JCTC Journal of Chemical Theory and Computation

Huge Disulfide-Linkage'S Electron Capture Variation Induced by α-Helix Orientation

Elise Dumont,* Pierre-François Loos, Adèle D. Laurent, and Xavier Assfeld

Equipe de Chimie et Biochimie Théoriques, UMR 7565 CNRS-UHP, Institut Jean Barriol (FR CNRS 2843), Faculté des Sciences et Techniques, Nancy-Université, B.P. 239, 54506 Vandœuvre-lès-Nancy, France

Received May 14, 2008

Abstract: An active site containing a Cys-X-X-Cys motif (CXXC), where X denotes any amino acid, is always found in the thiol-disulfide oxidoreductase superfamily. Because of its very high propensity for N-termini of α-helices, we examine the effect of this secondary structure on the disulfide-linked CXXC electron affinity. A Cys-Gly-Pro-Cys motif (CGPC) is chosen as an example, as it is the canonical motif found in thioredoxins. QM/MM calculations (MP2/6−31+G**:CHARMM) establish that the electron capture is strongly favored by an N-terminal α-helix, due to the positive electrostatic potential in the vicinity of the active site. The enhancement of adiabatic electron affinity accounts for ca. 0.9 eV for a 12-residues helix and rapidly converges as the number of alanine residues increases. A close agreement between a reference thioredoxin (Trx h) and the corresponding model peptide is found (respectively +2.20 and +2.12 eV), in parallel with experimental redox potentials [Iqbalsyah et al. *Protein Sci.* **2006**, *15*, 2026−2030]. This suggests a simple additive rule for geometrical and electrostatic effects. The electron affinity of the CXXC active site is first considered in an isolated way. Then, the strong modulation of the electrostatic field created by the α-helix can be added up. This simple partition scheme allows a proper quantification of the ease of attachment of a low-energy electron.

Introduction

Rationalizing the outcome of biomolecules under ionizing radiations is of the utmost importance for the development of radiation therapy.[1] These radiations massively lead to the *in situ* formation of low-energy secondary electrons. Strong experimental evidence for their highly specific attachment on disulfide linkages of protein has recently been reported.[2] In this Letter, hybrid calculations establish that electron capture is strongly favored by an N-terminal α-helix (ca. 1 eV), due to the positive electrostatic potential in the vicinity of the active site. Conversely, a C-terminal helix lowers the disulfide-linkage electron affinity. We find a close agreement between a reference thioredoxin (Trx h, PDB ID code 1TOF) and the corresponding model peptide (respectively +2.20 and +2.12 eV). This provides a simple way to predict the ease of attachment of a low-energy electron.

Disulfide bridges are essential for protein structure and reactivity. However, they are prone to damage, due to the relative weakness of covalent sulfur−sulfur bonds. Many questions arise concerning the following: the ease with which a disulfide linkage can capture an electron, the factors which govern the site of attachment, and the outcome of so-formed disulfide radical anion intermediates. Previous studies have brought some important partial answers: pulse radiolysis,[3] electron capture dissociation,[4] Coulomb-stabilization cleavage,[5] etc. Yet, we are still far from a sound understanding of the one-electron addition on disulfide-linkages. In order to gain more insights into this reaction in macromolecular systems, both geometrical and electrostatic contributions have to be considered and properly evaluated. Disulfide linkages are most often cyclic in proteins, and we recently demonstrated that disulfide electron affinity is governed by ring strain (1,2-dithiacycloalkanes)[6] or topological frustration (peptides).[7] The scope of this Letter is to assess the contribution of electrostatic effects occurring in larger biomolecules.

Enzymes of the thioredoxin superfamily are extensively studied due to their importance in many biological redox processes.[8] Their common active site is a CXXC motif forming a hairpin-like loop (β-turn), where X denotes any amino acid. Isolated CXXC peptides exhibit different redox potentials, and the whole tertiary structure (thioredoxin fold) induces an overall constant shift toward more reducing values.[9] More specifically, the key role of the N-terminal α-helix,[10] although well established: lowering of pK_a,[11] enhancement of redox potential,[12] etc., needs a more quantitative assessment for an in-depth understanding of proteins biochemistry. Recent experimental studies have used intermediate size designed peptides. Neat agreement between the many-complicated protein and a small peptide containing *solely* the α-helix has been reported. For instance, Doig and co-workers recently established that the redox potential for the model peptide NH_2−CAACAAAAKA-

* Corresponding author e-mail: Elise.Dumont@cbt.uhp-nancy.fr.

10.1021/ct800161m CCC: $40.75
Published on Web 07/23/2008

AAAKGY-NH_2 is very close to the characteristic value of thioredoxins.[12] This has been *a posteriori* related to the very high propensity of the CXXC motif to be located at the N-terminus end of α-helices. Theoretical calculations have also proved their usefulness for providing new insights into the role and action of terminal helices in a wide range of situations.[13]

Computational Details

QM/MM calculations have been performed (MP2/6−31+G**: CHARMM, see the Supporting Information) to study the effect of the α-helix on the disulfide bond's electron affinity.

The second-order Møller-Plesser perturbation theory (MP2) method has been used to ensure a proper description of a two center three electron bond. MP2 calculations, alongside with large basis sets including diffuse functions, have proved to give accurate results for treating electron capture by simple disulfides.[6,7] The choice for the Gaussian basis set (6−31+G**) is based on previous benchmark calculations on diethyl disulfide.[6]

The MM surrounding is described with the CHARMM force field using the CHARMM27 parameters for proteins.[14] The QM/MM calculations are performed with a modified version of the Gaussian 03 package[15] linked to the Tinker software[16] for the MM calculations. In the present case, the QM/MM frontiers are set at the C_α-C_β bonds of the two disulfide-linked cysteine residues. This frontier has been discussed in a previous article:[7] the key idea is to take advantage of the highly localized character of the disulfide bond density.

Results and Discussion

We consider a CGPC tetrapeptide, commonly found in thioredoxins, grafted onto an alanine homopolymers chain in an α-helix conformation. We test both possible orientations (N- and C- termini) for the helix, and we let its length n vary from 0 to 24 residues.

Geometrical parameters for both neutral and radical anionic forms are given in the Supporting Information. As observed for other cyclic disulfides,[6,7] geometries of radical anion compounds are highly malleable. The initial tetrapeptide Ace-CGPC-NHMe ($n = 0$) has a positive adiabatic electron affinity (EA_{ad} = 1.20 eV), which simply reflects the topological frustration of the neutral tetrapeptide.[7] This value is strongly enhanced by the presence of an N-terminal α-helix (ca. 1 eV). The variation of EA_{ad} as a function of n is displayed in Figure 1a.

A short alanine chain ($n = 1-4$) induces weak and irregular variations of electron affinity. They come from the subtle interplay of electronic and geometrical contributions, as structures are highly flexible. Disulfide linkage propensity to capture an electron increases monotonously with the number of Ala residues, ranging from +2.07 ($n = 8$) to +2.43 eV ($n = 24$). This enhancement simply corresponds to the usual representation of the α-helix, which creates an intense dipole (see the Supporting Information). The N-terminal orientation imposes a positively charged extremity of the associated electric dipole in the vicinity of the CXXC motif. The radical anion is preferentially stabilized, hence the higher electron affinities. Our calculations provide a quantitative assessment of electrostatic contributions, which account for ca. 1 eV. This proves that disulfide electron affinity is considerably modulated by the environment (tertiary structure); by comparison, mutation of intraloop residues X accounts for less than 0.3 eV.[7] Interestingly enough, a similar result has been recently obtained for redox potential of a CAAC motif placed at the N-terminal of a 12-residue-long α-helix.[12]

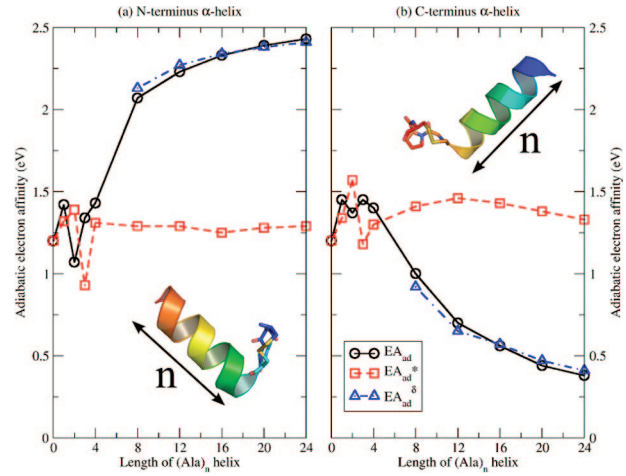


Figure 1. Electron affinity (in eV) of a disulfide-linked CGPC tetrapeptide grafted on a (a) N-terminus and (b) C-terminus α-helix, as a function of the number of Ala residues (n). The EA_{ad}, EA_{ad}*, and EA_{ad}^{δ} values are represented by black solid (circles), red dashed (square dots), and blue dashed lines (triangle dots), respectively. EA_{ad} and EA_{ad}* correspond to the QM/MM calculations with and without the electrostatic embedding respectively. EA_{ad}^{δ} corresponds to simplified calculations where the helix dipole is modeled with two point charges.

This evolution of adiabatic electron affinity EA_{ad} (Figure 1), once the CXXC motif is grafted onto the $(Ala)_n$ chain, arises from two contributions: (i) the outer electric field and (ii) the geometrical interplay between the α-helix and the β-turn fragments. The rapid and regular increase suggests the predominance of the dipole field effect: the positive charge remains almost fixed with respect to the disulfide-linkage barycenter, while the negative charge (C-terminus) moves away from it as the size of the helix increases. In order to ascertain this hypothesis and to gain further insights into the nature of the electrostatic modulation, we performed single point calculations where the partial charges of alanine residues were turned off. In this way, the electrostatic embedding of the α-helix is switched off: the corresponding electron affinities are denoted EA_{ad}*. Their values, given in the Supporting Information, are very close to the reference for isolated tetrapeptide (Figure 1a). This clearly establishes that the α-helix modulation almost exclusively comes from electrostatic contributions. The small deviations observed for $n = 1-4$ are caused by the limited structural reorganization. A simple yet efficient model for the α-helix dipole[13] is given by placing two point charges δ, respectively positive and negative, on the nitrogen and carbon termini atoms of the α-helix. Electron affinities EA_{ad}^{δ} were computed following this scheme, with all other helix point charges set to zero. A value of ±0.4 e accurately reproduces the helix effect (Figure 1a).

Conversely, a C-terminal helix tends to disfavor electron capture by the disulfide-linkage (Figure 1b). The previous interpretation is directly transferable and is further confirmed by the electrostatic potential (see Figure 2) . Quantitatively, a

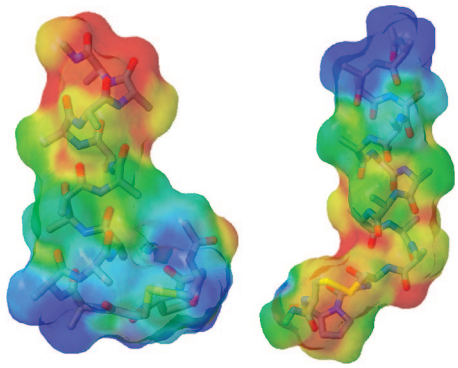


Figure 2. Electrostatic potential mapped on the van der Waals surface of a disulfide-linked CGPC tetrapeptide grafted on a 12-residues δ-helix (neutral form), at the N-terminus (left) or at the C-terminus (right). Red corresponds to negative values of the electrostatic potential, blue to positive ones.

noticeable weakening of the effect is observed: for the same number of alanine residues, the shifting toward lower electron affinities is roughly halved in comparison with the enhancement produced by an N-helix. Variations of the C-terminal domain is used in proteins of the thioredoxin-like family to reach a modulation of redox potential.[17]

If the previous results provide a reliable measure of α-helix electrostatic contributions, the comparison with a full protein structure is legitimate. We chose thioredoxin h, a protein from the eukaryotic green alga *C. reinhardtii*, as an example. It features a CGPC motif linked to the N-terminal group of the α_2 motif and the C-terminal group of the β_2 motif (C-terminal strand, see Figure 2). The difference between the isolated (capped) tetrapeptide Ace-CGPC-NHMe and the full protein structure comes from the specific environment, the highly characteristic thioredoxin fold. Our QM/MM calculations predict a value of +2.12 eV for the adiabatic electron affinity EA_{ad}, which is significantly higher than the +1.20 eV value obtained for the Ace-CGPC-NHMe tetrapeptide. It is noteworthy that this value (2.12 eV) is very close to the value obtained for the model peptide Ace-A_{12}-CGPC-NHMe (EA_{ad} = 2.23 eV). This agreement could seem rather surprising given the difference between the real full protein and the model peptide. However, a similar result was recently reported by Doig and co-workers[12] The modeled CAAC peptide, also grafted on a 12-residue-long N-terminus helix, was found to possess a redox potential of −220 mV, which is very close to the experimental value for the corresponding protein (−231 mV). This parallel ascertains the value of these designed peptides as model systems.

Conclusion

In the present study, the modulation created by an N- or C-terminal α-helix on disulfide electron affinity of a common disulfide-linked CGPC motif has been investigated. The electrostatic effects are of the utmost importance in proteins and lead to a dramatic modulation of disulfide-linkage electron affinity. A simple partition scheme based on a decomposition of geometrical and electrostatic effects provides a robust and simple way to investigate the propensity of thioredoxins to fix an electron on their unique disulfide bridge. This work sheds new light on the factors which govern the site of attachment of low-energy electrons. The analogy between studies on redox potentials and electron affinities is insightful and may contribute to the general understanding on disulfide reactivity.

Acknowledgment. The authors are grateful to Dr. Nicolas Ferré (Université Aix-Marseille I, France) for providing a local version of the link atom scheme. They also acknowledge financial support from the Jean Barriol Institute (FR CNRS 2843).

Supporting Information Available: Representation of the thioredoxin h structure, additional details of the QM/MM scheme, frontier orbitals and electrostatic potentials, scheme of the dipole created by an α-helix, and table of geometrical parameters and adiabatic electron affinities. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Sanche, L. *Eur. Phys. J. D* **2005**, *35*, 367–390.

(2) Weik, M.; Ravelli, R. B.; Silman, I.; Sussman, J. L.; Gros, P.; Kroon, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 623–628.

(3) Purdie, J. W.; Gillis, H. A.; Klassen, N. V. *Can. J. Chem.* **1973**, *51*, 3132–3142.

(4) Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 2857–2862.

(5) Sawicka, A.; Skurski, P.; Hudgins, R. R.; Simons, J. *J. Phys. Chem. B* **2003**, *107*, 13505–13511.

(6) Dumont, E.; Loos, P.-F.; Assfeld, X. *Chem. Phys. Lett.* **2008**, *458*, 276–280.

(7) Dumont, E.; Loos, P.-F.; Assfeld, X. *J. Phys. Chem. B* **2008**, in prepartion.

(8) Holmgren, A. *Annu. Rev. Biochem.* **1985**, *54*, 237–271.

(9) Siedler, F.; Rudolph-Bohner, S.; Doi, M.; Musiol, H.-J.; Moroder, L. *Biochemistry* **1993**, *32*, 7488–7495.

(10) Hol, W. G.; van Duijnen, P. T.; Berendsen, H. J. *Nature* **1978**, *273*, 443–446.

(11) Quiocho, T.; Warshel, A. *J. Mol. Biol.* **1995**, *253*, 799–812.

(12) Iqbalsyah, T. M.; Moutevelis, E.; Warwicker, J.; Errington, N.; Doig, A. J. *Protein Sci.* **2006**, *15*, 1945–1950.

(13) Aqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *88*, 2026–2030.

(14) MacKerel, A. D., Jr; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(15) *Gaussian 03, Revision C.02*; Frisch, M. J. et al. Gaussian, Inc.: Wallingford, CT, 2004.

(16) Ponder, J. W. *Tinker, version 4.2*; Washington University: St. Louis, MO, 2004.

(17) Pan, J. L.; Bardwell, J. C. A. *Protein Sci.* **2006**, *15*, 2217–2227.

CT800161M

JCTC Journal of Chemical Theory and Computation

Magnetostructural Dynamics with the Extended Broken Symmetry Formalism: Antiferromagnetic [2Fe-2S] Complexes

Nisanth N. Nair,* Eduard Schreiner,† Rodolphe Pollet,‡ Volker Staemmler, and Dominik Marx

Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany

Received March 13, 2008

Abstract: A general spin-projection framework is laid out which allows one to perform ab initio molecular dynamics simulations of antiferromagnetically coupled spin dimers. The method extends the well-established broken-symmetry formalism and is systematically and consistently improvable. It allows for accessing structure within the same spin-projection approximation as employed to compute the exchange coupling constant J of such complexes in their low-spin state as a function of time. The resulting time evolution of the exchange coupling, $J(t)$, can be analyzed most conveniently in terms of the corresponding power spectrum, $J(\omega)$, thus giving access to dynamical magnetostructural properties. The method has been implemented using a well-tested approximation to spin-projection and was applied to a minimal [2Fe-2S] model, i.e. to the $[Fe_2S_2(SH)_4]^{2-}$ complex at 300 K in vacuo. Thermal fluctuations at room temperature are found to change the antiferromagnetic coupling J by about 50% with respect to the average value, and the features of its power spectrum, $J(\omega)$, can be traced back to a coupling of J to particular vibrational modes.

1. Introduction

There is a vast amount of literature on the computation of exchange coupling constants in bi- and polynuclear transition metal complexes using both wave function-based and density-based methods.[1–9] Part of this interest is stimulated by the crucial importance of magnetically coupled transition metal centers in metalloproteins, such as nonheme iron proteins.[10–13] In particular, in iron−sulfur proteins prosthetic groups of the type [*n*Fe-*m*S] consist of a varying number of iron centers and sulfur atoms that are bridging the iron centers. In addition, the [*n*Fe-*m*S] complexes are bound to the backbone by sulfur atoms typically belonging to cysteine residues or by nitrogen atoms from histidines in the case of Rieske proteins. By virtue of the different oxidation states of the iron sites such cofactors serve as electron transfer relays and catalytic redox agents not only in bacteria but also in plants and higher animals.[10–13] Thus, particular combinations of oxidized (ferric) Fe(III) and reduced (ferrous) Fe(II) sites can establish different total oxidation states in a given biomolecular environment. In addition to tunable oxidation states, the magnetic structure can be rather involved, as the individual iron centers typically prefer a high-spin (HS) state, $S_{Fe} = 5/2$, in the oxidized state with a formal charge of +3 and a d^5 configuration.

A particularly important and well-studied family is the fully oxidized ferredoxin where the two HS metal centers in the inorganic [2Fe-2S] complex are magnetically coupled. This magnetic system consists of two paramagnetic transition metal centers, Fe^{3+}, both of which have an equal number of electrons, bridged by two diamagnetic centers, S^{2-}. In the ground state, the local spins at the metal sites are antiferromagnetically coupled to an overall low-spin (LS) state of this homovalent binuclear core.[14,15] To understand the

* Corresponding author e-mail: nisanth.nair@theochem.ruhr-uni-bochum.de.

† Present address: Theoretical and Computational Biophysics Group, Beckman Institute, Urbana, IL 61801.

‡ Permanent address: Laboratoire Francis Perrin, Commissariat a l'Énergie Atomique, 91191 Gif-sur-Yvette, France.

10.1021/ct800089x CCC: $40.75
Published on Web 06/26/2008

antiferromagnetic coupling it is worth looking at the exchange interactions between the two metal centers.[16–20] The two sites can interact either by "direct" exchange or by "superexchange". Direct exchange is a through-space interaction depending on the direct overlap of the orbitals at the two metal centers. It decays exponentially with the metal−metal distance and causes always a ferromagnetic alignment. Superexchange, on the other hand, is mediated through a diamagnetic bridge and yields antiferromagnetic coupling in most cases.

Two complementary pictures are frequently involved to explain the physical origin of superexchange.[4] In the Anderson model,[17,18] one assumes that an electron is hopping from one metal center to the other through the diamagnetic bridge. In the HS case, in which all unpaired electrons on the two metals have the same spin, such hopping is not possible because all orbitals with this spin are already occupied. However, in states with lower spin, this hopping leads to a lowering of the energy, thus to an antiferromagnetic coupling. The theoretical description of this effect requires the inclusion of "charge transfer" excitations, i.e. in our system $d^5d^5 \rightarrow d^4d^6$ excitations.[10,21] On the other hand, in the Kahn model,[22] one uses wave functions in which the partially occupied orbitals at the two metal centers, the "magnetic orbitals", are allowed to be nonorthogonal. They can be delocalized and overlap at the diamagnetic bridge. Again, in the HS state the Pauli exclusion principle prevents this overlap, but in states with lower spin it reduces the energy and leads to an antiferromagnetic coupling. Both models have in common that the strength of the superexchange coupling, i.e. the exchange coupling constant J, depends sensitively on the distance between the metal and the bridging atom in the "exchange pathway" as well as on the metal-bridge-metal angle.[19,20,23]

Accurate treatments of the true multireference nature of such antiferromagnetically coupled LS ground states require, in principle, sophisticated electronic structure methods.[5–9,19] A much simpler approach is to use a single-determinant spin-polarized (unrestricted) representation to describe their electronic structure, usually referred to as the broken-symmetry (BS) determinant, which describes an electronic configuration with unpaired electrons localized mainly on the two transition metal centers. A clear discussion of the physical content of such BS single-determinant wave functions can be found in ref 21. Obviously, the BS state does not represent the true antiferromagnetic ground state.[24,25] Moreover, this state features spin contamination since it is not an eigenstate of the total spin operator. Thus, using a single BS determinant imposes a severe approximation, although it can be used straightforwardly in the framework of ab initio molecular dynamics (AIMD) to take fluctuation effects into account.

Sophisticated quantum-chemical methods, which can be employed to correctly treat the multireference character of the ground-state wave function,[5–9,23,26] however impose severe limitations on the systems that are accessible and the questions that can be addressed. A viable alternative is to have recourse to exact[26,27] or approximate[29–31] spin-projection schemes, which offer a compromise in terms of cost versus accuracy and allow the calculation of the exchange coupling constant J directly. They rely on an approximate purification of the spin state[24,32] by carrying out *independent* total energy calculations for several spin configurations.[1,33] Such schemes have proven to perform particularly well to compute exchange coupling constants for binuclear homovalent complexes in static electronic structure calculations.[34–37]

Here, we extend the idea of using a two-determinant spin-projection technique formulated in a generalized framework, which is suitable for describing the antiferromagnetic coupling of homovalent binuclear iron complexes. In particular, this ansatz is the basis for implementing multideterminant Car−Parrinello AIMD techniques thus yielding the extended broken-symmetry (EBS) scheme. The EBS method allows one to investigate the interplay of structural changes and magnetic properties at finite temperature, thus extending the Goodenough-Kanamori rules[38–40] from the static limit to unconstrained dynamics. Along this line of thinking about magnetic interactions, a very useful tool is to analyze the power spectrum, $J(\omega)$, obtained from the dynamical evolution, $J(t)$, of the exchange coupling constant, J, that is computed "on the fly" along the EBS AIMD trajectories. A particular version of this scheme has already been applied to analyze the dynamical magnetostructural properties of a specific ferredoxin in the framework of a QM/MM generalization,[41] where the interplay between magnetic dynamics and vibrational/configurational motion can be understood in terms of $J(\omega)$.

2. Methodology

2.1. Getting Started: Heisenberg Model. On a phenomenological level the exchange coupling in homovalent magnetic dimers may be modeled by the Heisenberg (or Heisenberg−Dirac-van Vleck) spin Hamiltonian[42–44]

$$\hat{H} = -2J\hat{S}_A\hat{S}_B \tag{1}$$

where $\hat{S}_A$ and $\hat{S}_B$ are effective local spin operators at the two sites A and B, respectively. Usually, J is referred to as the exchange (magnetic) coupling constant between the two paramagnetic centers and the sign is taken in such a way that $J > 0$ implies ferromagnetic coupling, whereas $J < 0$ describes antiferromagnetic or ferrimagnetic coupling.

Defining the total (coupled) spin operator $\hat{S}$ in terms of the local spin operators $\hat{S}_A$ and $\hat{S}_B$

$$\hat{S} = \hat{S}_A + \hat{S}_B \tag{2}$$

the Hamiltonian (1) may be rewritten[45] as

$$\hat{H} = -J(\hat{S}^2 - \hat{S}_A^2 - \hat{S}_B^2) \tag{3}$$

and hence the energy of a pure spin state is in general given by

$$E^S = -J[\langle\hat{S}^2\rangle^S - \langle\hat{S}_A^2\rangle - \langle\hat{S}_B^2\rangle] = -J[S(S+1) - S_A(S_A+1) - S_B(S_B+1)] \tag{4}$$

where $S \in [S_{min}, S_{max}]$ is the total spin quantum number of the coupled system; S_A and S_B are the local spin quantum numbers, which are kept fixed for the sites A and B. Thus,

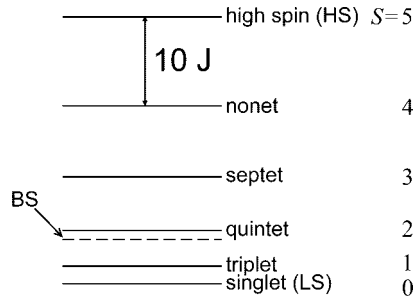


Figure 1. Sketch of a spin-ladder generated by Heisenberg's spin Hamiltonian (eq 1) for a d^5-d^5 binuclear system with an antiferromagnetic coupling. The position of the BS state within the ladder is indicated by the dashed line. Note: the BS state is not the $S = 0$ antiferromagnetic LS eigenstate of the coupled spins.

Hamiltonian (1) generates Heisenberg's spin-ladder of eigenstates (see Figure 1) with total spins S in the range from $S_{min} = |S_A - S_B|$ (low-spin, LS) to $S_{max} = S_A + S_B$ (high-spin, HS).

Thus, for given fixed local spin quantum numbers S_A and S_B at the two sites, eq 4 gives rise to Landé's interval rule

$$E^S - E^{S-1} = -2JS \tag{5}$$

which allows one to calculate the exchange coupling constant J knowing the energies of only two states, E^S and E^{S-1}. Moreover, it creates the possibility to access the energies of *all* spin states in the ladder (see Figure 1) if J and the total energy of *one* state is known. To calculate these quantities one needs the Clebsch-Gordan-coupled spin eigenstates $|S_A S_B; SM\rangle$ of the employed many-electron Hamiltonian. On the one hand, these exact spin eigenstates are not easily accessible computationally for many-electron systems such as molecular complexes or solids in the framework of realistic total energy electronic structure calculations. On the other hand, it appears viable to evaluate the expectation value $\langle \hat{S}^2 \rangle^S$ of the total spin operator in the total energy expression

$$E^S = -J[\langle \hat{S}^2 \rangle^S - S_A(S_A + 1) - S_B(S_B + 1)] \tag{6}$$

using approximately coupled states as obtained by approximate electronic structure calculations. Importantly, this will yield access, via eq 5, not only to J itself but also to any coupled state of the entire spin-ladder if only one of its states is known.

2.2. The Broken Symmetry Approach to *J*. In particular, the high-spin (HS) and broken-symmetry (BS) states may be used to estimate the coupling constant J using eq 6 in conjunction with Landé's rule.[28,29,45] The total energies of the S = HS and the S = BS states, i.e. E^{HS} and E^{BS}, are given by

$$E^{HS/BS} = -J[\langle \hat{S}^2 \rangle^{HS/BS} - S_A(S_A + 1) - S_B(S_B + 1)] \tag{7}$$

and can be easily calculated using a single Slater or Kohn−Sham (KS) spin-polarized (unrestricted) determinant. Here, it is assumed that the considered state is an eigenstate of the local $\hat{S}_A^2$ and $\hat{S}_B^2$ operators. From eq 7 it is clear that J can be readily expressed using the difference of the energies of the BS and HS states

$$J = \frac{E^{BS} - E^{HS}}{\langle \hat{S}^2 \rangle^{HS} - \langle \hat{S}^2 \rangle^{BS}} \tag{8}$$

which is the same formula written down earlier by Yamaguchi and co-workers.[46,47]

At this point it is necessary to address the role of the so-called "magnetic orbitals" in the calculation of the expectation value of the $\hat{S}^2$ operator.[21] Löwdin's formally *exact* expression[48] for the total spin of an n-electron system (with n^α and n^β spin-up and spin-down electrons respectively), i.e.

$$\langle \hat{S}^2 \rangle = -\frac{n(n-4)}{4} + \int \Gamma(\mathbf{r}_1\sigma_1, \mathbf{r}_2\sigma_2 | \mathbf{r}_1\sigma_2, \mathbf{r}_2\sigma_1) d\mathbf{x}_1 d\mathbf{x}_2 \tag{9}$$

where $\mathbf{x}_i = (\mathbf{r}_i, \sigma_i)$ is the combined spatial and spin coordinate of electron i and $\Gamma(\mathbf{r}_1\sigma_1, \mathbf{r}_2\sigma_2 | \mathbf{r}_1\sigma_2, \mathbf{r}_2\sigma_1)$ is the two-particle density matrix

$$\Gamma(\mathbf{x'}_1, \mathbf{x'}_2 | \mathbf{x}_1, \mathbf{x}_2) = \frac{n(n-1)}{2} \int \Psi^*(\mathbf{x'}_1, \mathbf{x'}_2, ..., \mathbf{x}_n) \times \Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n) d\mathbf{x}_3 ... d\mathbf{x}_n \tag{10}$$

can be rewritten in terms of off-diagonal elements of the two-particle density matrix[49]

$$\langle \hat{S}^2 \rangle = \left(\frac{n^\alpha - n^\beta}{2}\right)\left(\frac{n^\alpha - n^\beta}{2} + 1\right) + n^\beta + 2 \int \Gamma(\mathbf{r}_1\alpha, \mathbf{r}_2\beta | \mathbf{r}_1\beta, \mathbf{r}_2\alpha) d\mathbf{r}_1 d\mathbf{r}_2 \tag{11}$$

where $n^\alpha + n^\beta = n$, and it is assumed without loss of generality that $n^\alpha \geq n^\beta$.

In order to rewrite this equation in terms of local spin operators it is necessary to substitute $n^\alpha = n^\alpha_{mag} + n^\alpha_{nmag}$ and $n^\beta = n^\beta_{mag} + n^\beta_{nmag}$, where n^α_{mag} (n^β_{mag}) and n^α_{nmag} (n^β_{nmag}) is the number of unpaired and paired, i.e., "magnetic" and "nonmagnetic" α (β) electrons, respectively. Since $n^\alpha_{nmag} = n^\beta_{nmag}$, eq 11 simplifies to

$$\langle \hat{S}^2 \rangle = \left(\frac{n^\alpha_{mag} - n^\beta_{mag}}{2}\right)\left(\frac{n^\alpha_{mag} - n^\beta_{mag}}{2} + 1\right) + n^\beta_{mag} + \Theta \tag{12}$$

where

$$\Theta = n^\beta_{nmag} + 2 \int \Gamma(\mathbf{r}_1\alpha, \mathbf{r}_2\beta | \mathbf{r}_1\beta, \mathbf{r}_2\alpha) d\mathbf{r}_1 d\mathbf{r}_2 \tag{13}$$

which is an exact expression that does not rely on any orbital picture.

In the BS state $n^\alpha_{mag} = 2S_A$ and $n^\beta_{mag} = 2S_B$, whereas for the HS state $n^\alpha_{mag} = 2(S_A + S_B)$ and $n^\beta_{mag} = 0$. Thus, eq 8 may be reformulated using the explicit expression for the total spin operators to obtain a general formula for the coupling constant J

$$J = \frac{E^{BS} - E^{HS}}{4S_A S_B - \Theta^{BS} + \Theta^{HS}} = \frac{E^{BS} - E^{HS}}{S^2_{max} - S^2_{min} - \Theta^{BS} + \Theta^{HS}} \tag{14}$$

based on two total energies E^{BS} and E^{HS} together with the corresponding terms Θ^{BS} and Θ^{HS}, respectively. Note, for

a pure spin state $|S_A S_B; SM\rangle$ one finds $\langle \hat{S}^2 \rangle^S = S(S+1)$, i.e., $\Theta = 0$, which makes clear the well-known point that Θ is a spin correction term.

In order to understand the physical meaning of the spin correction terms Θ^{HS} and Θ^{BS}, it is useful to consider the integral in eq 11 from the point of view of unrestricted Hartree−Fock (UHF) theory, i.e., within a convenient orbital picture.[24] In this one-determinantal orbital approximation Θ can be expressed for the HS and BS state as

$$\Theta_{UHF} = n^{\beta}_{nmag} - \sum_{i}^{m}\sum_{j}^{m} f_i^{\alpha} f_j^{\beta} \left| \int_{ij}^{\alpha\beta} \right|^2 \tag{15}$$

where $f_i^{\sigma} = \{0, 1\}$ denotes the occupation number of the HF spin orbital ψ_i^{σ}, m is the number of orbitals, and $\int_{ij}^{\alpha\beta} = \langle \psi_i^{\alpha} | \psi_j^{\beta} \rangle$ is the overlap matrix. For a closed shell case, when the spatial parts of the α and β parts are identical $|\int_{ij}^{\alpha\beta}| = \delta_{ij}$ and the double sum in eq 15 reduces to the number of occupied states, i.e. the number of paired electrons, such that the two terms in eq 15 cancel each other, and thus $\Theta = 0$.

For open shell cases the α and β electrons of the lower-lying closed shells do not occupy the same spatial orbitals, since the interactions are different between electrons with same spin and electrons with opposite spin. Therefore, to stay within the UHF picture, the double sum will yield less electrons than the total number of β electrons, which implies immediately that the expectation value of $\hat{S}^2$ will be larger than the exact one. This gives rise to spin contamination. In cases like the iron sulfur systems considered in this work, where there are two radical sites of same spin, the double sum takes care of the nonorthogonality of the doubly occupied orbital of α and β spins and the overlap of magnetic orbitals. Thus, Θ is a "correction term" and takes into account the spin contamination and the overlap of magnetic orbitals when $\langle \hat{S}^2 \rangle^{HS/BS}$ is evaluated in expressions like eq 8.

Having discussed the correction Θ in terms of the UHF approximation, it should be stressed that this term can be alternatively approximated in the framework of density functional theory following for instance ref 49. In particular, using the exchange-only homogeneous electron gas leads to the expression

$$\Theta_{LSD} = -\int' [\rho^{\alpha}(\mathbf{r}) - \rho^{\beta}(\mathbf{r})] d\mathbf{r} \tag{16}$$

in the spirit of the local spin density (LSD) approximation,[49] where the prime indicates that the integration is carried out only in those regions of space where the spin density is negative, i.e. for $[\rho^{\alpha}(\mathbf{r}) - \rho^{\beta}(\mathbf{r})] < 0$; the spin density can of course be computed using a density functional beyond the LSD approximation. Alternatively, one might just use the spin-polarized KS orbitals φ_i instead of the unrestricted HF orbitals ψ_i to compute the orbital overlap matrix $\int_{ij}^{\alpha\beta}$ in eq 15 leading to

$$\Theta_{nonint} = n^{\beta}_{nmag} - \sum_{i}^{m}\sum_{j}^{m} f_i^{\alpha} f_j^{\beta} \langle \varphi_i^{\alpha} | \varphi_j^{\beta} \rangle^2 \tag{17}$$

which yields an approximation in the spirit of the (fictitious) noninteracting KS reference system concept,[49] keeping in mind that the KS determinant is not the wave function of the (real) interacting many-electron system. Both density functional based approximations to $\langle \hat{S}^2 \rangle$, and thus to Θ in expressions such as eq 14, have been shown to perform reasonably well for a set of atoms and small open-shell molecules,[49] see also ref 50 for further assessment and development. It is mentioned in passing that this general idea is open to more sophisticated spin correction treatments, possibly in connection with more elaborate density functionals as e.g. addressed in refs 50 and 51 or by incorporating spin symmetry into the KS treatment as done in the spin-restricted ensemble-referenced KS method.[52–54]

In order to proceed, it is useful to consider various approximations to simplify Θ for the HS and BS states, which in turn will lead to corresponding approximations to eq 8 in order to estimate J. First of all, a useful approximation is to neglect Θ in the HS state, $\Theta^{HS} = 0$, since spin contamination effects are typically rather small for HS states.[21] Thus, using either the unrestricted HF or noninteracting KS interpretation of Θ, eq 14 simplifies readily to

$$J = \frac{E^{BS} - E^{HS}}{S^2_{max} - S^2_{min} - n^{\beta}_{nmag} + \sum_i^m \sum_j^m f_i^{\alpha,BS} f_j^{\beta,BS} |\int_{ij}^{\alpha\beta,BS}|^2} \tag{18}$$

In the special case of having only one electron on each site and considering only the magnetic orbital contribution to the overlap double sum, this expression reduces to the formula

$$J = \frac{E^{BS} - E^{HS}}{1 + |\int_{mag}^{\alpha\beta,BS}|^2} \tag{19}$$

derived earlier in ref 33 along different lines. Next, the overlap contribution of the magnetic orbitals can be simplified further for systems with more than one electron per site by treating it in two limiting cases. We will assume in the following that the α and β nonmagnetic orbitals constituting the "closed shells" are identical, i.e. $\int_{ij,nmag}^{\alpha\beta,BS} = \delta_{ij}$. The first possibility is to assume no overlap of the magnetic orbitals, i.e. "strong localization". This leads to $\Theta^{BS} = 0$ and yields

$$J = \frac{E^{BS} - E^{HS}}{S^2_{max} - S^2_{min}} \text{ if, } \int_{ij,mag}^{\alpha\beta,BS} = 0 \tag{20}$$

Furthermore, if $S_{min} = 0$ as in the case of homovalent binuclear compounds, J is given by

$$J = \frac{E^{BS} - E^{HS}}{S^2_{max}} \tag{21}$$

which is the formula obtained by Noodleman[30] a long time ago by a very different approach. The opposite approximation, i.e. "strong delocalization", consists in assuming that $\int_{ij,mag}^{\alpha\beta,BS} = 1$ when $i = j$ and zero otherwise, i.e. $\Theta^{BS} = -n^{\beta}_{mag} = -2S_B$. This treatment leads to

$$J = \frac{E^{BS} - E^{HS}}{S^2_{max} - S^2_{min} + 2S_B}, \text{ if } \int_{ij,mag}^{\alpha\beta,BS} = 1 \tag{22}$$

Using again $n^{\alpha} = n^{\beta}$, i.e. $S_{min} = 0$ and $S_{max} = 2S_B$, the exchange coupling becomes

$$J = \frac{E^{BS} - E^{HS}}{S_{max}(S_{max}+1)} \tag{23}$$

which is another formula known from the literature[55] but derived here consistently with other approximations. In addition to just rederiving known formulas for computing J in a simple and unified manner, the route taken here allows one to understand on a common basis the various approximations that are inherent to quite different approaches used previously in calculations.

2.3. Accessing the Low-Spin Total Energy. Apart from providing a way to compute J from standard electronic structure calculations, the spin-ladder idea allows one to also express the *total energy* of every spin-eigenstate in terms of the HS and BS total energies. In particular, the energy of the (antiferromagnetic) LS ground state, i.e.

$$E^{LS} = -J[\langle\hat{S}^2\rangle^{LS} - \langle\hat{S}_A^2\rangle - \langle\hat{S}_B^2\rangle] = -J[S_{min}(S_{min}+1) - S_A(S_A+1) - S_B(S_B+1)] \tag{24}$$

can be rewritten formally using the energy of the single-determinant BS state from eqs 7 and 12, i.e.

$$E^{BS} = -J[S_{min}(S_{min}+1) - S_A(S_A+1) - S_B(S_B+1)] - J[2S_B + \Theta^{BS}] \tag{25}$$

as

$$E^{LS} = E^{BS} + J[S_{max} - S_{min} + \Theta^{BS}] \tag{26}$$

Finally, by substituting the general expression for J, eq 14, a general equation for the ground-state total energy can be obtained as a sum of suitably weighted total energies of the BS and HS states

$$E^{LS} = (1+c)E^{BS} - cE^{HS} \tag{27}$$

where c is given by

$$c = \frac{S_{max} - S_{min} + \Theta^{BS}}{S_{max}^2 - S_{min}^2 - \Theta^{BS} + \Theta^{HS}} \tag{28}$$

This constitutes a convenient expression of the total energy of the low-spin, antiferromagnetically coupled ground state of a spin dimer in terms of easily accessible total energies. This general energy expression can, of course, be used in conjunction with the various approximations derived in the previous section for Θ and thus for J, which yields simplified LS total energies.

2.4. Extension of the Broken Symmetry Approach: Two-Determinant Ab Initio Molecular Dynamics. In the last section it has been shown that the total energy and thus the potential energy surface of an antiferromagnetically coupled electronic ground state may be obtained from a weighted sum of the potential energy surfaces of the HS and BS states. Consequently, forces acting on the nuclei in the antiferromagnetic LS state can be represented as a suitably weighted sum of the corresponding sets of forces

$$\mathbf{F}_I^{LS} = -\nabla_{\mathbf{R}_I}\{(1+c)E^{BS}\} - \nabla_{\mathbf{R}_I}\{-cE^{HS}\} \tag{29}$$

where it must be stressed that Θ might be a complicated function of nuclear positions and a functional of orbitals or the spin density. A dramatic simplification is achieved by invoking the above-mentioned approximation of no overlap of the magnetic orbitals and no spin contamination according to eq 20 (or eq 21 for the $n^\alpha = n^\beta$ special case). This implies that

$$c \approx \frac{S_{max} - S_{min}}{S_{max}^2 - S_{min}^2} \tag{30}$$

so that $\mathbf{F}_I^{LS}$ is obtained as a linear combination from

$$\mathbf{F}_I^{LS} = (1+c)\mathbf{F}_I^{BS} - c\mathbf{F}_I^{HS} \tag{31}$$

since c is just a constant parameter in this limit. Effects due to spin contamination and overlap of magnetic orbitals will be discussed later in the manuscript.

At this stage, one can formulate AIMD very efficiently in the framework of a Car−Parrinello Lagrangian[56,57] generalized to two determinants

$$\begin{aligned} L_{CP}^{LS} = {} & \frac{1}{2}\sum_I^N M_I\dot{\mathbf{R}}_I^2 \\ & + \frac{1}{2}\sum_i \{(1+c)[\mu\langle\dot{\varphi}_i^{BS}|\dot{\varphi}_i^{BS}\rangle] \\ & - c[\mu\langle\dot{\varphi}_i^{HS}|\dot{\varphi}_i^{HS}\rangle]\} \\ & - \{(1+c)E^{BS}[\{\varphi_i^{BS}\},\{\mathbf{R}_I\}] \\ & - cE^{HS}[\{\varphi_i^{HS}\},\{\mathbf{R}_I\}]\} \\ & + \sum_{i,j}\{(1+c)[\Lambda_{ij}^{BS}(\langle\varphi_i^{BS}|\varphi_j^{BS}\rangle - \delta_{ij})] \\ & - c[\Lambda_{ij}^{HS}(\langle\varphi_i^{HS}|\varphi_j^{HS}\rangle - \delta_{ij})] \end{aligned} \tag{32}$$

by expressing all the contributions to the Lagrangian depending on the electronic structure as a linear combination of the BS and the HS states. Note that two independent sets of KS orbitals, $\{\varphi_i^{HS}\}$ and $\{\varphi_i^{BS}\}$, are used to built up the KS determinants underlying the HS and BS total energy expressions. Correspondingly, the equations of motion for N nuclei with mass $\{M_I\}$ are given by

$$M_I\ddot{\mathbf{R}}_I = -(1+c)(\nabla_{\mathbf{R}_I}E^{BS}) + c(\nabla_{\mathbf{R}_I}E^{HS}) \tag{33}$$

whereas the orbitals move according to

$$\begin{aligned} \mu\ddot{\varphi}_i^{BS} &= -(1+c)(\nabla_{\varphi_i^{BS}}E^{BS} + f_{constr}^{BS}) \\ \mu\ddot{\varphi}_i^{HS} &= c(\nabla_{\varphi_i^{HS}}E^{HS} + f_{constr}^{HS}) \end{aligned} \tag{34}$$

Here, μ is the ficticious mass of the orbitals, and f_{constr} are the forces due to the orthonormality constraints imposed on the orbitals. These equations show that within this approach the BS and the HS wave functions are simultaneously propagated under the constraint that the orthonormality condition, imposed by the Lagrange multipliers $\{\Lambda_{ij}\}$, is satisfied by the two sets of orbitals individually. Thus, the nuclear dynamics is performed on the ground-state LS energy hypersurface obtained by a projection according to eq 27. Note that if thermostats for the electronic wave functions are invoked,[57] the thermostat energy must be calculated according to the projection scheme as well.

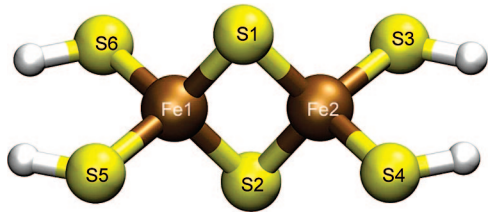


Figure 2. Minimum model system for binuclear iron−sulfur complexes with atomic labeling. Color scheme: iron atoms are brown, sulfur atoms are yellow, and hydrogen atoms are gray.

We will be referring to the above introduced approach as extended broken-symmetry (EBS) scheme. Most importantly, this EBS approach allows one to perform molecular dynamics in the antiferromagnetically coupled ground state of a binuclear complex. Of course, this method can be used immediately to optimize the structure of such complexes or to perform vibrational analysis in the harmonic approximation. Finally, when it comes to computing J it is worth noting that only such an approach is intrinsically consistent where the same approximate spin-projection scheme is used to optimize first the structure for which J is then obtained at no additional cost. It is noted in passing that the afore derived EBS scheme can be directly implemented in the framework of wave function-based or density functional-based methods.

