Ru(η^5-C_5H_5)(η^3-$CH_2CHCHC_6H_5$)LL′]$^{n+}$, such as the chloro acetonitrile molecule (L = Cl^-, L′ = CH_3CN, n =1) with maximum (Δ) and mean (δ) absolute deviations for the distances around the metal of 0.14 and 0.05 Å, respectively.[74] Here, the comparison is also made to the experimental structure of the Cp* analogue. For the carbonate complex mentioned above (LL′ = $CH_3OCO_2^-$, n = 1) the optimized geometry (Δ = 0.15 Å and δ = 0.07 Å) is also in good agreement with experimental data from the X-ray structure of the Cp* complex.[16] In addition, for these two complexes, the carbonate and the chloro acetonitrile molecules, the calculated Ru−C(allyl) distances are within 0.03−0.15 Å of the experimental values. On the other hand, the allyl coordination is also poorly described, at the B3LYP/($I_M I_L$) level, in the optimized geometry of the analogous dicationic bis(DMF) complex (DMF = dimethylformamide = Me_2-NCHO, complexed via the O-atom), although to a lesser extent than in the case of the bis(acetonitrile) molecule of Figure 1. In this case (L = L′ = DMF, n = 2) the Ru−C3 distance is overestimated by 0.390 Å, relative to that in the X-ray structure of the Cp* analogue,[17] while the coordination of the remaining ligands is well described by the calculations, with acceptable deviations (0.003−0.079 Å) for the Ru−X distances. These results suggest that the performance of B3LYP/($I_M I_L$) in the geometry optimization described is related to the charge on the cation, being much better for monocations than for dications. In the following discussion, we will focus only on the geometry optimization of the bis-(acetonitrile) complex, [Ru(η^5-C_5H_5)(η^3-$CH_2CHCHC_6H_5$)(CH_3-$CN)_2$]$^{2+}$, since this is the most puzzling case, and, in particular, on the problematic Ru−C3 distance.

Besides the limitations of the theory level, other reasons may be responsible for the mismatch between the calculated and the experimental Ru−C3 bond distance, e.g., solid-state packing effects and quality of the model. A close look at the X-ray structure seems to exclude the possibility that the presence of the counterions (PF_6^-) leads to a distortion of the allyl coordination geometry. The closest contact of the anion to the allyl ligand, 3.161 Å, is observed between the allyl carbon C2 and fluorine atoms and is rather long. All of the C3−F distances are longer than 3.6 Å. The second aspect concerns the adequacy of cyclopentadienyl (Cp) as a model of pentamethylcyclopentadienyl (Cp*), and it might be asked whether the stereochemical constraints imposed by the bulkier Cp*, or electronic effects associated with this ligand, affect the allyl coordination in a way that cannot be well reproduced by the calculations carried out with the simple model. Thus, the geometry of the real molecule, [Ru(η^5-C_5-Me_5)(η^3-$CH_2CHCHC_6H_5$)($CH_3CN)_2$]$^{2+}$, was also optimized at the same level of theory (B3LYP/($I_M I_L$)), in order to probe the quality of the model. The resulting geometry is very similar to the one obtained for the Cp model, with differences in the Ru−X bond lengths between 0.001 and 0.100 Å. In particular, the Ru−C3 distance in the Cp* complex (2.740 Å) is still 0.358 Å longer than the experimental one, although this value is slightly better than the one obtained with the Cp model (2.840 Å). This result demonstrates that the nature of the model used is not the main cause for the poor performance observed with B3LYP/($I_M I_L$).

The effect of allyl substituents on the calculated Ru−C(allyl) bonding distances was also analyzed by optimizing Ru(η^5-C_5H_5)(η^3-CH_2CHCHR)($CH_3CN)_2$]$^{2+}$ complexes using same conditions (B3LYP/($I_M I_L$)) with R = H and R = Me. For R = H, the C_s molecule obtained had two equal bonds between the metal and the terminal carbon atoms (2.298 Å), reasonably close (within 0.1 Å) to the usual Ru−C(allyl) bond lengths.[74] With R = Me, the optimized Ru−$C_{1/3}$ distances were 2.252 and 2.510 Å, revealing an asymmetric coordination of this ligand that is similar to that found when R = Ph (see above), with the longest Ru−C separation involving the substituted terminal carbon. This difference between the Ru−C1 and Ru−C3 bond lengths (0.26 Å) is, however, much less pronounced than for R = Ph (0.61 Å). This different behavior can hardly be explained solely by stereochemical factors, such as the size and shape of methyl and phenyl groups, specially taking into account that the "upright" conformation of phenyl minimizes repulsion with the closest acetonitrile ligand (see Figure 1).