2.5. Computational Setup. In the minimum model used here for [2Fe-2S] complexes, i.e. $[Fe_2S_2(SH)_4]^{2-}$ in isolation, the Fe_2S_2 core is capped by four SH^- groups according to Figure 2, which mimic cysteinyl ligands as found in typical protein environments.[10] This model complex has been studied first using Xα valence bond theory[58] and later using semiempirical as well as configuration interaction (CI) techniques.[59] The EBS method has been implemented into the CPMD simulation package[57,60] which was used for structure optimization as well as for AIMD simulations all performed within the spin-unrestricted Kohn−Sham density functional theory in its plane wave/pseudopotential formulation.

The PBE[61,62] exchange-correlation functional was chosen, and the core electrons were taken into account using Vanderbilt's ultrasoft pseudopotentials[63] which contain additional d-projectors in the case of sulfur as well as scalar relativistic corrections and semicore states for iron; note that other functionals of the generalized gradient approximation type could be used here instead. For the ultrasoft pseudopotentials a rather low plane wave cutoff of 30 Ry was sufficient to obtain convergence in the forces of the nuclei. The simulation cell was an unusually large cubic box with a constant edge length of 16.9 Å which turned out to be necessary in order to converge the representation of the density for such a 2-fold negatively charged complex. The charge of the system was decoupled from its periodic images using the Martyna-Tuckerman Poisson solver to impose cluster boundary conditions.[64]

The structure optimizations were performed for the single-determinant HS and BS states as well as for the projected LS state within the two-determinant EBS scheme. If not otherwise stated the approximations to eq 18 leading to eq 21 to compute J[30] and eq 27 together with eq 30 to calculate E^{LS} were used throughout this paper. The criteria were 10^{-4} and 10^{-7} a.u. for the forces on ions and the gradient of the wave function, respectively. The AIMD simulations were performed using Car−Parrinello propagation[56] along with separate Nosé-Hoover chain[65] thermostats for nuclei and electronic orbitals. The system was thermostatted at 300 K, and a time step of 0.145 fs was used for the integration of the equations of motion. The fictitious mass for the orbitals was 700 a.u., and the hydrogen (H) mass was substituted by the deuterium (D) mass in all dynamical calculations. The minimum energy structures from the structure optimization were taken as initial configurations and equilibrated for approximately 5 ps at 300 K thermostatting every degree of freedom individually ("massive thermostating") to ensure energy equipartition between all vibrational modes. Subsequently, 17 ps long trajectories were collected using one thermostat for the complete system.

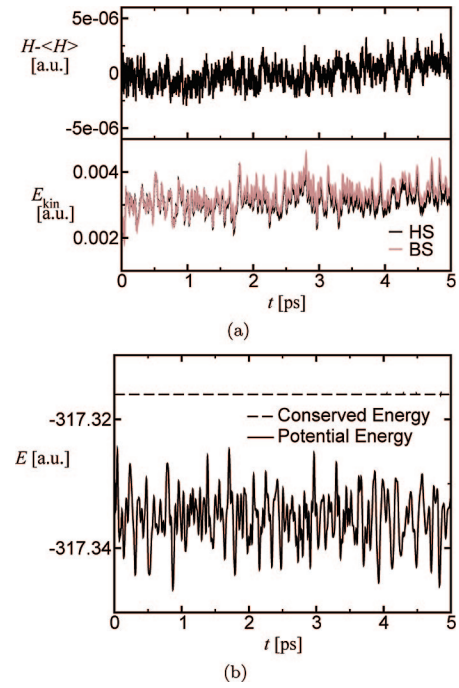


Figure 3. Energy conservation during a microcanonical AIMD simulation in the two-determinant EBS formalism. (a) Upper panel: instantaneous deviation of the conserved energy H from its average value, $H-\langle H\rangle$; lower panel: fictitious kinetic energy of the orbitals, E_{kin}, for the BS (brown) and HS (black) determinants. (b) Conserved energy (dashed line) on the scale of the EBS potential energy (solid line). All energies are reported in a.u.

3. Results and Discussion

3.1. Stability of EBS–CP Molecular Dynamics. To assess the stability of the two-determinant Car−Parrinello propagation, different energies were calculated from microcanonical EBS simulations of the minimum model $[Fe_2S_2(SH)_4]^{2-}$ for iron−sulfur complexes as shown in Figure 3. At variance with the simulation parameters described in section 2.5 and used later for the analysis of the magneto-structural dynamics of the complex, no thermostats were employed, hydrogen atoms were used instead of deuterium, and the integration time step was therefore 4 au (≈0.1 fs) in

Table 1. Structural Data for the Iron−Sulfur Complex in Different Spin States[a]

distance [Å]	HS	BS	EBS
Fe1−Fe2	2.960	2.676	2.622
S1−S2	3.488	3.488	3.467
Fe1−S1	2.284	2.197	2.172
Fe1−S2	2.297	2.194	2.170
Fe2−S1	2.290	2.202	2.177
Fe2−S2	2.279	2.198	2.176

angle [°]	HS	BS	EBS
S2−Fe1−S1	99.1	105.2	107.0
S2−Fe2−S1	99.5	104.9	105.6
Fe1−S1−Fe2	80.6	74.9	74.2
Fe1−S2−Fe2	80.6	75.1	74.2
rmsd (Fe_2S_2) [Å]	0.123	0.020	0.0
rmsd ($Fe_2S_2S_4$) [Å]	0.174	0.020	0.0
J [cm^{-1}]	−227	−390	−435

[a] Selected structural properties and the coupling constant J of $[Fe_2S_2(SH)_4]^{2-}$ in vacuo in the high-spin (HS), broken-symmetry (BS), and extended broken-symmetry (EBS) states resulting from structure optimization in the corresponding state. J is computed using eq 21 in all cases.

this case. All other parameters were identical to those described in section 2.5.

In Figure 3(a) the deviation from the average of the conserved energy,[57] H-$\langle H\rangle$, is plotted along with the ficticious kinetic energy of the electrons, E_{kin}, for both BS and HS determinants. On the simulated time scale of 5 ps the fluctuations are within 4.0×10^{-6} a.u., and there is no appreciable drift in conserved energy. This is affirmed by Figure 3(b), which shows the conserved and the potential energies on the same scale. Also the ficticious kinetic energy of the orbitals had no drastic shifts or oscillations and behaved well during the entire simulation (see the lower panel in Figure 3(a)).

3.2. Structure Optimization. The first point to be addressed is the impact of the two-determinant total energy scheme on the structure of the $[Fe_2S_2(SH)_4]^{2-}$ complex in vacuo, i.e. a comparison between structures optimized using the EBS scheme with eqs 30 and 31, compared with the structures optimized using the conventional single-determinant unrestricted BS and HS states. At the same time we determine the effect of these structural changes on the coupling constant J which is evaluated within the approximation eq 21 in all the three cases. Usually in literature, J is computed by optimizing the structure within the unrestricted BS approach, i.e. using E^{BS} only.

Concerning essential structural features such as iron−iron distance or iron−sulfur−iron angle the BS and the EBS optimized structures are found to be quite similar, whereas the complex is more open in the HS state, see Table 1. This is confirmed by the root-mean-square-deviations (rmsd) of the atomic positions relative to the EBS structure as compiled in Table 1.

The rmsd values of the bare core are 0.020 Å and 0.123 Å for the broken-symmetry and the high-spin state, respectively. If the ligand sulfur atoms S3-S6 are taken into account additionally, the rmsd value for the high-spin case increases to 0.174 Å, but the trend remains the same. Closer inspection of the data in Table 1 indicates that the main differences in the structure are due to the changes in bond lengths and angles in the Fe_2S_2 core, whereas the other bond lengths and angles are very similar. In a previous study[66] single point LS energies have been computed employing Noodleman's approach, i.e. eqs 27 and 30, at various Fe−Fe distances for a similar system, $[(Fe_2S_2)(SCH_3)_4]^{2-}$. The conclusion that the Fe−Fe distance becomes shorter upon going from the BS state to the spin-projected LS structure is in agreement with our observation.

3.3. Exchange Coupling Constant. More pronounced is the effect of the structural changes due to spin-projection on the value of the exchange coupling constant J when obtained consistently at the respective optimized structures. Clearly, the EBS structure is most compact and thus yields the largest, i.e. most negative, antiferromagnetic coupling constant, whereas $|J|$ is smallest for the least compact HS structure. Overall, the J value can change by nearly a factor of 2 depending on the structure used to compute it.

To our knowledge no converged quantum chemical J values are available even for this simple system. But for related systems[67] the absolute value of the coupling constant is smaller by a factor of 2−3 in experiment as compared to the present calculations. However, such an overestimation of $|J|$ is a well-known artifact of using the local density approximation (LDA), or functionals of the generalized gradient approximation (GGA) type,[68–70] and may be attributed to the self-interaction error (SIE), whereas using only Fock exchange underestimates J considerably, see the lucid discussion in ref 25. In standard KS density functional calculations using LDA or GGA functionals the SIE tends to delocalize the magnetic orbitals, leading to a stronger bonding and thus to shorter bond distances. In particular, such an error can result in an overestimation of the overlap of magnetic orbitals and thus of $|J|$.

The problem can be reduced using self-interaction correction (SIC) schemes,[71] for instance in simplified versions,[72–75] or DFT+U methods.[76–80] There are also recent efforts to improve the calculation of J values by exploiting the virtues of constrained DFT techniques[81] in the framework of magnetic complexes.[82,83] As an alternative approach it was demonstrated that accurate J values can also be obtained by the tuning the Fock exchange contribution in hybrid functionals such as B3LYP[84] at the expense of adjusting density functionals for computing a particular property. It is mentioned in passing that these and similar schemes can be combined with our EBS approach thus improving the approximate DFT description of the BS state as such.

A recent study advocates that formula eq 23 should yield very good results for the exchange coupling constant compared to experiment *if no SIC is applied*.[70] This idea can be analyzed within the present framework; see also ref 25 for a discussion of this proposal. In deriving eq 23 from eq 18 within the general framework laid out in section 2.2 the approximation of assuming full overlap of magnetic orbitals, i.e. strong delocalization, was imposed to evaluate the spin contamination in the BS state only, whereas strong localization and thus no overlap was assumed in the HS state as usual. Obviously this increases the denominator by the

Table 2. Structural Data for in Vacuo Optimized $[Fe_2S_2(SH)_4]^{2-}$ Complex Using Different Values for the Overlap of Magnetic Orbitals[a]

Θ^{BS}	0.000	−0.313	−1.250	−2.813	−5.000
$\int^{\alpha\beta,BS}_{ii,mag}$	0.00	0.25	0.50	0.75	1.00
c	0.200	0.185	0.143	0.079	0.000
distance [Å]					
Fe1−Fe2	2.622	2.626	2.637	2.655	2.676
S1−S2	3.467	3.470	3.475	3.481	3.488
Fe1−S1	2.172	2.174	2.179	2.188	2.197
Fe1−S2	2.170	2.172	2.177	2.185	2.194
Fe2−S1	2.177	2.178	2.183	2.193	2.202
Fe2−S2	2.176	2.180	2.185	2.190	2.198
angle [°]					
S2−Fe1−S1	107.0	106.0	105.8	105.5	105.2
S2−Fe2−S1	105.6	105.6	105.4	105.2	104.9
Fe1−S1−Fe2	74.2	74.2	74.4	74.6	74.9
Fe1−S2−Fe2	74.2	74.3	74.4	74.7	75.1
J [cm^{-1}]	−435	−426	−402	−366	−325

[a] Selected structural properties and coupling constant J of $[Fe_2S_2(SH)_4]^{2-}$ in vacuo using different values for the spin correction term Θ^{BS} while keeping $\Theta^{HS} = 0$. The corresponding values of overlap of magnetic orbitals $\int^{\alpha\beta,BS}_{ii,mag}$ when $\int^{\alpha\beta,BS}_{ij,mag} = 0.0$ for $i \neq j$. Note that $\Theta^{BS} = -5.0$ results in a structure identical to the BS structure (see Table 1); however, the J values are different as they were computed using eq 18 with different values of Θ^{BS}; see text for details.

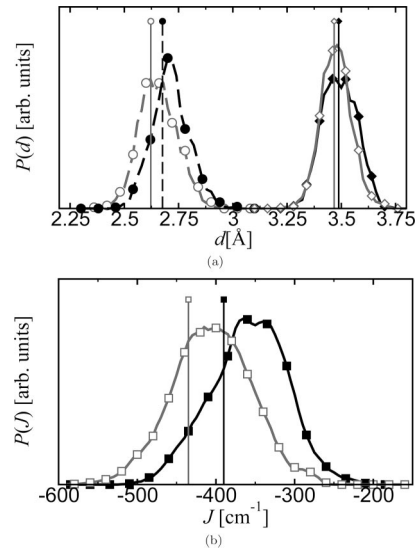


Figure 4. Probability distribution functions for selected structural quantities and J of the $[Fe_2S_2(SH)_4]^{2-}$ complex for BS and EBS-AIMD. Normalized probability distribution functions of (a) the Fe1−Fe2 (broken line, circles) and S1−S2 (solid line, diamonds) distance and (b) of the coupling constant J (solid line, squares) obtained from AIMD of the $[Fe_2S_2(SH)_4]^{2-}$ complex at 300 K in vacuo using the BS (black, filled symbols) and EBS (gray, open symbols) schemes. The corresponding values obtained for the optimized minimum energy structures are shown as vertical bars using the same labeling.

term $+S_{max}$ compared to eq 21 and thus decreases the value of $|J|$ accordingly. The assumption of full overlap results in $\Theta^{BS} = -n^{\beta}_{mag} = -S_{max}$ and $S_{min} = 0$ having $n^{\alpha} = n^{\beta}$. From eq 28 follows, $E^{LS} = E^{BS}$ as $c = 0$ in eq 27. Thus, the true multideterminantal LS ground state is effectively approximated by the single-determinant BS state. Besides, this approximation is intrinsically inconsistent because full overlap of the magnetic orbitals implies a closed shell electronic configuration, but an unrestricted BS determinant is used instead.

At this point it is worth estimating the effect of the spin correction term Θ^{BS} which includes the degree of overlap of magnetic orbitals on both structure and coupling constant J. These properties are easily accessible within the EBS approach on equal footing. In order to work out these trends qualitatively, we change the value of Θ from its maximum value of 0 to its minimum value of −5.0. The structure of the complex was independently optimized for each value of Θ^{BS} using the forces eq 29 with c according to eq 28 and J is computed using eq 14 correspondingly. For the sake of simplicity this can be viewed as a systematic change in the overlap of the magnetic orbitals $\int^{\alpha\beta,BS}_{ii,mag}$, assuming $\int^{\alpha\beta,BS}_{ij,mag} = 0$ for $i \neq j$; note that as usual Θ^{HS} was always kept zero.

The data compiled in Table 2 show that distances are more affected than angles. It is important to note here that the structure optimized using $\Theta^{BS} = -5.0$ (see Table 2) within the EBS scheme results in a structure which is identical to that optimized just using the BS state (see Table 1). But the J values of 325 and 390 cm^{-1} for this structure were obtained using eq 18 with $\Theta^{BS} = -5.0$ (in Table 2) and using eq 21 (in Table 1), respectively. It is apparent from this comparison that the larger the overlap term $\int^{\alpha\beta}$ the more open is the resulting structure and thereby the smaller the magnitude of J. Thus, the structure as well as the value of the coupling constant could be easily "tuned" by changing the overlap term, as effectively done in ref 70. However, the assumption of full overlap of magnetic orbitals seems less justified in the case of weakly coupled dimers than the other limiting case, which is the assumption of zero overlap.[25] Furthermore, imposing extreme delocalization is inconsistent with the basic assumption underlying the validity of the Heisenberg model in eq 1, which is the basis of the present and related spin-projection schemes.

3.4. Dynamical Effects on Structure and *J*. In order to relate the dynamics of the antiferromagnetic coupling constant J to the structural dynamics of the iron−sulfur complex within the BS and the EBS schemes, Car−Parrinello AIMD simulations were performed using both approaches. The same trend as obtained by structure optimizations is seen in both the average structural quantities and the J values sampled from molecular dynamics trajectories (see Figure 4(a) and Figure 4(b), respectively).

The average values of J are −356 and −403 cm^{-1} for the BS and EBS case, respectively, which is a 10−15% effect on the coupling constant when considering the two-determinant EBS structure consistent with the spin-projection formula used to compute J, instead of simply using the single-determinant BS structure. In addition, the rather broad probability distributions of J clearly demonstrate J is fluctuating with considerable amplitudes at 300 K covering the range from −600 to −300 cm^{-1} in case of the EBS approach. These results underline the need of considering dynamical behavior of such a spin-coupled complex when

studying its magnetic properties, since J can change by a factor of 2 just because of thermal fluctuations at room temperature!

The corresponding averages of the Fe−Fe and the S−S distances within the core are different for the conventional and the extended broken-symmetry schemes, indicating different minima on the free energy surface (FES) even at room temperature. Additionally, since the widths and thus the fluctuations are also different, the shapes of the underlying FES in the sampled regions differ as well. While the average distances of the two bridging sulfur atoms are almost the same as in the minimum energy structures (cf. the vertical bars), the Fe−Fe distances are shifted to larger values in both schemes compared to the respective optimized structures. Together with the fact that the most probable distances (determined by the minimum of the PES) and the average distances (determined by the minimum of the FES) are not equal, this implies that the observed effect is due to an asymmetric anharmonicity in the effective Fe−Fe interactions.

3.5. Dynamical Structural Properties. The different FES produced by the two methods should be mirrored in the *dynamical behavior* of this iron−sulfur complex as well. To investigate the dynamical behavior of a property **P**, it is appropriate to define a correlation function

$$G_{\mathbf{PP}}(t) = \frac{\langle \mathbf{P}(t_0)\cdot\mathbf{P}(t_0+t)\rangle_{t_0}}{\langle \mathbf{P}(t_0)\cdot\mathbf{P}(t_0)\rangle_{t_0}} \tag{35}$$

where $\langle ... \rangle_{t_0}$ implies the usual average over the reference time points t_0 along the generated trajectory. Its Fourier transform

$$G_{\mathbf{PP}}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G_{\mathbf{PP}}(t)\exp(-i\omega t)dt \tag{36}$$

represents the spectral density, or the power spectrum, of the corresponding property and can be used to define an "absorption coefficient"

$$A_{\mathbf{PP}}(\omega) = \omega^2 G_{\mathbf{PP}}(\omega) \tag{37}$$

which already includes the so-called harmonic quantum correction factor.[85]

To explore the structural dynamics of the complex, it is convenient to use the (nuclear) velocity autocorrelation function, $G_{\dot{\mathbf{R}}\dot{\mathbf{R}}}(t)$ where $\dot{\mathbf{R}}$ is a vector that contains the velocities of all or selected nuclei in the complex. The resulting spectra obtained from the EBS and the BS trajectories at 300 K in vacuo are shown in Figure 5.

At first glance they appear similar, and in both cases two distinct spectral regions can be identified. The high frequency region around 1750 cm^{-1} is assigned to S-D vibrations, whereas the region below about 500 cm^{-1} is comprised of iron−sulfur motion; recall that H was substituted by D for technical reasons. But on a closer look the spectra obtained by autocorrelating only the four atoms in the Fe_2S_2 core in Figure 5(b) are seen to display differences. In particular the EBS spectrum is broader at the high-frequency wings and more structured in the region between about 200 and 400 cm^{-1}.

The individual features of these spectra can only be understood after performing a spectral decomposition. To achieve this it is necessary to choose a basis in which the

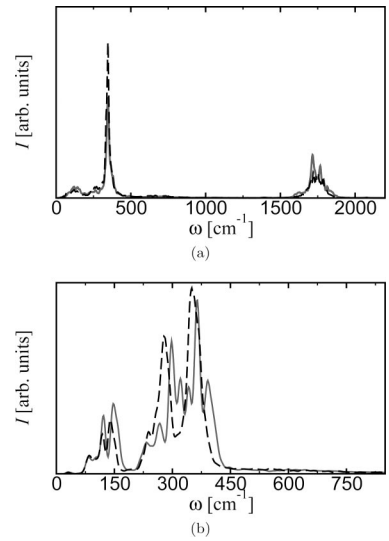


Figure 5. Spectra $A_{\dot{\mathbf{R}}\dot{\mathbf{R}}}(\omega)$ of the velocity autocorrelation function obtained according to eq 37 from AIMD simulations of the $[Fe_2S_2(SH)_4]^{2-}$ complex in vacuo for (a) the full system and (b) only the four atoms in the Fe_2S_2 core using the BS (dashed lines) and EBS (solid lines) methods.

dynamics can be expressed. Since the atoms of the Fe_2S_2 core are expected to affect the magnetic properties much more than the four peripheral SH^- groups, we will restrict the investigation to the dynamics of the Fe_2S_2 core in terms of normal modes of the Fe_2S_2 fragment. However, it is clear that the ligands also affect the internal motion of the core due to coupling effects.

Taking into account the approximate D_{2h} symmetry of the average structure of the Fe_2S_2 core from the AIMD simulation, it is appropriate to use corresponding symmetry-adapted normal modes as a basis. There exist six modes, $\{\mathbf{q}_\xi\}$ (see Figure 6), apart from rigid rotations and translations. Generally, a coupling of the modes is expected during a molecular dynamics trajectory. But by taking an appropriate linear combination of these modes, it is possible to generate a set of new modes, $\{\tilde{\mathbf{q}}_\xi\}$, with minimal coupling. The coefficients for the linear combination were obtained by minimizing $\int|A_{\xi,\xi'}(\omega)|d\omega$. The spectrum $A_{\xi,\xi'}(\omega)$ is calculated according to eqs 36 and 37 from the corresponding cross-correlation function

$$G_{\xi,\xi'}(t) = \frac{\langle q_\xi(t_0)q_{\xi'}(t_0+t)\rangle_{t_0}}{\langle q_\xi(t_0)q_{\xi'}(t_0)\rangle_{t_0}}; \xi \neq \xi', \xi = 1, \ldots, 6 \tag{38}$$

Here, $\{q_\xi(t)\}$ denote the projections of the mass-weighted nuclear positions of the Fe_2S_2 core, $\mathbf{R}_c(t)$, along the molecular dynamics trajectory on each of the six basis functions $\{\mathbf{q}_\xi\}$

$$q_\xi(t) = \mathbf{R}_c(t)\cdot\mathbf{q}_\xi \tag{39}$$

normalized to zero mean and unit variance where $\mathbf{R}_c(t)$ are in the frame of reference given by the basis.

The modes $\{\tilde{\mathbf{q}}_\xi\}$ obtained by this procedure were used to calculate the spectra $A_{\xi\xi}(\omega)$ according to eqs 36 and 37 from the corresponding autocorrelation functions $G_{\xi\xi}(t)$. In Figure 7 the individual spectral components $A_{\xi\xi}(\omega)$ are shown

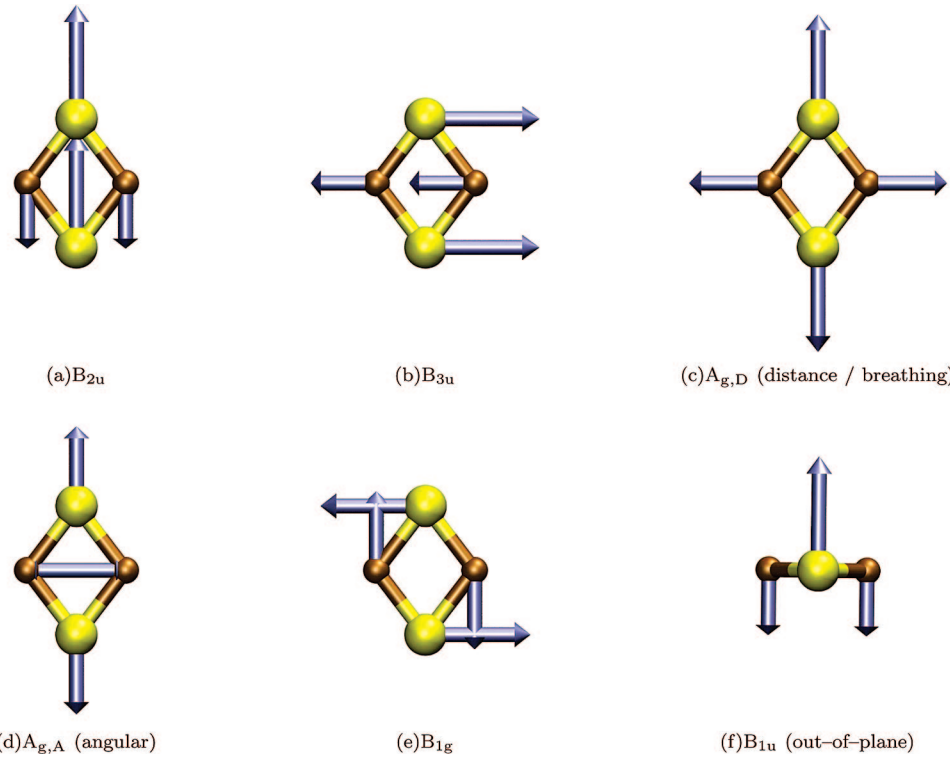


Figure 6. Ideal D_{2h} symmetry-adapted normal modes, $\{\mathbf{q}_\xi\}$, for the Fe_2S_2 core of the $[Fe_2S_2(SH)_4]^{2-}$ complex. The small brown and large yellow spheres represent iron and sulfur atoms, respectively. The blue arrows show the phases and relative magnitudes of the atomic displacements along each mode.

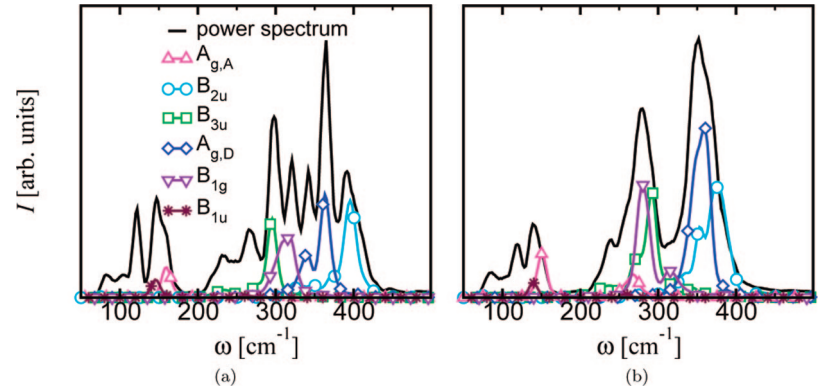


Figure 7. Spectral decomposition of the spectra $A_{\dot{R}\dot{R}}(\omega)$ of the Fe_2S_2 core obtained from AIMD of the $[Fe_2S_2(SH)_4]^{2-}$complex in vacuo using (a) the EBS and (b) the BS approaches. The spectral components, $A_{\xi\xi}(\omega)$, were smoothened by a convolution with a Gaussian function of 5 cm^{-1} width.

together with the full spectrum obtained for the Fe_2S_2 fragment. In the following, the modes $\tilde{\mathbf{q}}_\xi$ are referred to according to their largest $\mathbf{q}_\xi$ component which always exceeds 97%. Due to the loss of structural symmetry during the dynamics, *g* and *u* modes can couple to each other. Note, since a normalization to zero mean and unit variance was performed for the individual $\tilde{q}_\xi$, the relative intensities are preserved only within the individual spectra $A_{\xi\xi}(\omega)$.

First, the EBS vibrational spectrum is analyzed along these lines. Its spectral decomposition in Figure 7(a) shows that the out-of-plane vibration B_{1u} leads to a resonance at 145 cm^{-1}. The B_{3u} and the B_{1g} in-plane vibrations appear at 295 cm^{-1} and 318 cm^{-1}, respectively. The former shows an additional peak at 235 cm^{-1}. The B_{2u} mode spans a rather broad spectral range (300−430 cm^{-1}) having a sharp peak at 396 cm^{-1} in addition to a weak intensity region at 300−375 cm^{-1}. The latter could be attributed to a slight coupling to the $A_{g,D}$ vibration. The $A_{g,A}$ angular mode gives rise to a peak at 160 cm^{-1}, where a slight mixing with B_{1u} occurs. Also in the spectrum of the $A_{g,D}$ vibration two signals are observable, at 340 cm^{-1} and 364 cm^{-1}. Despite the decorrelation procedure, residual mixing of this motion with several other modes (B_{2u}, B_{3u}, B_{1g}, and $A_{g,A}$) could not be avoided, which is, however, weak.

In the BS scheme the spectrum is less complex (see Figure 7(b)) as that of EBS. All modes exhibit a red shift of their frequencies with respect to the EBS spectrum. The strongest shifts are observed for the B_{1g} and the B_{2u} vibrations, which are 35 and 20 cm^{-1}, respectively. Other modes are only slightly shifted (up to 5 cm^{-1}). Another striking difference

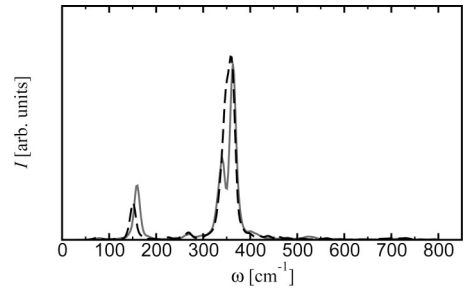


Figure 8. Spectra $A_{JJ}(\omega)$, calculated from the autocorrelation function of $J(t)$ from AIMD of the $[Fe_2S_2(SH)_4]^{2-}$ complex in vacuo using the BS (black dashed line) and EBS approaches (gray solid line).

is that the $A_{g,D}$ breathing mode exhibits a double peak in the EBS case (340 and 364 cm^{-1}) but has only a single peak (360 cm^{-1}) with a shoulder (350 cm^{-1}) in the BS scheme. It turns out that the mixing of the two A_g modes is slightly weaker compared to the EBS situation. Interestingly, the B_{2u} mode couples very strongly to the $A_{g,D}$ vibration in the BS case. This coupling could not be removed by the decorrelation procedure described above.

We note in passing that the remaining mixing of the individual vibrational modes is attributed to a deviation from the harmonic picture or to the interactions of the corresponding modes via the four SH^- ligands as noted in the beginning of this section. Indeed, principal component analysis[86] (PCA) of the trajectories shows, for example, modes in which the core motion shows $A_{g,A}$ character combined with symmetric and asymmetric ligand vibrations (data are not shown here but see ref 87 for full analysis). Ligand contributions are also responsible for the features up to 125 cm^{-1} and the peaks at 267 and 240 cm^{-1} in Figures 7(a) and 7(b), respectively, which are not fully explained by $\{\tilde{\mathbf{q}}_\xi\}$. Also in this case PCA could reveal modes which show signals in these particular regions.[87]

3.6. Dynamical Magnetostructural Properties. In a simple orbital picture the value of the exchange coupling constant J is related to the interactions among singly occupied orbitals, which are mainly located in the Fe_2S_2 core. Thus, changes of the structure due to the dynamical evolution of the four core atoms are expected to most directly affect J and, thereby, imprint a time-dependence, $J(t)$. This time-dependent exchange coupling $J(t)$ is most conveniently analyzed in Fourier space, i.e. in terms of its spectrum $A_{JJ}(\omega)$ calculated according to eqs 36 and 37 from the autocorrelation function $G_{JJ}(t)$ defined in eq 35.

The power spectra resulting from the two-determinant EBS and one-determinant BS dynamics depicted in Figure 8 feature two main resonances at about 150 and 350 cm^{-1} and an additional, small feature around 270 cm^{-1}. Moreover, the lower frequency peak at ≈150 cm^{-1} is slightly red-shifted in the BS case w.r.t. the EBS reference where the 350 cm^{-1} resonance is clearly split in addition.

To reveal the relationships between the structural dynamics of the complex unraveled in section 3.5 and the "magnetodynamics" embodied in $A_{JJ}(\omega)$, it is appropriate to analyze the frequency dependent correlations of $J(t)$ with the individual contributions of the different vibrational motion, $\tilde{q}_\xi(t)$

$$G_{J\xi}(t) = \frac{\langle \tilde{J}(t_0)\tilde{q}_\xi(t_0+t)\rangle_{t_0}}{\langle \tilde{J}(t_0)\tilde{q}_\xi(t_0)\rangle_{t_0}} \tag{40}$$

where $\tilde{J}$ indicates a zero mean and unity variance normalization of $J(t)$. The spectra calculated from these cross-correlation functions using eqs 36 and 37, $A_{J\xi}(\omega)$, clearly display frequency dependent correlations of magnetic and structural dynamics.

The spectrum of the exchange coupling constant, $A_{JJ}(\omega)$, and the spectral components of the individual vibrational modes, $A_{\xi\xi}$, are shown for the EBS and BS results in Figures 9(a) and 9(c), respectively. In both cases several vibrational modes can be identified in those frequency regions where $J(\omega)$ has large intensity. However, the more sophisticated cross-correlation spectra (see Figures 9(b) and 9(d)) clearly show that all features of $A_{JJ}(\omega)$ for both the EBS and the BS results can be explained using mainly the two A_g modes.

In particular, the peak at 160 cm^{-1} and the small feature at 267 cm^{-1} in $A_{JJ}(\omega)$ of the EBS simulation are both due to $A_{g,A}$ vibrations, while the $A_{g,D}$ symmetric motion is related to the peaks at 340 and 364 cm^{-1}. Other modes show relatively small intensities in the cross-correlation spectra thus indicating only minor contributions to the magneto-structural dynamics of J.

The influence of the individual modes on the coupling constant can be understood qualitatively by recalling the nature of the superexchange interaction,[23] in particular how the vibrations of the core atoms influence the exchange pathways from one iron atom to the other via the sulfur bridges. It is clear that increasing Fe−S distances as well as decreasing Fe−S−Fe angles will decrease the superexchange. Along the two A_g modes, see Figure 6(c),(d), both superexchange paths are always affected in a symmetric manner. For the $A_{g,D}$ mode both paths simultaneously become shorter or longer thus enhancing or weakening the superexchange interaction; note that the Fe−S−Fe angles remain constant along displacements according to this symmetry. On the other hand, along the $A_{g,A}$ motion the weakening and enhancing of the superexchange is affected by a symmetric increase and decrease in the Fe−S−Fe angles, respectively.

In the BS case the overall picture is similar; the two A_g modes contribute most to the dynamics of the exchange coupling. However, Figure 9(d) shows an additional cross-correlation of $J(t)$ with the B_{2u} vibrational component. This mode couples very strongly to the $A_{g,D}$ vibration. This coupling could not be removed by the decorrelation procedure described in the previous section (data not shown) indicating a deviation from the imposed harmonic picture or a coupling via ligand modes. Moreover, PCA results show that, unlike in the EBS case, the B_{2u} mode of the BS simulations deviates more strongly from the perfect symmetry (see ref 87 for full analysis). This deviation might be responsible for the mutual coupling of $A_{g,D}$ and B_{2u}, directly

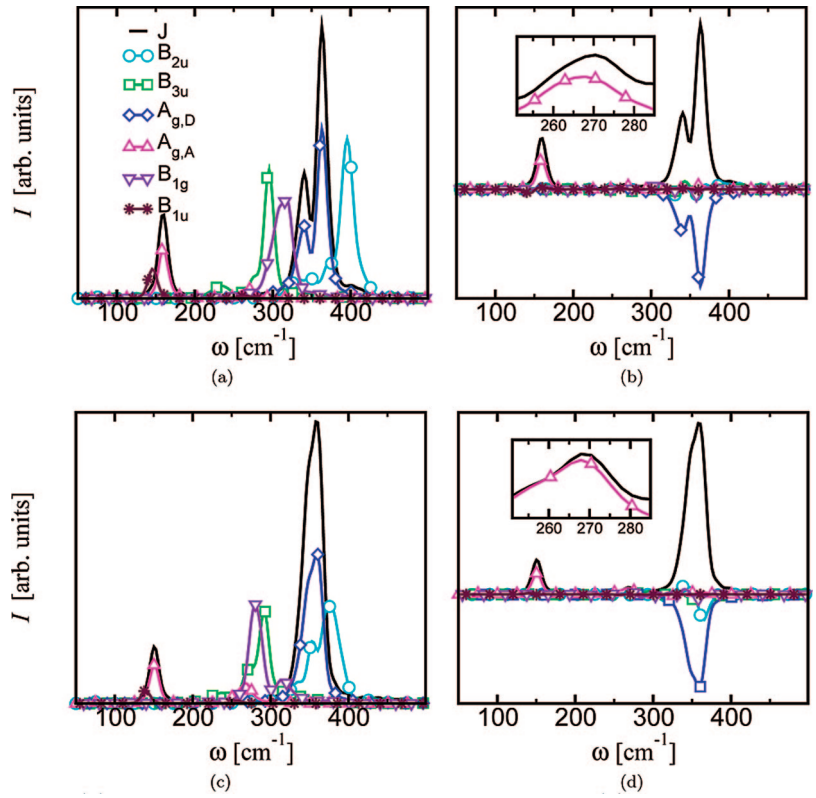


Figure 9. Spectra $A_{JJ}(\omega)$, together with the vibrational spectral components, $A_{\xi\xi}(\omega)$, of the Fe_2S_2 core dynamics obtained from AIMD of the $[Fe_2S_2(SH)_4]^{2-}$ complex in vacuo using (a) the EBS and (c) the BS approaches. Panels (b) and (d) show the corresponding frequency dependent cross correlations, $A_{J\xi}(\omega)$, between $J(t)$ and the individual vibrational components, $\tilde{q}_\xi(t)$, together with $A_{JJ}(\omega)$ for the EBS and BS approaches, respectively.

or via ligand vibrations, and causes an incomplete cancelation of the positive and negative contributions to the exchange interaction.

Overall, the BS and EBS schemes provide a qualitatively similar picture of the influence of the skeleton vibrations on the dynamics of the magnetic coupling constant J, i.e. on the dynamical magnetostructural properties of this iron−sulfur complex. In both approaches $A_{JJ}(\omega)$ shows a high and a low frequency domain, which can be ascribed to the $A_{g,D}$ and $A_{g,A}$ modes, respectively. Reinforcing the previous finding observed from the optimized structures, the exchange coupling J differs depending on the used scheme alike the average structures. Within the EBS scheme a more compact average structure is obtained, and thus the average $|J|$ value is about 10−15% larger (403 vs 356 cm^{-1}) in the two-determinant EBS approach compared to the one-determinant BS approximation. Moreover, apart from a slight blue shift of the EBS vibrations and thus the resonances in $A_{JJ}(\omega)$ with respect to their BS counterparts, the mixing of modes and their interplay with the ligands differ. In particular, coupling of the B_{2u} and $A_{g,D}$ modes is observed in the BS case. One manifestation of this different coupling pattern is the pronounced double peak feature at 340 and 364 cm^{-1} in the EBS case which appears as a single peak in the BS spectrum (at 360 cm^{-1} with a very weak shoulder at 350 cm^{-1}).

Conventionally, the magnetic exchange coupling in bridged binuclear transition metal complexes is discussed by means of the Goodenough-Kanamori rules[38–40] which relate the sign and the size of the coupling constant J to the symmetry of the magnetic orbitals, the coordination of the metal cations, and the structure of the diamagnetic bridge. In the present treatment we are able to go far beyond this static picture and correlate the structural changes of the Fe_2S_2 core with the fluctuations of the antiferromagnetic coupling at finite temperature. This opens the doorway to a direct comparison between calculated and measured magnetic properties of metalloproteins. It should also be mentioned that the Good-enough-Kanamori rules[38–40] give no prediction for the exchange coupling of two d^5 metal cations in tetrahedral coordination.

4. Summary, Conclusions, and Outlook

In order to understand the structural, but in particular the dynamical, properties of antiferromagnetically coupled bi-nuclear transition metal complexes, a novel *dynamical* method is introduced. This density functional based scheme is build upon an approximate spin-projection technique to represent the multireference low-spin ground state which extends the traditional description using broken-symmetry approaches or similar methods. Within this "extended broken-symmetry" (EBS) approach the total energy and the forces in the low-spin state are obtained as a particular weighted sum of the corresponding total energies and forces of the high-spin and broken-symmetry reference states which is achieved by an approximate spin-projection technique. Since this approach provides easy access to gradients, it allows for structure optimization and for the consistent

calculation of the exchange coupling constant J, at no additional cost, using the same spin-projection approximation for both calculations.

Most importantly, the proposed framework opens the doorway for an efficient implementation of ab initio molecular dynamics by propagating simultaneously two unrestricted Kohn−Sham determinants using Car−Parrinello techniques. The approach rests on theoretically solid ground as long as the system can be described by means of the Heisenberg spin Hamiltonian. The method is general, and the present implementation can be improved both systematically and consistently by using more sophisticated spin-projection approximations to eliminate the spin contaminations of the two Kohn−Sham reference determinants and by employing a better density functional based representation of these states.

To assess the value of the new scheme, it was compared to single-determinant treatments using a relevant but small binuclear iron−sulfur test system, the $[Fe_2S_2(SH)_4]^{2-}$ complex at 300 K in vacuo. In particular, the more compact Fe_2S_2 core leads to a $|J|$ value that is larger by about 10−15% in the EBS case. An interesting finding is the sensitivity of the minimum energy structure and thus of J on the approximation to the orbital overlap, i.e. the value of the overlap integrals, which establishes a unified connection of the outlined method to various schemes that can be found in the literature. From a very pragmatic point of view this dependence would offer the possibility to "fine tune" the structure of the system and thus $|J|$. However, it is clearly recommended to improve the underlying approximations (to the spin-projection and/or the representation of the reference states, see above) consistently within the general framework instead.

The present method allows one to carry out efficiently ab initio molecular dynamics simulations of antiferromagnetically coupled binuclear complexes within a two-determinant Car−Parrinello scheme. Analyzing trajectories obtained at room temperature it turns out that thermal fluctuations are able to change J by as much as about 50% with respect to the corresponding average value at 300 K. Since J is generated at no extra cost "on the fly" along the molecular dynamics trajectory, one can access readily its spectrum (or "density of states"), $A_{JJ}(\omega)$, by virtue of Fourier transforming its time-dependent evolution, $J(t)$. For the simple $[Fe_2S_2(SH)_4]^{2-}$ complex in vacuo the resulting power spectrum of J consists of two major peaks around 150 and 350 cm^{-1}, which can be understood in terms of two symmetry-adapted vibrational modes, $A_{g,A}$ and $A_{g,D}$, of the Fe_2S_2 core. This is to be expected qualitatively since J is known to depend sensitively on the distance between the metal cations and the bridging anions as well as on the bonding angle. But the present technique allows one to go beyond these qualitative rules in more complex situations, e.g., when environmental motion couples to the dynamics of the magnetic core as demonstrated in ref 41.

In the present study, the method has been assessed in much detail using a minimal model in vacuo. Clearly, for such a simple case more traditional methods such as harmonic analysis followed by evaluation of J along selected eigenvector displacements can be used to unravel the magneto-structural properties. At the same time it is evident that the method introduced can be used "as is" for much larger molecular systems, i.e. those which are accessible by present day density functional methods, where many coupled modes might be relevant for J thus rendering approaches difficult to apply where *all* relevant modes have to be identified a priori. Furthermore, the presented technique can be combined with suitable QM/MM coupling schemes to include extended environments via standard force field descriptions as demonstrated in ref 41. The potential of such an approach has been illustrated recently by studying the rather complex dynamical magnetostructural properties of the binuclear iron−sulfur cofactor in fully solvated ferredoxin from cyanobacterium *Anabaena* PCC7119. This investigation[41] uncovered that valuable information is encoded in the dynamics of the exchange coupling, $J(t)$, which can be analyzed conveniently when Fourier transformed into the corresponding spectrum $A_{JJ}(\omega)$. As suggested in ref 41 there is hope that experimental methods will be developed to follow the theoretical advances in this emerging field of dynamical magnetostructural properties.

Acknowledgment. We are grateful to Christian Boehme, Gerald Mathias, and Bernd Meyer for fruitful discussions and technical help. This work was supported in part by Deutsche Forschungsgemeinschaft (DFG) via grant MA 1547/7 and Fonds der Chemischen Industrie (FCI). The calculations were carried out using resources of Rechnerverbund-NRW, BOVILAB@RUB, and NIC.

Note Added after ASAP Publication. This article was released ASAP on June 26, 2008 with a minor error in the Acknowledgment. The correct version was posted on August 12, 2008.

References

(1) Noodleman, L.; Peng, C. Y.; Case, D. A.; Mouesca, J.-M. *Coord. Chem. Rev.* **1995**, *144*, 199–244.

(2) Lovell, T.; Himo, F.; Han, W.-G.; Noodleman, L. *Coord. Chem. Rev.* **2003**, *238* (239), 211–232.

(3) Yamanaka, S.; Yamaguchi, K. *Bull. Chem. Soc. Jpn.* **2004**, *77*, 1269–1286.

(4) Ciofini, I.; Daul, C. A. *Coord. Chem. Rev.* **2003**, *238* (239), 187−209.

(5) Calzado, C. J.; Cabrero, J.; Malrieu, J. P.; Caballol, R. *J. Chem. Phys.* **2002**, *116*, 2728–2747.

(6) Calzado, C. J.; Cabrero, J.; Malrieu, J. P.; Caballol, R. *J. Chem. Phys.* **2002**, *116*, 3985–4000.

(7) de Graaf, C.; Sousa, C.; de Moreira, P. R. I.; Illas, F. *J. Phys. Chem. A* **2001**, *105*, 11371–11378.

(8) Fink, K.; Fink, R.; Staemmler, V. *Inorg. Chem.* **1994**, *33*, 6219–6229.

(9) Broer, R.; Hozoi, L.; Nieuwpoort, W. C. *Mol. Phys.* **2003**, *101*, 233–240.

(10) Beinert, H.; Holm, R. H.; Münck, E. *Science* **1997**, *277*, 653–659.

(11) Rees, D. C. *Annu. Rev. Biochem.* **2002**, *71*, 221–246.

(12) Rees, D. C.; Howard, J. B. *Science* **2003**, *300*, 929–931.

(13) Johnson, D. C.; Dean, D. R.; Smith, A. D.; Johnson, M. K. *Annu. Rev. Biochem.* **2005**, *74*, 247–281.

(14) Palmer, G.; Dunham, W. R.; Fee, J. A.; Sands, R. H.; Iizuka, T.; Yonetani, T. *Biochim. Biophys. Acta* **1971**, *245*, 201–207.

(15) Sands, R. H.; Dunham, W. R. *Q. Rev. Biophys.* **1975**, *7*, 443–504.

(16) Kramers, H. A. *Physica* **1934**, *1*, 182–192.

(17) Anderson, P. W. *Phys. Rev.* **1950**, *79*, 350–356.

(18) Anderson, P. W. *Phys. Rev.* **1959**, *115*, 2–13.

(19) Kahn, O. *Molecular Magnetism*; VCH Publishers, Inc.: New York, 1993; pp 145−183.

(20) Hay, P. J.; Thibeault, J. C.; Hoffmann, R. *J. Am. Chem. Soc.* **1975**, *97*, 4884–4899.

(21) Neese, F. *J. Phys. Chem. Solids* **2004**, *65*, 781–785.

(22) Kahn, O.; Briat, B. *J. Chem. Soc. Faraday Trans. 2* **1976**, *72*, 268–281.

(23) Wang, C.; Fink, K.; Staemmler, V. *Chem. Phys.* **1995**, *201*, 87–94.

(24) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, Inc.: New York, 1996; pp 39−107.

(25) Adamo, C.; Barone, V.; Bencini, A.; Broer, R.; Filatov, M.; Harrison, N. M.; Illas, F.; Malrieu, J. P.; de Moreira, P. R. I. *J. Chem. Phys.* **2006**, *124*, 107101.

(26) Hübner, O.; Sauer, J. *J. Chem. Phys.* **2002**, *116*, 617–628.

(27) Mayer, I. *Adv. Quantum Chem.* **1980**, *12*, 189–262.

(28) Mayer, I.; Angelov, S. A. *Int. J. Quantum Chem.* **1980**, *18*, 783–796.

(29) Yamaguchi, K.; Yoshioka, Y.; Takatsuka, T.; Fueno, T. *Theor. Chim. Acta* **1978**, *48*, 185–206.

(30) Noodleman, L. *J. Chem. Phys.* **1981**, *74*, 5737–5743.

(31) Shoji, M.; Nishiyama, Y.; Maruno, Y.; Koizumi, K.; Kitagawa, Y.; Yamanaka, S.; Kawakami, T.; Okumura, M.; Yamaguchi, K. *Int. J. Quantum Chem.* **2004**, *100*, 887–906.

(32) Yamaguchi, K.; Jensen, F.; Dorigo, A.; Houk, K. N. *Chem. Phys. Lett.* **1988**, *149*, 537–542.

(33) Caballol, R.; Castell, O.; Illas, F.; de, P. R.; Moreira, I.; Malrieu, J. P. *J. Phys. Chem. A* **1997**, *101*, 7860–7866.

(34) Noodleman, L.; Lovell, T.; Liu, T.; Himo, F.; Torres, R. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 259–273.

(35) Hübner, O.; Sauer, J. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5234–5243.

(36) Shoji, M.; Koizumi, K.; Kitagawa, Y.; Yamanaka, S.; Okumura, T. K. M.; Yamaguchi, K. *Int. J. Quantum Chem.* **2005**, *105*, 628–644.

(37) Shoji, M.; Koizumi, K.; Kitagawa, Y.; Yamanaka, S.; Okumura, M.; Yamaguchi, K. *Int. J. Quantum Chem.* **2007**, *107*, 609–627.

(38) Goodenough, J. B. *J. Phys. Chem. Solids* **1958**, *6*, 287–297.

(39) Kanamori, J. *J. Phys. Chem. Solids* **1959**, *10*, 87–98.

(40) Ginsberg, A. P. *Inorg. Chim. Acta Rev.* **1971**, *5*, 45–68.

(41) Schreiner, E.; Nair, N. N.; Pollet, R.; Staemmler, V.; Marx, D. *Proc. Natl. Acad. Sci.* **2007**, *104*, 20725–20730.

(42) Heisenberg, W. *Z. Phys.* **1928**, *49*, 619–636.

(43) Dirac, P. A. M. *Proc. Roy. Soc. A.* **1929**, *123*, 714–733.

(44) van Vleck, J. H. *The Theory of Electric and Magnetic Susceptibilities*, 1st ed.; Oxford University Press: London, United Kingdom, 1932; pp 316−360.

(45) Note: there is a obvious typographical error in the corresponding equation, eq 1, of ref 41.

(46) Yamaguchi, K.; Fukui, H.; Fueno, T. *Chem. Lett.* **1986**, 625–628.

(47) Nishino, M.; Yamanaka, S.; Yoshioka, Y.; Yamaguchi, K. *J. Phys. Chem. A* **1997**, *101*, 705–712.

(48) Löwdin, P. O. *Phys. Rev.* **1955**, *97*, 1474–1489.

(49) Wang, J.; Becke, A. D., Jr *J. Chem. Phys.* **1995**, *102*, 3477–3480.

(50) Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **2007**, *126*, 214104.

(51) Theophilou, I.; Thanos, S.; Theophilou, A. K. *J. Chem. Phys.* **2007**, *127*, 234103.

(52) Filatov, M.; Shaik, S. *Chem. Phys. Lett.* **1998**, *288*, 689–697.

(53) Filatov, M.; Shaik, S. *Chem. Phys. Lett.* **1999**, *304*, 429–437.

(54) de Moreira, P. R. I.; Costa, R.; Filatov, M.; Illas, F. *J. Chem. Theory Comput.* **2007**, *3*, 764–774.

(55) Ruiz, E.; Cano, J.; Alvarez, S.; Alemany, P. *J. Comput. Chem.* **1999**, *20*, 1391–1400.

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

(57) Marx, D.; Hutter, J. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing (NIC): Forschungszentrum Jülich, Germany, 2000; *3*, pp 301−449.

(58) Norman, J. G.; Ryan, P. B.; Noodleman, L. *J. Am. Chem. Soc.* **1980**, *102*, 4279–4282.

(59) Cory, M. G.; Stavrev, K. K.; Zerner, M. C. *Int. J. Quantum Chem.* **1997**, *63*, 781–795.

(60) CPMD, Version 3.11, Copyright IBM Corp. 1990−2008, MPI für Festkörperforschung Stuttgart, 1997−2001.

(61) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(62) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(63) Vanderbilt, D. *Phys. Rev. B* **1990**, *41*, 7892–7895.

(64) Martyna, G. J.; Tuckerman, M. E. *J. Chem. Phys.* **1999**, *110*, 2810–2821.

(65) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(66) Sigfridsson, E.; Olsson, M. H. M.; Ryde, U. *Inorg. Chem.* **2001**, *40*, 2509–2519.

(67) Gillum, W. O.; Frankel, R. B.; Foner, S.; Holm, R. H. *Inorg. Chem.* **1976**, *15*, 1095–1100.

(68) Cabrero, J.; Calzado, C. J.; Maynau, D.; Caballol, R.; Malrieu, J. P. *J. Phys. Chem. A* **2002**, *106*, 8146–8155.

(69) Martin, R. L.; Illas, F. *Phys. Rev. Lett.* **1997**, *79*, 1539–1542.

(70) Ruiz, E.; Alvarez, S.; Cano, J.; Polo, V. *J. Chem. Phys.* **2005**, *123*, 164110.

(71) Perdew, J. P.; Zunger, A. *Phys. Rev. B* **1981**, *23*, 5048–5079.

(72) Akande, A.; Sanvito, S. *J. Chem. Phys.* **2007**, *127*, 034112.

(73) d'Avezac, M.; Calandra, M.; Mauri, F. *Phys. Rev. B* **2005**, *71*, 205210.

(74) VandeVondele, J.; Sprik, M. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1363–1367.

(75) Tavernelli, I. *J. Phys. Chem. A* **2007**, *111*, 13528–13536.

(76) Anisimov, V. I.; Zaanen, J.; Andersen, O. K. *Phys. Rev. B* **1991**, *44*, 943–954.

(77) Liechtenstein, A. I.; Anisimov, V. I.; Zaanen, J. *Phys. Rev. B* **1995**, *52*, R5467–R5470.

(78) Rohrbach, A.; Hafner, J.; Kresse, G. *J. Phys.: Condens. Matter* **2003**, *15*, 979–996.

(79) Kulik, H. J.; Cococcioni, M.; Scherlis, D. A.; Marzari, N. *Phys. Rev. Lett.* **2006**, *97*, 103001.

(80) Sit, P. H.-L.; Cococcioni, M.; Marzari, N. *J. Electroanal. Chem.* **2007**, *607*, 107–112.

(81) Dederichs, P. H.; Blügel, S.; Zeller, R.; Akai, H. *Phys. Rev. Lett.* **1984**, *53*, 2512–2515.

(82) Rudra, I.; Wu, Q.; van Voorhis, T. *J. Chem. Phys.* **2006**, *124*, 024103.

(83) Rudra, I.; Wu, Q.; van Voorhis, T. *Inorg. Chem.* **2007**, *46*, 10539–10548.

(84) Herrmann, C.; Yu, L.; Reiher, M. *J. Comput. Chem.* **2006**, *27*, 1223–1239.

(85) Ramírez, R.; López-Ciudad, T. P.; Kumar, P.; Marx, D. *J. Chem. Phys.* **2004**, *121*, 3973–3983.

(86) Leach, A. R. *Molecular modelling: principles and applications*, 2nd ed.; Pearson Education Limited: Harlow, United Kingdom, 2001; pp 458−508.

(87) Schreiner, E. Biochemical Aspects of Iron-Sulfur Systems: Magnetostructural Properties of Ferredoxins and Prebiotic Peptide Synthesis Involving Pyrite, Ph.D. Thesis, Ruhr-Universität Bochum, Germany, 2007.

CT800089X

JCTC Journal of Chemical Theory and Computation

Mechanism of Ultrafast Photodecay in Restricted Motions in Protonated Schiff Bases: The Pentadieniminium Cation

Jaroslaw J. Szymczak,* Mario Barbatti, and Hans Lischka*

Institute for Theoretical Chemistry, University of Vienna, Waehringerstrasse 17, A-1090 Vienna, Austria

Received May 6, 2008

Abstract: Ab initio surface-hopping dynamics simulations for the *trans*-penta-3,5-dieniminium cation (PSB3) are presented imposing different sets of mechanical restrictions in order to investigate the response of the molecular system to certain environmental degrees of hindrance. A general scheme for classification of photoisomerization mechanisms in conjugated chains based on the analysis of torsional angles is proposed allowing direct characterization of the different isomerization mechanisms proposed previously. On the basis of a statistical analysis of 300 trajectories a new photoisomerization mechanism—the Folding Table—was found. This mechanism and the One-Bond-Flip are almost entirely responsible for the photoisomerization process in PSB3.

1. Introduction

Protonated Schiff bases $CH_2(CH)_{n+1}NH_2^+$ (PSBn) constitute a very interesting and important class of molecular systems since they serve as models for studying the photoisomerization of retinal, which is involved in the primary process of vision and represents the driving force for a proton pump through cell membranes of *Halobacterium salinarium*.[1,2] The former process is initiated by a *cis-trans* photoisomerization of 11-*cis* retinal, the chromophore of rhodopsin.[3–7] A related photochemical reaction starts with UV excitation of all-*trans* retinal followed *trans-cis* photoisomerization to 13-*cis* retinal in bacteriorhodopsin.[8,9] These photoisomerization processes belong to the fastest photochemical reactions in nature[3] and have been studied extensively in experimental[1,10–19] and theoretical investigations.[20–27] Extensive quantum chemical calculations have been performed for the computation of photochemical reaction pathways starting from the Franck–Condon region and leading to the conical intersections where ultrafast, radiationless decay to the electronic ground state occurs.[21,28–31] Dynamics calculations[22–27,32,33] have given detailed insights into the actual course of these processes beyond the static analysis. Whereas most of these investigations have been performed for the isolated PSB system, several theoretical investigations have been carried out also with the inclusion of environmental effects.[30,34–41] Extensive calculations with realistic representation of the protein environment have been undertaken for the electronic ground state of rhodopsin[41] and bacteriorhodopsin[42] models including dynamics simulations using quantum chemical and force field methods. Dynamical simulations of the photoisomerization process considering environmental effects are even more challenging due to increased difficulties in performing the excited-state calculations. Combined quantum mechanical/molecular mechanics (QM/MM) approaches have been used for that purpose.[24,43] Because of the excessive computational cost of these simulations only a few trajectory runs could be performed limiting the statistical significance of the obtained results.

Most of the aforementioned theoretical investigations have led to the general picture that the photoisomerization proceeds in a relatively straightforward way by starting with in-plane skeletal relaxation followed by torsion around CC bonds holding double bond character in the ground state. An exception is the recent work of Send and Sundholm[21] based on single-reference time-dependent density functional (TD-DFT) and resolution-of-the-identity coupled cluster to second order (RI-CC2) methods giving a somewhat different picture. In spite of the interesting conclusions drawn in this

* Corresponding author e-mail: jaroslaw.szymczak@univie.ac.at (J.J.S.) and hans.lischka@univie.ac.at.

10.1021/ct800148n CCC: $40.75
Published on Web 07/26/2008

work the reliability of these methods is questionable as comparison of analogous results for smaller PSB chains with multireference approaches show.[25,37,44] In any case, it is clear that, when retinal or a retinal model is considered within the protein pocket, spatial requirements for the *cis-trans* isomerization become essential for the description of the isomerization.

In the present work we want to follow a simpler concept for inclusion of external effects on a PSB chain as compared to detailed QM/MM simulations. Extended dynamics simulations have been performed by mechanically restricting the motion of specific PSB sites by attributing artificially large nuclear masses to terminal hydrogen atoms (for details see below) in a similar way as implemented by Warshel.[27] This technique of mass restrictions has also been successfully used by us for aminopyrimidine[26,45,46] in order to simulate the adenine photodynamics. The advantage of this approach is that the restricted dynamics can be performed at the same computational cost as the nonrestricted approach and allows more freedom in the selection of quantum chemical methods in the dynamics and the calculation of statistically relevant sets of trajectories. These mass restrictions can be viewed as models for the link of retinal to the protein surroundings by the Schiff base bond and for cutting of the retinal end containing the β-ionone ring. One can also look at these restrictions in a more general way as certain general restraints on an unsaturated carbon chain.

Several mechanisms for restricted *cis-trans* isomerizations have been proposed in the literature so far. The simplest one, the torsion around one formal double-bond (one-bond-flip - OBF),[10,30,47] which is supposed to take place primarily in the nonrestricted isomerization, has been put aside because of excessive space requirements. The first alternative, proposed by Warshel based on mixed quantum-classical dynamics simulations,[27] is the bicycle-pedal motion (BP), which consists of the simultaneous torsion of two formal double bonds separated by a single bond. Later on, Warshel and Barboy[48] noted that the isomerization could also take place through a mechanism similar to the BP but with one of the formal double bonds performing only a partial torsional motion. We refer to this mechanism as the nonrigid BP (NRBP). Recently, Frutos et al.[24] have brought new evidence in favor of the NRBP mechanism by their dynamics simulations on retinal. In their investigation, however, only one single trajectory was computed, and it is not possible to draw statistical inferences. Liu and Asato[49] proposed another mechanism named concerted torsion or hula-twist (HT), which consists of the simultaneous torsion of two adjacent bonds followed by skeletal relaxation. The HT mechanism has been extensively discussed in the work of Ruiz et al.[50] and Norton and Houk.[51] Additionally, a new mechanism has been identified in the course of the present dynamics simulations and will be introduced in the current work. The so-called folding-table mechanism (FT) is a combination of three concerted torsions, which will be discussed in detail below.

Although several attempts have been made to verify by means of dynamics simulations[22,23,26,27,34,52,53] how the photoisomerization actually takes place, the somewhat loose definition of the different mechanisms makes the comparison of different data sets difficult. The situation is particularly critical because the freedom of motion in the dynamics simulation usually produces geometries that are far from ideal prototypes. Our present goal is to utilize the PSB3 system, considered as the simplest example in retinal modeling,[28,54,55] to establish clear definitions of these mechanisms, which can be directly applied to the classification of the dynamical processes occurring in molecular chains such as PSBn. Application of this classification scheme is not restricted to PSB3 chains but can in principle be used for any molecular chain presenting *cis-trans* isomerization.

2. Computational Details

The quantum chemical calculations were performed at the complete active space self-consistent field (CASSCF) and multireference configuration interaction (MRCI) levels. The CASSCF calculations were carried out using a CAS(6,6) averaging over the two lowest singlet states (denoted as SA-2-CASSCF(6,6)). The reference space for the MRCI calculations is based on the SA-2-CASSCF(6,6) calculation. For the MRCI calculations a CAS(4,5) reference space was chosen (denoted as MRCI(4,5)). This selection was chosen on the basis of a natural orbital occupation criterion moving the lowest CASSCF orbital (occupation 1.96) into the doubly occupied space. The CI expansion included either all single and double substitutions (MR-CISD) or only single excitations (MR-CIS) from the reference space. When double excitations are included, the generalized interacting space restriction[56] was adopted. The 3−21G and 6−31G* basis sets were used.[57,58]

On-the-fly ab initio nonadiabatic dynamics calculations were performed using Tully's surface hopping approach.[59,60] The nuclear motion is computed by solving Newton's equations based on the Born−Oppenheimer potential obtained by the respective quantum chemical method. The integration of the classical equations is performed by means of the velocity-Verlet algorithm[61] in time-steps of 0.5 fs. The total simulation time was 200 fs. The time-dependent wave function is expanded in the adiabatic representation, and the time-dependent electronic Schrödinger equation is integrated using the fifth-order Butcher algorithm.[62] In order to further improve the numerical integration of the time-dependent Schrödinger equation, a smaller time step $\Delta t' = \Delta t/m_s$ ($m_s = 20$) is used, with the relevant quantities interpolated from t to $t + \Delta t$. The obtained time-dependent adiabatic populations were corrected for decoherence effects[63] ($\alpha = 0.1$ hartree) and used for computing the surface hopping probabilities of a nonadiabatic transition according to the fewest-switches algorithms proposed by Tully[59,60] and Hammes-Schiffer and Tully.[64] At each time-step, a random event is used to decide whether the system will switch to another state. The momentum after frustrated hoppings was kept constant, and after actual hoppings it was readjusted along the nonadiabatic coupling vector. The initial conditions for the simulated trajectories are generated by means of a ground-state Wigner distribution by treating the nuclear coordinates and momenta within the quantum-harmonic-oscillator approximation.

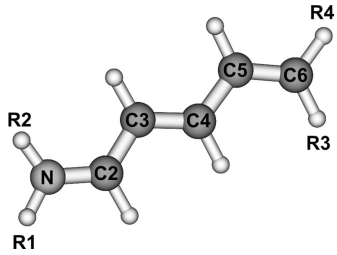


Figure 1. Numbering of atoms in PSB3 with indication of restrictions; nonrestricted PSB3: R(1−4) = H; PSB3 with two restrictions: R1 = R4 = "restricted hydrogen", R2 = R3 = H; PSB3 with 4 restrictions: R(1−4) = "restricted hydrogen".

As already mentioned above, in addition to the nonrestricted dynamics, restricted simulations were performed. The restrictions were imposed by increasing the masses of the terminal hydrogen atoms. Two sets of restrictions were constructed. In the first set, two hydrogen atoms (R1 and R4 in Figure 1, later on denoted as restricted hydrogens) were modified. The restricted hydrogens were positioned in *trans* orientation to each other. This choice is designed to model the situation where the restricted hydrogen atom located at the nitrogen end represents the binding of PSB3 to the protein, whereas the other restricted hydrogen atom stands for the rest of the retinal chain. After preliminary test calculations a mass of 1000 au was used. This choice practically fixes the positions of these hydrogen atoms during the dynamics. In the second set of restrictions, masses of 1000 au were given to all four terminal hydrogens. Using these two sets of restrictions, dynamics simulations starting at the *trans* isomers of PSB3 were performed.

A total of 300 trajectory calculations for the restricted PSB3 systems were performed (100 for each of the system with no, two, and four restrictions). The CASSCF and MRCI calculations have been performed using the COLUMBUS program system[65–67] and the methods for analytic computation of gradient and nonadiabatic coupling vectors.[68–72] The on-the-fly surface-hopping calculations were carried out by means of the program system NEWTON-X[25,73] using the quantum chemical data computed by COLUMBUS at each time step. The automatic classification of the observed motions was performed using an analysis program developed in our group.

3. Validation of the Theoretical Level

In on-the-fly dynamics calculations, energies, gradients, and nonadiabatic coupling vectors have to be computed at each time step leading to hundreds of thousands of individual quantum chemical calculations. Therefore, it is crucial to find a balance between the quality of the approach (choice of method and basis set) and the computational cost. Under these circumstances accuracies as they are achievable in conventional single-point calculations and geometry optimizations are out of reach. Thus, before starting such time-consuming dynamics simulations, static test calculations on the investigated systems were carried out with the aim of judging the performance of different approaches for calculating the excited-state energies and finding the optimal procedure. The main purpose of this part of the investigation is the assessment of the unpolarized 3−21G basis for use in the main body of dynamics calculations. This basis set, which has been examined and used in recent dynamics investigations of other groups[74–76] and in our group as well,[25] is a natural candidate for achieving substantial savings in computer time if the resulting energy surfaces are not distorted too much. However, it is also clear that with the drastic reduction in basis set size certain inadequacies are unavoidable. Additionally, the influence of the choice of the quantum chemical method has been investigated as well.[37] This assessment is performed in several steps. In step 1 the critical points on the energy surfaces relevant for the photoisomerization process−vertical excitation energies, the S_1 energy minimum (S_1min) under planarity restriction, and the structure for the minimum on the crossing seam (MXS) of the torsion around the central bond−were investigated. In step 2 approximate reaction paths starting from S_1min to the S_0/S_1 MXS have been examined. The reaction paths have been obtained by means of the method of linear interpolation of internal coordinates (LIIC) between these two structures using natural internal coordinates as defined by Fogarasi et al.[77] Finally, the lifetimes of dynamics calculations are compared with previous results.[23,24,26,27,33–35]

Figure 2a shows the differences in bond lengths along the *trans*-PSB3 chain for the S_1min structure employing the SA-2-CASSCF(6,6), MR-CIS(4,5), and MR-CISD(4,5) methods and the 6−31G* and 3−21G basis sets. Full Cartesian coordinates are given in the Supporting Information. Results for the optimized S_1min structures obtained with the smaller basis sets agree quite well with those obtained with the 6−31G* basis. Differences in bond distances are mostly smaller than 0.01 Å within the same method. The largest difference of 0.04 Å was observed for the C3−C4 bond with MR-CISD(4,5). However, all methods predict this bond to be the longest one, and, therefore, the torsion around it is expected to be favored. The second largest difference between both basis sets is the elongation of the N1−C2 bond by ∼0.015 Å for the CASSCF/3−21G calculation. This difference is also reflected in dynamics calculations for CASSCF/3−21G where a secondary channel of deactivation due to the torsion around the N1−C2 bond has been observed.[25] The use of the MR-CIS/3−21G method removed this artifact due to a reduction of the N1−C2 bond in comparison to CASSCF. The basis set differences in bond lengths observed for the MXS structure are of the same order as in the case of the S_1min structures (Figure 2b and the Supporting Information for Cartesian coordinates). In agreement with previous investigations[31,53–55] the geometry of the MXS is reached by a ninety degrees rotation around the central double bond and concomitant adjustment of bond lengths.

Figure 3a shows that the shapes of the LIIC curves are very well reproduced by the calculations employing the smaller basis set. Differences are observed primarily in the neighborhood of the S_1min structures and amount to about 0.3 eV in the maximum and result from differences in the quantum chemical method rather and not from the basis set. Furthermore, the energies of the aforementioned key points of the photoisomerization process relative to the ground-

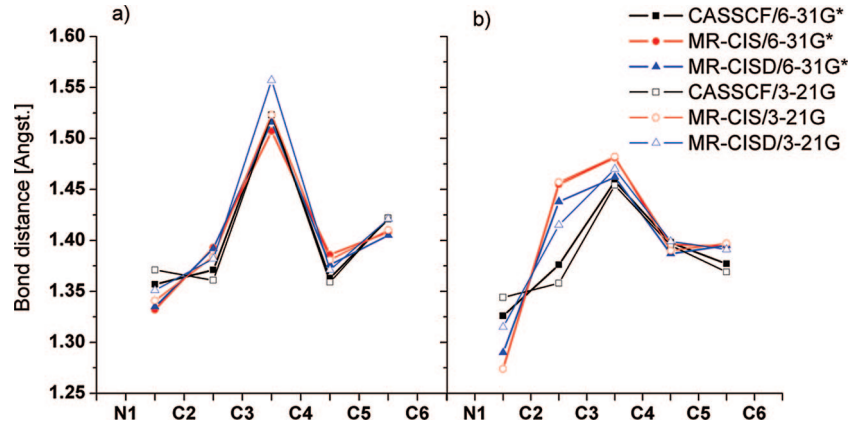


Figure 2. CN and CC bond lengths for unrestricted PSB3 in a) S_1 minimum and b) for the S_1/S_0 MXS determined for different approaches employing the 6−31G* and 3−21G basis sets.

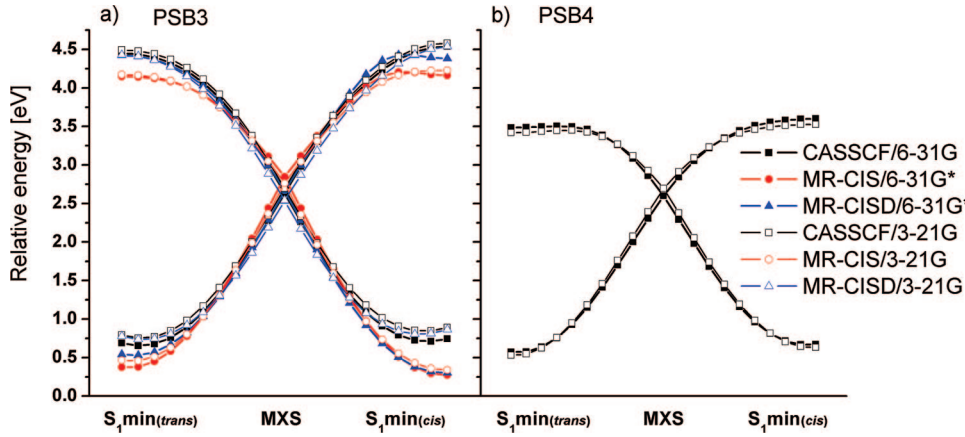


Figure 3. Energy paths connecting the minimum in the S_1 state (*trans* isomer) to the MXS and the MXS to the minimum in the S_1 state (*cis* isomer). a) PSB3 calculated at the (SA-2) CASSCF(6,6), MR-CIS(4,5)/CASSCF(6,6) and MR-CISD(4,5)/CASSCF(6,6) levels using the 6−31G* and 3−21G basis sets and b) PSB4 calculated at the SA-2-CASSCF(8,8) level using the 6−31G* and 3−21G basis sets. Energies are given relative to the optimized ground-state structure.

state minimum are quite the same for different basis sets (see Table 1). Typical differences are close to 0.1 eV or less for the CASSCF and MR-CIS methods. Slightly larger differences reflecting the corresponding changes in geometry for the same level are observed only for the MR-CISD method.

Moreover, the lifetime of PSB3 obtained in the current work is not significantly affected by the choice of the two examined basis sets. The average lifetime of 96 fs obtained from the trajectories with no restrictions and the 3−21G basis agree well with the results obtained with the 6−31G* basis set (98 fs for unrestricted dynamics).[26]

To verify our choice of computational method and basis set even further, especially in view of future applications to larger protonated Schiff bases, calculations on the 5-*cis* and 5-*trans* isomers of the hepta-3,5,7-trieniminium cation (PSB4) were performed as well. Similar to PSB3 a very good performance of the 3−21G basis set was observed. In the Supporting Information (Table SI1) vertical excitation energies and energies of key points for both isomers of PSB4 are collected. The calculations employing the CASSCF(8,8)/6−31G* and CASSCF(8,8)/3−21G methods show very good agreement. This is further assured by LIIC potential curves (Figure 3b) computed for both levels of theory.

Combining the all just-described experience, the MR-CIS/3−21G approach was chosen for performing the dynamics calculations of PSB3. Tests using the MR-CISD/3−21G approach did not show significant improvement, and the cost of such calculations is severe making it extremely difficult to obtain statistically significant samples. The present systematic survey gives us the confidence that the major features of the dynamics such as the structural and energetic changes and lifetimes are reasonably well reproduced. We expect—and test calculations on PSB4 presented here confirm this expectation—that the performance of the 3−21G basis will be of similar good quality also for the larger PSB members. However, it is also clear that great care has to be exercised when trying to extend this experience to other molecular systems.

4. Classification Scheme

The purpose of this subsection is to describe the methods used in this work to analyze the geometrical evolution of the PSB3 chain. The behavior of the dynamics of PSBn systems during the photodecay has been characterized in previous theoretical investigations as a two-step process in terms of initial skeletal stretching and subsequent torsional

Table 1. Relative Vertical Excitation Energies (VEE), Energies of S_1min, and MXS for *cis* and *trans* Isomers of PSB3

	CASSCF	MR-CIS	MR-CISD
	PSB3 (*trans*)		
S_0[a]			
6−31G*	−248.250193[b]	−248.349618[b]	−248.891836[b]
3−21G	−246.875623[b]	−246.959738[b]	−247.331737[b]
VEE			
6−31G*	4.84	4.39	4.61
3−21G	4.93	4.45	4.70
E(S_1min)			
6−31G*	4.45	4.15	4.42
3−21G	4.49	4.17	4.43
E(MXS)			
6−31G*	2.63	2.85	2.66
3−21G	2.70	2.76	2.53
	PSB3 (*cis*)		
S_0[a]			
6−31G*	0.15	0.16	0.14
3−21G	0.13	0.15	0.12
VEE			
6−31G*	4.68	4.22	4.45
3−21G	4.81	4.31	4.57
E(S_1min)			
6−31G*	4.54	4.16	4.38
3−21G	4.58	4.22	4.54

[a] Geometries obtained at the B3LYP/SV(P) level of theory. [b] Values in Hartrees.

motion around a double bond.[54,78,79] In the initial stage of the dynamics the system relaxes by adjusting the bonds length, elongating the double bonds, and shortening the single bonds.[54,78–80] After adjusting all bonds, the molecular system proceeds to the second step of the dynamics. The crossing seam is reached by skeletal torsions around one[54] or more bonds.[24,27,48] When the system then switches from the excited state to the ground state along a torsional mode, it can either continue or reverse the motion that led to the MXS and further relax to the final product geometry.

The initial phase of the quasi-planar CC bond length relaxation is relatively short[24,26,33] (10−15 fs) and seems to be straightforward, whereas the second phase of torsional modes is certainly much more involved and finally determines the outcome of the dynamics. Therefore, we concentrate in our analysis on the torsional phase. In order to get better insight into this process this step was analyzed by dividing it into two stages. The first stage concerns the torsions leading to the conical intersection, and the second stage relates to the continuation of the motion on the ground-state energy. Both stages are characterized by complicated torsional modes coupled to bending and stretching modes. To characterize the evolution of the torsional angles in the first stage differences $\Delta\theta_i = \theta_i^{hop} - \theta_i^0$ are computed for all skeletal torsional angles where θ_i^0 is taken from the ground-state geometry as given by the initial conditions of the trajectory and θ_i^{hop} is the angle at the time of the hopping. The dihedral angles used in the analysis are θ_1 = R1−N−C2−C3, θ_2 = N−C2−C3−C4, θ_3 = C2−C3−C4−C5, θ_4 = C3−C4−C5−C6, and θ_5 = C4−C5−C6−R4 (for numbering see Figure 1). Analysis of these angles in the course of the dynamics will lead to wide distributions. For examples see Figures 4 and 5, which are discussed below. From these distributions no obvious separation in different classes can

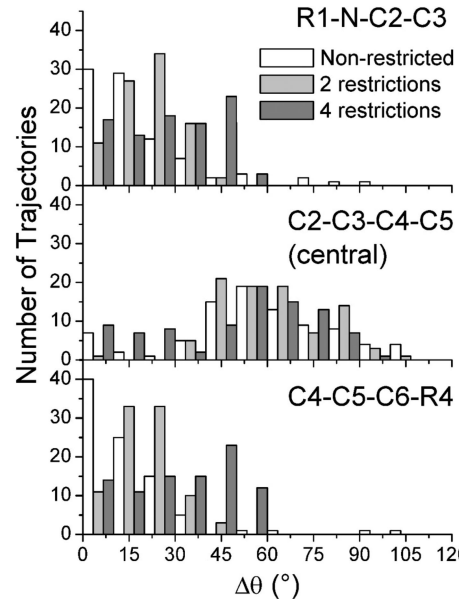


Figure 4. Histograms of changes in torsional angles of the formal double bonds ($\Delta\theta_i$, i = 1,3,5) at the moment of hopping relative to the planar ground-state structure.

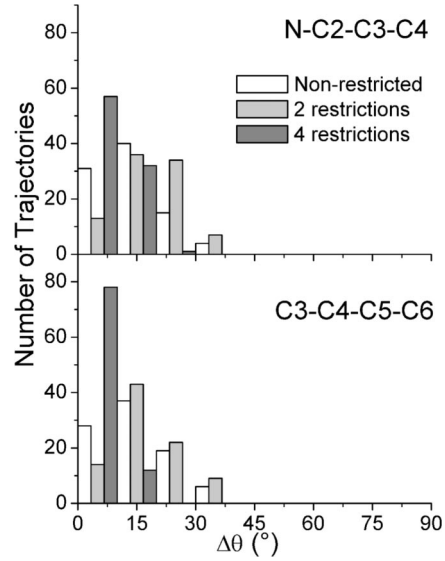


Figure 5. Histograms of changes in torsional angles of the formal single bonds ($\Delta\theta_i$, i = 2,4) at the moment of hopping relative to the planar ground-state structure.

be observed. In order to achieve a simple classification, $\Delta\theta_i$ values have been assigned to one of three groups in 30° ranges between 0 and 90°. These groups have been labeled according to their progress in the torsional mode as discussed below. Test calculations using different angle increments have shown that this division represents a good compromise between sufficiently fine granularity for the description of individual classes of structures and simplicity in terms of number of classes.

The purpose of this classification is to give a rough overview of an otherwise complicated and detailed situation. In spite of the admittedly arbitrariness of this procedure with respect to the selected angle thresholds we think that it is useful and allows for simple classification schemes. If the change in a given torsional angle was located between 0°

Table 2. Examples of $\{\Delta\theta\}$ Patterns Used in the Classification of the Types of Motions in PSB3[a]

	dihedral angle change $\Delta\theta_i$				
	N−C2	C2−C3	C3−C4	C4−C5	C5−C6
OBF	a	a	c	a	a
BP	a	c	a	c	a
NRBP	a	c	a	b	a
HT	a	a	c	c	a
FT	b	a	c	a	b

[a] For OBF, for example, only the pattern corresponding to the central torsion is shown. a - $0° \le \Delta\theta_i < 30°$. b - $30° \le \Delta\theta_i < 60°$. c - $60° \le \Delta\theta_i \le 90°$.

and 30°, this change has been considered as "not relevant" for the *trans-cis* isomerization and has been labeled as "*unchanged*" (denoted by the letter 'a'). When a given $\Delta\theta_i$ was situated in the interval between 30° and 60°, the change was classified as "*partial*" (denoted 'b'). The third group consisted of angles that had changed by more than 60°. This change was considered as "*significant*" (denoted 'c').

Using this criterion, any given PSB3 geometry can be classified according to its five dihedral changes $\Delta\theta_i$ or $\{\Delta\theta_i\}$ patterns, and the progress in the torsional modes can be analyzed quickly. Table 2 shows examples of $\{\Delta\theta\}$ patterns observed in the dynamics runs corresponding to each of the five types of isomerization mechanisms mentioned in the Introduction. The one-bond-flip (OBF)[10,30,47] is defined here as the process for which only one *significant* or *partial* change in a dihedral angle is found, whereas all other dihedral angles remain *unchanged*. The folding-table (FT) mechanism can be described as a combination of three torsions, with the main central torsion accompanied by two *partial* torsions on both sides separated by one *unchanged* dihedral angle each. The bicycle-pedal (BP)[27] mechanism is characterized by simultaneous *significant* changes of two dihedrals that are separated by one dihedral angle in between that stays *unchanged* during the motion. The nonrigid bicycle-pedal (NRBP) mechanism[48] represents a case similar to BP; however, the torsion around one of the bonds proceeds slower or starts later than the other. This behavior results in a *partial* change of the corresponding dihedral angle, which resembles a situation when one of the "pedals" is loose and the motion is no longer rigid. The last mechanism, the hula-twist[49–51] (HT) is described as a simultaneous *significant* change of two neighboring dihedral angles not accompanied by any other torsion.

The second stage of the torsional part of the dynamics—the behavior of the molecule after hopping to the ground state—has been examined from two points of view. The first one concerns the continuation of the motion leading to the conical intersection, and the second one its outcome in terms of the final product. For this purpose the following scheme was applied. The torsional motion of PSB3 is checked 20 fs after the hopping for changes in the dihedral angles with respect to the value at the hopping. Only torsional angles characteristic for the given type of motion are investigated. In the OBF case, for example, only the dihedral angle with *significant* or *partial* change is checked. If this change is larger than a threshold of $\delta = 20°$ or smaller than $-\delta$, then the motion is classified as *continued* or *reversed*, respectively. If this threshold is not reached within this time, the motion is examined again after each 5 fs. Previous calculations had shown that major changes of dihedral angles proceeded within a time range of about 50 fs.[24,39,43,52] Therefore we decided to continue this analysis for a somewhat longer time (60 fs after the hopping event), always trying to classify the motion either as *continued* or *reversed*. In addition to checking the motion toward the conical intersection with respect to continuation or reversal, the question is examined whether this motion is so pronounced that it also leads to a *cis* or *trans* structure within 60 fs. For this purpose all five dihedral angles θ_i are examined. The designated fragment is considered as *trans* (E) if the corresponding dihedral angle was in the range of $180° \pm 60°$ or *cis* (Z) if the angle fell into the range of $0° \pm 60°$. The geometry is classified in terms of $\{\theta_i\}$ patterns. The EEEEE and EEZEE patterns, for example, correspond to the all-*trans* and to the 3-*cis* structures, respectively. It is expected that the application of such a final product analysis should be especially important for the classes of restricted systems. Since the size of the PSB3 molecule is relatively small and full relaxation is not allowed in this case, the excess energy can cause additional torsions, usually not observed in the case of the nonrestricted dynamics. Therefore, especially with 4-restricted PSB3 the direct continuation of the motion leading to the conical intersection need not result in creation of the corresponding final product in many cases. Independently, the final structure is determined at the end of the simulation time.

5. Results and Discussion

5.1. Mechanism Leading to the Conical Intersection. The main mechanism leading to the conical intersection in unrestricted PSB3 is the OBF around the central C3−C4 bond.[22,26,30,54] The existence of external restrictions, however, modifies the character of this mechanism by enforcing additional torsions. The analysis of $\Delta\theta_i$ values at the time of hopping (Figures 4 and 5) indicates that the distribution of changes in torsional angles extends significantly beyond 30° only for bonds corresponding to double bonds in the ground state. Thus, practically only those bonds are involved in the initial phase of the isomerization process, while torsion around formal single bonds is of minor importance. As expected, the major changes in the torsional angles are observed for the central CC bond, which shows broad peaks at around 60° for all types of dynamics simulations, unrestricted, 2- and 4-fold restricted (Figure 4). This fact correlates well with the strong elongation of this bond in the planar S_1 minimum structure displayed in Figure 1. The fact that the conical intersection is on the average encountered at torsional angles significantly smaller than the 90° of the MXS has been already observed and discussed previously.[26,29]

All trajectories have been analyzed according to $\{\Delta\theta_i\}$ patterns presented in the Classification Scheme described above. This analysis is summarized in Table 3. In the case of the nonrestricted dynamics and the dynamics with two restrictions, the dominating fraction of trajectories follows the OBF to the conical intersection. In the nonrestricted case,

Table 3. Statistics of the Motions for all-*trans* PSB3 Leading to the Conical Intersection (in % of Trajectories)

system	OBF	FT	other motion	no hopping
no restrictions	91	0	4[a]	5
2 restrictions	75	9	14	2
4 restrictions	33	55	0	12

[a] All 4 trajectories showed $OBF_{(C3=C4)}$ accompanied by neighboring rotation slightly (~1°) larger than 30°.

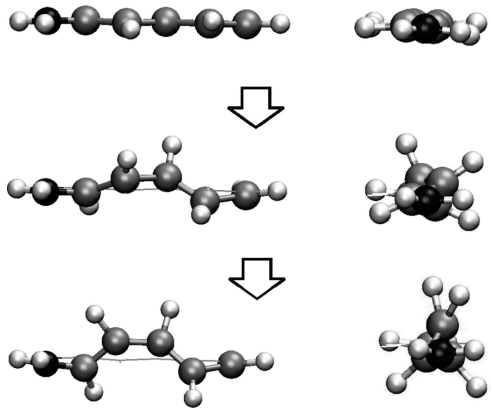


Figure 6. Characterization of the folding-table mechanism (snapshots taken from an actual trajectory). The line given in the second and third snapshot represents the initial chain.

the OBF is realized entirely by torsion around the central bond. Additional analysis of trajectories classified as "other motion" showed that all of them also feature the OBF mechanism, accompanied by *partial* neighboring torsion changes exceeding 30° slightly by ~1°. For the system with two restrictions, 75% of trajectories reach the conical intersection by rotating around the central bond. In both cases no additional isolated rotations at the ends of the PSB3 chain have been observed. Similar to the unrestricted system additional analysis of the "other motion" trajectories represent a situation similar to the one-bond-flip, or alternatively a folding-table process occurs, which is accompanied by a *partial* single bond rotation. This latter torsion angle change is, however, larger (average of 5° beyond 30°) than in the unrestricted dynamics, a fact which precludes them from being classified as an OBF or FT according to our criteria (see also Figures 4 and 5). A minor fraction of trajectories with two restrictions (9%) followed the FT mechanism.

With four restrictions the percentage of OBF drops to 33%. Since all terminal hydrogen atoms are fixed, torsion around the central bond is accompanied by other structural adjustments within the heavy-atom skeleton, while the possibility of the adjustments made by extended motion of the terminal heavy atom is ruled out. As a consequence of these restrictions the FT mechanism, illustrated in Figure 6, becomes dominant. The character of the main torsion resembles an OBF; however, one observes additional *partial* torsions at the terminal bonds on either side of PSB3. These partial torsions lead to a rise of the central CC bond above the original PSB3 plane.

It is worth noting that the other mechanisms that have been suggested to occur in restricted isomerizaton−BP, NRBP, and HT−have not been observed at all during the PSB3 dynamics.

Table 4. Motions after the Hopping (in % of Trajectories Given at the Time of Hopping (See Table *3*))

	OBF		FT		
system	cont.	rev.	cont.	rev.	no classification possible[a]
no restrictions	70	70	0	0	0/0
2 restrictions	65	32	62	38	3/0
4 restrictions	70	27	40	49	3/11

[a] Rrelative to all trajectories assigned to OBF/FT.

Table 5. Fraction of Trajectories (in % of *Continued* Motion Given in Table *4*) after the Hopping Leading to 3-*cis* or 3-*trans* PSB3 (See Text for Explanation)

	OBF			FT		
	cont. motion			cont. motion		
system	*cis*	*trans*	undeterm	*cis*	*trans*	undeterm
0R	89	5	6	-	-	-
2R	84	16	0	100	0	0
4R	22	13	65	32	5	63

5.2. Motion after Surface Hopping. The motion after hopping to the ground state is important since it will determine the final product formation. Several situations can be distinguished. The motion leading to the conical intersection is *continued* or *reversed*, or the molecular system stays in the vicinity of the conical intersection at least for the time of the analysis (60 fs, see the section on the Classification Scheme). Since in this work only the isolated PSB3 system is investigated, no energy transfer to the environment can occur, and the dynamics correspond to a vibrationally hot motion in the ground state during this stage. Therefore, additional vibrational modes may be activated by means of internal vibrational relaxation. In the course of the time period (200 fs) that was investigated here, such vibrational activation occurred mostly in the case when the *continued* motion was observed and new torsions around other bonds inactive in the earlier stages of the dynamics became important. On the other hand, the analysis presented below shows that when a *reversed* motion is observed, it reverses always with 100% efficiency.