The influence of the basis set on the optimized geometry of [Ru(η^5-C_5H_5)(η^3-$CH_2CHCHC_6H_5$)($CH_3CN)_2$]$^{2+}$ was tested through a series of calculations using the B3LYP functional and several basis sets. Three sets of basis (I_M, II_M, and III_M) were used to describe ruthenium, and another six (I_L, II_L, III_L, IV_L, V_L, and VI_L) for C, H, and N (see Computational Details). The Ru−C3 distances calculated with each basis set as well as the corresponding mean absolute deviation between the calculated and the experimental values of all Ru−X distances (δ) are summarized in Table 1. The basis sets selected are rather limited, but they include the typical basis set used in most computational studies recently published in all fields of chemistry. Besides, the option of using larger and better basis sets is not available for realistic medium sized systems and the computational capabilities of most users.

The results in Table 1 show a small dependence of the geometry obtained with the basis set used in the calculations. The Ru−C3 bond lengths differ at most by 0.075 Å, and the overall description of the coordination geometry of the entire complex is even more uniform, as reflected in the maximum difference of 0.018 Å for the δ values. However, a slight improvement occurs as better metal basis sets are used. Upon moving from I_M (LanL2DZ plus f-polarization) to II_M (SDD plus f-polarization), the Ru−C3 distance

Table 1. Calculated Ru−C3 Bond Distances (Å) and Mean Absolute Deviations of All Ru−X Distances (δ, Å) between the Optimized Geometry of $[Ru(\eta^5\text{-}C_5H_5)(\eta^3\text{-}CH_2CHCHC_6H_5)(CH_3CN)_2]^{2+}$ and the Experimental One[a]

basis set	Ru−C3/Å	δ/Å
exp.[b]	2.382	
(I_MI_L)	2.840 (0.458)	0.083
(II_MI_L)	2.794 (0.412)	0.065
(III_MI_L)	2.765 (0.383)	0.076
(II_MII_L)	2.810 (0.428)	0.066
(II_MIII_L)	2.827 (0.445)	0.068
(II_MIV_L)	2.833 (0.451)	0.071
(II_MV_L)	2.840 (0.458)	0.070
(II_MVI_L)	2.839 (0.457)	0.067

[a] The differences between each Ru−C3 value and the experimental one are given in parentheses (Å). [b] From the Cp^* analogue.[17]

shortens 0.046 and δ becomes smaller. The use of an all-electron basis set for Ru (III_M = 3-21G*) leads to another decrease in the Ru−C3 distance (0.029 Å), though the overall description of the complex geometry is slightly worse than the one obtained with II_M (δ is 0.011 Å higher). The minor improvement obtained in the optimized Ru−C3 distance, going from II_M to III_M, is not compensated by the increase in computational cost due to the large all-electron basis set on the second row transition metal. Therefore, in the discussion below, the II_M basis set (SDD ECP with valence triple-ζ and an added f-polarization function) will be kept to describe the Ru atom, and different basis sets will be tested for the ligands. As the size of the basis set for C, H, and N increases, the quality of the calculated geometry decreases. The addition of diffuse functions to the basis sets of the three allylic carbons going from (II_MI_L) to (II_MII_L) leads to Ru−C3 distances larger by 0.016 Å, while valence triple-ζ basis sets (III_L, IV_L, and V_L) lengthen this distance another 0.017−0.030 Å. The Ru−C3 distance obtained with the quadruple-ζ basis set (VI_L) is pratically equal to the one produced by V_L (within 0.001 Å). This effect is diluted in the overall description of the molecule, and the δ values are less sensitive to the basis set used for the ligands, as shown by a maximum difference of 0.006 Å.

The II_MI_L basis set (4-31G* for C, H, and N and the basis set described above for Ru) reveals the best compromise between accuracy and computational cost and was used for the following analysis of the effect of changing the functional and method within a wide range. The results are given in Table 2.