Table 4 presents the results of the analysis after the hopping with regard to motion continuation or reversal with a maximum assessment time of 60 fs. Further analysis determining whether the examined torsion leads at least temporarily within a 60 fs time span after the hopping to an associated *cis* or *trans* structure is presented in Table 5. In the nonrestricted system 70% of trajectories classified as OBF continued this torsion after hopping (Table 4). The results given in Table 5 show that the majority (89%) of *continued* motion resulted in the formation of *cis*-PSB3. A similar trend was observed for the doubly restricted dynamics, where 65% of OBF motions (Table 4) leading to the conical intersection also led to *cis*-PSB3 with an efficiency of 84% (Table 5). The situation changes when the fully restricted system is regarded. The ratio of the *continued* motions remains approximately the same (70%); however, only 22% of them end up in the *cis* structure. This drop is accompanied by a very large percentage of undetermined geometries (65% of *continued* OBF). This fact is closely related to the restriction imposed on the system, which simply makes the change of

Table 6. Summary of Photoproduct Analysis (in % of All Trajectories) after the Hopping (Table *5*) and at the End of the Dynamics (Values in Parentheses)

product conformation	corresponding motion	*trans* PSB3		
		0R	2R	4R
EEEEE	-	32(21)	44(39)	40 (55)
EEZEE	$OBF_{C3=C4,}$ FT	59(18)	51(46)	12(3)
other		(17)[a]		
not assigned		4(39)	3(13)	36 (30)
no hopping (excluded from analysis)		5	2	12

[a] Other products: EZZZE(2), EEZZE(9), EZEEE(5), EEEZE(1).

the central torsional angle beyond 120° extremely difficult. In this case the central torsion stays in the range between 60° and 120°. On the other hand, when the motion leading to the conical intersection is *reversed* after hopping, reversion proceeds completely in all cases regardless of the type of motion and system. The statistical analysis of the FT motion for the restricted systems resembles that for the OBF: 62% and 40% of the initial motion is *continued* after the hopping for the doubly and fully restricted system, respectively. In the case of the fully restricted system continuation always implies rotation around the central bond since it is almost impossible to perform an extended rotation around a terminal double bond. For the dynamics with two restrictions any other motion beyond that around the central bond was not observed either. Yields of the product formation in the FT case are comparable to that of the OBF (Table 5). However, it is worth mentioning that similar to the OBF motion in more than half of the cases with four restrictions it was not possible to determine the final product of the *continued* motion unambiguously.

As the last step of the analysis the final structure at the end of the dynamics run was determined independently of the analysis of the PSB3 motion in the different stages of the dynamics (see Table 6). It is interesting to note that the differences between the yields of product formation examined after the hopping to the ground state and at the end of the dynamics are substantial, especially for the unrestricted case. They reflect the changes due to the vibrationally hot dynamics in the ground state after the initial product assignment was made. This is also manifested by the large number of structures that could not be assigned to one of the types of torsional motion at the end of the dynamics since at least one of the dihedral angles was located between 60° and 120°, i.e. far away from planarity. Nevertheless, the trends of the product formation are preserved and vary systematically by reducing the amount of 3-*cis* (EEZEE) and increasing the all-*trans* product formation (EEEEE). Although restricted PSB3 in gas phase is certainly a poor model for quantitative comparisons with real processes taking place in bacteriorhodopsin, it is worth noting that the experimental quantum efficiency of the reaction following the photoexcitation is 0.64,[81] which is well comparable with the present results.

5.3. Lifetimes. The depopulation of the S_1 state as a function of time is illustrated in Figure 7. This figure shows the S_1 occupation defined as the fraction of trajectories in the S_1 state in each time step. The occupation remains constant for a certain time t_d after which it starts to decrease in an exponential way from there on.

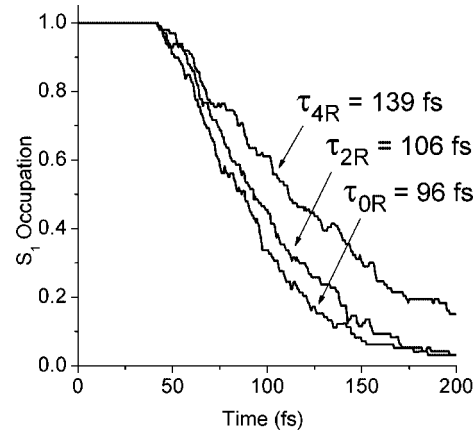


Figure 7. Average occupations of the excited state and lifetimes of trajectories for all studied systems.

The lifetimes $\tau = t_d + t_e$ were obtained by fitting the S_1-state occupation with the function $f(t) = \exp(-(t-t_d)/t_e)$ for all studied systems. The systematic increase of the lifetimes (96, 106, and 139 fs for no, two, and four restrictions, respectively) is consistent with the growing number of restrictions, which makes it increasingly difficult to reach the conical intersection. The change from no to two restrictions has a rather small influence on the lifetime, whereas the change from two to four restrictions is more significant. Currently obtained lifetimes agree well with our recent results[26] and fit well into the range of previous theoretical (80−200 fs)[24,27,33–35] and experimental (100−300 fs)[10,12,14,82,83] findings on Rh and bR and their models.

6. Conclusions

In this work the results of ab initio surface-hopping dynamics simulations for the *trans*-PSB3 cation have been presented using two sets of mechanical restrictions. In the first set the masses of two terminal hydrogen atoms located at opposite ends of the PSB3 chain were kept fixed. In the second set all four terminal hydrogen atoms were restricted. Our main goal has been to analyze and characterize the actual motions occurring in the dynamics of the restricted PSB3 system and compare them to different idealized mechanisms (one-bond-flip (OBF), bicycle-pedal (BP), nonrigid bicycle-pedal (NRBP), hula-twist (HT), and folding-table (FT)).

At the beginning, a survey of important sections of the ground- and excited-state energy surfaces for *cis*- and *trans*-PSB3 and PSB4 including conical intersections has been performed showing that relatively small basis sets can reproduce essential features of these surfaces and the nonadiabatic dynamics on them quite well. This finding reduces the computational cost of the on-the-fly dynamics approach substantially allowing the computation of a relatively large number of trajectories (300 in total) under several conditions. Thus, at least a reliable global picture of the processes occurring in the dynamics of PSB3 system can be given in this way.

The analysis of trajectories shows that for the first stage of the dynamics (the motion leading to the conical intersection) the nonrestricted and doubly restricted cases are dominated by the OBF mechanism. This mechanism, however, decreases in importance with the increase of the number of restricted atoms. In the dynamics with two and four restrictions, a mechanism not previously reported in the literature termed folding-table (FT) was observed. This mechanism is related to the OBF and consists of the torsion around the central bond together with simultaneous partial torsions (about 45°) around the two other formal double bonds. Although the FT mechanism is statistically not important for the doubly restricted case, it dominates in the case of four restrictions. Besides OBF and FT no other of the standard mechanisms were observed. In the doubly restricted system, a substantial number of events (14%) was found that connects the central torsion with a torsion at the neighboring bond by slightly more than 30°.

The second stage of the dynamics—the continuation after hopping to the ground state—gave interesting information on the continuation or reversal of the process leading to the conical intersection and on the possibility of product formation. A major percentage of the trajectories continues the motion which has led to the conical intersection. For no and two restrictions, in a large fraction of cases the motion is even continued up to the formation of the *cis* product for OBF and FT. This drastically changes for the four-restricted case where even the initial continuation of the motion leads only in very few cases to the *cis* product. After having reached a specific structure, the vibrationally hot motion continues, and at the end of the dynamics simulation the originally observed structure was lost again. This happens especially in the unrestricted case. The degree of restriction on the terminal hydrogen atoms has an influence on lifetimes as well. The change from no to two restrictions is rather small, whereas the change from two to four restrictions is more significant.

The application of the classification of torsional motions along a chain of conjugated π-bonds is general and straightforward. They allow easy automation and analysis of dynamics runs without significant human interference and should be applicable to other cases beyond the present application to protonated Schiff bases.

Acknowledgment. This work was supported by the Austrian Science Fund within the framework of the Special Research Program F16 (Advanced Light Sources) and Project P18411-N19. We are grateful for technical support and computer time at the Linux PC cluster Schrödinger III of the computer center of the University of Vienna and Wroclaw Centre for Networking and Supercomputing at Wroclaw University of Technology.

Supporting Information Available: Cartesian coordinates of the optimized geometries (minima and MXS structures) used in this study for all methods applied and vertical excitation energies and energies of key points for both isomers of PSB4. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Wald, G. *Science* **1968**, *162*, 230–&.

(2) Oesterhelt, D.; Stoeckenius, W. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 2853–2857.

(3) Birge, R. R. *Biochim. Biophys. Acta* **1990**, *1016*, 293 327.

(4) Palings, I.; Pardoen, J. A.; Vandenberg, E.; Winkel, C.; Lugtenburg, J.; Mathies, R. A. *Biochemistry* **1987**, *26*, 2544–2556.

(5) Birge, R. R. *Annu. Rev. Biophys. Bioeng.* **1981**, *10*, 315–354.

(6) Schoenlein, R. W.; Peteanu, L. A.; Mathies, R. A.; Shank, C. V. *Science* **1991**, *254*, 412–415.

(7) Wang, Q.; Schoenlein, R. W.; Peteanu, L. A.; Mathies, R. A.; Shank, C. V. *Science* **1994**, *266*, 422–424.

(8) Haupts, U.; Tittor, J.; Oesterhelt, D. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 367–399.

(9) Neutze, R.; Pebay-Peyroula, E.; Edman, K.; Royant, A.; Navarro, J.; Landau, E. M. *Biochim. Biophys. Acta-Biomembranes* **2002**, *1565*, 144–167.

(10) Kukura, P.; McCamant, D. W.; Yoon, S.; Wandschneider, D. B.; Mathies, R. A. *Science* **2005**, *310*, 1006–1009.

(11) Schenkl, S.; van Mourik, F.; Friedman, N.; Sheves, M.; Schlesinger, R.; Haacke, S.; Chergui, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4101–4106.

(12) Kobayashi, T.; Saito, T.; Ohtani, H. *Nature* **2001**, *414*, 531–534.

(13) Kukura, P.; McCamant, D. W.; Mathies, R. A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 461–488.

(14) McCamant, D. W.; Kukura, P.; Mathies, R. A. *J. Phys. Chem. B* **2005**, *109*, 10449–10457.

(15) Terentis, A. C.; Ujj, L.; Abramczyk, H.; Atkinson, G. H. *Chem. Phys.* **2005**, *313*, 51–62.

(16) Atkinson, G. H.; Ujj, L.; Zhou, Y. D. *J. Phys. Chem. A* **2000**, *104*, 4130–4139.

(17) Jager, F.; Ujj, L.; Atkinson, G. H.; Sheves, M.; Livnah, N.; Ottolenghi, M. *J. Phys. Chem.* **1996**, *100*, 12066–12075.

(18) Ujj, L.; Volodin, B. L.; Popp, A.; Delaney, J. K.; Atkinson, G. H. *Chem. Phys.* **1994**, *182*, 291–311.

(19) Mathies, R. A.; Lin, S.; Fodor, S. P. A.; Pollard, W. T.; Ames, J.; Gebhard, R.; Vandenberg, E.; Winkel, C.; Lugtenburg, J. *Biophys. J.* **1988**, *53*, A245—A245.

(20) Garavelli, M.; Negri, F.; Olivucci, M. *J. Am. Chem. Soc.* **1999**, *121*, 1023–1029.

(21) Send, R.; Sundholm, D. *J. Phys. Chem. A* **2007**, *111*, 8766–8773.

(22) Vreven, T.; Bernardi, F.; Garavelli, M.; Olivucci, M.; Robb, M. A.; Schlegel, H. B. *J. Am. Chem. Soc.* **1997**, *119*, 12687–12688.

(23) Weingart, O.; Schapiro, I.; Buss, V. *J. Phys. Chem. B* **2007**, *111*, 3782–3788.

(24) Frutos, L. M.; Andruniow, T.; Santoro, F.; Ferre, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7764–7769.

(25) Barbatti, M.; Granucci, G.; Persico, M.; Ruckenbauer, M.; Vazdar, M.; Eckert-Maksic, M.; Lischka, H. *J. Photochem. Photobiol. A: Chem.* **2007**, *190*, 228.

(26) Barbatti, M.; Ruckenbauer, M.; Szymczak, J. J.; Aquino, A. J. A.; Lischka, H. *Phys. Chem. Chem. Phys.* **2008**, *10*, 482–494.

(27) Warshel, A. *Nature* **1976**, *260*, 679–683.

(28) Sumita, M.; Saito, K. *Chem. Phys. Lett.* **2006**, *424*, 374–378.

(29) Migani, A.; Robb, M. A.; Olivucci, M. *J. Am. Chem. Soc.* **2003**, *125*, 2804–2808.

(30) Cembran, A.; Bernardi, F.; Olivucci, M.; Garavelli, M. *J. Am. Chem. Soc.* **2004**, *126*, 16018–16037.

(31) Migani, A.; Sinicropi, A.; Ferre, N.; Cembran, A.; Garavelli, M.; Olivucci, M. *Faraday Discuss.* **2004**, *127*, 179–191.

(32) Birge, R. R.; Hubbard, L. M. *Biophys. J.* **1981**, *34*, 517–534.

(33) Weingart, O.; Schapiro, I.; Buss, V. *J. Mol. Model.* **2006**, *12*, 713–721.

(34) Saam, J.; Tajkhorshid, E.; Hayashi, S.; Schulten, K. *Biophys. J.* **2002**, *83*, 3097–3112.

(35) Warshel, A.; Chu, Z. T. *J. Phys. Chem. B* **2001**, *105*, 9857–9871.

(36) Wanko, M.; Hoffmann, M.; Strodet, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(37) Aquino, A. J. A.; Barbatti, M.; Lischka, H. *ChemPhysChem* **2006**, *7*, 2089–2096.

(38) Burghardt, I.; Cederbaum, L. S.; Hynes, J. T. *Faraday Discuss.* **2004**, *127*, 395–411.

(39) Rohrig, U. F.; Guidoni, L.; Laio, A.; Frank, I.; Rothlisberger, U. *J. Am. Chem. Soc.* **2004**, *126*, 15328–15329.

(40) Rohrig, U. F.; Guidoni, L.; Rothlisberger, U. *ChemPhysChem* **2005**, *6*, 1836–1847.

(41) Sekharan, S.; Sugihara, M.; Buss, V. *Angew. Chem., Int. Ed.* **2007**, *46*, 269–271.

(42) Hoffmann, M.; Wanko, M.; Strodel, P.; Konig, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Soc.* **2006**, *128*, 10808–10818.

(43) Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2003**, *85*, 1440–1449.

(44) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(45) Barbatti, M.; Lischka, H. *J. Phys. Chem. A* **2007**, *111*, 2852–2858.

(46) Barbatti, M.; Lischka, H. *J. Am. Chem. Soc.* **2008**, in press, DOI:10.1021/ja800589p.

(47) Migani, A.; Olivucci, M. *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy*; World Scientific Publishing Company: 2004.

(48) Warshel, A.; Barboy, N. *J. Am. Chem. Soc.* **1982**, *104*, 1469–1476.

(49) Liu, R. S. H.; Asato, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 259–263.

(50) Ruiz, D. S.; Cembran, A.; Garavelli, M.; Olivucci, M.; Fuss, W. *Photochem. Photobiol.* **2002**, *76*, 622–633.

(51) Norton, J. E.; Houk, K. N. *Mol. Phys.* **2006**, *104*, 993–1008.

(52) Logunov, I.; Schulten, K. *J. Am. Chem. Soc.* **1996**, *118*, 9727–9735.

(53) Weingart, O.; Buss, V.; Robb, M. A. *Phase Transitions* **2005**, *78*, 17–24.

(54) Garavelli, M.; Celani, P.; Bernardi, F.; Robb, M. A.; Olivucci, M. *J. Am. Chem. Soc.* **1997**, *119*, 6891–6901.

(55) Weingart, O.; Migani, A.; Olivucci, M.; Robb, M. A.; Buss, V.; Hunt, P. *J. Phys. Chem. A* **2004**, *108*, 4685–4693.

(56) Bunge, A. *J. Chem. Phys.* **1970**, *53*, 20–28.

(57) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(58) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939–947.

(59) Tully, J. C. *J. Chem. Phys.* **1990**, *93*, 1061–1071.

(60) Tully, J. C. *Faraday Discuss.* **1998**, *110*, 407–419.

(61) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637–649.

(62) Butcher, J. *J. Assoc. Comput. Mach.* **1965**, *12*, 124–135.

(63) Granucci, G.; Persico, M. *J. Chem. Phys.* **2007**, *126*, 134114–11.

(64) Hammes-Schiffer, S.; Tully, J. C. *J. Chem. Phys.* **1994**, *101*, 4657–4667.

(65) Lischka, H.; Shepard, R.; Brown, F. B.; Shavitt, I. *Int. J. Quantum Chem.* **1981**, *S.15*, 91–100.

(66) Lischka, H.; Shepard, R.; Pitzer, R. M.; Shavitt, I.; Dallos, M.; Muller, T.; Szalay, P. G.; Seth, M.; Kedziora, G. S.; Yabushita, S.; Zhang, Z. Y. *Phys. Chem. Chem. Phys.* **2001**, *3*, 664–673.

(67) Lischka, H.; Shepard, R.; Shavitt, I.; Pitzer, R. M.; Dallos, M.; Mueller, T.; Szalay, P. G.; Brown, F. B.; Ahlrichs, R.; Boehm, H. J.; Chang, A.; Comeau, D. C.; Gdanitz, R.; Dachsel, H.; Ehrhardt, C.; Ernzerhof, M.; Hoechtl, P.; Irle, S.; Kedziora, G.; Kovar, T.; Parasuk, V.; Pepper, M. J. M.; Scharf, P.; Schiffer, H.; Schindler, M.; Schueler, M.; Seth, M.; Stahlberg, E. A.; Zhao, J.-G.; Yabushita, S.; Zhang, Z.; Barbatti, M.; Matsika, S.; Schuurmann, M.; Yarkony, D. R.; Brozell, S. R.; Beck, E. V.; Blaudeau, J.-P. *COLUMBUS, an ab initio electronic structure program, release 5.9.1*; 2006.www.univie.ac.at/columbus.

(68) Shepard, R.; Lischka, H.; Szalay, P. G.; Kovar, T.; Ernzerhof, M. *J. Chem. Phys.* **1992**, *96*, 2085–2098.

(69) Shepard, R. In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific: Singapore, 1995; Vol. 1, p 345.

(70) Lischka, H.; Dallos, M.; Shepard, R. *Mol. Phys.* **2002**, *100*, 1647–1658.

(71) Lischka, H.; Dallos, M.; Szalay, P. G.; Yarkony, D. R.; Shepard, R. *J. Chem. Phys.* **2004**, *120*, 7322–7329.

(72) Dallos, M.; Lischka, H.; Shepard, R.; Yarkony, D. R.; Szalay, P. G. *J. Chem. Phys.* **2004**, *120*, 7330–7339.

(73) Barbatti, M.; Granucci, G.; Lischka, H.; Ruckenbauer, M.; Persico, M. *NEWTON-X: a package for Newtonian dynamics close to the crossing seam, version 0.14b*; 2007.. www.univie.ac.at/newtonx.

(74) Worth, G. A.; Hunt, P.; Robb, M. A. *J. Phy. Chem. A* **2003**, *107*, 621–631.

(75) Groenhof, G.; Bouxin-Cademartory, M.; Hess, B.; deVisser, S. P.; Berendsen, H. J. C.; Olivucci, M.; Mark, A. E.; Robb, M. A. *J. Am. Chem. Soc.* **2004**, *126*, 4228–4233.

(76) Groenhof, G.; Schafer, L. V.; Boggio-Pasqua, M.; Goette, M.; Grubmuller, H.; Robb, M. A. *J. Am. Chem. Soc.* **2007**, *129*, 6812–6819.

(77) Fogarasi, G.; Zhou, X. F.; Taylor, P. W.; Pulay, P. *J. Am. Chem. Soc.* **1992**, *114*, 8191–8201.

(78) Vreven, D. L.; Welch, L.; Reichel, F. D. *Invest. Ophthalmol. Visual Sci.* **1997**, *38*, 4221–4221.

(79) González-Luque, R.; Garavelli, M.; Bernardi, F.; Merchán, M.; Robb, M. A.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9379–9384.

(80) Sinicropi, A.; Migani, A.; De Vico, L.; Olivucci, M. *Photochem. Photobiol. Sci.* **2003**, *2*, 1250–1255.

(81) Govindjee, R.; Balashov, S. P.; Ebrey, T. G. *Biophys. J.* **1990**, *58*, 597–608.

(82) Mathies, R. A.; Cruz, C. H. B.; Pollard, W. T.; Shank, C. V. *Science* **1988**, *240*, 777–779.

(83) Kandori, H.; Furutani, Y.; Nishimura, S.; Shichida, Y.; Chosrowjan, H.; Shibata, Y.; Mataga, N. *Chem. Phys. Lett.* **2001**, *334*, 271–276.

CT800148N

JCTC Journal of Chemical Theory and Computation

A Combined QM/MM Poisson−Boltzmann Approach

Seth A. Hayik, Ning Liao, and Kenneth M. Merz, Jr.*

Department of Chemistry, Quantum Theory Project, University of Florida, P.O. Box 118435, Gainesville, Florida 32611-8435

Received September 24, 2007

Abstract: A method of solving the mixed quantum mechanical/molecular mechanical (QM/MM) Hamiltonian in solution, using the Poisson−Boltzmann (PB) equation to calculate partial charges and solvation free energies, is presented. This method combines a linear scaling divide and conquer semiempirical algorithm with the PB equation in a QM/MM framework, allowing only a specified region's charges to be polarized by the solvent while using fixed charges from a MM force field for the remaining system. This can save time over a full QM implementation, only requiring self-consistency to be achieved in a small QM region, while giving comparable results. The solvation free energy of pentapeptides capped with an acetyl group (ACE) at the N-terminus and an N-methylamine group at the C-terminus (NME) was used to study the accuracy of this method as well as several small protein systems. The solvation free energies for the QM/MM implementation compare well with a full QM treatment of the same system, giving reasonable representations for the solvation free energy of the entire system independent of the QM region's size. In the case of the pentapeptides, the average error was only 4.9 kcal/mol with the smallest QM region. This mixed method will allow an accurate description of solvation effects in an area of interest, such as an active site, using mixed QM/MM Hamiltonians. Possible applications for this method include protein−ligand binding and reaction mechanism studies.

Introduction

The effect of solvent on biological molecules is known to be very significant,[1–5] and in order to model biomacromolecules accurately one has to include this effect to properly account for electrostatic interactions. These solvation effects are often necessary in molecular dynamics simulations and protein ligand binding studies to get accurate free energies of macromolecular systems where a difference of 1 or 2 kcal/mol is substantial.[6–8] Computationally, the effect of solvent can be included explicitly, by including water molecules surrounding the solute,[9] or implicitly by adding terms to the Hamiltonian using a continuum method.[10] Explicitly modeling solvent is expensive because it adds more atoms to the system, whose numbers increase greatly with a protein's overall size and typically requires an extended equilibration of the system. To counter this potential increase in cost and sampling, the generalized Born[11] method (GB) is sometimes used in QM/MM and pure MM studies to provide an estimate of solvent effects on the system[12–15] with a lower cost than some other more accurate methods, such as the Poisson−Boltzmann approach. Inclusion of solvation effects allows a more accurate representation of the system, since most biological systems do not exist in a vacuum. Both explicit and implicit methods have been proven to be useful, and each has their own strengths and weaknesses and are essential for accurately modeling biomolecules.[1–5]

Continuum solvation methods were designed to cheaply and accurately approximate the statistical sampling of solvent configurations at the interface between a solute and solvent.[10,16–18] There are many continuum solvent models available, but this paper will focus on the Poisson−Boltzmann (PB) finite difference method as originally implemented in Delphi by Nicholls and Honig,[19] extended to QM methods by Friesner, Goddard, Honig, and co-workers,[20] adapted for the linear scaling semiempirical program DivCon by Gogonea and Merz,[2] and further elaborated on by Liao and Merz.[21] This method was chosen because it is suitable for use on biological macromolecules in a quantum mechanical frame-

* Corresponding author e-mail: merz@qtp.ufl.edu.

10.1021/ct700245a CCC: $40.75
Published on Web 07/01/2008

work both in terms of cost and accuracy.[2] These methods have been effectively used to account for solvation in previous studies of binding free energy of small molecules and provide an accurate estimate for solvation free energies.[6,22,23] The PB method is an efficient solvation model, making it useful in calculations involving large macromolecules, and has been shown to give good solvation energies when used with a quantum mechanical method.[20,24,25] Combined with DivCon's linear scaling QM capabilities the PB method presents an efficient and accurate way to calculate the solvation free energy of an entire macromolecular system with a reasonable amount of time invested in the calculation.

The mixed quantum mechanical(QM)/molecular mechanical(MM) method was originally formulated in 1976 by Warshel and Levitt who were exploring the catalytic mechanism of lysozyme.[26] Interest in the QM/MM approach was revitalized in 1990 via application and validation work reported by Field, Bash, and Karplus on the basic QM/MM strategy and its application to chemical problems.[27] Mixed QM/MM methods are widely used now in a number of programs such as AMBER[28] and CHARMM[29] and can help reduce the cost of large calculations while allowing important areas of the protein to be treated with a QM Hamiltonian.[26,27,30] This has proven extremely useful in the study of reaction mechanisms when incorporated with molecular dynamics[31–33] and also for protein ligand studies in some cases.[22,34] Another advantage of these methods is that they can be used to save parametrization time in the case of nonstandard residues such as metal ions or drug molecules. In these instances the nonstandard residue can be included in the QM region so that the step of generating accurate force field parameters for the residue(s) in questions does not have to be undertaken. A third advantage of a mixed QM/MM method is that the QM region can properly represent the polarization effects of the surrounding MM and solvent regions through the QM-MM interactions. This polarization changes the QM region's electronic properties and can have a large impact on a reaction pathway or ligand binding.[34]

In our QM/MM implementation the linear scaling semiempirical QM program DivCon[35,36] has been integrated into AMBER.[28] This allows the solvation energy for an entire protein to be calculated utilizing atomic charges from semiempirical calculations in the QM region and fixed MM point charges in the surrounding protein. These can be Mulliken, CM1,[37] or CM2[38] charges, but due to accuracy CM1 or CM2 are typically used in DivCon calculations. This results in a method that can predict solvation free energies at a reduced cost compared to full QM calculations by centering specifically on a region of interest, an active site for example, for the expensive calculations and using the less expensive MM force calculations and charges on the rest of the protein. Our approach also takes advantage of DivCon's linear scaling, which allows a large QM region to be chosen in a QM/MM calculation without becoming exceedingly computationally expensive.

Methods

QM/MM Implementation. In the QM/MM method the entire system is first divided into two subsystems. One region (QM) contains a limited number of atoms to be treated with the chosen QM potential, while the second region (MM) is treated with an appropriate force field. In a protein, covalent bonds are cut at the interface between these two regions, and these bonds are treated by the addition of a hydrogen link atom to the QM region in order to ensure a closed shell calculation. This is a widely used method and is reliable given proper placement of the link atom away from polarized bonds.[30] In the AMBER implementation the interactions with atoms in the MM region are included, except those bonded to a QM atom (called the MM link pair), so that artificially high forces and polarization effects are not introduced into the calculation. For MM atoms directly bonded to QM atoms, electrostatic interactions are replaced by those of the link atom and the VDW interactions remain unchanged. A representative view of the QM region cut is shown in Figure 1. In our applications the peptide bonds are included in the QM region to avoid cutting these polarizable bonds.

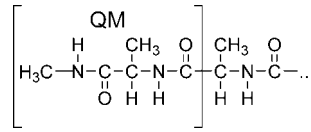


Figure 1. Representative view of the QM/MM region and its cut off points. This also shows an example of the "1 peptide" designation used in Table 1 (ACE cap plus one Ala residue).

The Hamiltonian of the system is divided into three terms describing the different interactions and is written as

$$H = H_{QM} + H_{MM} + H_{QM/MM} \tag{1}$$

where H_{QM} is the QM Hamiltonian, H_{MM} is the MM Hamiltonian, and $H_{QM/MM}$ is the Hamiltonian describing the interactions between the two subsystems. This Hamiltonian allows the QM region to interact with its surroundings, which allows the MM region to affect the electronic structure of the QM region, thereby, affecting reaction barrier heights, binding free energies, and many other properties.

The Divide and Conquer (D&C) Algorithm. In our implementation the linear-scaling D&C algorithm can be used to solve a semiempirical Hamiltonian for the large molecules of interest.[35,36] In this method the QM region is divided into subsystems, a Fock matrix for the entire system is built, and then each subsystem's Fock matrix is individually diagonalized giving the equation

$$F^{\alpha}C^{\alpha} = C^{\alpha}E^{\alpha}, \alpha = 1, ..., n_{sub} \tag{2}$$

This leads to a global density matrix, $P\mu\nu^{total}$, that is a sum over the density matrix of the subsystems, $P^{\alpha}_{\mu\nu}$

$$P^{total}_{\mu\nu} = \sum_{\alpha} P^{\alpha}_{\mu\nu} \tag{3}$$

where

$$P^{\alpha}_{\mu\nu} = D^{\alpha}_{\mu\nu} \sum_{i=1}^{M^{\alpha}} n^{\alpha}_{i} (c^{\alpha}_{\mu i}) * c^{\alpha}_{\nu i} \tag{4}$$

In eq 4 $D^{\alpha}_{\mu\nu}$ is equal to 0 or $1/n_{\mu\nu}$, where n is the number of times a basis function appears in a subsystem. With the

Table 1. Solvation Free Energy of Pentapeptides with Varying QM Region Size[a]

residue	DivCon	5 peptides	4 peptides	3 peptides	2 peptides	1 peptide
ALA	−30.5	−30.4	−29.0	−28.8	−28.7	−28.6
ARG	−572.6	−572.6	−572.9	−572.0	−570.2	−560.1
ASN	−55.9	−55.9	−55.7	−57.7	−58.9	−61.0
ASP	−698.2	−698.4	−694.2	−695.0	−698.0	−701.6
CYS	−29.1	−29.1	−29.0	−30.5	−31.1	−32.5
GLU	−653.9	−653.9	−652.2	−652.9	−653.3	−653.5
GLN	−69.3	−69.4	−71.5	−72.3	−70.5	−71.9
GLY	−39.1	−39.1	−37.7	−37.1	−36.5	−35.5
HIS	−81.0	−80.9	−74.5	−70.8	−67.9	−64.3
ILE	−24.9	−24.9	−24.9	−25.0	−24.1	−24.3
LEU	−24.2	−24.2	−24.1	−24.7	−24.4	−24.8
LYS	−624.1	−624.0	−619.6	−616.8	−607.4	−602.9
MET	−34.7	−34.6	−35.0	−33.9	−34.3	−34.3
PHE	−32.6	−32.6	−32.6	−32.7	−33.0	−32.0
PRO	−39.5	−39.4	−36.8	−34.1	−30.0	−27.4
SER	−41.1	−41.2	−39.6	−41.5	−43.3	−45.2
THR	−52.6	−52.6	−53.5	−51.5	−52.0	−52.2
TRP	−53.4	−53.4	−50.8	−52.3	−49.8	−48.3
TYR	−46.4	−46.5	−46.9	−47.6	−47.4	−48.4
VAL	−22.1	−22.1	−22.8	−21.9	−22.9	−23.7

[a] All energies in kcal/mol. Number of peptides refers to number of peptides included in the QM region. For example with glycine, 5 peptides refers to 5 glycine and the two capping residues (ACE and NME) in the QM region. Four peptides refers to 4 peptides and the ACE capping residue. "DivCon" refers to an AM1 calculation on the entire pentapeptide done in DivCon.

density matrix in hand we can directly calculate the energy using standard expressions. Our implementation of the divide and conquer algorithm is explained in detail in the extant literature,[35,36,39] hence, we will focus on the specifics of the QM/MM Poisson−Boltzmann implementation.

Poisson−Boltzmann Combined with DivCon. Implicit solvent models, like those based on the Poisson−Boltzmann approach, involve a trade off between accuracy and computational efficiency. In the PB model, the system is divided into regions with different dielectric constants, separated by the solvent accessible surface of the solute molecule. The solute is then described as a collection of discrete point charges, which are calculated by the QM/MM method. The potential of the electrostatic field (called the reaction field) generated by the polarized solvent due to the introduction of the solute is obtained by solving the Poisson−Boltzmann equation

$$\nabla[\varepsilon(\mathrm{r})\varphi\,\nabla(\mathrm{r})] - \kappa(\mathrm{r})2\sinh(\varphi(\mathrm{r})) = -4\rho\pi(\mathrm{r}) \qquad (5)$$

where $\rho(\mathrm{r})$ is the charge distribution of the solute, calculated either via semiempirical QM methods using Mulliken or CM1[37]/CM2[38] charge models for the QM region or carried over from the AMBER force field used for the MM region. $\phi(\mathrm{r})$ is the electrostatic potential that needs to be determined, and $\varepsilon(\mathrm{r})$ is the dielectric constant distribution in space, which is normally defined as ε_{out} for the solvent and ε_{in} for the solute. $\kappa(\mathrm{r})$ is a modified Debye−Huckel parameter that reflects the salt concentration and temperature.

The PB solver in our semiempirical QM program, DivCon, uses the Finite Difference Method (FDM)[40−42] to solve the PB equation. After solving the PB equation, the electrostatic potential is added into the QM Hamiltonian used in the treatment of the solute. In this way, the polarization of the solute by the reaction field is incorporated into the QM calculation. The polarized solute now yields a new set of atomic point charges, which will "repolarize" the surrounding solvent generating a new electrostatic potential upon solution of the PB equation again. This iterative Self Consistent Reaction Field (SCRF) calculation is continued until convergence is reached. Upon convergence the electrostatic potential is used to calculate the electrostatic or polar part of the solvation free energy, while the nonpolar part is calculated from the solvent accessible surface area of the solute. A more detailed discussion of this method and its use within DivCon will be reported on in a future publication.[21]

Solvation Free Energy in a QM/MM Calculation. A QM/MM implementation of the PB model within DivCon allows specific regions of the protein to be polarized by solvent dynamically, while ignoring the solvent's influence on regions of lesser importance. The implementation described is briefly mentioned by Murphy et al.[43] but is further elaborated upon and applied to our own program in the following. A similar method using DFT methods is discussed by Li et al.[44] where the protein and solvent are continua. They explore the redox potential of two-sulfur, two-iron clusters and find that including the protein field with the solvent makes a large contribution toward the accurate description of the potential in the QM region. The QM/MM PB implementation described herein will, in general, reduce the overall computational expense versus a full QM treatment, because the Hartree−Fock equations are solved for a much smaller region. The current implementation also presents the opportunity to use the Divide and Conquer linear scaling semiempirical method for the calculation of both the protein and solvent while including the effects of all of the MM atoms on the QM region.

After the normal QM/MM setup of the protein system, the point charges of the protein are built up from the QM and MM regions. The MM charges are treated as fixed atomic point charges and do not fluctuate during the SCRF process. These are obtained from the AMBER atom types and passed into DivCon through the QM/MM interface. The QM charges are calculated from the density matrix of the

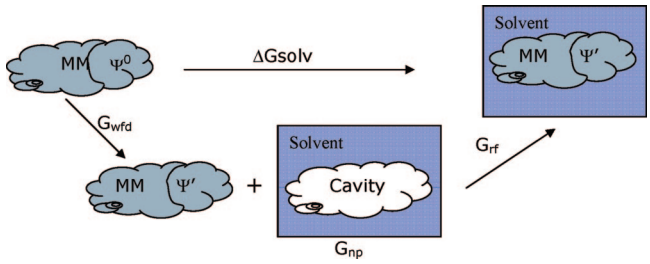


Figure 2. The terms involved in obtaining the free energy of solvation.

QM region using AM1[45] and are updated every SCRF cycle. DivCon allows the use of Mulliken, CM1,[37] or CM2[38] charges.

The entire collection of atomic point charges, from both the QM and MM regions, are then passed to the PB solver in DivCon, together with the coordinates of the protein. The dielectric constants are assigned based on the grid point location either in the solvent or protein as determined by the solvent accessible surface of the protein. The solver then solves the PB equation numerically based on the finite difference method and obtains the electrostatic potential. The resulting electrostatic potential is used to generate a set of virtual surface charges located on the solvent accessible surface of the protein. This set of virtual surface charges represents the electrostatic field (reaction field) produced by the polarized solvent. The Hamiltonian of the QM region in vacuum (H^0) is then perturbed by the interaction of the virtual surface charges with the solute electrons and nuclei and integrated over the solvent accessible surface of the protein, giving eq 6, the new potential energy operator. In this equation, $\sigma(r')$ is the virtual surface charge located at r', and a' is an area of the solute surface (S):

$$H = H^0 + \int_s da' \frac{\sigma(r')}{|r - r'|} \tag{6}$$

The perturbed Hamiltonian results in a new set of atomic point charges in the QM region, which reflects the polarization of the QM solute. This updated set of QM charges, together with the fixed MM charges, is then fed into the PB solver for another cycle of the SCRF calculation and this continues until the polarization between solute and solvent achieves convergence. The converged reaction field (represented by the virtual surface charges) and solute charge distribution (wave functions in the QM representation) are used to calculate the polar part of the solvation free energy (G_{rf}):[2]

$$G_{rf} = \frac{1}{2}\int_v dr\rho(r)\int_s dr' \frac{\sigma(r')}{|r - r'|} \tag{7}$$

A schematic view of a solvation free energy calculation, along with the terms represented in the above equations, is illustrated in Figure 2. During the dissolution of the protein into solvent, the solvent is polarized and generates a reaction field. The solute is also polarized by this reaction field, hence, the wave function of the QM portion (MM portion is fixed) is modified. We describe the energy difference between the two states (the vacuum and dissolved states) of the solute via the wave function distortion (wfd) term given by G_{wfd}. The electrostatic interaction between the reaction field and the dissolved state of the solute is called the reaction field (rf) energy (G_{rf}). The G_{wfd} and G_{rf} constitute the polar part of the solvation free energy. There is also an entropy penalty associated with making a cavity in the solvent, which is usually combined together with the nonpolar (np) interaction between the solute and solvent. These two terms are both proportional to the solvent accessible surface area and are lumped together in the G_{np} term. Further technical details of how to calculate these terms are given in ref 2.

In summary, the QM/MM PB method makes it possible to study the protein polarized by solvent only in the area of interest without the expense of having to perform the SCRF calculations on the entire protein while still yielding a reliable solvation free energy for the entire macromolecule.

Results and Discussion

We evaluated the QM/MM PB method by comparing QM/MM calculations of various sizes of the QM region with full QM SCRF calculations. The Poisson−Boltzmann solver within our semiempirical QM program, DivCon, has been described and validated,[2,21] so we will compare our QM/MM results with full QM results from this approach. Results comparing 65 small molecule's experimental solvation free energy[46,47] to calculated solvation free energy in DivCon can be found in the Supporting Information. DivCon allows the internal and external dielectric values to be set to user-defined values. For these calculations we will use the values of 1.0 for the internal dielectric (the QM and MM regions) and 80.0 for external dielectric value, the surrounding "water" environment. A grid spacing of 2.5 gridpts/Å, and a probe radius of 1.40 Å is used throughout. These parameters were found to give the best cost to performance benefit in these calculations. We use 1.0 for the internal dielectric to avoid double counting the polarization in the QM region. Since the QM atoms are already polarized by the environment through the PB approach, a higher dielectric to approximate the polarization effect, as is utilized in fixed point charge PB approaches, is unnecessary. Pentapeptides, capped with acetyl (ACE) and N-methylamine (NME) on their respective termini, were calculated, and their solvation free energy was compared to a full QM calculation using the AM1[45] Hamiltonian on the same conformation. The QM region cut was made between the carbonyl carbon and the α carbon of the next alanine or between the nitrogen and the α carbon of the previous alanine so that the peptide bond was not cut. An example of the QM/MM cut and a demonstration of the "1 peptide" designation found below can be seen in Figure 1.

These structures were minimized with 10 steps of steepest decent followed by a maximum of 4990 steps of conjugate gradient using the AMBER minimizer and the ff99 force field. All calculations were done as single point calculations at the minimized geometry, and the AM1 Hamiltonian is used throughout. This level of minimization was found to be sufficient because, on average, the rmsd between the AM1 and the molecular mechanical force field was found to be only 0.3 Å for these small systems. Moreover, since the absolute solvation free energy was of less interest than the

relative solvation free energies of the different methods, this level of minimization was deemed appropriate. The ending point of the minimization is unimportant for the purpose of this study; all we required was a low energy structure that we could make comparisons from. We are not comparing directly to the experimental conformation for this peptide, so sampling and conformational issues were not a concern. The minimized coordinates for all of the pentapeptide systems are available in the Supporting Information.

The CM1 charge model[37] was used for the partial charges in the PB calculations. In these calculations a subset of the pentapeptides was tested with the CM2 charge model,[38] and a similar trend between CM1 and CM2 solvation free energies was observed. Hence, we will only show the CM1 results. The solvation free energy of a subset of pentapeptides was also examined using the PM3 Hamiltonian, and as with the CM1/CM2 models there was no appreciable difference in performance between AM1 and PM3. Hence, the QM level we use throughout will be AM1.

Increasing QM Size. Results from increasing QM region size on the pentapeptides used above can be found in Table 1. For this, the number of atoms included in the QM region was gradually decreased so the accuracy of the QM/MM PB method could be determined. As the QM region gets smaller it is expected that the accuracy, compared to a full AM1 treatment of the entire pentapeptide, should decrease. Since only the QM atomic charges are allowed to fluctuate, the number of atoms that are allowed to polarize in the solvent field decreases, and the number of fixed charges provided by the force field increases potentially leading to less accurate solvation free energies. This means that the amount of quantum information in the SCRF calculations shrinks, and a gradual shift out of agreement with full AM1 calculations might be expected. For these systems the solvation energy remains accurate with the 1 peptide systems, with the smallest QM region, having an unsigned average difference of only 4.9 kcal/mol. Not unexpectedly, the size of the QM region has a greater impact on charged residues than aliphatic ones. Charged or highly polar residues, such as proline and histidine, are more polarizable and as these residues/atoms are assigned fixed MM charges the polarization experienced by the SCRF procedure is diminished, hence, the solvation free energy is expected to change the most as the number of QM residues is decreased. This is certainly true for the positively charged ARG and LYS residues, while the negatively charged GLU and ASP residues are affected to a lesser extent.

Table 2 shows the unsigned differences between the QM/MM calculation and the full QM DivCon solvation free energy. As expected, the accuracy of the QM/MM PB results decreases as the QM region decreases. This demonstrates the importance of the polarizable charges in the QM region against the fixed MM charges. As the MM region grows the solvation energy becomes less accurate relative to the full QM calculations but can still get pertinent information regarding the solvation free energy. Applying this to a protein could give valuable insights into active site polarization while still producing reliable solvation free energies.

Table 2. Unsigned Differences between Full QM and QM/MM Solvation Free Energy Calculations[a]

	5 peptide difference	4 peptide difference	3 peptide difference	2 peptide difference	1 peptide difference
ALA	0.0	1.5	1.6	1.7	1.9
ARG	0.0	0.2	0.6	2.4	12.6
ASN	0.0	0.3	1.7	3.0	5.1
ASP	0.1	4.0	3.2	0.2	3.4
CYS	0.0	0.1	1.4	2.0	3.4
GLU	0.0	1.8	1.1	0.7	0.4
GLN	0.0	2.2	2.9	1.1	2.5
GLY	0.0	1.4	2.0	2.6	3.6
HIS	0.1	6.6	10.2	13.1	16.7
ILE	0.0	0.0	0.1	0.8	0.6
LEU	0.1	0.1	0.4	0.1	0.5
LYS	0.0	4.4	7.3	16.6	21.2
MET	0.1	0.3	0.8	0.4	0.4
PHE	0.0	0.1	0.1	0.4	0.6
PRO	0.1	2.7	5.4	9.5	12.1
SER	0.1	1.5	0.4	2.2	4.1
THR	0.0	0.9	1.1	0.5	0.3
TRP	0.0	2.6	1.1	3.7	5.1
TYR	0.1	0.6	1.2	1.0	2.0
VAL	0.0	0.7	0.2	0.8	1.6
average	0.0	1.6	2.1	3.2	4.9

[a] Unsigned differences in solvation energy between QM/MM PB calculations and full QM PB calculations, in kcal/mol. All energies are compared to an entire pentapeptide in DivCon solvation free energy using the AM1 Hamiltonian.

Solvation Free Energies of Small Proteins. This QM/MM method is intended to be used on biomolecular systems, more specifically proteins. While the previous test shows the method's ability to perform on small systems, tests on protein systems need to be performed to gauge its efficacy on the systems of interest. In order to examine this, a series of solvation free energy calculations was done on small proteins Hepatocyte nuclear factor 1-alpha (DCoH),[48] SARS-coronavirus accessory protein (Orf7a),[49] bovine pancreatic trypsin inhibitor (BPTI),[50] and T4 lysozyme[51] containing 1284, 1034, 892, and 2603 total atoms, respectively (PDB ids: 1usm, 1xak, 4pti, 182l). First, the system was minimized within DivCon using the AM1 Hamiltonian and the LBFGS[52] method. After this the solvation free energy of the entire minimized system was calculated with the AM1 Hamiltonian. Following these calculations, a small QM region was picked and expanded in 2−3 Å increments giving QM sizes as small as 33 atoms and as large as 1734 atoms. The results of these calculations are given in Table 3 with differences between full QM solvation energy and QM/MM solvation energies shown in Table 4. The proteins give unsigned average differences from the full AM1 QM calculations of 19.13 kcal/mol, 19.67 kcal/mol, 11.02 kcal/mol, and 9.08 kcal/mol for DCoH, Orf7a, BPTI, and lysozyme, respectively. These are in good agreement with the expected solvation free energies given from full QM calculations. There is only a weak to nonexistent trend between including more atoms in the QM region and increased accuracy in the solvation free energy as might be expected for these systems. The solvation free energy differences are generally acceptable in most cases and show that even reasonably sized QM regions (less than 100 atoms) in a QM/MM calculation should be expected to give solvation free energies in good accord with full QM calculations. Also, supplied in the Supporting Information,

Table 3. Solvation Free Energy of Increasing QM Size for Four Different Protein Systems[a].

	solvation free energy			
Å (no. of QM atoms)	1usm	1xak	4pti	182l
3(33,40,62,40)	−452.62	−400.19	−780.44	−1411.92
5(59,98,107,54)	−467.23	−399.38	−779.70	−1412.87
8(187,161,176,182)	−462.01	−399.72	−805.40	−1391.85
10(236,243,308,262)	−459.19	−402.66	−802.21	−1403.46
12(421,281,373,342)	−464.64	−402.19	−777.16	−1399.31
15(696,405,490,630)	−458.71	−416.09	−818.12	−1387.74
18(959,525,621,925)	−495.69	−416.11	−804.70	−1385.38
21(1169,685,676,1290)	−456.99	−396.71	−801.81	−1401.54
24(1260,859,789,1734)	−442.26	−386.32	−798.18	−1417.58
27(913)	-	−392.52	-	-
AMBER Full QM	−443.24	−381.52	−803.50	−1401.69

[a] Solvation free energies for the different proteins examined in this study. The left column is the distance expanded from a specific atom. The values in parentheses are the number of QM atoms, not counting link atoms, in the QM/MM calculations for 1usm, 1xak, 4pti, and 182l, respectively. All energies are in kcal/mol.