The results summarized in Table 2 clearly indicate the poor capability of all DFT functionals to describe the coordination geometry of the allyl ligand in the bis(acetonitrile) complex, overestimating the Ru−C3 bond length by amounts from 0.23 to 0.50 Å, though some differences worth noticing may be found. PW91 performs much better as a correlation functional than LYP or P86, both when pure DFT functionals (see the results obtained with BLYP, and BPW91 or PW91) and hybrid functionals are used (demonstrated by the results of B3LYP, B3PW91, and mPW1PW91). In fact, Adamo and Barone's hybrid functional resulting from modified Perdew−Wang exchange (mPW1PW91) yields one of the best results in Table 2, only surpassed by the hybrid functional of

Table 2. Calculated Ru−C3 Bond Distances (Å) and Mean Absolute Deviations for All Ru−X Distances (δ, Å) between the Experimental Structure[a] and the Optimized Geometry of $[Ru(\eta^5\text{-}C_5H_5)(\eta^3\text{-}CH_2CHCHC_6H_5)(CH_3CN)_2]^{2+}$, Obtained from HF, MP2, and Several DFT Functionals, with the (II_MI_L) Basis Set[b]

method	Ru−C3/Å	δ/Å
B3LYP	2.794 (0.412)	0.065
B3PW91	2.688 (0.306)	0.047
mPW1PW91	2.640 (0.258)	0.043
PBE1PBE	2.615 (0.233)	0.042
BP86	2.738 (0.356)	0.058
BLYP	2.884 (0.502)	0.088
BPW91	2.759 (0.377)	0.058
PW91	2.684 (0.302)	0.050
PBE	2.683 (0.301)	0.048
HF	2.978 (0.596)	0.123
MP2	2.314 (−0.068)	0.043

[a] From the Cp^* analogue.[17] [b] The differences between each Ru−C3 value and the experimental one are given in parentheses (Å).

Perdew, Burke, and Ernzerhof (PBE1PBE). This observation refers to the calculated Ru−C3 distance as well as to the overall description of the geometry of the molecule, accounted by parameter δ. Another interesting point is the similarity in the performance of the two nonempirical functionals, PW91 and PBE,[15] with practically identical results for the Ru−C3 distance and similar δ values.

It can be added that the change in performance of the functional with the complexity of the basis set, observed for B3LYP (see discussion on the results of Table 1, above), applies to other functionals and basis sets, being thus quite general. For example, with the hybrid functional, mPW1PW91, the Ru−C3 bond length increases 0.008 Å going from the I_L to the II_L basis set, with the addition of diffuse functions for the three allylic carbon atoms.

Similar calculations (geometry optimization of the Ru complex) were carried out with the ADF program, testing two functionals (PW91 and BP98) with large Slater basis sets associated with that program. The results obtained with the basis set TZ2P showed a very long Ru−C3 bond (2.656 Å with PW91, 2.676 Å with BP86). Changing the basis set size from TZ2P to the larger QZ4P for C, H, and N and keeping TZ2P for Ru lead to an increase of 0.01 or 0.04 Å in the Ru−C3 distance with PW91 and BP86, respectively. On the other hand, when the metal basis set is also changed from QZ2P to QZ4P (QZ4P for all atoms), there is a small improvement on the calculated Ru−C3 distance that shortens 0.001 Å with PW91 (0.04 Å with BP86) relative to the previous values. They are still longer than the TZ2P bond lengths. Importantly, very long Ru−C3 distances (2.66−2.72 Å) were obtained in all ADF optimizations, in agreement with the other DFT calculations discussed above.

With respect to the wavefunction based methods, HF performs miserably, but the electronic energy corrected MP2 yields the best result shown in Table 2. More relevantly, this method provides the only result acceptable for a semiquantitative study of the case study complex, though MP2 underestimates the Ru−C3 bond length by 0.068 Å. The coordination of the η^3-allyl ligand is well reproduced by the MP2/(II_MI_L) calculation with deviations of 0.002 and

−0.049 Å for the other two carbon atoms, C1 and C2, respectively. However, as a value of δ (0.043 Å) equal to the one obtained with mPW1PW91 hints, the description of the coordination geometry of the nitriles and of the cyclopentadienyl is poorer with MP2 than with that hybrid functional. The two Ru−N distances are underestimated by 0.05 Å with MP2 and only by 0.02 Å and 0.03 Å with mPW1PW91. The maximum deviation on the Ru−C(Cp) bonds with the hybrid functional corresponds to an underestimation of 0.02 Å, while for MP2 it is 0.10 Å. In fact, MP2 is at its worst when describing the Ru−C(Cp) bonds, and the 0.10 Å underestimation corresponds to the maximum deviation between the optimized and the experimental Ru−X distances (Δ). For all the DFT functionals tested the worse optimized bond length (Δ) is the Ru−C3 distance.

Interestingly, for MP2, basis set convergence, in size, seems to be practically achieved with (II_MI_L). In fact, an optimization at the MP2/(II_MIII_L) level yielded a geometry similar to the one obtained at the MP2/(II_MI_L) level, with the calculated Ru−C3 distance (2.301 Å) only 0.013 Å shorter than the one obtained with the smaller basis set (see Table 2). Unfortunately, our computational limitations prevent us from testing higher level wavefunction based methods or larger basis sets as well as preclude any mechanistic investigation even at the MP2/(II_MI_L) level. In fact, the MP2 geometry optimization of $[Ru(\eta^5\text{-}C_5H_5)(\eta^3\text{-}CH_2CHCHC_6H_5)(CH_3CN)_2]^{2+}$ was about 30 times longer than the average of any of the DFT calculations.

The nature of the potential energy surface (PES) of $[Ru(\eta^5\text{-}C_5H_5)(\eta^3\text{-}CH_2CHCHC_6H_5)(CH_3CN)_2]^{2+}$ on the region corresponding to the elongation of the Ru−C3 bond was explored, in order to try to understand the results discussed above. A B3LYP/(II_MI_L) calculation with the Ru−C3 distance fixed at the experimental value (2.38 Å) yielded a molecule only 3 kcal mol^{-1} less stable than the one obtained with the unconstrained calculation ($d_{Ru-C3} = 2.79$ Å). This reflects a reasonably flat PES with respect to Ru−C3 stretching, indicating a weak interaction with probably a strong component of dispersion forces. It is well-known that DFT occasionally fails in the description of that sort of interactions,[9,75,76] and, in addition, a recent benchmark study[77] demonstrated that MP2 performs better than DFT in the description of weak interactions.

4. Summary

We report the failure of five pure and four hybrid DFT functionals, including some of the most commonly used by the chemistry community, in the description of the geometry of a Ru(IV) allyl complex, $[Ru(\eta^5\text{-}C_5H_5)(\eta^3\text{-}CH_2CHCHC_6H_5)(CH_3CN)_2]^{2+}$, especially with respect to the Ru−C bond distance for the substituted carbon atom of η^3-allyl ligand. MP2 was found to perform much better and yield a reasonable result, even with a poor basis set, probably due to a better account of a weak Ru−C interaction.

Acknowledgment. The authors acknowledge the COST working group D24/0014/02. M.J.C. thanks FCT for financial support (POCI/QUIM/58925/2004). The authors acknowledge Prof. Jeremy N. Harvey for important suggestions and many helpful discussions.

Supporting Information Available: Tables of atomic coordinates for all the optimized species. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989; p 1.

(2) Niu, S.; Hall, M. B. *Chem. Rev.* **2000**, *100*, 353.

(3) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651.

(4) Ziegler, T. *J. Chem. Soc., Dalton Trans.* **2002**, 642.

(5) Silaghi-Dumitrescu, R.; Silaghi-Dumitrescu, I. *Chemtracts* **2005**, *18*, 595.

(6) Siegbahn, P. E. M. *Adv. Chem. Phys.* **1996**, *93*, 333.

(7) Ziegler, T.; Autschbach, J. *Chem. Rev.* **2005**, *105*, 2695.

(8) Diedenhofen, M.; Wagener, T.; Frenking, G. *Computational Organometallic Chemistry*; Cundari, T. R., Ed.; Marcel Dekker: New York, 2001; p 1.

(9) Harvey, J. N. *Annu. Rep. Prog. Chem. Sect. C* **2006**, *102*, 203.

(10) For a general account and a list of references see, for example: Foresman, J. B.; Frisch Æ. *Exploring Chemistry with Electronic Structure Methods*, 2nd ed.; Gaussian, Inc.: Pittsburgh, PA, 1996; pp 141−161.

(11) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986; pp 20−22.

(12) Becke, A. D. J. *Chem. Phys.* **1993**, *98*, 5648.

(13) Lee, C.; Yang, W.; Parr, G. *Phys. Rev. B* **1988**, *37*, 785.

(14) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200.

(15) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. *J. Chem. Phys.* **2005**, *123*, 062201.

(16) Hermatschweiler, R.; Fernández, I.; Breher, F.; Pregosin, P. S.; Veiros, L. F.; Calhorda, M. J. *Angew. Chem., Int. Ed.* **2005**, *44*, 4397

(17) Fernández, I.; Hermatschweiler, R.; Breher, F.; Pregosin, P. S.; Veiros, L. F.; Calhorda, M. J. *Angew. Chem., Int. Ed.* **2006**, *45*, 6386.

(18) *Gaussian 03, Revision C.02*; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Wallingford, CT, 2004.