Table 4. Solvation Free Energy Differences for Four Proteins[a]

Å (no. of QM atoms)	1usm	1xak	4pti	182l
3(33,40,62,40)	−9.38	−18.67	23.06	−10.23
5(59,98,107,54)	−23.99	−17.86	23.80	−11.18
8(187,161,176,182)	−18.77	−18.20	−1.90	9.84
10(236,243,308,262)	−15.95	−21.14	1.29	−1.77
12(421,281,373,342)	−21.40	−20.67	26.34	2.38
15(696,405,490,630)	−15.47	−34.57	−14.62	13.95
18(959,525,621,925)	−52.45	−34.59	−1.20	16.31
21(1169,685,676,1290)	−13.75	−15.19	1.69	0.15
24(1260,859,789,1734)	0.98	−4.80	5.32	−15.89
27(913)	-	−11.00	-	-

[a] Differences in solvation free energies for the proteins examined. The left column is the distance expanded from a specific atom in Angstroms. The values in parentheses are the number of QM atoms, not counting link atoms, in the QM/MM calculations for 1usm, 1xak, 4pti, and 182l, respectively. All energies are in kcal/mol.

timings for 1usm for both standard and divide and conquer calculations can be found.

Conclusion

We have presented a QM/MM implementation of the SCRF Poisson−Boltzmann solver available in DivCon to calculate solvation free energies of large molecules within a semiempirical framework. A full QM version of this Poisson−Boltzmann solver has been validated by Liao and Merz,[21] and we find that it is also applicable using a QM/MM approach as implemented in the Sander program of the AMBER suite of programs. This method allows the solvation free energy of an entire protein to be calculated, including only part of the system in the quantum region, thus allowing only these charges to be polarized by the solvent while keeping the remaining system at fixed charges given by the force field. Using a divide and conquer technique, a large quantum region may be selected due to DivCon's linear scaling nature allowing the link atoms to be farther from the active site than other methods may allow due to computational expense. This also allows the solvation free energy to be calculated using Mulliken, CM1, or CM2 charges for the QM region. This method might be applied, in particular, to drug design due to its ability to allow polarization and charge transfer effects in a region of interest while including solvation terms and finding the solvation free energy of a solute. It could also be used to determine pK_a and thermodynamic quantities and generate improved force fields for nonstandard ligands/residues and many other applications relevant to biological systems.

Acknowledgment. We thank the NIH (GM066859) for financial support of this research.

Supporting Information Available: Average time per SCF cycle in vacuum and solution for the 1usm protein along with the number of SCF cycles required for convergence in vacuum and solution, experimental versus calculated solvation free energies from DivCon for 39 neutral, 19 anions, and 17 cations along with the coordinates of the 20 pentapeptides used, and coordinates for all the pentapeptides. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Friesner, R. A. Modeling Polarization in Proteins and Protein-ligand Complexes: Methods and Preliminary Results. *Adv. Protein Chem.* **2005**, *72*, 79–104.

(2) Gogonea, V.; Merz, K. M. Fully quantum mechanical description of proteins in solution. Combining linear scaling quantum mechanical methodologies with the Poisson-Boltzmann equation. *J. Phys. Chem. A* **1999**, *103* (26), 5171–5188.

(3) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268* (5214), 1144–1149.

(4) Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301–32.

(5) Schaefer, P.; Riccardi, D.; Cui, Q. Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J. Chem. Phys.* **2005**, *123* (1), 014905.

(6) Simonson, T.; Archontis, G.; Karplus, M. Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* **2002**, *35* (6), 430–7.

(7) Simonson, T. Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 243–52.

(8) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. 3rd, Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47* (12), 3032–47.

(9) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(10) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78* (1−2), 1–20.

(11) 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*, 6127–6129.

(12) Bashford, D.; Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–52.

(13) Tsui, V.; Case, D. A. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **2000**, *122* (11), 2489–2498.

(14) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55* (2), 383–394.

(15) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of the Poisson-Boltzmann equation. *J. Phys. Chem. B* **1997**, *101* (7), 1190–1197.

(16) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99* (8), 2161–2200.

(17) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94* (7), 2027–2094.

(18) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.* **2002**, *15* (6), 377–92.

(19) Nicholls, A.; Honig, B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* **1991**, *12* (4), 435–445.

(20) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; William, A.; Goddard, I.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875–82.

(21) Liao, N.; Merz, K. M. Manuscript in preparation.

(22) Grater, F.; Schwarzl, S. M.; Dejaegere, A.; Fischer, S.; Smith, J. C. Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J. Phys. Chem. B* **2005**, *109* (20), 10474–83.

(23) Archontis, G.; Simonson, T.; Karplus, M. Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol.* **2001**, *306* (2), 307–27.

(24) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (18), 10037–41.

(25) Smith, P. E.; Pettitt, B. M. Modeling Solvent in Biomolecular Systems. *J. Phys. Chem.* **1994**, *98* (39), 9700–9711.

(26) Warshel, A.; Levitt, M. Theoretical studies of enzymatic reactions: Dielectric, electrostatic and steric stabilization of a carbonium ion in the reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(27) 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* (6), 700–733.

(28) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowly, M.; Walker, R. C.; Zhang, W.; Wang,B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*, University of California: San Francisco, 2006.

(29) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4* (2), 187–217.

(30) Senn, H. M.; Thiel, W. QM/MM methods for biological systems. In *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*; Reiher, M., Ed.; Springer: Berlin, Heidelberg, 2007; Vol *268*, pp 173−290.

(31) Li, G. H.; Cui, Q. Mechanochemical coupling in myosin: A theoretical analysis with molecular dynamics and combined QM/MM reaction path calculations. *J. Phys. Chem. B* **2004**, *108* (10), 3342–3357.

(32) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. Quantum chemical studies of intermediates and reaction pathways in selected enzymes and catalytic synthetic systems. *Chem. Rev.* **2004**, *104* (2), 459–508.

(33) Ranaghan, K. E.; Ridder, L.; Szefczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J. Transition state stabilization and substrate strain in enzyme catalysis: ab initio QM/MM modelling of the chorismate mutase reaction. *Org. Biomol. Chem.* **2004**, *2* (7), 968–980.

(34) Hensen, C.; Hermann, J. C.; Nam, K.; Ma, S.; Gao, J.; Holtje, H. D. A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem.* **2004**, *47* (27), 6673–80.

(35) Dixon, S. L.; Merz, K. M. Semiemprical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104* (17), 6643–6649.

(36) Dixon, S. L.; Merz, K. M. Faster, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* **1997**, *107* (3), 879–893.

(37) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class-Iv Charge Models - a New Semiempirical Approach in Quantum-Chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9* (1), 87–110.

(38) Li, J. B.; Zhu, T. H.; Cramer, C. J.; Truhlar, D. G. New class IV charge model for extracting accurate partial charges from wave functions. *J. Phys. Chem. A* **1998**, *102* (10), 1820–1831.

(39) Yang, W. T.; Lee, T. S. A Density-Matrix Divide-and-Conquer Approach for Electronic-Structure Calculations of Large Molecules. *J. Chem. Phys.* **1995**, *103* (13), 5674–5678.

(40) Davis, M. E.; Mccammon, J. A. Solving the Finite-Difference Linearized Poisson-Boltzmann Equation - a Comparison of Relaxation and Conjugate-Gradient Methods. *J. Comput. Chem.* **1989**, *10* (3), 386–391.

(41) Holst, M.; Kozack, R. E.; Saied, F.; Subramaniam, S. Treatment of electrostatic effects in proteins: multigrid-based Newton iterative method for solution of the full nonlinear Poisson-Boltzmann equation. *Proteins* **1994**, *18* (3), 231–45.

(42) Warwicker, J.; Watson, H. C. Calculation of the electric potential in the active site cleft due to α helix dipoles. *J. Mol. Biol.* **1982**, *157* (4), 671–9.

(43) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.* **2000**, *21* (16), 1442–1457.

(44) Li, J.; Nelson, M. R.; Peng, C. Y.; Bashford, D.; Noodleman, L. Incorporating protein environments in density functional theory: A self-consistent reaction field calculation of redox potentials of [2Fe2S] clusters in ferredoxin and phthalate dioxygenase reductase. *J. Phys. Chem. A* **1998**, *102* (31), 6311–6324.

(45) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(46) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. J. Group Contributions to the Thermodynamic Properties of Non-ionic Organic Solutes in Dilue Aqueous Solution. *J. Solution Chem.* **1981**, *10* (8), 563–595.

(47) Wolfenden, R.; Anderson, L.; Cullis, P. M.; Southgate, C. C. Affinities of Amino Acid Side Chains for Solvent Water. *Biochemistry* **1981**, *20*, 849–855.

(48) Tahirov, T. H.; Inagaki, E. Unpublished PDB entry 1usm. http://www.rcsb.org(accessed Oct 3, 2007).

(49) Nelson, C. A.; Lee, C. A.; Fremont, D. H.; Joachimiak, A. Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein. *Structure* **2005**, *13*, 75–85.

(50) Huber, R.; Kukla, D.; Ruehlmann, A.; Epp, O.; Formanek, H.; Deisenhofer, J.; Steigemann, W. The Geometry of the Reactive Site andof the Peptide Groups in Trypsin, Trypsinogen and its Complexes and Inhibitors. *Acta Crystallogr.* **1983**, *39*, 480.

(51) Morton, A.; Matthews, B. W. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34* (27), 8576–88.

(52) Liu, D. C.; Nocedal, J. On the limited memorty BFGS method for large scale optimization. *Math. Prog.* **1989**, *45*, 503–528.

CT700245A

JCTC Journal of Chemical Theory and Computation

Multireference Model Chemistries for Thermochemical Kinetics

Oksana Tishchenko,* Jingjing Zheng, and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received March 5, 2008

Abstract: By combining the generalized valence bond ansatz of correlated participating orbitals (CPO) with the complete-active-space prescription for selecting configurations and with the use of multireference second order perturbation theory (MRMP2) for including dynamical correlation, we define three levels of multireference (MR) theoretical model chemistries for electronic structure calculations of chemical reaction energies and barrier heights. The three levels differ in their choice of which orbitals are considered to be participating; the choices are called nominal (*nom*-CPO), moderate (*mod*-CPO), and extended (*ext*-CPO). Combining any of these three choices with a method for treatment of dynamical correlation energy and a one-electron basis set yields a theoretical model chemistry. Unlike the full-valence choice of active orbitals, the CPO choices lead to active spaces that contain the orbitals needed to include important static correlation effects on chemical reactions but do not increase with the size of the nonparticipating portion of the system, and hence they remain viable computational options even for many large and complex reacting systems. The accuracies of the new levels, combined with the MG3S basis set (a partially augmented, multiply polarized valence triple-ζ basis with appropriately tight *d* functions for 3*p*-block elements) and with the fully augmented correlation-consistent aug-cc-pVTZ basis set, are assessed against a previously presented database of barrier heights for diverse reaction types. We find that *nom*-CPO level captures the bulk of the static correlation energy, and MRMP2/*nom*-CPO calculations have an average error of only 1.4 kcal/mol in barrier heights, which may be compared to 5.0 kcal/mol for single-reference MP2 theory, 2.5 kcal/mol for CCSD, and 4.1 and 1.0 kcal/mol for the B3LYP and M06−2X density functionals, respectively. The accuracy of MRMP2/CPO for transition structure bond lengths and donor−acceptor distances is excellent, with a mean unsigned error of only 0.007 Å as compared to 0.018 Å for CCSD, 0.019 Å for M06−2X, and 0.039 Å for MP2 and B3LYP. We also introduce a new multireference diagnostic, called the *M* diagnostic, that allows one to measure the importance of static correlation in a given reagent or transition state.

1. Introduction

The concept of theoretical model chemistry[1–4] is central to much recent progress in computational chemistry. A theoretical model chemistry is "an approximate but well-defined mathematical procedure of simulation".[2] Examples of popular theoretical model chemistries include HF/4−31G,[5] MP3/6−31G(d),[6] AM1,[7] B3LYP/6−31G(d),[8] Weizmann-1,[9] G3S,[10] MPW1K/6−31+G(d,p),[11] MR-G2(MP2),[12] BMC-CCSD,[13] ROCBS-QB3,[14] and innumerable others. With few exceptions,[12] essentially all theoretical model chemistries that are based on wave functions involve single-reference methods, that is, they are post-Hartree−Fock correlation methods based on a single-configuration reference state. Multireference methods, which are based on a multiconfiguration reference state, can provide increased accuracy for treating

* Corresponding author e-mail: truhlar@umn.edu (D.G.T.), o_t@t1.chem.umn.edu (O.T.).

10.1021/ct800077r CCC: $40.75
Published on Web 07/01/2008

many open-shell systems and transition states, especially when excited configurations are nearly degenerate with the ground state.[15] However, multireference methods have not been widely used as theoretical model chemistries and have been difficult to test systematically because the best choice of a reference state is system dependent; thus the precise reference state to be used for simulating a given process is not well defined by the method. The goals of the present article are, first, to define, in a general way, three levels of reference states suitable for calculating transition states, and, second, to test multireference calculations based on these reference states against a representative database[16] of diverse barrier heights.

The three levels of reference states that we define are complete active space self-consistent-field (CASSCF)[17–21] (equivalent to the fully optimized reaction space (FORS))[22] wave functions with three different schemes for choosing the number of active electrons and the number of active orbitals. (Note that the CAS and FORS prescriptions then specify inclusion of all configuration state functions that can be created by various ways of occupying these orbitals.) The three theoretical model chemistries that we consider are all multireference Møller−Plesset second order perturbation theory[23,24] (MRMP2) based on the CASSCF reference states. The CASSCF wave function includes static correlation, with the precise amount included depending on the choice of active space, that is, on the choice of active orbitals and the number of active electrons; the perturbation treatment includes dynamical correlation energy to second order. For a given active space the performance of MRMP2 should be very similar to that of other perturbation methods based on CASSCF, such as CASPT2[25] and various other versions of multireference second order perturbation theory.[26,27] One could imagine also using the three levels of reference states that we define here with post-CASSCF methods that include dynamical correlation to a higher order, such as multireference configuration interaction theory[28,29] or multireference coupled cluster theory;[30–44] however, we will not pursue this here.

2. Multireference Model Chemistries

A CASSCF wave function is a complete-active-space configuration interaction (CASCI) wave function in which all orbitals and all configurational coefficients are simultaneously optimized. A CASCI wave function is one in which a set of "inactive" orbitals are doubly occupied in all configurations and the remaining n electrons are distributed in all possible ways, consistent with a given symmetry and spin, over a set of m "active" orbitals. For singlet states with $m = n/2$, CASSCF reduces to single-configuration Hartree−Fock, but as n and m increase with $m > n/2$, the number of configurations increases rapidly. For example, with (n/m) denoting n active electrons distributed over m active orbitals, the number of determinants in singlet (10/14), (14/14), and (14/15) calculations are[45] 1.0×10^6, 1.2×10^7, and 4.1×10^7, respectively, which puts such calculations near the limit of doability.

One scheme for selecting n and m is to set n equal to the number of valence electrons and m equal to the number of valence orbitals (e.g., the number of valence orbitals is 1 for H and 4 for B through Ne or for Al through Ar).[12,22] This is called a full-valence CAS,[46,47] and it is unaffordable for all but the smallest molecules. (For example, a full-valence CAS for propane is (20,20).) In addition, it is not necessarily the best strategy. Consider formic acid, for which a full-valence CAS is (18,14). This yields only five unoccupied orbitals to correlate nine doubly occupied orbitals. In generalized valence bond (GVB) theory, one needs one correlating orbital to correlate each doubly occupied orbital, and no correlating orbital is required for a singly occupied orbital.[48] (Note that when we call an orbital singly or doubly occupied or unoccupied, we refer to its occupancy in the dominant configuration. Since all distributions of active electrons among active orbitals are included in the configuration list, any orbital in the active space is typically doubly occupied in some configurations and singly occupied or unoccupied in others).

Based on experience and the GVB analogy, we adopt the following guiding principle: To properly correlate the orbitals in the active space, there should be one correlating (i.e., unoccupied) orbital for every doubly occupied orbital in the active space. Furthermore, to keep the size of the active space manageable, we adopt another principle: only the orbitals that participate most strongly in bond breaking and bond forming are included in the active space. The combination of these two principles is called the correlated participating orbitals (CPO) scheme.

We will identify three choices of CPO, based on how we define participating orbitals: nominal (*nom*-CPO), moderate (*mod*-CPO), and extended (*ext*-CPO). In *nom*-CPO, the participating orbitals are those that describe bonds that break or form during the reaction. In *mod*-CPO, the participating orbitals are the nominal ones plus unshared pairs in p orbitals geminal to bonds that are broken or formed. In *ext*-CPO, the participating orbitals are the moderate ones plus unshared pairs in s orbitals or hybrid orbitals that are geminal to bonds being broken or formed. (Although the orbitals are delocalized, no confusion arises with identifying the most strongly bound lone pairs with s orbitals and the less strongly bound ones with p orbitals; this is commonly used[49] language.) Thus definitions of the three levels may be more clear when we give examples below.

For participating orbitals that are σ or π bonding orbitals, we take the correlating orbitals to be the corresponding σ^* and π^* antibonding orbitals, respectively. For participating orbitals that are $2p$ lone pairs, the correlating orbital is a $2p'$ orbital: $2p'_x$ for $2p_x$, $2p'_y$ for $2p_y$, and $2p'_z$ for $2p_z$. (In this notation, an orbital with a prime indicates an orbital that occupies approximately the same space as the corresponding orbital without a prime.) For $3p$ lone pairs, we use molecular orbitals of mixed $3p'/d$ character, one for each $3p$ orbital. (These are simply labeled $3d$ in Table 1.) For participating orbitals that are $2s$ lone pairs, the correlating orbital is a $2s'$ orbital. For participating orbitals with $3s$ lone pair characters, we make an exception to the 1:1 rule, and we add any $3d$ orbitals that have not already been added. We use $3d$ orbitals rather than $3s'$ because the latter are much higher in energy.

The key point is that one uses a small number of reaction-specific orbitals that gradually transform from reactants to products and result in a smooth potential energy profile. Variationally optimizing a CASSCF wave function with a CPO scheme for prescription of the choice of active orbitals will not necessarily lead to the global minimum of the energy functional at all molecular geometries. Consider a unimolecular dissociation reaction in which a single bond is broken, that is, X-A → X + A. At those geometries where the X-A distance is significantly stretched, using the σ and σ^* orbitals of the bond being broken as the active orbitals (i.e., using the *nom*-CPO recipe) will likely lead to the global minimum of the CASSCF energy; however, at geometries close to the equilibrium geometry of the X-A bond, it most likely will not, especially if one or both molecular fragments involves π-bonds or unshared electron pairs. Therefore, care must be taken to make sure that one gets the orbitals corresponding to the chosen CPO scheme at all molecular geometries of interest; this can often be accomplished by using an appropriate initial orbital set for the SCF process. It is important to ensure that SCF process converges to orbitals that make the calculated potential energy surface continuous with continuous derivatives.

In this article we apply the CPO levels to the reactions of the DBH24 database. DBH24[16] is a representative[50] database of 24 barrier heights, consisting of the forward and reverse barrier heights ($V_f^{\ddagger}$ and $V_r^{\ddagger}$, respectively) of 12 diverse reactions. There are also 12 energies of reaction (ΔV) for these reactions. These are all defined to be zero-point exclusive; that is that they refer to differences in potential energy, not to 0 K enthalpies. In the present article, all 12 reactions are written in the exoergic or ergoneutral direction; thus $\Delta V \leq 0$, and $V_f^{\ddagger} \leq V_r^{\ddagger}$.

The present MRPT2 calculations are all nonrelativistic, but the experimental results include spin−orbit coupling, which is a relativistic effect. Thus we added the spin−orbit energy to all reagents for which it is nonzero; these values are (in kcal/mol): OH, −0.20;[51a] SH, −0.54;[51a] Cl, −0.84;[51b] thus they lower the energies of these open-shell species. The spin−orbit energies are assumed to be negligible at the saddle points, which is a general effect (for many reactions like those considered here) discussed elsewhere.[51c] All results in this article already have the nonzero spin−orbit energies of OH, SH, and Cl included.

The three levels of reference wave function yielded by the prescriptions for the three CPO choices are completely specified for the reactions in DBH24 in Table 1. This table also lists the number N_d of determinants for each case. In Table 1, the z axis coincides with the bond that breaks, and the x axis is perpendicular to the plane defined by the three nuclei that are the reaction centers (in each reaction **R1**-**R12** the three reaction centers can be identified as the key atoms that are involved in bond breaking and bond formation, e.g. O, H_t, C in **R1**, H, H_t, O in **R2**, H, H_t, S in **R3**, etc., where H_t denotes the transferred hydrogen). In definitions of participating orbitals in this table, a pipe is used to indicate an orbital at the reactant (left) and the corresponding orbital at the product (right) (only for the orbitals that change their characters during a reaction).

Reactions **R1**-**R3** are hydrogen transfer reactions; the *nom*-CPO active space for these reactions involves the bonding and antibonding orbitals for the bond that undergoes breaking/formation and the SOMO(s). Reactions **R5** and **R6** are heavy atom transfer reactions; the *nom*-CPO active space for for these reactions is constructed in the same way as for reactions **R1**-**R3**. Reactions **R7**-**R9** are S_N2 reactions of the $X^- + CH_3Y \rightarrow XCH_3 + Y^-$ type; the *nom*-CPO active space for these reactions involves the bond that undergoes breaking/formation, the orbital χ at an X/Y group with the unshared pair of electrons that donates/accepts the electrons, and a correlating orbital for the orbital χ; this correlating orbital is either a $2p'$ orbital (if χ is on a second-row atom) or a $3d_{z^2}$ orbital (if χ is on a third-row atom). Reactions **R10** and **R11** can be considered as either dissociation or addition reactions; the *nom*-CPO active space for these reactions consists of the bonding and antibonding orbital pair that describe the bond that undergoes breaking/formation (π_{NN}/σ_{NH} and π_{CC}/σ_{CH}) and the SOMO ($1s_H$). Reactions **R4** and **R12** involve a change of several bonds in the same reaction step; therefore, the *nom*-CPO active space is larger in these cases: in the case of **R4** it involves the $1s_H$ and all $2p_O$ and $2p_N$ orbitals and is the same as the *mod*-CPO active space; in the case of **R12** it involves the $1s_H$, $2s_N$, $2s_C$, $2p_N$, and $2p_C$ orbitals and is the same as the *mod*-CPO and *ext*-CPO active spaces.

3. Computational Details

Molecular orbitals were optimized by using multiconfigurational self-consistent field wave functions of the fully optimized reaction space FORS[22] type which is equivalent to CASSCF.[17–21] Dynamical electron correlation was included by using multireference second-order Møller−Plesset theory (MRMP2).[23,24] In MRMP2 calculations, all orbitals were correlated except the 1*s* orbitals on nonhydrogenic atoms. Molecular geometries of the reactants, products, and reaction saddle points were optimized using numerical MRMP2 gradients. When optimizing geometries of reactants or products we set them at least 25 Å apart.

nom-CPO calculations were carried out with two basis sets: the larger aug-cc-pVTZ basis[52,53] and the efficient MG3S basis.[54] MG3S is the same as 6-311+G(3*d*2*f*,2*df*,2*p*)[55] for H−Si and is of similar size but improved[56] for P−Ar. The sizes of these basis sets for the transition states in DBH24 are tabulated elsewhere;[57] the average numbers of basis functions for the transition states in DBH24 are 158 primitive Gaussians and 108 contracted functions for MG3S and 228 primitive functions and 165 contracted functions for aug-cc-pVTZ. For all levels of theory other than *nom*-CPO we employed the MG3S basis.

All multireference calculations were performed using GAMESS[58] codes, and all single-reference calculations were performed using *Gaussian*.[59] M05 and M06 calculations were performed with MN-GFM.[60]

4. Results

Tables 2, 3, and 4 give results for, respectively, the MRMP2/*nom*-CPO, MRMP2/*mod*-CPO, and MRMP2/*ext*-CPO model chemistries. $\Delta(V_f^{\ddagger})$ is the error in the forward barrier height, $\Delta(V_r^{\ddagger})$ is the error in the reverse barrier height, and $\Delta(\Delta V)$

Table 1. Active Orbitals and the Number of Determinants (N_d) in FORS/CASSCF Wave Functions for Reactions in DBH24

	nom-CPO		*mod*-CPO		*ext*-CPO	
reaction	MOs	N_d, n/m	MOs	N_d, n/m	MOs	N_d, n/m
R1: OH + CH_4 → CH_3 + H_2O	$\sigma_{OH}\|\sigma_{CH}$, $\sigma^*_{OH}\|\sigma^*_{CH}$, $2p_{zO}\|2p_{zC}$ (SOMO)	9 3/3	σ_{OH}, σ^*_{OH}, $2p_{zO}\|2p_{zC}$ (SOMO) $2p_{xO}$, $2p_{yO}$, $2p'_{xO}$, $2p'_{yO}$	1225 7/7	σ_{OH}, σ^*_{OH}, $2p_{zO}\|2p_{zC}$ (SOMO) $2p_{xO}$, $2p_{yO}$, $2p'_{xO}$, $2p'_{yO}$, $2s_O$, $2s'_O$	15876 9/9
R2: H + OH → O + H_2	$\sigma_{OH}\|\sigma_{HH}$, $\sigma^*_{OH}\|\sigma^*_{HH}$, $2p_{xO}$ (SOMO), $1s_H\|2p_{zO}$ (SOMO)	16 4/4	$\sigma_{OH}\|\sigma_{HH}$, $\sigma^*_{OH}\|\sigma^*_{HH}$, $2p_{xO}$ (SOMO), $1s_H\|2p_{zO}$ (SOMO), $2p_{yO}$,$2p'_{yO}$	225 6/6	$\sigma_{OH}\|\sigma_{HH}$, $\sigma^*_{OH}\|\sigma^*_{HH}$, $2p_{xO}$ (SOMO), $1s_H\|2p_{zO}$ (SOMO), $2p_{yO}$,$2p'_{yO}$, $2s'_O$, $2s_O$	3136 8/8
R3: H + H_2S → H_2 + HS	$\sigma_{SH}\|\sigma_{HH}$, $\sigma^*_{SH}\|\sigma^*_{HH}$, $1s_H\|3p_{zS}$ (SOMO)	9 3/3	$\sigma_{SH}\|\sigma_{HH}$, $\sigma^*_{SH}\|\sigma^*_{HH}$, $1s_H\|3p_{zS}$ (SOMO) $3p_{xS}$, $3p_{yS}$, $3d_{xyS}$, $3d_{y^2S}$	1225 7/7	$\sigma_{SH}\|\sigma_{HH}$, $\sigma^*_{SH}\|\sigma^*_{HH}$, $1s_H\|3p_{zS}$ (SOMO) $3p_{xS}$, $3p_{yS}$, $3d_{xyS}$, $3d_{y^2S}$, $3s_S$, $3d_{xzS}$, $3d_{yzS}$, $3d_{(x^2-z^2)S}$	152460 9/11
R4: H + N_2O → OH + N_2	$2p_{xO}$, $2p_{yO}$, $2p_{xN_a}$, $2p_{yN_a}$, $2p_{xN_b}$, $2p_{yN_b}$, $2p'_{xO}$, $2p'_yO$, $1s_H\|2p_{zO}$ (SOMO), $\sigma_{NO}\|\sigma_{NN}$, σ^*_{NO},$\|\sigma^*_{NN}$	213444 11/11	$2p_{xO}$, $2p_{yO}$, $2p_{xN_a}$, $2p_{yN_a}$, $2p_{xN_b}$, $2p_{yN_b}$, $2p'_{x}O$, $2p'_{yO}$, $1s_H\|2p_{zO}$ (SOMO), $\sigma_{NO}\|\sigma_{NN}$, σ^*_{NO},$\|\sigma^*_{NN}$	213444 11/11	$2p_{xO}$, $2p_{yO}$, $2p_{xN_a}$, $2p_{yN_a}$, $2p_{xN_b}$, $2p_{yN_b}$, $2p'_{xO}$, $2p'_{yO}$, $1s_H\|2p_{zO}$ (SOMO), $\sigma_{NO}\|\sigma_{NN}$, σ^*_{NO},$\|\sigma^*_{NN}$, $2s_{N_a}$, $2s_{N_b}$, $2s_O$, $2s'_{N_a}$, $2s'_{N_b}$, $2s'_O$	213444 11/11
R5: H + ClH → HCl + H	σ_{ClH}, σ^*_{ClH}, $1s_H$	9 3/3	σ_{ClH}, σ^*_{ClH}, $1s_H$, $3p_{yCl}$, $3d_{xyCl}$, $3d_{(x^2-y^2)Cl}$	1225 7/7	σ_{ClH}, σ^*_{ClH}, $1s_H$, $3s_{Cl}$, $3p_{xCl}$, $3p_{yCl}$, $3d_{xyCl}$, $3d_{(x^2-y^2)Cl}$ $3d_{xzCl}$, $3d_{yzCl}$, $3d_{z^2Cl}$	152460 9/11
R6: CH_3 + FCl → CH_3F + Cl	$\sigma_{FCl}\|\sigma_{CF}$, σ^*_{FCl},$\|\sigma^*_{CF}$, $2p_{zC}\|3p_{zCl}$ (SOMO)	9 3/3	$\sigma_{FCl}\|\sigma_{CF}$, σ^*_{FCl},$\|\sigma^*_{CF}$, $2p_{zC}\|3p_{zCl}$ (SOMO), $2p_{xF}$, $2p_{yF}$, $2p'_{xF}$, $2p'_{yF}$, $3p_{xCl}$, $3p_{yCl}$, $3d_{xyCl}$, $3d_{(x^2-y^2)Cl}$	213444 11/11	$\sigma_{FCl}\|\sigma_{CF}$, σ^*_{FCl},$\|\sigma^*_{CF}$, $2p_{zC}\|3p_{zCl}$ (SOMO), $2p_{xF}$, $2p_{yF}$, $2p'_{xF}$, $2p'_{yF}$, $3p_{xCl}$, $3p_{yCl}$, $3d_{xyCl}$, $3d_{(x^2-y^2)Cl}$, $2s_F$, $2's_F$, $3s_{Cl}$, $3d_{xzCl}$, $3d_{yzCl}$, $3d_{z^2Cl}$	472780880 15/17
R7: $Cl^-\cdots CH_3$ Cl → $ClCH_3\cdots Cl^-$	σ_{CCl}, σ^*_{CCl}, $3p_{zCl}$, $3d_{z^2Cl}$	36 4/4	σ_{CCl}, σ^*_{CCl}, $3p_{zCl}$, $3d_{z^2Cl}$, $3p_{xCla}$, $3p_{xClb}$, $3p_{yCla}$, $3p_{yClb}$, $3d_{xzCla}$, $3d_{xzClb}$, $3d_{yzCla}$, $3d_{yzClb}$	853776 12/12	σ_{CCl}, σ^*_{CCl}, $3p_{zCl}$, $3p_{xCla}$, $3p_{xClb}$, $3p_{yCla}$, $3p_{yClb}$, $3s_{Cla}$, $3s_{Clb}$, + 10 × d_{Cl}	5712638724 16/19
R8: $F^-\cdots CH_3Cl$ → $FCH_3\cdots Cl^-$	$\sigma_{CCl}\|\sigma_{CF}$, $\sigma^*_{CCl}\|\sigma^*_{CF}$, $2p_{zF}\|3p_{zCl}$, $3p_{zF}\|3d_{z^2Cl}$	36 4/4	$\sigma_{CCl}\|\sigma_{CF}$, $\sigma^*_{CCl}\|\sigma^*_{CF}$, $2p_{zF}\|3p_{zCl}$, $3p_{zF}\|3d_{z^2Cl}$, $3p_{xCl}$, $2p_{xF}$, $3p_{yCl}$, $2p_{yF}$, $3d_{xzCl}$, $2p'_{xF}$, $3d_{yzCl}$, $2p'_{yF}$	853776 12/12	$\sigma_{CCl}\|\sigma_{CF}$, $\sigma^*_{CCl}\|\sigma^*_{CF}$, $2p_{zF}\|3p_{zCl}$, $3p_{zF}\|3d_{z^2Cl}$, $3p_{xCl}$, $2p_{xF}$, $3p_{yCl}$, $2p_{yF}$, $3d_{xzCl}$, $2p'_{xF}$, $3d_{yzCl}$, $2p'_{yF}$ $2s_F$, $3s_{Cl}$, $3s_F$, $3d_{xyCl}$, $3d_{(x^2-y^2)Cl}$	590976100 16/17
R9: OH^- + CH_3F → $HOCH_3$ + F^-	$\sigma_{CF}\|\sigma_{CO}$, σ^*_{CF},$\|\sigma^*_{CO}$, $2p_{zO}\|2p_{zF}$, $2p'_{zO}\|2p'_{zF}$	36 4/4	$\sigma_{CF}\|\sigma_{CO}$, σ^*_{CF},$\|\sigma^*_{CO}$, $2p_{zO}\|2p_{zF}$, $2p'_{zO}\|2p'_{zF}$, $2p_{xF}$, $2p_{yF}$, $2p_{xO}$, $2p_{yO}$, $2p_{x'F}$, $2p_{y'F}$, $2p_{x'O}$, $2p_{y'O}$	853776 12/12	$\sigma_{CF}\|\sigma_{CO}$, σ^*_{CF},$\|\sigma^*_{CO}$, $2p_{zO}\|2p_{zF}$, $2p'_{zO}\|2p'_{zF}$ $2p_{xF}$, $2p_{yF}$, $2p_{xO}$, $2p_{yO}$, $2p_{x'F}$, $2p_{y'F}$, $2p_{x'O}$, $2p_{y'O}$ $2s_F$, $2s_O$, $2s'_F$, $2s'_O$	165636900 16/16
R10: H + N_2 → HN_2	$2p_{zNa}\|\sigma_{NH}$, $2p_{zNb}$, $1s_H\|\sigma^*_{NH}$	9 3/3	$2p_{zNa}\|\sigma_{NH}$, $2p_{zNb}$, $1s_H\|\sigma^*_{NH}$ $2\pi_{xNN}$, $2\pi_{yNN}$, $2\pi_x{}^*_{NN}$, $2\pi_y{}^*_{NN}$	1225 7/7	$2p_{zNa}\|\sigma_{NH}$, $2p_{zNb}$, $1s_H\|\sigma^*_{NH}$ $2\pi_{xNN}$, $2\pi_{yNN}$, $2\pi_x{}^*_{NN}$, $2\pi_y{}^*_{NN}$ $2s_{Na}$, $2s_{Nb}$, $3s_{Na}$, $3s_{Na}$	213444 11/11
R11: H + C_2H_4 → CH_3CH_2	$2p_{zCa}\|\sigma_{CH}$, $2p_{zCb}$, $1s_H\|\sigma^*_{CH}$	9 3/3	$2p_{zCa}\|\sigma_{CH}$, $2p_{zCb}$, $1s_H\|\sigma^*_{CH}$	9 3/3	$2p_{zCa}\|\sigma_{CH}$, $2p_{zCb}$, $1s_H\|\sigma^*_{CH}$	9 3/3
R12: HCN → HNC	$\sigma_{CH}\|\sigma_{NH}$, σ^*_{CH},$\|\sigma^*_{NH}$ $2p_{xC}$, $2p_{yC}$, $2p_{xN}$, $2p_{yN}$, $2p'_{xC}$, $2p'_{yC}$, $2p'_{xN}$, $2p'_{yN}$	63504 10/10	$\sigma_{CH}\|\sigma_{NH}$, σ^*_{CH},$\|\sigma^*_{NH}$, $2p_{xC}$, $2p_{yC}$, $2p_{xN}$, $2p_{yN}$, $2p'_{xC}$, $2p'_{yC}$, $2p'_{xN}$, $2p'_{yN}$	63504 10/10	$\sigma_{CH}\|\sigma_{NH}$, σ^*_{CH},$\|\sigma^*_{NH}$, $2p_{xC}$, $2p_{yC}$, $2p_{xN}$, $2p_{yN}$, $2p'_{xC}$, $2p'_{yC}$, $2p'_{xN}$, $2p'_{yN}$	63504 10/10

is the error in the energy of reaction. Tables 5 and 6 give results for, respectively, single-reference Møller−Plesset second-order perturbation theory[61] (MP2) and coupled cluster theory with single and double excitations[62] (CCSD), respectively. The MP2 and CCSD calculations for open-shell cases are based on unrestricted Hartree−Fock reference functions. All results in Tables 2–6 correspond to consistently optimized structures at the level specified (the present article contains no single-point energies).

Note that in some cases, namely **R4**, **R11**, and **R12**, the model chemistry definitions yield the same size active space for *nom*-CPO and *mod*-CPO. In the case of **R12**, the active

Table 2. Barrier Heights and Errors for Reactions in DBH24 Calculated with MRMP2/*nom*-CPO with MG3S and aug-cc-pVTZ Basis Sets

reaction	(*n*/*m*)	$V_f^{\ddagger}$	$V_f^{\ddagger}$ (best est)	$V_r^{\ddagger}$	$V_r^{\ddagger}$ (best est)	$\Delta(V_f^{\ddagger})$	$\Delta(V_r^{\ddagger})$	$\Delta(\Delta V)$
			MG3S					
R1: OH + CH_4 → CH_3 + H_2O	(3/3)	6.23	6.7	20.87	19.6	−0.5	1.3	−1.7
R2: H + OH → O + H_2	(4/4)	10.6	10.7	14.78	13.1	−0.1	1.7	−1.8
R3: H + H_2S → H_2 + HS	(3/3)	3.31	3.6	22.85	17.3	−0.3	5.6	−5.8
R4: H + N_2O → OH + N_2	(11/11)	18.83	18.1	80.12	83.2	0.7	−3.1	3.8
R5: H + ClH → HCl + H	(3/3)	16.24	18.0	16.24	18.0	−1.8	−1.8	0.0
R6: CH_3 + FCl → CH_3F + Cl	(3/3)	4.56	7.4	61.49	60.5	−2.9	1.0	−3.8
R7: $Cl^-\cdots CH_3Cl$ → $ClCH_3\cdots Cl^-$	(4/4)	14.64	13.6	14.64	13.6	1.0	1.0	0.0
R8: $F^-\cdots CH_3Cl$ → FCH_3+ Cl^-	(4/4)	3.29	2.9	30.10	29.6	0.4	0.5	−0.1
R9: OH^- + CH_3F→ $HOCH_3\cdots F^-$	(4/4)	−1.13	−2.8	18.38	17.3	1.7	1.1	0.6
R10: HN_2 → H + N_2	(3/3)	14.96	14.7	9.61	10.7	0.3	−1.1	1.4
R11: H + C_2H_4 → CH_3CH_2	(3/3)	2.44	1.7	39.07	41.8	0.7	−2.7	3.4
R12: HNC → HCN	(10/10)	32.14	33.1	48.74	48.2	−1.0	0.6	−1.5
MUE(DBH24)						0.9	1.8	2.0
MUE(DBHS22)						1.0	1.6	1.8
MUE(DBHS20)						1.0	1.9	2.4
MUE(DBHS12)						0.5	2.1	2.6
			aug-cc-pVTZ					
R1: OH + CH_4 → CH_3 + H_2O	(3/3)	5.63	6.7	20.40	19.6	1.1	−0.8	−1.9
R2: H + OH → O + H_2	(4/4)	9.75	10.7	13.70	13.1	0.9	−0.6	−1.6
R3: H + H_2S → H_2 + HS	(3/3)	2.13	3.6	20.23	17.3	1.5	−2.9	−4.4
R5: H + ClH → HCl + H	(3/3)	14.75	18.0	14.75	18.0	3.3	3.3	0.0
R6: CH_3 + FCl → CH_3F + Cl	(3/3)	3.93	7.4	58.95	61.0	3.5	2.1	−1.4
R7: $Cl^-\cdots CH_3Cl$ → $ClCH_3\cdots Cl^-$	(4/4)	13.42	13.6	13.42	13.6	0.2	0.2	0.0
R8: $F^-\cdots CH_3Cl$ → $FCH_3\cdots Cl^-$	(4/4)	3.68	2.9	25.88	29.6	−1.1	3.8	4.9
R9: OH^- + CH_3F→ $HOCH_3$ + F^-	(4/4)	−1.78	−2.8	16.64	17.3	−1.7	−1.0	0.7
R10: HN_2 → H + N_2	(3/3)	9.80	10.7	13.58	14.7	0.9	1.1	0.2
R11: H + C_2H_4 → CH_3CH_2	(3/3)	1.67	1.7	39.78	41.8	0.1	2.0	1.9
R12: HNC → HCl	(10/10)	31.15	33.1	47.72	48.2	2.0	0.4	−1.5
MUE(DBHS22)						1.4	1.4	1.7
MUE(DBHS12)						1.1	1.3	1.9

Table 3. Barrier Heights and Errors for Reactions in DBH24 Calculated with MRMP2/*mod*-CPO/MG3S

reaction	(*n*/*m*)	$V_f^{\ddagger}$	$V_f^{\ddagger}$ (best est)	$V_r^{\ddagger}$	$V_r^{\ddagger}$ (best est)	$\Delta(V_f^{\ddagger})$	$\Delta(V_r^{\ddagger})$	$\Delta(\Delta V)$
R1: OH + CH_4 → CH_3 + H_2O	(7/7)	7.17	6.7	19.2	19.6	0.5	−0.4	0.9
R2: H + OH → O + H_2	(6/6)	10.61	10.7	16.18	13.1	−0.1	3.1	−3.2
R3: H + H_2S → H_2 + HS	(7/7)	4.33	3.6	21.49	17.3	0.7	4.2	−3.5
R4: H + N_2O → OH + N_2	(11/11)	18.83	18.1	79.92	83.2	0.7	−3.3	4.0
R5: H + ClH → HCl + H	(7/7)	18.78	18.00	18.78	18.0	0.8	0.8	0.0
R6: CH_3 + FCl → CH_3F + Cl	(11/11)	4.35	7.4	61.77	61.0	−3.1	0.8	−3.8
R9: OH^- + CH_3F → $HOCH_3\cdots F^-$	(12/12)	−0.04	−2.8	21.34	17.3	2.7	4.0	−1.3
R10: HN_2 → H + N_2	(7/7)	12.36	10.7	16.25	14.7	1.6	1.6	−0.1
R11: H + C_2H_4 → CH_3CH_2	(3/3)	2.44	1.7	39.07	41.8	0.7	−2.7	3.4
R12: HNC → HCN	(10/10)	32.14	33.1	48.74	48.2	−1.0	0.6	−1.6
MUE(DBHS20)						1.2	2.1	2.2
MUE(DBHS12)						0.8	2.1	2.1

Table 4. Barrier Heights and Errors for Reactions in DBH24 Calculated with MRMP2/*ext*-CPO/MG3S

reaction	(*n*/*m*)	$V_f^{\ddagger}$	$V_f^{\ddagger}$ (best est)	$V_r^{\ddagger}$	$V_r^{\ddagger}$ (best est)	$\Delta(V_f^{\ddagger})$	$\Delta(V_r^{\ddagger})$	$\Delta(\Delta V)$
R1: OH + CH_4 → CH_3 + H_2O	(9/9)	5.63	6.7	20.33	19.6	−1.1	0.7	−1.8
R2: H + OH → O + H_2	(8/8)	10.80	10.7	15.06	13.1	0.1	−2.0	−1.9
R3: H + H_2S → H_2 + HS	(9/11)	4.59	3.6	19.93	17.3	1.0	2.6	−1.6
R10: HN_2 → H + N_2	(11/11)	11.07	10.7	15.33	14.7	0.4	0.6	−0.3
R11: H + C_2H_4 → CH_3CH_2	(3/3)	2.44	1.7	39.07	41.8	0.7	−2.7	3.4
R12: HNC → HCN	(10/10)	32.14	33.1	48.74	48.2	−1.0	0.6	−1.5
MUE(DBHS12)						0.7	1.5	1.8

space is also the same size for *ext*-CPO. Note also that some of the larger calculations are omitted for *mod*-CPO and *ext*-CPO due to the large number of determinants or difficulty converging the geometry optimization, and one of the *nom*-CPO/aug-cc-pVTZ calculations is also omitted; thus the errors in these cases are for subsets of DBH24. These subsets are called DBHS22, DBHS20, and DBHS12, and these mean unsigned errors (MUEs) are also given for other methods (when available) for comparison. (Subset 22 (S22) is missing **R4**; subset 20 (S20) is missing **R7** and **R8**; and subset 12 (S12) is missing **R4**-**R9**.)

Tables 2–6 give separate MUEs for forward and reverse barrier heights as well as the MUEs for energies of reaction. The final MUE is the average over those for forward and reverse barriers, and these are given in Table 7.

5. Discussion

Before discussing the calculated barrier heights and energies of reaction, we analyze the extent of multireference character in the reagents and transition states. For this purpose we have defined a diagnostic quantity M in terms of the eigenvalues,

Table 5. Barrier Heights and Errors for Reactions in DBH24 Calculated with MP2/MG3S

reaction	$V_f^‡$	$V_f^‡$ (best est)	$V_r^‡$	$V_r^‡$ (best est)	$\Delta(V_f^‡)$	$\Delta(V_r^‡)$	$\Delta(\Delta V)$
R1: OH + CH_4 → CH_3 + H_2O	9.16	6.7	25.48	19.6	2.5	5.9	−3.4
R2: H + OH → O + H_2	18.50	10.7	17.47	13.1	7.8	4.4	3.4
R3: H + H_2S → H_2 + HS	7.72	3.6	18.81	17.3	4.1	−1.5	2.6
R4: H + N_2O → OH + N_2	36.25	18.1	87.57	83.2	18.1	4.4	13.8
R5: H + ClH → HCl + H	24.50	18.0	24.50	18.0	6.5	6.5	0.0
R6: CH_3 + FCl → CH_3F + Cl	15.82	7.4	72.94	61.0	8.4	11.9	−3.5
R7: $Cl^-\cdots CH_3Cl$ → $ClCH_3\cdots Cl^-$	14.46	13.6	14.46	13.6	0.9	0.9	0.0
R8: $F^-\cdots CH_3Cl$ → $FCH_3\cdots Cl^-$	3.75	2.9	31.19	29.6	0.9	1.6	−0.7
R9: OH^- + CH_3F → $HOCH_3$ + F^-	−2.84	−2.8	17.51	17.3	−0.1	0.2	−0.2
R10: HN_2 → H + N_2	9.89	10.7	28.04	14.7	−0.8	13.4	−14.2
R11: H + C_2H_4 → CH_3CH_2	9.93	1.7	46.04	41.8	8.2	4.3	3.9
R12: HNC → HCN	35.31	33.1	53.21	48.2	2.2	5.0	−2.9
MUE(DBH24)					5.0	5.0	4.1
MUE(DBHS22)					3.8	5.1	3.2
MUE(DBHS20)					5.9	5.7	4.8
MUE(DBHS12)					4.3	5.7	5.1

Table 6. Barrier Heights and Errors for Reactions in DBH24 Calculated with CCSD/MG3S

reaction	$V_f^‡$	$V_f^‡$ (best est)	$V_r^‡$	$V_r^‡$ (best est)	$\Delta(V_f^‡)$	$\Delta(V_r^‡)$	$\Delta(\Delta V)$
R1: OH + CH_4 → CH_3 + H_2O	9.94	6.7	21.19	19.6	3.2	1.6	1.6
R2: H + OH → O + H_2	11.36	10.7	16.96	13.1	0.7	3.9	−3.2
R3: H + H_2S → H_2 + HS	5.10	3.6	21.69	17.3	1.5	4.4	−2.9
R4: H + N_2O → OH + N_2	19.89	18.1	92.04	83.2	1.8	8.8	−7.1
R5: H + ClH → HCl + H	21.10	18.0	21.10	18.0	3.1	3.1	0.0
R6: CH_3 + FCl → CH_3F + Cl	9.29	7.4	67.10	61.0	1.9	6.1	−4.2
R7: $Cl^-\cdots CH_3Cl$ → $ClCH_3\cdots Cl^-$	14.94	13.6	14.94	13.6	1.3	1.3	0.0
R8: $F^-\cdots CH_3Cl$ → $FCH_3\cdots Cl^-$	3.64	2.9	34.09	29.6	0.7	4.5	−3.7
R9: OH^- + CH_3F → $HOCH_3$ + F^-	−0.90	−2.8	19.72	17.3	1.9	2.4	−0.5
R10: HN_2 → H + N_2	11.55	10.7	16.47	14.7	0.8	1.8	−0.9
R11: H + C_2H_4 → CH_3CH_2	3.09	1.7	44.02	41.8	1.4	2.3	−0.9
R12: HNC → HCN	34.33	33.1	48.89	48.2	1.2	0.7	0.5
MUE(DBH24)					1.6	3.4	2.1
MUE(DBHS22)					1.6	2.9	1.7
MUE(DBHS20)					1.7	3.5	2.2
MUE(DBHS12)					1.5	2.4	1.7

$n(i)$ or $n(j)$, of the first-order density matrix of the CASSCF wave functions. These eigenvalues are usually called the natural occupation numbers, and n(MCDONO), n(SOMO), and n(MCUNO) are respectively the natural orbital occupation numbers of the most correlated doubly occupied natural orbital (MCDONO), a singly occupied natural orbital (SOMO), and the most correlating unoccupied natural orbital (MCUNO). The MCDONO is defined as the "doubly occupied" (recall that this refers to the dominant configuration) natural orbital with the smallest $n(i)$, and the MCUNO is the "unoccupied" (again, in the dominant configuration) natural orbital with the largest occupation number. Let n_{SOMO} be the number of singly occupied molecular orbitals in the dominant configuration of the ground electronic state (label them $j = 1,..., n_{SOMO}$). Then the multireference diagnostic is defined as

$$M = \frac{1}{2}\left(2 - n(\mathrm{MCDONO}) + \sum_{j=1}^{n_{SOMO}} |n(j) - 1| + n(\mathrm{MCUNO})\right) \quad (1)$$

Note that one enumerates n_{SOMO} at a molecular geometry that represents a stationary structure on the potential energy surface, e.g., for dissociation of the hydrogen molecule, n_{SOMO} would be zero rather than two. Since $0 \le n(i \text{ or } j) \le 2$, the upper bound on the numerator is $2 + n_{SOMO}$, and therefore one must always have $0 \le M \le 1$ for closed-shell systems. For systems with open shells, the upper bound on M depends on n_{SOMO}, but since the occupancies of singly occupied natural orbitals are close to 1 (for reactions in DBH24, the largest deviation from unity is 0.011) and since we do not include a correlating orbital for SOMOs, the maximal value of M is not significantly larger than 1. For a Hartree−Fock wave function, M would be zero, and larger M for the CASSCF wave function shows larger static correlation.

For systems involving participating π bonds, such as, e.g., reaction **R11**, the MCDONO is often the same as the highest-energy molecular orbital (HOMO), and the MCUNO is the same as the lowest-energy unoccupied molecular orbital (LUMO), but this need not be the case in general.

In order to place the M values in perspective, it is interesting to compute M values for some prototype cases. Therefore, Figure 1 illustrates the manner in which the M values of the H_2, ethylene, and ozone molecules change as functions of the internuclear distances. The ozone molecule is a widely known example of the "multireference" system.[65] At the experimental equilibrium nuclear configurations of these molecules, the M values are 0.024, 0.072, and 0.225, respectively. As bonds stretch, the M value in the ozone case increases much more rapidly as compared to the other cases, and the M value for breaking a double bond increases more rapidly than that for breaking a single bond. Notice that the M value for C_2H_4 in Table 8 is 0.086, which is 19% larger than the 0.072 quoted above; there are two reasons for this difference. First of all, the value in Table 8 is for a geometry optimized at the MRMP2/*nom*-CPO/MG3S level, whereas Figure 1 is based on the experimental geometry which has a C−C bond length of 1.339 Å (vs 1.331 Å for Table 8).

Table 7. MUEs for MRMP2, MP2, and CCSD for Reactions in DBH24

method	MUE
DBH24	
MRMP2/*nom*-CPO/MG3S	1.4
MP2/MG3S	5.0
CCSD/MG3S	2.5
B3LYP/MG3S	4.1
M06−2X/MG3S	1.0
M05−2X/MG3S	1.7
DBHS22	
MRMP2/*nom*-CPO/MG3S	1.3
MRMP2/*nom*-CPO/aug-cc-pVTZ	1.4
MP2/MG3S	4.5
CCSD/MG3S	2.3
B3LYP/MG3S	3.8
M06−2X/MG3S	1.1
M05−2X/MG3S	1.6
DBHS20	
MRMP2/*nom*-CPO/MG3S	1.5
MRMP2/*mod*-CPO/MG3S	1.7
MP2/MG3S	5.8
CCSD/MG3S	2.6
B3LYP/MG3S	4.2
M06−2X/MG3S	1.0
M05−2X/MG3S	1.8
DBHS12	
MRMP2/*nom*-CPO/MG3S	1.3
MRMP2/*nom*-CPO/aug-cc-pVTZ	1.2
MRMP2/*mod*-CPO/MG3S	1.5
MRMP2/*ext*-CPO/MG3S	1.1
MP2/MG3S	5.0
CCSD/MG3S	2.0
B3LYP/MG3S	3.1
M06−2X/MG3S	1.2
M05−2X/MG3S	1.6

Second, and more important, is the dependence on the active space. The value in Table 8 is for reaction **R11**, for which the *nom*-CPO active space includes only the π and π^* orbitals for ethylene, but Figure 1 is for breaking a double bond, and so the active space includes σ_{CC} and σ^*_{CC} as well.

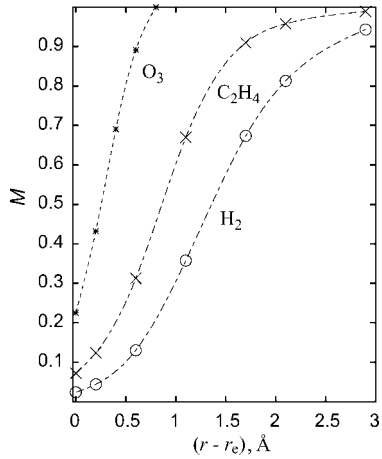


Figure 1. Mutlireference diagnostics *M* for H_2, ethylene, and ozone in their ground electronic states calculated with FORS(2/2), FORS(4/4), and FORS(12/12), respectively, as functions of the HH, CC stretching, and OO symmetric stretching coordinates, respectively. The equilibrium values of those distances that are varied are as follows: $r_{eHH} = 0.741$ Å,[63] $r_{eCC} = 1.339$ Å,[64] $r_{eOO} = 1.278$ Å.[64] The values of the other coordinates (held fixed) are $r_{eCH} = 1.086$ Å,[64] ∠HCH = 117.6°,[64] ∠HCC = 121.2°,[64] ∠OOO = 116.8°.[64]

Another interesting comparison is the reaction of H with H_2O_2 to give H_2O + OH or HO_2 + H_2. In a recent study[44] we found much greater multireference effects in the former than in the latter. Now we calculate *M* values of 0.099 and 0.064, respectively, for these reactions at the *mod*-CPO level. The fact that the former channel has a higher *M* value is consistent with our previous experience.

The multireference diagnostics for all CPO/MG3S calculations are given in Tables 8–10. Several interesting features emerge. First, all *M* values for reagents (a "reagent" is a reactant or product) are less than or equal to 0.086, and all *M* values for transition states are less than or equal to 0.106. Second, in all reactions except **R7** and **R8** and sometimes **R10**-**R12**, *M* is larger for the transition state than for either reagent. The average transition state *M* value is 0.054 for *nom*-CPO, 0.059 for *mod*-CPO, and 0.057 for *ext*-CPO. These are very consistent, especially consider that the latter two values are for subsets of the reactions. The chief exceptions to consistency of *M* from level to level are reaction **R6**, where *M* at the saddle point is surprisingly 0.03 lower at the higher *mod*-CPO level than at *nom*-CPO, and reaction **R10**, where *M* at the reactant is 0.05 lower at the *nom*-CPO than the average *M* at *mod*-CPO and *ext*-CPO levels. The differences in *M* for different levels for the cases with level dependence are due to two kinds of effects: (i) The MCDONO/MCUNO pair of the largest active spaces is not the same as the orbital that describes a breaking/forming bond; and thus this pair is "inactive" in the *nom*-CPO case, while it is "active" at higher *mod*-CPO and *ext*-CPO levels; in this case the *M* values for *nom*-CPO will be lower than those that at the higher levels (examples are the reactants in **R1**, **R3**, and **R10**). (ii) The inclusion of "extra" orbitals in *mod*-CPO and/or *ext*-CPO levels can result in a change of the natural orbital occupation numbers of the MCDONO and MCUNO; in this case the *M* values for *nom*-CPO can be either lower or higher than those that at the higher levels (the chief example is reaction **R6**).

Now we consider the energetics. Tables 2–4 show that the mean errors in MRMP2/CPO calculations are smaller for the forward (exothermic) reactions than for the reverse. This is reasonable since, by Hammond's postulate,[66] we expect the transition state of an exothermic reaction to resemble the reactants more than the products. Thus errors in dynamical correlation energies are more likely to cancel for this direction of reaction. Interestingly, the error in the MP2 barrier heights in Table 5 does not show this difference in accuracy between forward and reverse barrier heights. The comparison of MRMP2 to MP2 is particularly interesting in that they both correspond to second-order perturbation theory but with different reference states. Tables 2–5 and 7 show that the errors in MP2 barrier heights are, on average, about $3^1/_2$ times larger than those for any of the MRMP2/CPO model chemistries. One might expect MP2 to be relatively better for energies of reaction than for barrier heights because the reagents usually have smaller static correlation effects (as shown by the *M* values) than the transition states. This expectation is born out, but Tables 2–5 show that MP2 still has errors 2−3 times larger than MRMP2 even for energies of reaction.

Table 8. Mutlireference Diagnostics for Reactions in DBH24 Calculated with *nom*-CPO/MG3S

reaction	saddle point	reactants	products
R1: $OH + CH_4 \rightarrow CH_3 + H_2O$	0.025	0.018	0.021
R2: $H + OH \rightarrow O + H_2$	0.041	0.024	0.025
R3: $H + H_2S \rightarrow H_2 + HS$	0.029	0.023	0.025
R4: $H + N_2O \rightarrow OH + N_2$	0.092	0.060	0.082
R5: $H + ClH \rightarrow HCl + H$	0.050	0.023	0.023
R6: $CH_3 + FCl \rightarrow CH_3F + Cl$	0.106	0.063	0.025
R7: $Cl^-\cdots CH_3Cl \rightarrow ClCH_3\cdots Cl^-$	0.020	0.028	0.028
R8: $F^-\cdots CH_3\ Cl \rightarrow FCH_3\cdots Cl^-$	0.024	0.028	0.024
R9: $OH^- + CH_3F \rightarrow HOCH_3 + F^-$	0.032	0.025	0.023
R10: $HN_2 \rightarrow H + N_2$	0.072	0.022	0.064
R11: $H + C_2H_4 \rightarrow CH_3CH_2$	0.096	0.086	0.018
R12: HNC → HCN	0.063	0.052	0.063

Table 9. Mutlireference Diagnostics for Reactions in DBH24 Calculated with *mod*-CPO/MG3S

reaction	saddle point	reactants	products
R1: $OH + CH_4 \rightarrow CH_3 + H_2O$	0.029	0.026	0.024
R2: $H + OH \rightarrow O + H_2$	0.040	0.027	0.024
R3: $H + H_2S \rightarrow H_2 + HS$	0.038	0.033	0.029
R4: $H + N_2O \rightarrow OH + N_2$	0.092	0.060	0.082
R5: $H + ClH \rightarrow HCl + H$	0.051	0.026	0.026
R6: $CH_3 + FCl \rightarrow CH_3F + Cl$	0.077	0.047	0.025
R9: $OH^- + CH_3F \rightarrow HOCH_3 + F^-$	0.036	0.030	0.025
R10: $HN_2 \rightarrow H + N_2$	0.063	0.077	0.061
R11: $H + C_2H_4 \rightarrow CH_3CH_2$	0.096	0.086	0.018
R12: HNC → HCN	0.063	0.052	0.063

Table 10. Mutlireference Diagnostics for Reactions in DBH24 Calculated with *ext*-CPO/MG3S

reaction	saddle point	reactants	products
R1: $OH + CH_4 \rightarrow CH_3 + H_2O$	0.028	0.026	0.025
R2: $H + OH \rightarrow O + H_2$	0.038	0.026	0.024
R3: $H + H_2S \rightarrow H_2 + HS$	0.049	0.043	0.041
R10: $HN_2 \rightarrow H + N_2$	0.070	0.081	0.061
R11: $H + C_2H_4 \rightarrow CH_3CH_2$	0.096	0.086	0.018
R12: HNC → HCN	0.063	0.052	0.063

The comparison to CCSD calculations is interesting from a methodological point of view. Both MRMP2 and CCSD are limited in a formal sense to double excitations, but MRMP2 generates a subset of the triple and higher excitations from the dominant configuration by taking double excitations from other configurations in the CASSCF wave function, and CCSD (a single-reference method) generates all disconnected triple excitations by the coupled-cluster exponential excitation operator. Both methods also generate a subset of quadruple and higher excitations but different subsets. As a consequence of the triple and higher excitations, CCSD should be less sensitive to multireference effects than MP2 (and Tables 5–7 show that it is more accurate for barrier heights), but the present study is the first systematic comparison to MRMP2. Tables 2–4, 6, and 7 show that MRMP2 has an MUE in barrier heights that is a factor of ~1.5 lower than CCSD. This is very welcome because, for systems with a large number N_{atoms} of atoms, the cost of MRMP, like MP, scales as N_{atoms}^5, whereas CCSD scales as $Natoms^6$. We note that using a full-valence active space would cause much worse scaling of MRMP2, but the size of the active spaces does not increase with N_{atoms} for the CPO schemes because they are defined in terms of participating orbitals rather than all valence orbitals. This, in fact, was one of the motivations of defining the CPO schemes the way that we did.

Comparison of Table 2 to Table 5 shows no systematic improvement of MRMP2 over MP2 for the three S_N2 reactions (reactions **R7**-**R9**). The relatively good performance of MP2 for S_N2 reactions is well-known from many previous studies in the literature. Tables 8 and 9 show that S_N2 reactions have smaller than average M values.

It is interesting to center attention on the subset of reactions with larger M values. Reactions **R4**, **R6**, **R10**, and **R12** are the only reactions with $M > 0.04$. Comparing Tables 2, 3, and 5 for these reactions shows especially large errors in MP2 for **R4** and **R10**, which are greatly diminished by MRMP2. However reaction **R6** has large errors by all methods, and reaction **R12** has smaller than average errors for all methods.

Whereas MP2 has a pronounced systematic error, overestimating 21 of the 24 barriers, MRMP2 does not have such a systematic error. Thus one can find a general scaling factor[67,68] for the correlation contribution in MP2, but the optimum general scale factor for MRMP2 is close to unity. In fact, in the notation of the scaling external correlation method,[69] the optimum scale factor F for MRMP2/*nom*-CPO/MG3S is 1.03, and it lowers the MUE by only 0.03 kcal/mol as compared to $F = 1.00$ (no scaling). Hence we do not recommend developing general scaling features for MRMP2.

Next we turn our attention to the comparison of *nom*-CPO, *mod*-CPO, and *ext*-CPO to each other. Table 7 shows no systematic improvement as the size of the active space is improved. While this might be interpreted as a negative result by some, we take this as very encouraging. It shows that a minimal definition of participating orbitals already captures the bulk of the static correlation effect on chemical reaction

Table 11. Saddle Point Geometries (in Å) and Errors for Selected Reactions in DBH24

	geometry			
R2: H + OH → O + H_2	$r_{HH'}$	r_{OH}	$r_{OH'}$	MUE
best est[78]	0.894	1.215	2.109	
MRMP2/*nom*-CPO/MG3S	0.887	1.231	2.118	0.011
MRMP2/*mod*-CPO/MG3S	0.910	1.203	2.113	0.011
MRMP2/*ext*-CPO/MG3S	0.909	1.200	2.109	0.010
MRMP2/*nom*-CPO/aug-cc-pVTZ	0.884	1.239	2.122	0.016
MP2/MG3S	0.859	1.227	2.086	0.054
CCSD/MG3S	0.910	1.190	2.100	0.036
B3LYP/MG3S	0.907	1.202	2.109	0.030
M05−2X/MG3S	0.933	1.167	2.100	0.036
M06−2X/MG3S	0.921	1.179	2.100	0.036
R5: H + ClH → HCl + H	r_{ClH}	$r_{ClH'}$	$r_{HH'}$	MUE
best est[79]	1.480	1.480	2.960	
MRMP2/*nom*-CPO/MG3S	1.482	1.482	2.964	0.003
MRMP2/*mod*-CPO/MG3S	1.490	1.490	2.980	0.013
MRMP2/*nom*-CPO/aug-cc-pVTZ	1.482	1.482	2.964	0.003
MP2/MG3S	1.468	1.468	2.936	0.016
CCSD/MG3S	1.487	1.487	2.975	0.010
B3LYP/MG3S	1.490	1.490	2.979	0.013
M05−2X/MG3S	1.485	1.485	2.969	0.006
M06−2X/MG3S	1.486	1.486	2.973	0.008
R7: $Cl^-\cdots CH_3Cl \rightarrow ClCH_3\cdots Cl^-$	r_{ClC}	$r_{Cl'C}$	$r_{ClCl'}$	MUE
best est[80]	2.305	2.305	4.610	
MRMP2/*nom*-CPO/MG3S	2.302	2.302	4.604	0.004
MRMP2/*nom*-CPO/aug-cc-pVTZ	2.277	2.277	4.553	0.038
MP2/MG3S	2.288	2.288	4.576	0.023
CCSD/MG3S	2.319	2.319	4.637	0.018
B3LYP/MG3S	2.355	2.355	4.710	0.067
M05−2X/MG3S	2.308	2.308	4.616	0.004
M06−2X/MG3S	2.300	2.300	4.600	0.007
R12: HNC → HCN	r_{CH}	r_{NH}	r_{CN}	MUE
best est[81]	1.183	1.387	1.186	
MRMP2/*nom*-CPO/MG3S	1.194	1.379	1.195	0.009
MRMP2/*nom*-CPO/aug-cc-pVTZ	1.193	1.383	1.196	0.008
MP2/MG3S	1.173	1.417	1.186	0.013
CCSD/MG3S	1.179	1.402	1.186	0.006
B3LYP/MG3S	1.190	1.388	1.180	0.005
M05−2X/MG3S	1.177	1.434	1.170	0.023
M06−2X/MG3S	1.172	1.438	1.174	0.025

barrier heights. This means that the MRMP2/*nom*-CPO method is likely to be very useful even for large systems.

We can compare the present method to the MR-G3(MP2) and MR-G3/MP2 model chemistries by Sølling et al.[12] These are multilevel model chemistries; the former is based on MRCI calculations based on full-valence CASSCF, and the latter on CASPT2 (which is essentially the same as MRMP2), again based on full-valence CASSCF. Sølling et al. did not study barrier heights, but for heats of formation they found that the mean unsigned error in the MRCI-based model is 1.2 kcal/mol and that in the CASPT2-based model is 1.6 kcal/mol. The latter may approximately be compared to our mean unsigned errors in reaction energies, which are 1.8−2.6 kcal/mol, depending on the reaction subset, for *nom*-CPO, 2.1−2.2 kcal/mol for *mod*-CPO, and 1.8 kcal/mol for *ext*-CPO. Given the differences in the test data, the results are comparable, although our calculations do not include empirical high-level corrections, whereas the MR-G3(MP2) and MR-G3/MP2 methods each have four empirical parameters. We note that the present methods lead to continuous potential energies, whereas methods with high-level corrections do not.[70,71]

Additionally, we compare MRMP2 to density functional theory. We have shown elsewhere[16] (based on single-point calculations with the MG3S basis set and QCISD/MG3 geometries) that the most popular density functional (B3LYP)[8] has an MUE for the reactions in Table 7 of 4.3 kcal/mol. Table 7 shows that using consistently optimizing geometries reduces this to 4.1 kcal/mol. However one can do better with more recently developed density functionals; for example Table 7 shows that the M06−2X functional[72] leads to an error of only 1.0 kcal/mol. As mentioned above, the *M* diagnostic for the reactions in DBH24 shows that their multireference character is only small (≤0.05) to modest (0.05−0.10) and in one case large (≥0.10). In any event one expects larger errors in both MRMP2 and DFT when *M* gets larger. It would be interesting to test MRMP2 against density functional theory for reactions with larger multireference character such as reactions with biradical transition states, reactions involving coordinately unsaturated transition metals, or reactions of ozone. Such extensions would probably show different trends from those in Table 7, but the extension to transition metals might also require some refinement of the *nom*-, *mod*-, and *ext*-definitions. The mean unsigned error

Table 12. Saddle Point Geometries (in Å) for Some Other Reactions in DBH24

	geometry		
R8: $F^- \cdots CH_3Cl \rightarrow FCH_3 \cdots Cl^-$	r_{FC}	r_{ClC}	r_{FCl}
ref 82	2.030	2.121	4.151
ref 83	2.019	2.124	4.143
MRMP2/*nom*-CPO/MG3S	1.959	2.191	4.150
MP2/MG3S	2.018	2.098	4.116
CCSD/MG3S	2.038	2.114	4.152
B3LYP/MG3S	2.157	2.077	4.234
M05−2X/MG3S	2.051	2.105	4.156
M06−2X/MG3S	2.036	2.096	4.133

	geometry		
R9: $OH^- + CH_3F \rightarrow HOCH_3 + F^-$	r_{OC}	r_{CF}	r_{OF}
ref 83	2.000	1.753	3.753
MRMP2/*nom*-CPO/MG3S	2.045	1.772	3.817
MP2/MG3S	1.981	1.740	3.721
CCSD/MG3S	1.978	1.754	3.732
B3LYP/MG3S	2.051	1.771	3.821
M05−2X/MG3S	1.976	1.742	3.718
M06−2X/MG3S	1.965	1.734	3.699

in the 12 *nom*-CPO/MG3S barrier heights for which the saddle point *M* value is less than 0.05 is 1.3 kcal/mol, and that for the 10 *nom*-CPO/MG3S barrier heights for which the saddle point *M* value is greater than 0.05 is 1.4 kcal/mol, which is essentially the same. For some other methods though, the MUE increases in passing from the M < 0.05 subset to the for the M > 0.05 subset, in particular from 2.5 to 7.5 kcal/mol for MP2, from 2.3 to 2.8 kcal/mol for CCSD, from 1.4 to 2.0 kcal/mol for M05-2X, and from 4.0 to 4.2 kcal/mol for B3LYP (the MUE is 1.0 kcal/mol for both subsets for M06-2X). The more uniform performance of MRMP2 as compared to several other methods is encouraging and is encougaging and supports the hope that MRMP2 might remain accurate for large *M* values, although this clearly needs to be confirmed by actual tests, whereas MP2, CCSD, and hybrid DFT would all be expected to become less accurate in that case.

Table 2 allows us to compare the performance of two basis sets, MG3S and aug-cc-pVTZ; both are augmented or partially augmented with diffuse functions, are multiply polarized, and have triple-split quality for valence orbitals, but aug-cc-pVTZ is much larger (and hence much more expensive). Table 2 shows that aug-cc-pVTZ gives slightly more accurate (on average) reaction energies but significantly less accurate barrier heights. Four of the five reactions (**R3** and **R5** to **R8**) that involve S or Cl have significantly less accurate forward barrier heights with aug-cc-pVTZ than with MG3S. This might be related in part to a known deficiency[73–76] of the aug-cc-pVTZ basis set for 3*p*-block elements that can be remedied by adding tighter *d* functions. To test this possibility we repeated three of the reactions including S or Cl with the aug-cc-pV(T+d)Z basis set, but the barriers did not change by more than 1.0 kcal. This deficiency (i.e., lack of tight *d* functions for 3*p* elements) does not affect the MG3S basis set because MG3S includes the improved 3*d* functions suggested by Curtiss et al.[56] for P, S, and Cl. The present comparison confirms the high efficiency of the MG3S basis set, which we have employed with satisfactory results for a variety of problems in our group.

Finally we consider transition state geometries. It is harder to test theory for transition state geometries than for transition state energetics because so few accurate transition state geometries are known.[77] Table 11 compares the present transition state geometries to accurate results[78–81] for the only four cases in DBH24 for which we judge the best available estimates of the transition state geometries to be accurate (converged with respect to the inclusion of electron correlation energy) to ≤0.005 Å. The reactions in Table 1 all involve making a bond, X-Y, and breaking a bond, Y-Z. In each case we give the three key bond distances, X-Y, Y-Z, and X-Z, where Y-Z is called the donor−acceptor distance. We also give the mean unsigned error of these three distances as compared to the accurate values. Table 11 shows an average MUE (averaged over all four reactions with the MG3S basis set) in transition state bond lengths and donor−acceptor distances of 0.007 Å for MRMP2/*nom*-CPO, 0.018 Å for CCSD, 0.019 Å for M06−2X, and 0.039 Å for MP2 and B3LYP. The decided superiority of MRMP2 is evident but has never been systematically demonstrated in previous work. Extending the active space to *mod*-CPO (two examples) or *ext*-CPO (one example) does not improve the results, but there is little room for improvement.

There are two additional reactions for which reasonably accurate values of transition state geometries have been published.[82,83] We do not believe that the convergence of these geometries to 0.005 Å has been demonstrated, and therefore we did not include them in Table 11, and we do not interpret deviations from the present calculations as necessarily indicating errors in the present calculations, but the comparisons are still interesting and are given in Table 12. The deviations of MRMP2 from the previous calculations are larger than for the reactions in Table 12, and the various results for the $OH^- + CH_3F$ nucleophilic substitution reaction differ greatly (a range of 0.080 Å for the O−C distance and a range of 0.122 Å for the donor−acceptor distance). We conclude that further study of the transition state geometries in Table 12 would be very interesting.

6. Concluding Remarks

We have defined three systematic choices of active space that, unlike full-valence spaces, do not increase in size, for a given reaction type, as the reagents increase in size, and we have tested them with the MRMP2[23,24] wave function method against a database of diverse barrier heights and a smaller database of accurately known transition state geometries. The smallest of the three systematically defined active spaces, called *nom*-CPO, has an average size of only 4.6 active electrons in 4.6 active orbitals for the reactions studied (ranging from 3 active electrons in 3 active orbitals to 11 active electrons in 11 active orbitals), but it captures the bulk of the static correlation energy, which is quite significant for these reactions. Thus the mean unsigned error in barrier heights for MRMP2/*nom*-CPO is 1.4 kcal/mol, whereas single-reference MP2 has a mean unsigned error of 5.0 kcal/mol for the same reactions. Even CCSD, which is well-known to be less sensitive than MP2 to multireference effects, has a mean unsigned error of 2.5 kcal/mol for these reactions. MRMP2 also gives a remarkable improvement in

accuracy for transition state geometries with a mean unsigned error in forming and breaking bond lengths and donor−acceptor distances of only 0.007 Å, as compared to 0.039 Å for MP2 and 0.018 Å for CCSD. The present results confirm earlier indications that MRMP2 can provide accurate barrier heights[84–87] and transition state geometries[44] for chemical reactions. The present work provides systematic demonstration on a diverse set of reactions, and, by defining three levels of active spaces, it paves the way for systematic applications of multireference theoretical model chemistries in the future.

We defined a multireference diagnostic called M that approximately measures the amount of multireference character in the reagents and transition states. It ranges from 0.020 to 0.106 for the reactions in our diverse barrier height database. This establishes a range of validation for the new MRMP2 model chemistries, and further validation is recommended for applications to reactions with $M > 0.1$.

Acknowledgment. This work was supported in part by the United States Department of Energy, Office of Basic Energy Sciences under grant no. DE-FG02-86ER13579.

Supporting Information Available: Optimized molecular geometries in Cartesian coordinates. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Pople, J. A. In *Energy, Structure and Reactivity: Proceedings of the 1972 Boulder Summer Research Conference on Theoretical Chemistry*; Smith, D. W., McRae, W. B., Eds.; Wiley: New York, 1973; p 51.

(2) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.

(3) Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213.

(4) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 1009.

(5) Pople, J. A.; Radom, L.; Hehre, W. J. *J. Am. Chem. Soc.* **1971**, *93*, 289.

(6) DeFrees, D. J.; Raghavachari, K.; Schlegel, H. B.; Pople, J. A. *J. Am. Chem. Soc.* **1982**, *104*, 5576.

(7) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(8) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(9) Martin, J. M. L. De Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843.

(10) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 1125.

(11) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4813.

(12) Sølling, T. I.; Smith, D. M.; Radom, L.; Freitag, M. A.; Gordon, M. S. *J. Chem. Phys.* **2001**, *115*, 8758.

(13) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 1643.

(14) Wood, G. P. F.; Radom, L.; Petersson, G. A.; Barnes, E. C.; Frisch, M. J.; Montgomery, J. A. *J. Chem. Phys.* **2006**, *125*, 194106.

(15) Truhlar, D. G. *J. Comput. Chem.* **2007**, *28*, 73.

(16) Zheng, J.; Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 569.

(17) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. *Chem. Phys.* **1980**, *48*, 157.

(18) Siegbahn, P. E. M.; Almlöf, J.; Heiberg, A.; Roos, B. O. *J. Chem. Phys.* **1981**, *74*, 2384.

(19) (a) Werner, H.-J.; Knowles, P. J. *J. Chem. Phys.* **1985**, *82*, 5053. (b) Knowles, P. J.; Werner, H.-J. *Chem. Phys. Lett.* **1985**, *115*, 259.

(20) Roos, B. O. *Adv. Chem. Phys.* **1987**, *69*, 399.

(21) Schmidt, M. W.; Gordon, M. S. *Annu. Rev. Phys. Chem.* **1998**, *49*, 233.

(22) (a) Cheung, L. M.; Sundberg, K. R.; Ruedenberg, K. *Int. J. Quantum Chem.* **1979**, *16*, 1103. (b) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. *Chem. Phys.* **1982**, *71*, 41. FORS was originally defined to denote what whould be labeled a "full-valence" CASSCF calculation in the more popular nomenclature, but it is now generally accepted as being synonymous with CASSCF. See also ref 21.

(23) Hirao, K. *Chem. Phys. Lett.* **1992**, *190*, 374.

(24) Hirao, K. *Int. J. Quantum Chem. Symp.* **1992**, *26*, 517.

(25) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483.

(26) Wolinski, K.; Sellers, H. L.; Pulay, P. *Chem. Phys. Lett.* **1987**, *140*, 225.

(27) Kozlowski, P. M.; Davidson, E. R. *J. Chem. Phys.* **1994**, *100*, 3672.

(28) Brown, F. B.; Shavitt, I.; Shepard, R. *Chem. Phys. Lett.* **1984**, *105*, 363.

(29) Werner, H.-J. *Adv. Chem. Phys.* **1987**, *69*, 1.

(30) Jeziorski, B.; Monkhorst, H. *J. Phys. Rev. A* **1981**, *24*, 1668.

(31) Balkova, A.; Kucharski, S. A.; Meissner, L.; Bartlett, R. J. *J. Chem. Phys.* **1991**, *95*, 4311.

(32) Piecuch, P. Paldus, J. *Theor. Chim. Acta* **1992**, *83*, 69.

(33) Piecuch, P.; Oliphant, N.; Adamowicz, L. *J. Chem. Phys.* **1993**, *99*, 1875.

(34) Balkova, A.; Bartlett, R. J. *J. Chem. Phys.* **1994**, *101*, 8972.

(35) Ghose, K. B.; Piecuch, P.; Adamowicz, L. *J. Chem. Phys.* **1995**, *103*, 9331.

(36) Chattopadhyay, S.; Mahapatra, U. S.; Datta, B.; Mukherjee, B. *Chem. Phys. Lett.* **2002**, *357*, 426.

(37) Pittner, J. *J. Chem. Phys.* **2003**, *118*, 10876.

(38) Li, X.; Paldus, J. *J. Chem. Phys.* **2003**, *119*, 5320.

(39) Li, X.; Paldus, J. *J. Chem. Phys.* **2003**, *119*, 5334.

(40) Eckert-Maksic, M.; Vazdar, M.; Barbatti, M.; Lischka, H.; Maksic, Z. B. *J. Chem. Phys.* **2006**, *125*, 64310.

(41) Li, X.; Paldus, J. *J. Chem. Phys.* **2006**, *125*, 164107.

(42) Li, X.; Paldus, J. *J. Phys. Chem. A* **2007**, *111*, 11189.

(43) Evangelista, F. A.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* **2006**, *125*, 154113.

(44) Ellingson, B. A.; Theis, D. P.; Tishchenko, O.; Zheng, J.; Truhlar, D. G. *J. Phys. Chem. A* **2007**, *111*, 13554.

(45) Gan, Z.; Alexeev, Y.; Gordon, M. S.; Kendall, R. A. *J. Chem. Phys.* **2003**, *119*, 47.

(46) Frenking, G.; Koch, W.; Gauss, J.; Cremer, D. *J. Am. Chem. Soc.* 1988*110*, 8007.

(47) Martin, J. M. L.; Taylor, P. R. *J. Phys. Chem.* **1994**, *98*, 6105.

(48) Hunt, W. J.; Hay, P. J.; Goddard, W. A. *J. Chem. Phys.* **1972**, *57*, 738.

(49) Walch, S. P.; Rohlfing, C. M.; Melius, C. F.; Baushclicher, C. W., Jr. *J. Chem. Phys.* **1988**, *88*, 6273.

(50) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 8996.

(51) (a) Herzberg, G. *Spectra of Diatomic Molecules*; 2nd ed.; D. Van Nostrand: Princeton, 1950; pp 540−541, 560−561. (b) Moore, C. National Bureau of Standards (U.S.) Circular 467; 1952. (c) Roberto-Neto, O.; Coitino, E. L.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 4568.

(52) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. J. Chem. Phys. 1992, *96*, 6796.

(53) Woon, D. E.; Dunning, T. H., Jr *J. Chem. Phys.* **1993**, *98*, 1358.

(54) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.

(55) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.

(56) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(57) Zheng, J.; Gour, J. R.; Lutz, J. J.; Wloch, M.; Piecuch, P.; Truhlar, D. G. *J. Chem. Phys.* **2008**, *128*, 44108.

(58) 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.

(59) 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.01*; Gaussian, Inc.: Wallingford, CT, 2004.

(60) Zhao, Y.; Truhlar, D. *MN-GFM - version 3.0*; University of Minnesota: Minneapolis, 2007.

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

(62) Purvis, G. D., III; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.

(63) Huber, K. P.; Herzberg, G. *Molecular Spectra and Molecular Structure. IV. Constants of Diatomic Molecules*; Van Nostrand Reinhold Co.: New York, 1979.

(64) Herzberg, G., *Electronic spectra and electronic structure of polyatomic molecules*; Van Nostrand: New York, 1966.

(65) (a) Laidig, W. D.; Schaefer, H. F.,., III *J. Chem. Phys.* **1981**, *74*, 3411 (b) Xantheas, S. S.; Atchity, G. C.; Elbert, S. T.; Ruedenberg, K. J. *Chem. Phys.* **1991**, *94*, 8054. (c) Borowski, P.; Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *95*, 2107. (d) Tsueneda, T.; Nakano, H.; Hirao, K. J. *Chem. Phys.* **1995**, *103*, 6520. (e) Szalay, P. G. J. *Phys. Chem.* **1996**, *100*, 6288. (f) Leininger, M. L., III *Chem. Phys.* **1997**, *107*, 9059. (g) Li, X.; Paldus, J. J. *Chem. Phys.* **1999**, *110*, 2844. (h) Vaval, N.; Pal, S. J. *Chem. Phys.* **1999**, *111*, 4051. (i) Xie, D.; Guo, H.; Peterson, K. A. J. *Chem. Phys.* **2000**, *112*, 8378. (j) Ljubic, I.; Sabljic, A. J. *Phys. Chem. A* **2002**, *106*, 4745. (k) Ajitha, D.; Hirao, K.; Pal, S. *Collect. Czech. Chem. Commun.* **2003**, *68*, 47.

(66) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334.

(67) Gordon, M. S.; Truhlar, D. G. *J. Phys. Chem. A* **1986**, *108*, 5412.

(68) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 3898.

(69) Brown, F. B.; Truhlar, D. G. *Chem. Phys. Lett.* **1985**, *117*, 307.

(70) Fast, P. L.; Sanchez, M. L.; Corchado, J. C.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 11679.

(71) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407.

(72) Zhao, Y.; Truhlar, D. G. Theor. Chem. Acc., in press; published online athttp://dx.doi.org/10.1007/s00214-007-310-x.

(73) Dunning, T. H.; Petersen, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244.

(74) Wang, N. X.; Wilson, A. K. *J. Phys. Chem. A* **2003**, *107*, 6720.

(75) Wilson, A. K.; Dunning, T. H., Jr. *J. Phys. Chem. A* **2004**, *108*, 3129.

(76) Bell, R. D.; Wilson, A. K. *Chem. Phys. Lett.* **2004**, *394*, 105. Chem. Phys. Lett. 1999, *306*, 407.

(77) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 2936.

(78) Bian, W.; Werner, H.-J. *J. Chem. Phys.* **2000**, *112*, 220.

(79) Peterson, K. A.; Dunning, T. H., Jr. *J. Phys. Chem. A* **1997**, *101*, 6280.

(80) Parthiban, S.; Olivereira, G. D.; Martin, J. M. L. *J. Phys. Chem. A* **2001**, *105*, 895.

(81) van Mourik, T.; Harris, G. J.; Polyansky, O. L.; Tennyson, J.; Császár, A.; Knowles, P. *J. Chem. Phys.* **2001**, *115*, 3707.

(82) Botschwina, P.; Horn, M.; Seeger, S.; Oswald, R. *Ber. Bunsenges. Phys. Chem.* **1997**, *101*, 387.

(83) Gonzales, J. M.; Pak, C.; Cox, R. S.; Allen, W. D.; Schaefer, H. F.; Császár, A. G.; Tarczay, G. *Chem. Eur. J.* **2003**, *9*, 2173.

(84) Roberto-Neto, O.; Machado, F. B. C.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *111*, 10046.

(85) Kobayashi, Y.; Kamiya, M.; Hirao, K. *Chem. Phys. Lett.* **2000**, *319*, 695.

(86) Tishchenko, O.; Vinckier, C.; Nguyen, M. T. *J. Chem. Phys. A* **2004**, *108*, 1268.

(87) Tishchenko, O.; Vinckier, C.; Ceulemans, A.; Nguyen, M. T. *J. Chem. Phys. A* **2005**, *109*, 6099.

CT800077R

JCTC Journal of Chemical Theory and Computation

Anchoring the Absolute Proton Affinity Scale

Gábor Czakó,[†] Edit Mátyus,[†] Andrew C. Simmonett,[‡] Attila G. Császár,*[,†] Henry F. Schaefer III,[‡] and Wesley D. Allen*[,‡]

Laboratory of Molecular Spectroscopy, Institute of Chemistry, Eötvös University, P.O. Box 32, H-1518 Budapest 112, Hungary, and Center for Computational Chemistry and Department of Chemistry, University of Georgia, Athens, Georgia 30602

Received March 10, 2008

Abstract: Converged first-principles proton affinities (PA) of ammonia and carbon monoxide have been determined by the focal-point analysis (FPA) approach, thus fixing the high and low ends of the molecular proton affinity scale. The electronic structure computations employed the all-electron (AE) coupled-cluster (CC) method up to single, double, triple, quadruple, and pentuple excitations. Aug-cc-pCV*X*Z [*X* = 2(D), 3(T), 4(Q), 5, and 6] correlation-consistent (cc) Gaussian basis sets for C, N, and O were used in conjunction with the corresponding aug-cc-pV*X*Z (*X* = 2−6) sets for H. Our FPA study supersedes previous computational work by accounting for (a) electron correlation beyond the "gold standard" CCSD(T) level; (b) the nonadditivity of core electron correlation effects; (c) scalar relativity; (d) diagonal Born−Oppenheimer corrections (DBOC); (e) anharmonicity of zero-point vibrational energies, based on accurate AE-CCSD(T)/cc-pCVQZ internal coordinate quartic force fields and fully variational vibrational computations; and (f) thermal corrections to enthalpies by direct summation over rovibrational energy levels. Our final proton affinities at 298.15(0.0) K are $\Delta_{pa}H^{\circ}(NH_3) = 852.6(846.4) \pm 0.3$ kJ mol^{-1} and $\Delta_{pa}H^{\circ}(CO) = 592.4(586.5) \pm 0.2$ kJ mol^{-1}. These values have better accuracy and considerably lower uncertainty than the best previous recommendations and thus anchor the proton affinity scale of molecules for future use.

I. Introduction

Modern mass spectrometry (MS) is an advanced and highly versatile experimental technique that allows studies of fundamental energetic quantities in the gas phase, including proton affinities (PA) and the related gas-phase basicities (GB). PAs and GBs have special relevance for MS fragmentation processes in tandem mass spectrometry (MS/MS) experiments. MS/MS fragments yield not only essential structural data for a wide variety of compounds but also information about dissociation processes. Among these fragmentation processes, those of protonated peptides and proteins are of particular importance because they provide the basis of protein identification by mass spectrometry in proteomics studies.

The "mobile proton model" has been developed for understanding peptide fragmentation in MS experiments.[1–10] In essence, this relatively simple but now widely accepted model states that upon ion activation the proton(s) added to a peptide will migrate to various sites prior to fragmentation and will thereby trigger charge-directed cleavages. Proton migration is crucial in inducing fragmentation and, as indicated by simple quantum chemical computations, the thermodynamically most stable protonated forms are not the preferred fragmenting structures.[11–13] The ease or difficulty of proton migration depends on the PA (or GB) values of the protonation sites; for example, peptides containing the most basic amino acids, such as arginine (R) or lysine (K), require higher internal energy (more efficient ion activation) to fragment. Statistical evaluations of the MS/MS data of a large number of protonated peptides clearly show that the extent of fragmentation for particular amide bonds can be related to the PAs of the amino acids.[14–16]

* Corresponding author e-mail: wdallen@uga.edu (W.D.A.) and csaszar@chem.elte.hu (A.G.C.).

[†] Eötvös University.

[‡] University of Georgia.