(19) te Velde, G.; Bickelhaupt, F. M.; van Gisbergen, S. J. A.; Guerra, C. F.; Baerends, E. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931.

(20) Guerra, C. F.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391.

(21) *ADF2005.01*; SCM, Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands. http://www.scm.com (accessed Nov 20, 2006).

(22) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(23) Burke, K.; Perdew, J. P.; Wang, Y. *Electronic Density Functional Theory: Recent Progress and New Directions*; Dobson, J. F., Vignale, G., Das, M. P., Eds.; Plenum: 1998; p 1.

(24) Perdew, J. P. *Electronic Structure of Solids ' 91*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; p 11.

(25) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.

(26) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1993**, *48*, 4978.

(27) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533.

(28) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(29) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.

(30) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.

(31) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.

(32) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(33) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503.

(34) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 275.

(35) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 281.

(36) Head-Gordon, M.; Head-Gordon, T. *Chem. Phys. Lett.* **1994**, *220*, 122.

(37) Saebo, S.; Almlof, J. *Chem. Phys. Lett.* **1989**, *154*, 83.

(38) Dunning, T. H., Jr.; Hay, P. J. *Modern Theoretical Chemistry*; Schaefer, H. F., III, Ed.; Plenum: New York, 1976; Vol. 3, p 1.

(39) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270.

(40) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284.

(41) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 2299.

(42) Häussermann, U.; Dolg, M.; Stoll, H.; Preuss, H.; Schwerdtfeger, P.; Pitzer, R. M. *Mol. Phys.* **1993**, *78*, 1211.

(43) Kuechle, W.; Dolg, M.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1994**, *100*, 7535.

(44) Leininger, T.; Nicklass, A.; Stoll, H.; Dolg, M.; Schwerdtfeger, P. *J. Chem. Phys.* **1996**, *105*, 1052.

(45) Binkley, J. S.; Pople, J. A. Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.

(46) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J. Hehre, W. J. *J. Am. Chem. Soc.* **1982**, *10*, 2797.

(47) Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Defrees, D. J.; Pople, J. A. Binkley, J. S. *J. Am. Chem. Soc.* **1982**, *104*, 5039.

(48) Dobbs, K. D. Hehre, W. J. *J. Comput. Chem.* **1986**, *7*, 359.

(49) Dobbs, K. D. Hehre, W. J. *J. Comput. Chem.* **1987**, *8*, 861.

(50) Dobbs, K. D. Hehre, W. J. *J. Comput. Chem.* **1987**, *8*, 880.

(51) Ehlers, A. W.; Böhme, M.; Dapprich, S.; Gobbi, A.; Höllwarth, A.; Jonas, V.; Köhler, K. F.; Stegmann, R.; Veldkamp, A.; Frenking, G. *Chem. Phys. Lett.* **1993**, *208*, 111.

(52) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.

(53) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(54) Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, *27*, 209.

(55) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163.

(56) McClean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.

(57) Krishnan, R.; Binkley, J. S.; Seeger, R. Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(58) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033.

(59) Hay, P. J. *J. Chem. Phys.* **1977**, *66*, 4377.

(60) Raghavachari, K.; Trucks, G. W. *J. Chem. Phys.* **1989**, *91*, 1062.

(61) Binning, R. C.; Curtiss, L. A. *J. Comput. Chem.* **1995**, *103*, 6104.

(62) McGrath, M. P.; Radom, L. *J. Chem. Phys.* **1991**, *94*, 511.

(63) Schaefer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.

(64) Schaefer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.

(65) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(66) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(67) Versluis, L.; Ziegler, T. *J. Chem. Phys.* **1988**, *88*, 322.

(68) Fan, L.; Ziegler, T. *J. Chem. Phys.* **1991**, *95*, 7401.

(69) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(70) van Lenthe, E.; Ehlers, A.; Baerends, E. J. *J. Chem. Phys.* **1999**, *110*, 8943.

(71) Snijders, J. G.; Baerends, E. J. *At. Data Nucl. Data Tables* **1982**, *26*, 483.

(72) van Lenthe, E.; Baerends, E. J. *J. Comput. Chem.* **2003**, *24*, 1142.

(73) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41.

(74) Hermatschweiler, R.; Fernández, I.; Pregosin, P. S.; Watson, E. J.; Albinati, A.; Rizzato, S.; Veiros, L. F.; Calhorda, M. J. *Organometallics* **2005**, *24*, 1809.

(75) Kohn, W.; Meir, Y.; Makarov D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.

(76) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.

(77) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.

CT600341R