10.1021/ct800082r CCC: $40.75
Published on Web 07/02/2008

As eloquently pointed out by Paizs and Suhai in their recent review[17] on peptide fragmentation, a significant simplification embodied in the mobile proton model is that it focuses mostly on the step(s) prior to dissociation. To overcome this limitation, they proposed the "pathways in competition" (PIC) description involving a detailed energetic and kinetic characterization of the major fragmentation pathways, requiring the computation of parts of the potential energy surfaces (PES). While electronic structure computations of ever increasing accuracy have started to appear for amino acids, peptide models, and small peptides,[18–24] for oligopeptides of practical interest computations at higher levels of electronic structure theory are still not routine, and a full, reliable mapping of the PES is time-consuming even for relatively small dipeptides. For example, several fragmentation pathways for the loss of ammonia have been found for a small protonated dipeptide.[25] The complexity of fragmentation pathways and the practical limitations of rigorous theoretical computations justify the use of experimental MS/MS data and their relatively simple interpretation based on relative proton affinities of possible protonation sites. Therefore, anchoring the PA scale is not only of theoretical but also of practical importance.

Several studies,[26–33] including elaborate reviews and critical compilations, have been published that attempted to fix the absolute proton affinity scale of organic compounds. To accomplish this goal, the PAs of molecules at both the high and low end of the scale must be pinpointed. The purpose of the present study is to determine benchmark first-principles PAs of ammonia (NH_3) and carbon monoxide (CO). These molecules have been chosen because they are small species amenable to highly sophisticated methods of electronic structure and nuclear motion theory, and they have extremely different proton affinities at the low (CO) and the high (NH_3) ends of the absolute PA scale. Among other things, very precise proton affinities of NH_3 and CO will facilitate the determination of ion energetics from measurements of equilibrium constants for reversible proton-transfer reactions

$$AH^+ + B \rightleftharpoons BH^+ + A \qquad (1)$$

The proton affinities at 298.15(0.00) K, $\Delta_{pa}H^o_{298}(\Delta_{pa}H^o_0)$, of NH_3 and CO have been studied frequently, and numerous data with varying uncertainties are available for both species. Tables 1 and 2 and their footnotes contain a compilation and assessment of literature data for the PAs of NH_3 and CO, respectively.

In 1984, a $\Delta_{pa}H^o_{298}(NH_3)$ value of 853.5 kJ mol^{-1} was recommended by Lias and co-workers[32] based on critical evaluation of different measurements. In the 1990s, the best ab initio computations gave the 298.15 K PA of ammonia as 853.6 kJ mol^{-1} (ref 34, with no uncertainty reported), 853.1 ± 1.3 kJ mol^{-1} (ref 35), and 854.0 ± 1.3 kJ mol^{-1} (ref 36). In 1998, in the most recent compilation of PAs, Hunter and Lias[31] selected the proton affinity of NH_3 computed by Smith and Radom,[34] considering its close match with experimental measurements and the earlier recommendation.[32] At the beginning of this decade, more sophisticated ab initio results were computed for PA(NH_3), which translate to 853.2 kJ mol^{-1} (ref 37) and 853.1 (ref 38) at 298.15 K. The best experimental $\Delta_{pa}H^o_{298}$(CO) was obtained by Traeger[39] in 1985 as 594 ± 3 kJ mol^{-1} by means of dissociative photoionization of formic acid. This value was recommended by Hunter and Lias in their 1998 critical evaluation of literature data.[31] The best previously computed values of $\Delta_{pa}H^o_{298}$(CO) are 593.3 ± 2.1 kJ mol^{-1} published by Komornicki and Dixon[40] and 593.0 kJ mol^{-1} reported by Smith and Radom.[34]

Despite the large amount of information for the PAs of these two molecules, the attendant uncertainties are still considerably larger than what can be achieved from state-of-the-art computations on molecules of this size.[41–55] For small molecules and radicals (at present up to 5−6 "heavy" atoms), first-principles computations of thermochemical quantities are often more accurate than experimental measurements; frequently, the uncertainties of the best computed values can only be surpassed by comprehensively incorporating both empirical and theoretical data in schemes such as the Active Thermochemical Tables (ATcT).[56] Previous computations of energetic quantities other than PAs,[41–55] employing variants of the focal-point analysis (FPA) approach,[57,58] clearly prove the effectiveness of the sophisticated first-principles methods employed in this study.

The present FPA study pushes the ab initio treatment of PAs to new heights by means of the following advances: (a) electron correlation beyond the "gold standard" CCSD(T) level is taken into account by performing coupled cluster computations complete through quadruple excitations and, for NH_3 and NH_4^+, even considering the effect of connected pentuples; (b) FPA limits are determined with all electrons correlated (AE), avoiding any additivity assumptions regarding valence and core correlation; (c) relativistic shifts are evaluated by first-order perturbation theory applied to the mass-velocity and one-electron Darwin terms (MVD1);[59,60] (d) the electronic structure computations go beyond the clamped nucleus assumption by appending diagonal Born−Oppenheimer corrections (DBOC);[61–64] (e) anharmonic zero-point vibrational energies (ZPVEs) have been determined by computing new AE-CCSD(T)/cc-pCVQZ internal coordinate quartic force field representations of the ground-state potential energy surfaces and executing variational vibrational motion computations using exact kinetic energy operators; and (f) the enthalpy increments needed to convert the 0 K computational results to finite temperatures, in the present case 298.15 K, have been obtained by explicitly summing over (ro)vibrational energy levels, obtained mostly through our variational vibrational motion computations, within standard statistical mechanical expressions.

II. Computational Methods

The (aug-)cc-p(C)V*X*Z [*X* = 2(D), 3(T), 4(Q), 5, and 6] families of correlation-consistent, atom-centered Gaussian basis sets[65–68] were employed in this study. The orbital contraction schemes of these basis sets range from [11*s*6*p*2*d*] → [5*s*4*p*2*d*] to [22*s*16*p*10*d*8*f*6*g*4*h*2*i*] → [13*s*12*p*10*d*8*f*6*g*4*h*2*i*] for N, C, O, and from [5*s*2*p*] → [3*s*2*p*] to [11*s*6*p*5*d*4*f*3*g*2*h*] → [7*s*6*p*5*d*4*f*3*g*2*h*] for H. These atomic-orbital basis sets give superior performance in approaching the complete basis set

Table 1. Literature Data for the Proton Affinity of NH_3 (in kJ mol^{-1})

$\Delta_{pa}H^{\circ}_{298}$ ($\Delta_{pa}H^{\circ}_{0}$)	authors and references	comments	$\Delta_{pa}H^{\circ}_{298}$ ($\Delta_{pa}H^{\circ}_{0}$)	authors and references	comments
		Measurements			
851.9 ± 5.4	Ceyer et al. (1979)[110]	*a*	851.4 ± 3.3	Szulejko and McMahon (1993)[27]	*c*
867.8 ± 6.7	Meot-Ner and Sieck (1991)[111]	*b*			
		Ab Initio Computations			
860.2 ± 4.2	Eades et al. (1980)[112]	*d*	852.8	Smith and Radom (1995)[118]	*k*
853.5	DeFrees and McLean (1986)[113]	*e*	853.1 ± 1.3 (846.8)	Martin and Lee (1996)[35]	*l*
853.5 (847.3)	Pople and Curtiss (1987)[114]	*f*	854.0 ± 1.3	Peterson et al. (1998)[36]	*m*
858.6	Del Bene and Shavitt (1990)[115]	*g*	849.3	Seo et al. (2001)[119]	*n*
853.5 (847.3)	Curtiss et al. (1991)[116]	*h*	853.2 (847.0)	Dixon et al. (2001)[37]	*o*
852.3	Del Bene (1993)[117]	*i*	853.1	Parthiban and Martin (2001)[38]	*p*
853.6 (847.4)	Smith and Radom (1993)[34]	*j*			
		Reviews and Evaluations			
853.5 ± 8	Lias et al. (1984)[32]	*q*	853.6	Hunter and Lias (1998)[31]	*r*

a Examination of the photoionization threshold for the appearance of the ammonium ion generated from the ammonia van der Waals dimer gave PA(NH_3) = 203.6 ± 1.3 kcal mol^{-1}. *b* Originally reported as 208.3 kcal mol^{-1} at 600 K, or PA(NH_3) = 207.4 ± 1.6 kcal mol^{-1} at 300 K. This value was soon challenged in ref 27. *c* From temperature-dependent proton transfer equilibrium experiments, originally reported as PA(NH_3) = 203.5 ± 0.8 kcal mol^{-1}. *d* Originally reported as 205.6 ± 1 kcal mol^{-1}. The proton affinity of ammonia was computed at different levels of theory. The SCF calculations were performed using both Gaussian-type orbitals (GTO) and Slater-type orbitals (STO). The STO basis was used in CI calculations with all single and double excitations excluding the top virtual orbital ($\epsilon > 10\ E_h$); correction was made for quadruple excitations. The ZPVE contribution was obtained from SCF harmonic frequencies computed with the GTO basis and scaled to experiment. *e* Computed at the MP4/6-311++G(3*df*,3*pd*) level; originally reported as 204.0 kcal mol^{-1}. *f* Obtained from MP perturbation theory through fourth order with a combination of isogyric reactions and basis set additivity approximations; originally reported as 204.0(202.5) kcal mol^{-1}. *g* Computed at the CISD+Q/6-31+G(2*d*,2*p*) level; originally reported as 213.2 kcal mol^{-1}, at 0 K and without a ZPVE correction. With a MP2/6-31+G(*d*,*p*) harmonic ZPVE correction, the PA estimate becomes 205.2 kcal mol^{-1}. *h* From G2 computations; originally reported as 204.0(202.5) kcal mol^{-1}. *i* Computed at the CCSD + T(CCSD)/aug-cc-pVTZ level; originally reported as 203.7 kcal mol^{-1}. *j* Proton affinities of 31 molecules (including NH_3 and CO) were computed at the G2 level of theory. G2 corresponds to QCISD(T)/6-311+G(3*df*,2*p*)//MP2/6-31G(*d*) computations with zero-point vibrational and empirical "higher-level" corrections. In this case, the higher-level corrections cancel, and thus the G2 proton affinities are purely ab initio. *k* The proton affinities of several molecules, including NH_3 and CO, were computed using the G2(MP2,SVP) procedure, which is similar to G2(MP2) but uses the split-valence plus polarization (SVP) 6-31G(*d*) basis set for the QCISD(T) computations. ZPVE and enthalpy corrections were obtained from scaled HF/6-31G(*d*) vibrational frequencies. *l* Based on CCSD(T)/(aug)-cc-pVXZ (X = 2-5) computations. The largest basis set employed was aug′-cc-pV5Z, where the diffuse functions on hydrogen were removed. Frozen-core CCSD(T) energies were extrapolated to the CBS limit; core correlation and ZPVE corrections were appended. Anharmonic ZPVEs were obtained from a CCSD(T)/cc-pVTZ quartic force field by means of VPT2 with the G_0 term included. Not considered were the following: (a) electron correlation beyond the CCSD(T) level, (b) relativistic shifts, (c) DBOC terms, and (d) nonadditivity of the core correlation effect. With a thermal contribution of 1.50 kcal mol^{-1}, a final result of $\Delta_{pa}H_{298}(NH_3)$ = 203.9 ± 0.3 kcal mol^{-1} was originally reported. We have subtracted off the thermal correction to obtain the 0 K proton affinity listed in the table. *m* Based on computations up to the frozen-core CCSD(T)/aug-cc-pVQZ and CCSD(T)/cc-pV5Z levels; originally reported as 204.1 ± 0.3 kcal mol^{-1}. CCSD(T)/aug-cc-pwCVTZ// CCSD(T)/aug-cc-pVTZ core-valence correlation effects were included. ZPVE contributions were taken from experiment and from scaled theoretical results. *n* Proton affinities of several molecules, including NH_3 and CO, were obtained at the multicoefficient QCISD (MC-QCISD) level of theory. For $\Delta_{pa}H_{298}(NH_3)$, both MP2/6-31G(*d*,*p*) and full multilevel (ML) optimum geometries gave the same results. Zero-point and thermal energies were obtained from MP2/6-31G(*d*,*p*) frequencies scaled by 0.9676. *o* Originally reported as 202.43 kcal mol^{-1} at 0 K; our thermal correction has been used to obtain the corresponding 298.15 K value. A vibrationless proton affinity of 886.05 ± 0.04 kJ mol^{-1} was inferred as the CBS limit of frozen-core CCSD(T). A core correlation correction of −0.54 kJ mol^{-1} was then appended, which may have a sign error considering that our CBS AE-CCSD(T) vibrationless proton affinity is 886.61 kJ mol^{-1}. The approximate ZPVE correction of Dixon et al. deviates from the more accurate result of this study by 1 kJ mol^{-1}. *p* From W2 theory, originally reported as 203.9 kcal mol^{-1}. The corresponding W1 values are 853.9(847.6) kJ mol^{-1} at 298.15(0.0) K. This paper also gives a 298.15 K value of 849.8 kJ mol^{-1} from the G3 method. *q* Recommendation based on a critical evaluation of different measurements. The uncertainty is a general value stated in the Introduction. *r* The recommended proton affinity was based on a close match between the computed value of Smith and Radom[34] and the previously recommended[32] value.

(CBS) limit in a systematic fashion during traditional electronic structure computations. When (aug-)cc-pCV*X*Z sets were used to effectuate the correlation of core electrons for C, N, and O, the corresponding (aug-)cc-pV*X*Z functions were utilized for H.

Reference electronic wave functions were determined by the single-configuration restricted Hartree−Fock (RHF) method.[69] Electron correlation was accounted for by the coupled-cluster (CC) method[70,71] including all single and double (CCSD), triple (CCSDT), and quadruple (CCSDTQ) excitations.[72] The CCSD(T)[73] and CCSDT(Q)[74,75] methods, which include perturbative (T) and (Q) terms for connected triple and quadruple excitations, respectively, were also used extensively. For NH_3 and NH_4^+, the CCSDTQ(P) method[74] was also employed with the aug-cc-pCVDZ basis set. All electrons were included, unless otherwise noted, in the active space for the correlation energy computations.

In the spirit of the FPA approach,[52–55,57,58] the aug-cc-pCV*X*Z sequences of electronic energies were extrapolated to determine CBS limits. For extrapolation of the Hartree−Fock energies, two-[76,77] and three-parameter[78] exponential functions of the cardinal number X were used

$$E_X^{HF} = E_{CBS}^{HF} + a(X+1)e^{-9\sqrt{X}} \quad (2)$$

and

$$E_X^{HF} = E_{CBS}^{HF} + ae^{-bX} \quad (3)$$

For the molecules considered here, the two extrapolation formulas, when applied with the largest possible X values, gave RHF energies and PAs in agreement to better than 0.04 and 0.001 kJ mol^{-1}, respectively. The perhaps slightly more accurate[76,79] two-parameter results are reported in Tables 3 and 4 as the CBS RHF proton affinities for NH_3 and CO,

Table 2. Literature Data for the Proton Affinity of CO (in kJ mol^{-1})

$\Delta_{pa}H^{\circ}_{298}$ ($\Delta_{pa}H^{\circ}_{0}$)	authors and references	comments	$\Delta_{pa}H^{\circ}_{298}$ ($\Delta_{pa}H^{\circ}_{0}$)	authors and references	comments
		Measurements			
594 ± 3	Traeger (1985)[39]	*a*	591.6 ± 4.2	Adams et al. (1989)[28]	*b*
		Ab Initio Computations			
(597.5)	Del Bene et al. (1982)[120]	*c*	592.4(586.5) ± 2.1	Martin et al. (1993)[104]	*j*
599.6	Ikuta (1984)[121]	*d*	594.1	Botschwina et al. (1993)[124]	*k*
591.6	Dixon et al. (1984)[122]	*e*	597.1	Smith and Radom (1995)[118]	*l*
599.1	DeFrees and McLean (1986)[113]	*f*	594.1	Mladenović and Schmatz (1998)[98]	*m*
593.3 ± 2.1	Komornicki and Dixon (1992)[40]	*g*	(587.8) ± 2.1	van Mourik et al. (2000)[125]	*n*
(589.3)	Ma et al. (1992)[123]	*h*	599.9	Seo et al. (2001)[119]	*o*
593.0 (587.1)	Smith and Radom (1993)[34]	*i*			
		Reviews and Evaluations			
594 ± 6	Lias et al. (1984)[32]	*p*	594 ± 3	Hunter and Lias (1998)[31]	*r*
593.7	Szulejko and McMahon (1993)[27]	*q*			

a From examination of the dissociative photoionization of a number of formyl compounds (in this case HCOOH) to yield the formyl cation HCO^+. *b* The enthalpy of the reaction $HN_2O^+ + CO \rightarrow HCO^+ + N_2O$ was obtained from a van't Hoff plot of the measured equilibrium constant at different temperatures, yielding PA(CO) = 141.4 ± 1 kcal mol^{-1} at 300 K. *c* PA at 0 K is 142.8 kcal mol^{-1} as originally reported. SCF geometries and frequencies were used along with partial fourth-order perturbation theory and the 6-31G** basis. *d* The PA value at 0 K computed at the MP3/6-31G** level is originally reported as 149.7 kcal mol^{-1}. This value does not contain the ZPVE correction. In this table a value of 143.3 kcal mol^{-1} is given as originally cited by Komornicki and Dixon.[40] *e* Originally reported as 141.4 kcal mol^{-1}. Geometries and frequencies were determined at the CI(SD) level using a TZP basis set. *f* PA(CO) = 143.2 kcal mol^{-1} was computed using MP4-SDTQ fourth-order perturbation theory with an extensive one-particle basis including *f* and *d* double polarization functions on the heavy atoms and hydrogen, respectively. CI(D)/ 6-31G(*d*) optimum geometries and scaled SCF frequencies were used. *g* PA(CO) was computed at the SCF, MP2, CCSD, and CCSD(T) levels of theory using three different basis sets, with account of BSSE. Empirical geometrical parameters and vibrational frequencies were used. A final value of 141.8 ± 0.5 kcal mol^{-1} was originally reported. *h* Based on QCISD(T)/6-311+G(3*df*,2*p*) energies and scaled MP2/6-31G(*d*) harmonic ZPVE estimates. *i* See comment *j* of Table 1. *j* Reported as 141.59(140.17) kcal mol^{-1}, based on frozen-core CCSD(T)/cc-pVQZ computations. *k* Originally reported as 142.0 kcal mol^{-1}. This value was based on an equilibrium CBS CCSD(T) proton affinity of 147.6 kcal mol^{-1}, a ZPVE contribution of −7.0 kcal mol^{-1}, a difference in the mean vibrational energies between HCO^+ and CO of −0.1 kcal mol^{-1}, a difference in the translational energies of 0.9 kcal mol^{-1}, and a $\Delta(PV)$ term of 0.6 kcal mol^{-1}. *l* See comment *k* of Table 1. *m* Originally reported as 142.0 kcal mol^{-1}, based on frozen-core CCSD(T)/cc-pVQZ (no *f* functions on hydrogens) energies and variational ZPVE estimates. *n* The 0 K value, 140.49 kcal mol^{-1} with an uncertainty estimate of ±0.5 kcal mol^{-1}, was reported in Table 6 as $D_0(HCO^+)$. *o* See comment *n* of Table 1. The MC-QCISD//ML value is listed in the table; at the MC-QCISD//MP2/6-31G(*d*,*p*) level, the PA is 598.7 kJ mol^{-1}. *p* Selected value from experimental measurements, taken from Table 2 of the compilation. The uncertainty is a general value stated in the Introduction. *q* PA(CO) = 141.9 kcal mol^{-1} was chosen because it was midway between the Traeger[39] experimental value and the ab initio calculation of Komornicki and Dixon[40] and because it agreed with the previous assessment from the NIST proton affinity tables. *r* Recommended value is the proton affinity originally measured by Traeger.[39]

Table 3. Focal-Point Analysis of the All-Electron Nonrelativistic Born−Oppenheimer Proton Affinity (ΔE_e, kJ mol^{-1}) of NH_3 at 0 K[a,b]

	ΔE_e(RHF)	δ[CCSD]	δ[CCSD(T)]	δ[CCSDT]	δ[CCSDT(Q)]	δ[CCSDTQ]	δ[CCSDTQ(P)]	ΔE_e[CCSDTQ(P)]
aug-cc-pCVDZ	898.68	−11.59	−2.95	−0.08	−0.30	+0.04	−0.005	883.80
aug-cc-pCVTZ	903.28	−13.04	−4.00	+0.04	−0.33	[+0.04]	[−0.005]	[885.98]
aug-cc-pCVQZ	903.87	−13.08	−4.18	+0.08	[−0.33]	[+0.04]	[−0.005]	[886.39]
aug-cc-pCV5Z	904.03	−13.06	−4.24	[+0.09]	[−0.33]	[+0.04]	[−0.005]	[886.52]
aug-cc-pCV6Z	904.06	−13.10	−4.26	[+0.10]	[−0.33]	[+0.04]	[−0.005]	[886.48]
CBS[c,d]	[904.06]	[−13.15]	[−4.30]	[+0.11]	[−0.33]	[+0.04]	[−0.005]	[886.43]

a Based on AE-CCSD(T)/aug-cc-pCVQZ reference structures. *b* For the all-electron coupled cluster computations, the symbol δ denotes the increments in the proton affinity, ΔE_e, with respect to the preceding level of theory. Brackets signify increments obtained from basis set extrapolations or additivity approximations. *c* The complete basis set RHF, CCSD, and CCSD(T) entries were obtained from aug-cc-pCV(5,6)Z energies using the two-parameter extrapolation formulas given in eqs 2 and 4. The bracketed CCSDT entries result from direct extrapolation of aug-cc-pCV(T,Q)Z increments (rather than individual energies). *d* Application of the three-parameter formula (eq 3) with aug-cc-pCV(Q,5,6)Z energies for the extrapolation to the RHF CBS limit also results in 904.06 kJ mol^{-1}.

respectively. The CCSD and CCSD(T) electron correlation energies ($\varepsilon_X \equiv E_X^{CC} - E_X^{HF}$) were extrapolated using a two-parameter polynomial formula[80]

$$\varepsilon_X = \varepsilon_{CBS} + bX^{-3} \quad (4)$$

The δ[CCSDT] correlation increments in the focal-point analyses were extrapolated likewise.

The program packages ACESII,[81,82] MRCC (interfaced to ACESII),[83,84] and MOLPRO[85,86] were used for the electronic structure computations. Analytic gradient techniques[87−89]were utilized to obtain optimum geometric structures at the all-electron CCSD(T) level with the cc-pVQZ, cc-pCVQZ, and aug-cc-pCVQZ basis sets. The AE-CCSD(T)/aug-cc-pCVQZ structures were adopted for all electronic structure computations involved in the FPA analyses. The DBOC corrections were computed at the frozen-core CCSD/aug-cc-pVDZ level within the formalism of ref 62 utilizing a private version of the conjoined ACESII and MRCC program packages. Relativistic effects were evaluated by first-order perturbation theory applied to the mass-velocity and one-electron Darwin terms (MVD1),[59] as implemented in ACESII. For this purpose, AE-CCSD(T)/aug-cc-pCVTZ wave functions were employed.

The full quartic force fields of NH_3 and NH_4^+ were determined in internal coordinates at the AE-CCSD(T)/cc-

Table 4. Focal-Point Analysis of the All-Electron Nonrelativistic Born-Oppenheimer Proton Affinity (ΔE_e, kJ mol^{-1}) of CO at 0 K[a]

	ΔE_e(RHF)	δ[CCSD]	δ[CCSD(T)]	δ[CCSDT]	δ[CCSDT(Q)]	δ[CCSDTQ]	ΔE_e(CCSDTQ)
aug-cc-pCVDZ	597.14	+17.12	+0.04	−0.03	+0.29	−0.23	614.34
aug-cc-pCVTZ	603.37	+14.28	−0.17	−0.19	+0.28	[−0.23]	[617.33]
aug-cc-pCVQZ	604.07	+13.50	−0.23	−0.17	[+0.28]	[−0.23]	[617.22]
aug-cc-pCV5Z	604.14	+13.06	−0.27	[−0.17]	[+0.28]	[−0.23]	[616.81]
aug-cc-pCV6Z	604.16	+12.83	−0.29	[−0.17]	[+0.28]	[−0.23]	[616.58]
CBS[b]	[604.16]	[+12.52]	[−0.32]	[−0.17]	[+0.28]	[−0.23]	[616.25]

[a] See footnotes a, *b*, and *c* of Table 3. [b] Application of the three-parameter formula (eq 3) with aug-cc-pCV(Q,5,6)Z energies for the extrapolation to the RHF CBS limit yields the same value as given in the table.

Table 5. Auxiliary Corrections (DBOC, MVD1, and ZPVE) and Final Proton Affinities, All in kJ mol^{-1}, for the NH_3/NH_4^+ and CO/HCO^+ Systems

	DBOC[a]	relativistic MVD1[b]	ZPVE[c]	final $\Delta_{pa}H_0^o$
NH_3	7.038	−76.251	89.17	846.40
NH_4^+	7.225	−76.112	128.87	
CO	10.637	−176.656	12.94	586.51
HCO^+	11.052	−176.533	42.14	

[a] Frozen-core CCSD/aug-cc-pVDZ//AE-CCSD(T)/aug-cc-pCVQZ level. [b] AE-CCSD(T)/aug-cc-pCVTZ//CCSD(T)/aug-cc-pCVQZ level. [c] For NH_3 and NH_4^+ accurate anharmonic ZPVEs were computed in this study by variational vibrational methods, as described in the text. The ZPVEs for CO and HCO^+ were taken from refs 97 and 98, respectively.

pCVQZ level of theory, using high-precision energy points from the MOLPRO package and carefully validated higher-order finite-difference procedures built into the code INTDIF2005.[90,91] The force fields were transformed by INTDER2005[92–94] into a representation with Simons−Parr−Finlan (SPF)[95] bond-stretching coordinates for use in the variational vibrational procedures.

The variational vibrational computations were performed with a recently developed program called DEWE,[96] which employs a discrete variable representation (DVR) of the Eckart−Watson (EW) Hamiltonian and involves an exact transformation from normal to internal coordinates, thus allowing the exact inclusion of an arbitrary potential. For NH_3 and NH_4^+, the AE-CCSD(T)/cc-pCVQZ quartic force fields in SPF coordinates were used to obtain accurate zero-point vibrational energies (ZPVEs), as well as low-lying vibrational band origins (VBOs) used for evaluating thermal enthalpy increments via direct summation. Variational ZPVEs were taken from the literature for carbon monoxide[97] and the HCO^+ cation.[98]

III. Results and Discussion

The primary focal-point analyses of the proton affinities of NH_3 and CO are presented in Tables 3 and 4, respectively. The auxiliary data for the DBOC, relativistic, and ZPVE corrections are given in Table 5. The absolute total energies of the NH_3, NH_4^+, CO, and HCO^+ species, from which the relative energies are derived, are tabulated in Supporting Information (Tables S1−S4) for this paper. The less stable HOC^+ isomer, with a relative energy and isomerization barrier of 166 and 321 kJ mol^{-1} above HCO^+, respectively,[98] was not considered in this study.

III.1. Reference Structures. For all species, AE-CCSD(T)/aug-cc-pCVQZ equilibrium structures were used as reference geometries in our final FPA computations. Accordingly, for NH_3 we employed [r_e(N−H), θ_e(H−N−H)] = (1.0115 Å, 106.71°), which are in almost perfect agreement with both a longstanding empirical structure[99] (1.0116 Å, 106.7°), and the optimum parameters (1.0109 Å, 106.81°) given by the highest current levels of ab initio theory.[100] Our AE-CCSD(T)/aug-cc-pCVQZ bond distance in NH_4^+ (T_d point-group symmetry) is 1.0208 Å. The corresponding bond length in CO is 1.1293 Å, while for the linear HCO^+ cation, r_e(C−H) = 1.0925 Å and r_e(C−O) = 1.1066 Å. These AE-CCSD(T)/aug-cc-pCVQZ equilibrium distances are also in excellent agreement with the best experimental bond lengths, namely, r_e(C−O) = 1.1283 Å for carbon monoxide[101] and [r_e(C−H), r_e(C−O)] = (1.0916, 1.1056) Å for HCO^+.[102] As expected,[103] the close matching of experimental and computed structures requires the correlation of all electrons, as demonstrated also by previously reported[104] frozen-core CCSD(T)/cc-pVQZ results: r_e(C−O) = 1.1314 Å for CO and [r_e(C−H), r_e(C−O)] = (1.0935, 1.1086) Å for HCO^+.

FPA energies and corresponding proton affinities were also determined at our AE-CCSD(T)/cc-pVQZ equilibrium structures, as reported in the Supporting Information (Tables S5−S10). These alternative results show that when the reference structures have an accuracy better than about 0.001 Å and 0.5°, the precise choice has a rather small influence, less than 0.15 kJ mol^{-1}, on the computed PA values. Nevertheless, unlike most previous studies, the level of precision sought here requires the choice of the reference structures to be taken into account in computing proton affinities and ascribing uncertainties to them.

III.2. Nonrelativistic Born−Oppenheimer Proton Affinities of NH_3 and CO. The present FPA analysis starts at the RHF/aug-cc-pCVDZ level, which yields 898.68 and 597.14 kJ mol^{-1} for the vibrationless proton affinities of NH_3 and CO, respectively. Enlarging the basis set to aug-cc-pCV6Z increases the Hartree−Fock PAs of NH_3 and CO by 5.38 and 7.02 kJ mol^{-1}, respectively. The differences between the aug-cc-pCV6Z and CBS Hartree−Fock proton affinities are minuscule, less than 0.005 kJ mol^{-1} for both molecules.

As expected, the electron correlation energies and the lowest-order FPA increments exhibit considerably slower basis set convergence. The aug-cc-pCV6Z and extrapolated (CBS) δ[CCSD] increments deviate by 0.05 and 0.31 kJ mol^{-1} for the proton affinities of NH_3 and CO, respectively. The level of convergence of the CCSD increment in the CO case largely determines the eventual uncertainty in the FPA value of $\Delta_{pa}H_0^o$(CO). Consistent with the foundations of the

FPA approach, all of the higher-order correlation increments converge rapidly to their respective CBS limits; for example, the changes in the δ[CCSD(T)] increments beyond aug-cc-pCV6Z are only about 0.03 kJ mol^{-1}.

The final CBS estimates of the coupled-cluster correlation energy increments for $\Delta_{pa}H_0^o(NH_3)$ are −13.15, −4.19, and −0.28 kJ mol^{-1} for the full treatments of single and double (SD), triple (T), and quadruple (Q) excitations, in order. The same increments for $\Delta_{pa}H_0^o(CO)$ are +12.52, −0.48, and +0.05 kJ mol^{-1}, in order. For both NH_3 and CO, the total correlation contribution to the proton affinity is about 2.0%. Interestingly, the effect of higher-order electron correlation past CCSD on the proton affinity of carbon monoxide is 1 order of magnitude smaller than in the NH_3 case, contrary to general expectations for an electron-dense, multiply bonded system such as CO/HCO$^+$. From the sequences of CCSD, CCSDT, and CCSDTQ values, it appears that full inclusion of pentuple excitations would decrease both $\Delta_{pa}H_0^o(NH_3)$ and $\Delta_{pa}H_0^o(CO)$ on the order of 0.01 kJ mol^{-1}. The actual CCSDTQ(P)/aug-cc-pCVDZ increment of −0.005 kJ mol^{-1} computed for the proton affinity of NH_3 supports this expectation.

Our final values from Tables 3 and 4, with conservative uncertainty estimates, for the all-electron nonrelativistic proton affinities (without ZPVE) at 0 K are 886.43 ± 0.10 and 616.25 ± 0.20 kJ mol^{-1} for NH_3 and CO, respectively.

III.3. Relativistic Effects. Relativistic effects on the proton affinities were computed by applying first-order perturbation theory to the mass-velocity and one-electron Darwin terms (MVD1).[59,60] Detailed previous studies, *e.g.*, refs 60 and 105, suggest that for systems such as those investigated here, the accuracy of MVD1 relative energy corrections is excellent, as compared to those from more complicated multicomponent methods. Employing the AE-CCSD(T)/aug-cc-pCVTZ level of theory, the relativistic energy shifts for the PAs of NH_3 and CO are −0.14 and −0.12 kJ mol^{-1}, respectively. Similar results, −0.15 and −0.11 kJ mol^{-1}, are obtained from the lower-level CCSD/aug-cc-pCVDZ method, suggesting that these quantities do not contribute any significant amount to the uncertainties of our final PAs.

III.4. DBOC Contributions. The effects of computing electronic wave functions beyond the clamped nucleus formalism can be estimated from diagonal Born−Oppenheimer corrections (DBOCs).[106] The DBOC contributions to the proton affinities at the frozen-core CCSD/aug-cc-pVDZ level of theory are −0.19 kJ mol^{-1} (NH_3) and −0.42 kJ mol^{-1} (CO). For comparison, the corresponding HF/aug-cc-pVDZ numbers are −0.14 kJ mol^{-1} (NH_3) and −0.36 kJ mol^{-1} (CO). Although the DBOCs for the total energies of the individual species are 1 order of magnitude smaller than their relativistic counterparts (Table 5), the DBOC shifts in the proton affinities are sizable, especially for carbon monoxide. To achieve the level of accuracy sought in this study for proton affinities, the standard Born−Oppenheimer approximation is clearly not sufficient.

III.5. Zero-Point Vibrational Energies. The ZPVEs of all four species were obtained from variational nuclear motion computations. The PESs of NH_3 and NH_4^+ were represented by newly determined quartic internal coordinate force fields obtained at the AE-CCSD(T)/cc-pCVQZ level of theory. The force constants are tabulated in the Supporting Information (Tables S11 and S12), in terms of SPF coordinates for the bond stretches with reference distances corresponding to the optimized equilibrium ones.

The converged variational, anharmonic ZPVEs of NH_3 and NH_4^+ are 89.17 and 128.87 kJ mol^{-1}, respectively. For comparison, the harmonic ZPVEs are 90.69 kJ mol^{-1} (NH_3) and 131.02 kJ mol^{-1} (NH_4^+). Our anharmonic ZPVE for ammonia is consistent with earlier variational ZPVEs of 88.87 kJ mol^{-1},[107] 89.08 kJ mol^{-1},[108] and 89.25 kJ mol^{-1}.[109] It is likely that our computation on the 5-atomic NH_4^+ molecular ion is the first variational determination of the low-lying vibrational levels of this cation. Therefore, the ZPVEs and the first several vibrational band origins (VBOs) are reported for NH_3 and NH_4^+ in the Supporting Information (Table S13). Our computed anharmonic (harmonic) ZPVE contribution to the PA of ammonia is −39.70 (−40.33) kJ mol^{-1}.

Because our vibrational computations employed a uniform, converged variational method and potentials from the same level of electronic structure theory, the uncertainty in our ZPVE contribution to the proton affinity should be considerably smaller than the uncertainties in the individual ZPVEs. In support of this contention, the AE-CCSD(T)/cc-pCVQZ harmonic ZPVEs for NH_3 and NH_4^+ are 1.52 and 2.15 kJ mol^{-1} too large, respectively, compared to the corresponding variational anharmonic values, but the harmonic NH_3 − NH_4^+ ZPVE difference is in error by only −0.63 kJ mol^{-1}. In comparison to previous ZPVE contributions (Δ_{ZPVE}) computed for PA(NH_3),[35–38] our value is within 0.05 kJ mol^{-1} of the second-order vibrational perturbation theory (VPT2) result of Martin and Lee,[35] who included the leading G_0 term in their analysis and employed slightly less accurate quartic force fields from the frozen-core CCSD(T)/cc-pVTZ level of theory; however, previous Δ_{ZPVE} estimates of –38.66 kJ mol^{-1} (ref 37) and –39.25 kJ mol^{-1} (determined by the standard W1 protocol from scaled B3LYP harmonic frequencies)[38] are significantly different. Overall, we are confident that the uncertainty in our ZPVE term for PA(NH_3) is no larger than 0.15 kJ mol^{-1}.

The variationally computed ZPVE values for CO and HCO$^+$ were taken from the literature. We chose ZPVE(CO) = 12.94 kJ mol^{-1} from an experimentally derived RKR potential,[97] which is confirmed by the 12.96 kJ mol^{-1} value we obtained variationally from the SPF quartic force field given by AE-CCSD(T)/cc-pCVQZ theory. In 1998, Mladenović and Schmatz[98] performed variational rovibrational computations using a new analytic global PES of HCO$^+$/HOC$^+$ determined at the frozen-core CCSD(T)/cc-pVQZ level of theory; this work yielded ZPVE(HCO$^+$) = 42.14 kJ mol^{-1}. Thus, the ZPVE contribution to the proton affinity of CO is −29.20 kJ mol^{-1}, with an uncertainty not larger than 0.05 kJ mol^{-1}. This value is considerably different from that used by Komornicki and Dixon,[40] −28.5 kJ mol^{-1}, based on experimental fundamental frequencies.

III.6. Final Proton Affinities at 0 K. The final proton affinities of (NH_3, CO) at 0 K are obtained by summing the vibrationless all-electron nonrelativistic proton affinities

(886.43, 616.25) kJ mol^{-1}, the relativistic energy shifts (−0.14, −0.12 kJ mol^{-1}), the DBOC corrections (−0.19, −0.42 kJ mol^{-1}), and the ZPVE contributions (−39.70, −29.20 kJ mol^{-1}). Accordingly, we determine $\Delta_{pa}H_0^o(NH_3)$ = 846.4 ± 0.3 and $\Delta_{pa}H_0^o(CO)$ = 586.5 ± 0.2 kJ mol^{-1}. The uncertainties ascribed to these values correspond to 2σ and arise mostly from the reference geometry and ZPVE effects for NH_3 and the CBS extrapolation of the δ[CCSD] correlation increment for CO. We estimate that all other sources of error collectively contribute no more than 0.1 kJ mol^{-1} to the uncertainties of our proton affinities. The best previous theoretical $\Delta_{pa}H_0^o(NH_3)$ value (846.8 kJ mol^{-1}), computed by Martin and Lee,[35] agrees well with our improved result. For $\Delta_{pa}H_0^o(CO)$, the G2 computation (587.1 kJ mol^{-1}) of Smith and Radom[34] is within 0.6 kJ mol^{-1} of our converged ab initio result.

III.7. Final Proton Affinities at 298.15 K. Proton affinities, as universally employed, are quantities defined at a finite temperature, usually 298.15 K. Therefore, the ab initio PA values determined in the previous subsection, referring to 0 K, need to be converted to 298.15 K.

The proton affinity of a neutral molecule A is defined as the enthalpy change for the isogyric reaction $AH^+ \rightarrow A + H^+$. The proton affinity at 0 K can be obtained by performing quantum chemical computations for the electronic energies and zero-point vibrational energies of the A and AH^+ species. To compute PAs at a nonzero temperature ($\Delta_{pa}H_T^o$), the heat capacities of the species have to be taken into account as

$$\Delta_{pa}H_T^o = \Delta_{pa}H_0^o + \int_0^T C_p(A)dT - \int_0^T C_p(AH^+)dT + \frac{5}{2}RT \quad (5)$$

The most important term in eq 5 is the last one, which is the translational enthalpy of H^+. The two integrals in eq 5 cancel if one assumes that A and AH^+ are classical rigid rotors of the same type (linear or nonlinear). Therefore, some studies have considered only the $5/2RT$ term (6.20 kJ mol^{-1} at 298.15 K) as the thermal contribution to PAs.

Vibrational enthalpy effects can be treated by evaluating partition functions via direct summation of variationally computed vibrational energy levels (Table S13). In this way, we obtained vibrational enthalpy contributions of +0.06 and −0.35 kJ mol^{-1} to the 298.15 K proton affinities of NH_3 and CO, respectively. These corrections are certainly not negligible for our target accuracy, especially in the carbon monoxide case. The large effect on PA(CO) is due to the disparity between the relatively low bending frequency (830.7 cm^{-1}) of HCO^+ and the stretching fundamental of CO (2143.3 cm^{-1}).

To evaluate rotational enthalpy contributions, we employed the usual rigid-rotor analytic formulas for rotational energy levels of linear molecules and symmetric and spherical tops in terms of equilibrium rotational constants. The rigid-rotor rotational energies were appended to our variationally computed vibrational levels in the direct summations for the rotational−vibrational partition functions. We found that the rotational contribution to the 298.15 K proton affinity of NH_3 is −0.02 kJ mol^{-1}, whereas the corresponding rotational effect for CO is negligible (<10^{-3} kJ mol^{-1}).

Adding the translational enthalpy of H^+ (6.20 kJ mol^{-1}) to our vibrational and rotational corrections yields total thermal contributions of +6.24 and +5.85 kJ mol^{-1} for the 298.15 K proton affinities of NH_3 and CO, respectively. Therefore, we arrive at the final values $\Delta_{pa}H_{298}^o(NH_3)$ = 852.6 ± 0.3 kJ mol^{-1} and $\Delta_{pa}H_{298}^o(CO)$ = 592.4 ± 0.2 kJ mol^{-1}.

IV. Conclusions

The highest levels of electronic structure theory currently feasible have been employed in focal-point analyses to systematically converge on the 0 K proton affinities of NH_3 and CO, two molecules fixing the high and low ends of the absolute PA scale. Moreover, thermal contributions to these proton affinities have been evaluated by direct summation of partition functions over computed (ro)vibrational energy levels. The current study is another methodological milestone for ab initio quantum chemistry because we pinpoint proton affinities to 0.2−0.3 kJ mol^{-1}, roughly an order of magnitude more precise than previous PA measurements and critical evaluations. In the process, a number of valuable observations are made regarding the performance of state-of-the-art theoretical methods: (1) each step in the coupled cluster series CCSD → CCSDT → CCSDTQ → CCSDTQ(P) reduces the electron correlation error in the PAs by at least 90%, and CCSDTQ appears sufficient to converge within 0.01 kJ mol^{-1} of the full configuration interaction (FCI) limit; (2) somewhat surprisingly, the coupled cluster convergence to the FCI proton affinity is slower for NH_3/NH_4^+ than for the electron-dense, multiply bonded CO/HCO^+ system; (3) the perturbative CCSDT(Q) method reproduces the full CCSDTQ effect quite well for PA(NH_3) but not for PA(CO); (4) conventional CCSD(T) computations with the aug-cc-pCV6Z basis appear to be within 0.05 kJ mol^{-1} of the corresponding CBS limit for PA(NH_3), but the aug-cc-pCV6Z incompleteness error is still 0.3 kJ mol^{-1} for PA(CO); (5) one of the largest sources of uncertainty in the first-principles determination of highly accurate PAs of polyatomic molecules is clearly the ZPVE correction; for example, use of the harmonic approximation in evaluating the effect of ZPVE on PA(NH_3) engenders a 0.63 kJ mol^{-1} error; (6) the Born−Oppenheimer approximation is not satisfactory in predicting PA(CO) to our target accuracy, as DBOCs shift this quantity by 0.42 kJ mol^{-1}; (7) neglecting terms other than the translation enthalpy of H^+ in determining the 298.15 K proton affinity of CO causes a 0.35 kJ mol^{-1} error; and (8) MVD1 relativistic effects shift both PA(NH_3) and PA(CO) downward by just over 0.1 kJ mol^{-1} and are the smallest auxiliary corrections considered here.

The final 298.15 K results determined in this study are $\Delta_{pa}H_{298}^o(NH_3)$ = 852.6 ± 0.3 kJ mol^{-1} and $\Delta_{pa}H_{298}^o(CO)$ = 592.4 ± 0.2 kJ mol^{-1}. These values supersede all previous determinations due to the rigor and precision with which they have been pinpointed. In particular, our proton affinities improve on the values recommended in the 1998 critical evaluation of Hunter and Lias:[31] 853.6 kJ mol^{-1} for NH_3 based on the computations of Smith and Radom,[34] and 594 ± 3 kJ mol^{-1} for CO given by the experiments of Traeger.[39] It is not likely that the uncertainties in our proton affinities

can be significantly lowered anytime soon, and thus these PAs should anchor the proton affinity scale for the foreseeable future.

Acknowledgment. The authors are indebted to Drs. Á. Somogyi and B. Ruscic for comments on the manuscript. The research performed in Budapest was supported by the Hungarian Scientific Research Fund (OTKA T47185 and T72885). The research at the University of Georgia was supported by the US National Science Foundation, grant CHE-045144. The collaboration between Budapest and Athens was aided by an NSF-MTA-OTKA travel fund (Grant INT-0312355). The computing resources of the NERSC Facility of the Lawrence Berkeley National Laboratory were used to determine the quartic force fields. Professors J. Gauss and J. Stanton are thanked for providing a private version of ACESII that has the capability of computing DBOC energies at the CCSD level.

Supporting Information Available: Compilation of all total energies utilized in the focal point analyses of the proton affinities and tabulation of the quartic force fields and lowest-lying variational vibrational levels of NH_3 and NH_4^+. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Biemann, K.; Martin, S. A. *Mass Spectrom. Rev.* **1987**, *6*, 1–75.

(2) Tang, X. J.; Thibault, P.; Boyd, R. K. *Anal. Chem.* **1993**, *65*, 2824–2834.

(3) Tang, X. J.; Boyd, R. K. *Rapid Commun. Mass Spectrom.* **1992**, *6*, 651–657.

(4) Tsang, C. W.; Harrison, A. G. *J. Am. Chem. Soc.* **1976**, *98*, 1301–1308.

(5) Harrison, A. G.; Yalcin, T. *Int. J. Mass Spectrom. Ion Proc.* **1997**, *165*, 339–347.

(6) Burlet, O.; Yang, C. Y.; Gaskell, S. J. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 337–344.

(7) Cox, K. A.; Gaskell, S. J.; Morris, M.; Whiting, A. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 522–531.

(8) Jones, J. L.; Dongre, A. R.; Somogyi, Á.; Wysocki, V. H. *J. Am. Chem. Soc.* **1994**, *116*, 8368–8369.

(9) Dongre, A. R.; Somogyi, Á.; Wysocki, V. H. *J. Mass Spectrom.* **1996**, *31*, 339–350.

(10) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *J. Am. Chem. Soc.* **1999**, *121*, 5142–5154.

(11) McCormack, A. L.; Somogyi, Á.; Dongre, A. R.; Wysocki, V. H. *Anal. Chem.* **1993**, *65*, 2859–2872.

(12) Somogyi, Á.; Wysocki, V. H.; Mayer, I. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 704–717.

(13) Komáromi, I.; Somogyi, Á.; Wysocki, V. H. *Int. J. Mass Spectrom.* **2005**, *241*, 315–323.

(14) Huang, Y.; Triscari, J. M.; Pasa-Tolic, L.; Anderson, G. A.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *J. Am. Chem. Soc.* **2004**, *126*, 3034–3035.

(15) Tabb, D. L.; Smith, L. L.; Breci, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R. *Anal. Chem.* **2003**, *75*, 1155–1163.

(16) Kapp, E. A.; Schutz, F.; Reid, G. E.; Eddes, J. S.; Moritz, R. L.; O'Hair, R. A. J.; Speed, T. P.; Simpson, R. J. *Anal. Chem.* **2003**, *75*, 6251–6264.

(17) Paizs, B.; Suhai, S. *Mass Spec. Rev.* **2005**, *24*, 508–548.

(18) Császár, A. G.; Perczel, A. *Prog. Biophys. Mol. Biol.* **1999**, *71*, 243–309.

(19) Császár, A. G. *J. Am. Chem. Soc.* **1992**, *114*, 9568–9575.

(20) Császár, A. G. *J. Phys. Chem. A* **1996**, *100*, 3541–3551.

(21) Demaison, J.; Császár, A. G.; Kleiner, I.; Møllendal, H. *J. Phys. Chem. A* **2007**, *111*, 2574–2586.

(22) Kasalová, V.; Allen, W. D.; Schaefer, H. F.; Czinki, E.; Császár, A. G. *J. Comput. Chem.* **2007**, *28*, 1373–1383.

(23) Czinki, E.; Császár, A. G. *Chem. Eur. J.* **2003**, *9*, 1008–1019.

(24) Allen, W. D.; Czinki, E.; Császár, A. G. *Chem. Eur. J.* **2004**, *10*, 4512–4517.

(25) Paizs, B.; Suhai, S.; Hargittai, B.; Hruby, V. J.; Somogyi, Á. *Int. J. Mass Spectrom.* **2002**, *219*, 203–232.

(26) Meot-Ner (Mautner) , M. *Int. J. Mass Spectrom.* **2003**, *227*, 525–554.

(27) Szulejko, J. E.; McMahon, T. B. *J. Am. Chem. Soc.* **1993**, *115*, 7839–7848.

(28) Adams, N. G.; Smith, D.; Tichy, M.; Javahery, G.; Twiddy, N. D.; Ferguson, E. E. *J. Chem. Phys.* **1989**, *91*, 4037–4042.

(29) Ruscic, B.; Schwarz, M.; Berkowitz, J. *J. Chem. Phys.* **1989**, *91*, 6772–6779.

(30) Traeger, J. C.; Kompe, B. M. *Org. Mass. Spectrom.* **1991**, *26*, 209–214.

(31) Hunter, E. P. L.; Lias, S. G. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413–656.

(32) Lias, S. G.; Liebman, J. F.; Levin, R. D. *J. Phys. Chem. Ref. Data* **1984**, *13*, 695–808.

(33) Traeger, J. C. *Rapid Comm. Mass. Spectrom.* **1996**, *10*, 119–122.

(34) Smith, B. J.; Radom, L. *J. Am. Chem. Soc.* **1993**, *115*, 4885–4888.

(35) (a) Martin, J. M. L.; Lee, T. J. *Chem. Phys. Lett.* **1996**, *258*, 136–143. (b) Martin, J. M. L.; Lee, T. J. *Chem. Phys. Lett.* **1996**, *258*, 129–135.

(36) Peterson, K. A.; Xantheas, S. S.; Dixon, D. A.; Dunning, T. H., Jr. *J. Phys. Chem. A* **1998**, *102*, 2449–2454.

(37) Dixon, D. A.; Feller, D.; Peterson, K. A. *J. Chem. Phys.* **2001**, *115*, 2576–2581.

(38) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014–6029.

(39) Traeger, J. C. *Int. J. Mass Spectrom. Ion Proc.* **1985**, *66*, 271–282.

(40) Komornicki, A.; Dixon, D. A. *J. Chem. Phys.* **1992**, *97*, 1087–1094.

(41) Ruscic, B.; Boggs, J. E.; Burcat, A.; Császár, A. G.; Demaison, J.; Janoschek, R.; Martin, J. M. L.; Morton, M.; Rossi, M. J.; Stanton, J. F.; Szalay, P. G.; Westmoreland, P. R.; Zabel, F.; Bérces, T. *J. Phys. Chem. Ref. Data* **2005**, *34*, 573–656.

(42) Karton, A.; Rabinovich, E.; Martin, J. M. L. *J. Chem. Phys.* **2006**, *125*, 144108.

(43) Császár, A. G.; Szalay, P. G.; Leininger, M. L. *Mol. Phys.* **2002**, *100*, 3879–3883.

(44) Császár, A. G.; Leininger, M. L.; Burcat, A. *J. Phys. Chem. A* **2003**, *107*, 2061–2065.

(45) Császár, A. G.; Leininger, M. L.; Szalay, V. *J. Chem. Phys.* **2003**, *118*, 10631–10642.

(46) Peterson, K. A.; Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 7410–7415.

(47) Feller, D.; Peterson, K. A.; de Jong, W. A.; Dixon, D. A. *J. Chem. Phys.* **2003**, *118*, 3510–3522.

(48) Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vazquéz, J.; Stanton, J. F. *J. Chem. Phys.* **2004**, *121*, 11599–11613.

(49) Boese, A. D.; Oren, M.; Atasolyu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129–4141.

(50) Bomble, Y. J.; Vázquez, J.; Kállay, M.; Michauk, C.; Szalay, P. G.; Császár, A. G.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2006**, *125*, 064108.

(51) Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2008**, *128*, 114111.

(52) East, A. L. L.; Allen, W. D. *J. Chem. Phys.* **1993**, *99*, 4638–4650.

(53) Gonzales, J. M.; Pak, C.; Cox, R. S.; Allen, W. D.; Schaefer, H. F.; Császár, A. G.; Tarczay, G. *Chem. Eur. J.* **2003**, *9*, 2173–2192.

(54) Schuurman, M. S.; Muir, S. R.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* **2004**, *120*, 11586–11599.

(55) Császár, A. G.; Tarczay, G.; Leininger, M. L.; Polyansky, O. L.; Tennyson, J.; Allen, W. D. In *Spectroscopy from Space*; Demaison, J., Sarka, K., Eds.; Kluwer: Dordrecht, 2001; pp 317−339.

(56) Ruscic, B.; Pinzon, R. E.; Morton, M. L.; Von Laszevski, G.; Bittner, S. J.; Nijsure, S. G.; Amin, K. A.; Minkoff, M.; Wagner, A. F. *J. Phys. Chem. A* **2004**, *108*, 9979–9997.

(57) Allen, W. D.; East, A. L. L.; Császár, A. G. In *Structures and Conformations of Non-Rigid Molecules*; Laane, J., Dakkouri, M., van der Veken, B., Oberhammer, H., Eds.; Kluwer: Dordrecht, 1993; pp 343−373.

(58) Császár, A. G.; Allen, W. D.; Schaefer, H. F. *J. Chem. Phys.* **1998**, *108*, 9751–9764.

(59) Cowan, R. D.; Griffin, D. C. *J. Opt. Soc. Am.* **1976**, *66*, 1010–1014.

(60) Tarczay, G.; Császár, A. G.; Klopper, W.; Quiney, H. M. *Mol. Phys.* **2001**, *99*, 1769–1794.

(61) Handy, N. C.; Yamaguchi, Y.; Schaefer, H. F. *J. Chem. Phys.* **1986**, *84*, 4481–4484.

(62) Gauss, J.; Tajti, A.; Kállay, M.; Stanton, J. F.; Szalay, P. G. *J. Chem. Phys.* **2006**, *125*, 144111.

(63) Kutzelnigg, W. *Mol. Phys.* **1997**, *90*, 909–916.

(64) Valeev, E. F.; Sherrill, C. D. *J. Chem. Phys.* **2003**, *118*, 3921–3927.

(65) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(66) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(67) Wilson, A. K.; van Mourik, T.; Dunning, T. H., Jr. *J. Mol. Struct. (THEOCHEM)* **1996**, *388*, 339–349.

(68) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1995**, *103*, 4572–4585.

(69) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(70) Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256–4266.

(71) Crawford, T. D.; Schaefer, H. F. *Rev. Comput. Chem.* **2000**, *14*, 33–136.

(72) Kállay, M.; Surján, P. R. *J. Chem. Phys.* **2001**, *115*, 2945–2954.

(73) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(74) Kállay, M.; Gauss, J. *J. Chem. Phys.* **2005**, *123*, 214105.

(75) Bomble, Y.; Kállay, M.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2005**, *123*, 054101.

(76) Karton, A.; Martin, J. M. L. *Theor. Chem. Acc.* **2006**, *115*, 330–333.

(77) Klopper, W.; Kutzelnigg, W. *J. Mol. Struct. (THEOCHEM)* **1986**, *28*, 339–356.

(78) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104–6114. Ibid. **1993**, *98*, 7059−7071.

(79) Tasi, G.; Császár, A. G. *Chem. Phys. Lett.* **2007**, *438*, 139–143.

(80) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639–9646.

(81) The Mainz-Austin-Budapest version of ACES II, Stanton, J. F.; Gauss, J.; Watts, J. D.; Szalay, P. G.; Bartlett, R. J. with contributions from Auer, A. A.; Bernholdt, D. B.; Christiansen, O.; Harding, M. E.; Heckert, M.; Heun, O.; Huber, C.; Jonsson, D.; Jusélius, J.; Lauderdale, W. J.; Metzroth, T.; Michauk, C.; O'Neill, D. P.; Price, D. R.; Ruud, K.; Schiffmann, F.; Varner, M. E.; Vázquez, J. and the integral packages MOLECULE (Almlöf, J.; Taylor, P. R.), PROPS (Taylor, P. R.), and ABACUS (Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.).

(82) For the latest version, see: http://aces2.de (accessed on April 26, 2008).

(83) MRCC, a string-based quantum chemical program suite written by M. Kállay. See also ref 72.

(84) For the latest version, see: http://mrcc.hu (accessed on April 26, 2008).

(85) MOLPRO, version 2006.1, a package of ab initio programs, written by Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.

(86) For the latest version, see: http://www.molpro.net (accessed on April 26, 2008).

(87) Scuseria, G. E. *J. Chem. Phys.* **1991**, *94*, 442–447.

(88) Lee, T. J.; Rendell, A. P. *J. Chem. Phys.* **1991**, *94*, 6229–6236.

(89) Watts, J. D.; Gauss, J.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 8718–8733.

(90) INTDIF2005 is an abstract program written by Wesley D. Allen for Mathematica to perform general numerical differentiations to high orders of electronic structure data.

(91) DeKock, R. L.; McGuire, M. J.; Piecuch, P.; Allen, W. D.; Schaefer, H. F.; Kowalski, K.; Kucharski, S. A.; Musial, M.; Bonner, A. R.; Spronk, S. A.; Lawson, D. B.; Laursen, S. L. *J. Phys. Chem. A* **2004**, *108*, 2893–2903.

(92) INTDER2005 is a general program written by Wesley D. Allen, which performs sundry vibrational analyses and higher order nonlinear transformations among force field representations. See program descriptions in Sarka, K.; Demaison, J. In *Computational Molecular Spectroscopy*, Jensen, P., Bunker, P. R., Eds.; Wiley: Chichester, 2000, p 255.

(93) Allen, W. D.; Császár, A. G. *J. Chem. Phys.* **1993**, *98*, 2983–3015.

(94) Allen, W. D.; Császár, A. G.; Szalay, V.; Mills, I. M. *Mol. Phys.* **1996**, *89*, 1213–1221.

(95) Simons, G.; Parr, R. G.; Finlan, J. M. *J. Chem. Phys.* **1973**, *59*, 3229–3234.

(96) Mátyus, E.; Czakó, G.; Sutcliffe, B. T.; Császár, A. G. *J. Chem. Phys.* **2007**, *127*, 084102.

(97) Telle, H.; Tell, U. *J. Mol. Spectrosc.* **1981**, *85*, 248–252.

(98) Mladenović, M.; Schmatz, S. *J. Chem. Phys.* **1998**, *109*, 4456–4470.

(99) Duncan, J. L.; Mills, I. M. *Spectrochim. Acta* **1964**, *20*, 523–546.

(100) Heckert, M.; Kállay, M.; Tew, D. P.; Klopper, W.; Gauss, J. *J. Chem. Phys.* **2006**, *125*, 044108.

(101) Huber, K. P.; Herzberg, G. *Constants of Diatomic Molecules*; van Nostrand Reinhold: New York, 1979.

(102) Puzzarini, C.; Tarroni, R.; Palmieri, P.; Carter, S.; Dore, L. *Mol. Phys.* **1996**, *87*, 879–898.

(103) Császár, A. G.; Allen, W. D. *J. Chem. Phys.* **1996**, *104*, 2746–2748.

(104) Martin, J. M. L.; Taylor, P. R.; Lee, T. J. *J. Chem. Phys.* **1993**, *99*, 286–292.

(105) Feller, D.; Dixon, D. A. *J. Chem. Phys.* **2001**, *115*, 3484–3496.

(106) Császár, A. G.; Allen, W. D.; Yamaguchi, Y.; Schaefer, H. F. In *Computational Molecular Spectroscopy*; Jensen, P., Bunker, P. R., Eds.; Wiley: Chichester, 2000; pp 15−68.

(107) Léonard, C.; Carter, S.; Handy, N. C. *Chem. Phys. Lett.* **2003**, *370*, 360–365. Also personal communication (2003).

(108) Yurchenko, S. N.; Carvajal, M.; Jensen, P.; Lin, H.; Zheng, J.; Thiel, W. *Mol. Phys.* **2005**, *103*, 359–378.

(109) Gatti, F.; Iung, C.; Leforestier, C.; Chapuisat, X. *J. Chem. Phys.* **1999**, *111*, 7236–7243.

(110) Ceyer, S. T.; Triedemann, P. W.; Mahan, B. H.; Lee, Y. T. *J. Chem. Phys.* **1979**, *70*, 14–17.

(111) Meot-Ner, M.; Sieck, L. W. *J. Am. Chem. Soc.* **1991**, *113*, 4448–4460.

(112) Eades, R. A.; Scanlon, K.; Ellenberger, M. R.; Dixon, D. A.; Marynick, D. S. *J. Phys. Chem.* **1980**, *84*, 2840–2842.

(113) DeFrees, D. J.; McLean, A. D. *J. Comput. Chem.* **1986**, *7*, 321–333.

(114) Pople, J. A.; Curtiss, L. A. *J. Phys. Chem.* **1987**, *91*, 155–162.

(115) Del Bene, J. E.; Shavitt, I. *J. Phys. Chem.* **1990**, *94*, 5514–5518.

(116) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221–7230.

(117) Del Bene, J. E. *J. Phys. Chem.* **1993**, *97*, 107–110.

(118) Smith, B. J.; Radom, L. *J. Phys. Chem.* **1995**, *99*, 6468–6471.

(119) Seo, Y.; Kim, Y.; Kim, Y. *Chem. Phys. Lett.* **2001**, *340*, 186–193.

(120) Del Bene, J. E.; Frisch, M. J.; Raghavachari, K.; Pople, J. A. *J. Phys. Chem.* **1982**, *86*, 1529–1535.

(121) Ikuta, S. *Chem. Phys. Lett.* **1984**, *109*, 550–553.

(122) Dixon, D. A.; Komornicki, A.; Kraemer, W. P. *J. Chem. Phys.* **1984**, *81*, 3603–3611.

(123) Ma, N. L.; Smith, B. J.; Radom, L. *Chem. Phys. Lett.* **1992**, *197*, 573–580.

(124) Botschwina, P.; Horn, M.; Flügge, J.; Seeger, S. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 2219–2230.

(125) van Mourik, T.; Dunning, T. H., Jr.; Peterson, K. A. *J. Phys. Chem. A* **2000**, *104*, 2287–2293.

CT800082R

JCTC Journal of Chemical Theory and Computation

Accelerating Density Functional Calculations with Graphics Processing Unit

Koji Yasuda*

Graduate School of Information Science, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan

Received March 27, 2008

Abstract: An algorithm is presented for graphics processing units (GPUs), which execute single-precision arithmetic much faster than commodity microprocessors (CPUs), to calculate the exchange-correlation term in ab initio density functional calculations. The algorithm was implemented and applied to two molecules, taxol and valinomycin. The errors in the total energies were about 10^{-5} a.u., which is accurate enough for practical usage. If the exchange-correlation term is split into a simple analytic model potential and the correction to it, and only the latter is calculated with the GPU, the energy error is decreased by an order of magnitude. The resulting time to compute the exchange-correlation term is smaller than it is on the latest CPU by a factor of 10, indicating that a GPU running the proposed algorithm accelerates the density functional calculation considerably.

I. Introduction

Streaming processors such as graphics processing units (GPUs) are attracting attention as high-performance computing devices. They are potentially several times faster than commodity central processing units (CPUs) for calculating single-precision floating-point numbers because they are specialized for computation-intensive, highly data-parallel computations. They are designed such that more transistors are devoted to data processing than to data caching and flow control. Most transistors in CPUs are devoted to cache memory and control logic to reduce memory latency and pipeline stalls. The high performance of GPUs makes them an attractive way to accelerate scientific computation. GPU computation has been successfully applied to various types of applications, including gravitational dynamics,[1] molecular dynamics,[2] quantum chemistry,[3–5] and quantum Monte Carlo.[6]

However, the streaming processors have several shortcomings. They are optimized for applications in which numerical accuracy is less important than performance. Most of them do not support double-precision floating-point numbers natively. Some of them support double-precision numbers but do not exhibit high performance. The total amount of working memory (registers and caches) on a chip is small. To avoid the pipeline stalls many parallel threads must be run concurrently. Consequently the usefulness of streaming processors for certain applications is not clear a priori. Moreover, to obtain high performance, they require a fine-grade parallel algorithm that can run with few hardware resources. Such an algorithm would differ from the usual, coarse-grained parallel algorithm for conventional CPUs.

A potential area for applying GPUs is structural biology. Advances in structural biology have stimulated the investigation of the electronic structure of biological molecules. Investigations have traditionally used density functional theory (DFT) for the ground-state and time-dependent DFT for the excited states.[7] It is now becoming possible to perform all-electron ab initio calculations of macromolecules for certain molecular structures. Along with this there is a growing demand for much faster ab initio methods, as current algorithms and computers are not sufficient for elucidating the free energy surface or the relaxation of the excited state of a macromolecule from the first principle. There is thus an opportunity for high-throughput processors like GPUs to play an active role.

A density functional calculation consists of many complicated procedures, but only a few of them account for most of the computational time. Table 1 shows, for example, the computational time of the major steps for the molecules taxol and valinomycin. The Gaussian03 program,[8] the generalized

* Corresponding author e-mail: yasudak@is.nagoya-u.ac.jp.

10.1021/ct8001046 CCC: $40.75
Published on Web 07/04/2008

Table 1. Computational Time for Major Steps in Density Functional Calculation Using PW91 Functional[9] and Default Parameters of Gaussian03 Program[a]

	user time in seconds (percentage)			
	taxol ($C_{47}H_{51}NO_{14}$)		valinomycin ($C_{54}H_{90}N_6O_{18}$)	
procedure	3−21G	6−31G	3−21G	6−31G
Coulomb	125 (6.9)	475 (13.5)	204 (6.1)	589 (11.3)
Fock Diag	82 (4.6)	116 (3.3)	201 (6.0)	225 (4.3)
Ex-Cor	1598 (88.4)	2922 (83.1)	2947 (87.8)	4383 (84.3)
Grid	190 (10.5)	287 (8.2)	451 (13.4)	540 (10.4)
χ	91 (5.0)	174 (4.9)	154 (4.6)	234 (4.5)
ρ	458 (25.3)	868 (24.7)	775 (23.1)	1205 (23.2)
v_{xc}	837 (46.3)	1561 (44.4)	1533 (45.7)	2363 (45.4)
total	1807 (100)	3516 (100)	3355 (100)	5200 (100)

[a] Grid: generation of points and weights; χ: calculation of AOs and gradient; ρ: calculation of density and gradient; v_{xc}: calculation of AO matrix elements of exchange-correlation potential.

gradient approximation of the PW91 functional,[9] and the 3−21G and 6−31G basis sets[10] were used. About 90−95% of the total computational time was used to solve the self-consistent field (SCF) equation. About 10% of it was used to evaluate the Coulomb potential, 85% to evaluate the exchange-correlation potential, and 5% to diagonalize the Fock matrix. The density functional calculation can be accelerated considerably if we execute these time-consuming steps on a GPU.

An algorithm for GPUs that can be used to evaluate the two-electron integrals and the Coulomb potential was previously reported.[3] Quite recently Ufimtsev and Martinez[5] proposed a similar algorithm for evaluating two-electron integrals and presented promising results. This paper describes an algorithm for GPUs to evaluate the exchange-correlation term, which is the most time-consuming step in the density functional calculation. Since this step is essentially numerical quadrature with a three-dimensional grid, it is suitable for parallel processing. Some ideas to avoid the limitations of GPUs, few hardware resources and low numerical accuracy, are also described. The algorithm was implemented and applied to two molecules, taxol and valinomycin, and the total energy was accurate enough for practical usage. The computational time was an order of magnitude lower than that of latest commodity CPUs. GPUs thus offer a considerable speedup of the density functional calculations.

II. Algorithm

The major computational task in the density functional calculation is to evaluate the exchange-correlation energy E_{xc} and the matrix elements of the exchange-correlation potential v_{xc} in terms of the basis functions (AOs), $\chi_k(r)$, for a given σ-electron density, $\rho_\sigma(r)$.[11]

$$E_{xc} = \int \varepsilon(\rho_\alpha, \rho_\beta, \gamma_{\alpha\alpha}, \gamma_{\alpha\beta}, \gamma_{\beta\beta}) d^3r \quad (1)$$

$$v_{xc} = \int \left[\frac{\partial \varepsilon}{\partial \rho_\alpha} \chi_k \chi_l + \left(\frac{2\,\partial \varepsilon}{\partial \gamma_{\alpha\alpha}} \nabla \rho_\alpha + \frac{\partial \varepsilon}{\partial \gamma_{\alpha\beta}} \nabla \rho_\beta \right) \cdot \nabla (\chi_k \chi_l) \right] d^3r$$

$$\gamma_{\sigma\sigma'} = \nabla \rho_\sigma(r) \cdot \nabla \rho_{\sigma'}(r) \quad (2)$$

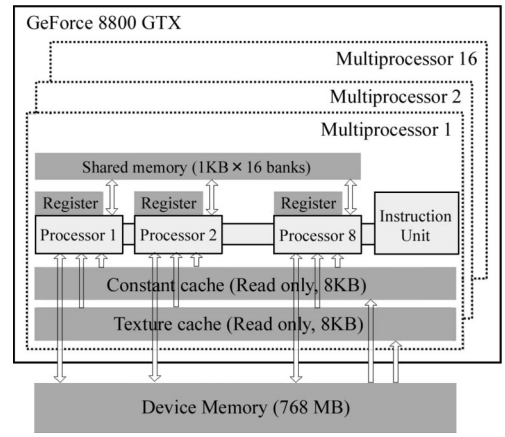


Figure 1. GeForce 8800 GTX chip has 16 multiprocessors, each containing a constant cache, a texture cache, a shared memory, and 8 processors.

The analytic integration of eqs 1 and 2 is impossible because ε is a complicated function of the electron density and gradient. We can evaluate them accurately by using the three-dimensional quadrature points (r_i) and the weights (w_i). For example, the exchange-correlation energy is given by

$$E_{xc} = \sum_i w_i \varepsilon(r_i),$$
$$\varepsilon(r_i) = \varepsilon(\rho_\alpha(r_i), \rho_\beta(r_i), \gamma_{\alpha\alpha}(r_i), \gamma_{\alpha\beta}(r_i), \gamma_{\beta\beta}(r_i)) \quad (3)$$

Any functions of position operator $\hat{r}$, like the electron density and ε, share the eigenfunctions of $\delta(r\text{-}r_0)$. Hence, the spatial grid is a natural choice and is suitable for evaluating ε and its derivatives, $\partial\varepsilon/\partial\rho_\sigma$ and $\partial\varepsilon/\partial\gamma_{\sigma\sigma'}$. The quadrature consists of four distinct steps. (i) First, the quadrature points and weights are generated. (ii) Then the electron density and the gradient on these points are calculated. (iii) These values are converted into the function ε and its derivatives, and (iv) finally the exchange-correlation energy and the matrix elements of the exchange-correlation potential are evaluated. Usually, most of the computational effort is spent on the second and fourth steps, the evaluation of the electron density, and the matrix elements of the potential. Their costs are proportional to the number of grid points and to the square of the number of AOs. To accelerate the quadrature, we must at least execute these time-consuming steps on a GPU with the architecture described below.

A. Architecture of GeForce 8800. The GPU used was NVIDIA's GeForce 8800 GTX.[12] As shown in Figure 1, it consists of 16 sets of SIMD multiprocessors with on-chip shared memory. Each multiprocessor is composed of eight processors. At any given clock cycle, each processor in each multiprocessor executes the same instruction but for different data. Each multiprocessor has four kinds of on-chip memory: (i) 1024 local registers per processor, (ii) memory shared by all the processors (the amount available to each multiprocessor is 16 KB divided into 16 banks), (iii) read-only constant cache that is shared by all the processors to speed up reads from the constant memory space (the amount available is 64 KB with a cache of 8 KB per multiprocessor), and (iv) read-only texture cache that is shared by all the processors to speed up reads from the texture memory space

(8 KB per multiprocessor). There is also an off-chip 768 MB memory device that is not cached. Its latency is about 400−600 clock cycles. The latency of the registers and the shared and global memories is hidden if enough independent threads are launched. It is recommended that more than 196 threads be run concurrently on each multiprocessor. The threads can be synchronized on a multiprocessor but not on different multiprocessors.

This GPU follows most of the IEEE-754 standard for single-precision binary floating-point arithmetic. The IEEE 754 compliant, single-precision floating-point number a is represented by[13]

$$\begin{aligned} a &= \pm 2^e m, \\ -125 &\le e \le 128, \\ 1 &\le m < 2 \end{aligned} \tag{4}$$

where exponent e is an integer, and mantissa m is a 24-bit binary decimal. The relative error in the single-precision number is $2^{-23} \approx 1.2 \times 10^{-7}$, so the absolute error is about $2^{-23} \times |a|$. The results of the addition and the multiplication are rounded off. The throughput of the instructions is four cycles for floating-point addition and multiplication and for fused addmul operation.

B. Quadrature Grid. The quadrature grid used is the standard pruned one of 75 radial shells and 302 angular points centered at each atom. The number of grid points is about 7×10^3 per atom. The nearby points in the same cube, called a "grid box", are grouped. The cube is 4 a.u. in each dimension. The weights of the quadrature are generated with the fuzzy Voronoi cell method of Becke.[14] Grid generation requires little computational time compared to generation of the weights (5−13% of time, Table 1). Note that the Gaussian03 program recalculates them at every SCF iteration. The computational effort is reduced by calculating them only once on the host computer and then storing them in the external storage.

C. Basis Functions on Points. We need the values of the AOs, $\chi_k(r_i)$, and their gradient, $\nabla\chi_k(r_i)$, at grid points r_i. The contracted Cartesian Gaussian basis function centered at A is given by[15]

$$\begin{aligned} \chi_k(x,y,z) &= (x-A_x)^{a_x}(y-A_y)^{a_y}(z-A_z)^{a_z}\sum c_A \exp(-aR_A^2), \\ R_A^2 &= (x-A_x)^2 + (y-A_y)^2 + (z-A_z)^2 \end{aligned} \tag{5}$$

where c_A is the contraction coefficient, a is the primitive Gaussian exponent, and a_x, a_y, and a_z are non-negative integers. Their evaluation requires about 5% of the computational time (Table 1). The calculation of sufficiently small terms is skipped in the Gaussian03 program as well as in the present algorithm. The absolute minimum of eq 5 and the gradient is calculated for each grid box, and only those AOs with a minimum greater than threshold λ_{AO} are kept. They are called "significant AOs". Every batch of points has a different set of significant AOs (30−700 in the examples studied). Thus, except for small molecules, it is not realistic to keep all the calculated $\chi_k(r_i)$ and $\nabla\chi_k(r_i)$ in the external storage for later use. They have to be recalculated at every SCF iteration. The wide range in the number of significant AOs means that constructing an efficient algorithm requires special care.

From the significant AOs the GPU calculates the density and the density gradient at the grid points as well as the AO matrix elements of the exchange-correlation potential. Because of the huge number of terms and the slow communication speed between the host memory and GPU, it is best to calculate $\chi_k(r_i)$ and $\nabla\chi_k(r_i)$ on the GPU as well. However, the internal memory (4096 words per multiprocessor) is not large enough to store the calculated AOs, even for a small batch of points, as there are potentially 700×4 words for each grid point. It is in principle possible to store some of them on a large, external graphic memory at the expense of long latency. However it was found to be better to deal with every 32 AOs in turn in order to keep as many as possible in the shared memory. Trial and error indicated that the best implementation is that 128 threads on a multiprocessor calculate 4 primitive Gaussians of the 32 significant AOs on 32 points in parallel.

Anderson et al. reported an optimized quantum Monte Carlo (QMC) algorithm[6] for a different GPU (NVIDIA 7800 GTX), in which the basis function evaluation on spatial points is the one of the rate-limiting steps. The computational effort of QMC differs from that of DFT considerably. About 73% is focused on basis function evaluation and 11% on dense matrix multiplication. On the other hand in DFT only 5% is used for the basis function evaluation (Table 1). Most of the effort is used to generate the electron density and the exchange-correlation potential, which are essentially the matrix multiplication processes. Another noticeable difference is that the proposed algorithm is a linear-scaling one, because it screens the grid points and the basis functions. In the QMC algorithm, localization procedures which lead to sparser matrices are ignored. Since the matrices with the proposed algorithm are much smaller than the QMC ones, they must be kept on faster shared memory to avoid memory latency. Moreover the lack of enough shared memory means that some of the basis function values must be recalculated. However, the low latency of the shared memory more than compensates for this extra recalculation.

D. Electron Density and Gradient. The electron density, $\rho_\sigma(r_i)$, and gradient, $\nabla\rho_\sigma(r_i)$, at the grid point are calculated for a given first-order reduced density matrix (1-RDM) of the σ-electron P^σ.

$$\rho_\sigma(r_i) = \sum_{k,l} P^\sigma_{kl}\chi_k(r_i)\chi_l(r_i) \tag{6}$$

$$\nabla\rho_\sigma(r_i) = 2\sum_{k,l} P^\sigma_{kl}\chi_k(r_i)\nabla\chi_l(r_i) \tag{7}$$

The summation runs over the significant AOs. As shown in Table 1, this step consumes 20−25% of the computational time, mostly for the calculation of the matrix-vector products, $A_l(r_i) = \Sigma_k P^\sigma_{kl}\chi_k(r_i)$. The $\rho_\sigma(r_i)$ and $\nabla\rho_\sigma(r_i)$ are calculated on a GPU as follows. First, the host computer sends the grid points, significant AOs, and 1-RDM elements to the external graphic memory. The GPU calculates the values of a set of 32 significant AOs, χ_k, on 32 points, r_i, concurrently. It then reads a 32×32 submatrix of P_{kl} from the external graphic memory and multiplies it by $\chi_k(r_i)$ to form product $A_l(r_i)$.

Next it calculates $\chi_l(r_i)$ and $\nabla\chi_l(r_i)$, which are multiplied by $A_l(r_i)$, to form the density and the gradient. Once all AOs have been processed, the results are sent back to the host computer. The communication time to send the 1-RDM elements from the host computer to the GPU dominates the total communication time, which is about 10% of the GPU computation time in the example below.

E. Functional on Points. The electron density and gradient at the grid points are then converted into the exchange-correlation potential, ε, and its derivatives, $\partial\varepsilon/\partial\rho_\sigma$ and $\partial\varepsilon/\partial\gamma_{\sigma\sigma'}$. They are converted into the total exchange-correlation energy [eq 3] and the quantities f and $\mathbf{g}$.

$$f(r_i) = \frac{w_i}{2}\frac{\partial\varepsilon(r_i)}{\partial\rho_\alpha} \tag{8}$$

$$g(r_i) = w_i\left(\frac{2\,\partial\varepsilon(r_i)}{\partial\gamma_{\alpha\alpha}}\nabla\rho_\alpha + \frac{\partial\varepsilon(r_i)}{\partial\gamma_{\alpha\beta}}\nabla\rho_\beta\right) \tag{9}$$

The host computer executes this step with double-precision accuracy, because the computational cost scales linearly with the number of points, so it requires little computational effort.

F. Matrix Elements of Exchange-Correlation Potential. Finally, the GPU calculates

$$A_{kl} = \sum_i B_{ki}\chi_l(r_i) \tag{10}$$

$$B_{ki} = f(r_i)\chi_k(r_i) + \mathbf{g}(r_i)\cdot\nabla\chi_k(r_i) \tag{11}$$

The sum runs over the batch of grid points, while the indices k and l run over the significant AOs. The AO matrix elements of the exchange-correlation potential, eq 2, are given by $v_{xc} = A_{kl}+A_{lk}$. This last step, which is essentially the evaluation of the matrix-matrix product, eq 10, consumes 40−45% of the computational time (Table 1). For a given batch of points, the GPU first evaluates a set of 64 AOs and the gradient for 32 points concurrently. They are converted into the matrix B_{ki} and stored in the internal shared memory. The matrix product between B_{ki} and $\chi_l(r_i)$ is calculated, and A_{kl} is updated. Because of the memory limitation, A_{kl} is kept in the external graphic memory. After all points are processed, the calculated matrix elements are sent back to the host computer. This communication time dominates the total communication time; it takes about 25% of the GPU computational time.

III. Accuracy Requirement

At first glance, the most serious issue with the present GPU is the numerical accuracy: it supports only single-precision floating-point numbers natively, and ab initio calculation obviously requires double-precision arithmetic. For example, the total electronic energy of a molecule should be calculated within an accuracy of 10^{-4} a.u., which is 1/10th the thermal energy of room temperature. Since the total electronic energy is about 40 au per carbon atom, the roundoff error of the total energy of a large molecule surpasses it. It is in principle possible to carry out multiprecision addition and multiplication by software.[16] However, since this is expensive, most operations should be done with single-precision accuracy.

Since the roundoff error should be unbiased and nearly random, the sum of the N single-precision numbers with an absolute value is about s and contains the relative error of $2^{-23}sN^{-1/2}$. It decreases as the number of terms increases. Single-precision arithmetic can be used when this error estimate is smaller than the tolerance.

Table 2. Exponent and Coefficient Values of Model Exchange-Correlation Potential [Eq *23*].

element	C		N		O		H
ζ	3.0	1.0	3.0	1.0	3.0	1.0	2.0
d	−0.198	−2.035	−0.847	−1.783	−1.616	−1.402	−1.658

The two strategies previously proposed[3] for calculating the Coulomb potential on a GPU are to use the GPU only in the early SCF iterations and to calculate the small electron repulsion integrals (ERIs) by GPU. The absolute errors of these ERIs are small. The host computer calculates the rest of the large ERIs with double-precision accuracy. These two strategies were found to work well. The accuracy requirement for the exchange-correlation term should be less severe than for the Coulomb term because the former is generally an order of magnitude smaller than the latter.

The exchange-correlation matrix elements should allow lesser accuracy than the density and the gradient because of the variational principle and the conservation of the number of electrons. The small error in the matrix elements induces a small change in the SCF density. However, since the SCF density satisfies the energy variational principle, $\delta E/\delta\rho(r) = \mu$, the first-order energy change always vanishes.

A deviation in the density or the gradient induces energy error of the first-order through eq 1. This error is reduced as follows. Two model potentials, $v_{local}(r)$ and $v_{gc}(r)$, that respectively mimic the local and gradient-dependent parts of the exchange-correlation potential at the ground-state density are used.

$$v_{local}(r) \approx \frac{\partial\varepsilon}{\partial\rho_\alpha} \tag{12}$$

$$v_{gc}(r) = \nabla\cdot\mathbf{V} \tag{13}$$

$$\mathbf{V}(r_i) \approx -\left(\frac{2\,\partial\varepsilon}{\partial\gamma_{\alpha\alpha}}\nabla\rho_\alpha + \frac{\partial\varepsilon}{\partial\gamma_{\alpha\beta}}\nabla\rho_\beta\right) \tag{14}$$

The total exchange-correlation energy is given as the sum of the reference energy, E_1, and the correction to it, E_2.

$$E_{xc} = E_1 + E_2 \tag{15}$$

$$E_1 = \int\{v_{local}(r) + v_{gc}(r)\}\rho(r)d^3r = Tr\{(v_{local} + v_{gc})P\} \tag{16}$$

$$E_2 = \sum_i w_i\{\varepsilon(r_i) - v_{local}(r_i)\rho(r_i) + \mathbf{V}(r_i)\cdot\nabla\rho(r_i)\} \tag{17}$$

The total 1-RDM and the density are denoted as $P = P^\alpha + P^\beta$ and $\rho = \rho_\alpha + \rho_\beta$, respectively. The equivalence of eqs 1 and 15 can be seen from Green's theorem.

$$\int(v_{gc}\rho + \mathbf{V}\cdot\nabla\rho)d^3r = 0 \tag{18}$$

The model potentials, $v_{local}(r)$ and $v_{gc}(r)$, are chosen so that the AO matrix elements can be calculated analytically. Thus,

we can evaluate E_1 analytically with double-precision accuracy, so correction E_2 simply needs to be approximated by numerical quadrature with single-precision accuracy. Assuming that the model potential approximates well the true exchange-correlation potential and that the numerical quadrature is accurate, the first-order error of E_{xc} with respect to $\rho(r_i)$ and $\nabla\rho(r_i)$ vanishes because

$$\delta\varepsilon(r_i) \approx v_{local}(r_i)\delta\rho(r_i) - \mathbf{V}(r_i)\cdot\delta(\nabla\rho(r_i)) \tag{19}$$

Note that the calculation of $v_{local}(r_i)$ and $v_{gc}(r_i)$ requires little computational effort. Vector field $\mathbf{V}$ represents the electronic field of the imaginary charge distribution, $v_{gc}(r)$.

We can also separate exchange-correlation potential v_{xc} into reference term v_1 and the correction to it, v_2.

$$v_{xc} = v_1 + v_2 \tag{20}$$

$$v_1 = (k)v_{local} + v_{gc}|l| \tag{21}$$

$$v_2 = \sum_i w_i \left[\left(\frac{\partial\varepsilon(r_i)}{\partial\rho_\alpha} - v_{local}(r_i) \right) \chi_k(r_i)\chi_l(r_i) + \left(\frac{2\,\partial\,\varepsilon(r_i)}{\partial\gamma_{\alpha\alpha}} \nabla\rho_\alpha(r_i) + \frac{\partial\varepsilon(r_i)}{\partial\gamma_{\alpha\beta}} \nabla\rho_\beta(r_i) + \mathbf{V}(r_i) \right) \cdot \nabla(\chi_k(r_i)\chi_l(r_i)) \right] \tag{22}$$

The reference term can be calculated analytically, so only the correction to it needs to be calculated with the GPU. The error in v_{xc} is lower because the former term is calculated exactly.

Here, the model potential v_{local} is expanded with the spherical Gaussian functions centered on each atom A, while v_{gc} is ignored.

$$v_{local} = \sum_A \sum_i d_i \exp(-\zeta_i(r-A)^2) \tag{23}$$

The exponent and coefficient of the model potential used in this paper are summarized in Table 2. They were determined as follows. The self-consistent ground-state density of a molecule, $CH_3CH(CHO)NH_2$, was calculated using the PW91 functional[9] and 3−21G basis set.[10] The AO matrix elements of the local part of the exchange-correlation potential were calculated from this density. The linear coefficients of the model potential, d_i, were then determined by minimizing the error

$$L = \mathrm{Tr}\{P(v - v_{local})P(v - v_{local})\} \tag{24}$$

where P is the ground-state 1-RDM. The same potential was used for the same elements in a molecule. The potentials for elements C, N, and O were expanded with the two spherical Gaussians, while that of H was expanded with only one Gaussian. The exponents ζ_i were determined from the dimensional argument and were not optimized.

IV. Results and Discussion

The algorithm described above was implemented using the beta release of the CUDA toolkit[12] to develop the GPU code, which was then linked with a modified Gaussian03 program.[8] In this section the accuracy of the total SCF energy and the exchange-correlation potential calculated with the GPU are first examined. The density functional calculations for taxol ($C_{47}H_{51}NO_{14}$) and valinomycin ($C_{54}H_{90}N_6O_{18}$) were performed using the 3−21G and 6−31G basis sets[10] and the PW91 functional[9] The electron density, the density gradient, and the matrix elements of the exchange-correlation potential were calculated with the GPU, while the rest were calculated on the host computer with double-precision accuracy. The host computer had a Pentium4 (2.8 GHz) CPU, 2 GB of memory, and the GeForce 8800 GTX graphic card; it ran Scientific Linux version 4.3.[17] An INTEL Fortran compiler (version 9) was used to compile the program.

Table 3. Error in Total SCF Energy[a]

	energy error in a.u.			
	taxol ($C_{47}H_{51}NO_{14}$)		valinomycin ($C_{54}H_{90}N_6O_{18}$)	
λ_{AO}	3−21G	6−31G	3−21G	6−31G
	Without Model Potential			
10^{-3}	5.58[−5]	3.75[−5]	8.00[−5]	5.39[−5]
10^{-6}	5.81[−5]	5.30[−5]	6.92[−5]	5.82[−5]
10^{-15}	5.81[−5]	5.03[−5]	6.91[−5]	6.01[−5]
	With Model Potential			
10^{-3}	−8.80[−7]	−6.94[−6]	−1.71[−6]	−5.17[−6]
10^{-6}	5.98[−6]	5.17[−6]	8.20[−6]	6.08[−6]
10^{-15}	6.61[−6]	5.06[−6]	8.09[−6]	6.39[−6]

[a] λ_{AO} is the threshold for significant AOs. Reference values were calculated with double-precision accuracy, and $\lambda_{AO} = 10^{-15}$. Numbers in square brackets indicate powers of ten.

Table 4. Computational Time for Exchange-Correlation Terms[a]

	user time in seconds			
	taxol ($C_{47}H_{51}NO_{14}$)		valinomycin ($C_{54}H_{90}N_6O_{18}$)	
λ_{AO}	3−21G	6−31G	3−21G	6−31G
	GPU			
10^{-3}	36.4	45.5	66.1	75.0
10^{-6}	46.1	61.4	83.9	100.4
10^{-15}	73.7	110.1	140.8	193.8
	Host			
10^{-3}	667.9	955.2	932.1	1371
10^{-6}	1513	2385	2704	4106
10^{-15}	4449	6822	10577	15739
	Original Gaussian03			
	1598	2922	2947	4383

[a] GPU: electron density, gradients, and AO matrix elements of v_{xc} were calculated on GPU. Host: the same program was executed on host computer with double-precision accuracy.

Table 3 summarizes the errors in the total energy for various AO cutoffs and model potentials. The algorithm for GPU was also executed on the host computer with double-precision accuracy to calculate the reference values. The fast multipole method (FMM), the special Coulomb engine (FofCou), and the tight SCF convergence criterion were used. As shown in the table the errors were about 6×10^{-5} a.u. on average when the model exchange-correlation potential was not used. They were of the same order of magnitude as the default SCF convergence criterion for single-point calculation. The error remained almost constant as the AO cutoff was increased up to $\lambda_{AO} = 10^{-6}$, indicating that this threshold is high enough for the molecules studied. When the model exchange-correlation potential was used, the energy error decreased by an order of magnitude, to about 6

Chart 1

```
GPU kernel for density
  arrays on shared mem: chi[64][32], rdmchi[32][32]
read a grid point from gmem
initialize ρ and ∇ρ
loop over AO batch
  chi = 0
  loop over primitives in batch
    read primitive information from gmem
    calculate basis functions on a grid point
    increase chi
  loop end
  loop over AO batch
    read RDM elements P from gmem
    calculate matrix-vector product, rdmchi = P× chi
    loop over primitives in batch
      read primitive information from gmem
      calculate primitive functions and their derivatives
      multiply them by rdmchi
      increase ρ and ∇ρ
    loop end
  loop end
loop end

GPU kernel for potential
  arrays on shared mem: chi[32][32], fchi[64][32]
loop over grid points
  read a grid point, f, and g of Eqs. (8) and (9) from gmem
  loop over primitives in a batch
    read primitive information from gmem
    calculate primitive functions and their derivatives
    multiply them by f or g
    increase fchi
  loop end
  32 threads concurrently transpose matrix fchi
  loop over AO batch
    chi = 0
    loop over primitives in batch
      read primitive information from gmem
      calculate primitive functions on a grid point
      increase chi
    loop end
    read A_kl of Eq. (10) from gmem
    calculate matrix product A_kl = A_kl + chi × fchi
    write A_kl to gmem
  loop end
loop end
```

$\times 10^{-6}$ a.u. on average. With the proposed algorithm, the GPU calculated total energies accurate enough for practical use. Note that the reference energy itself changed when the model potential was used. This difference, which is about 10^{-5} a.u., is attributed to the grid discretization error.

Finally, the computational time to calculate the exchange-correlation terms is compared. The results are summarized in Table 4. "Host" denotes the execution time of the same GPU algorithm on the host computer with double-precision accuracy. The execution time of the original Gaussian03 program is also shown. As shown in the table the GPU executed the same program about 40 times faster than the host computer (Pentium4, 2.8 GHz). The benchmark result for the same molecule[18] indicates that a standard commodity processor (INTEL Core2 Duo, 3.0 GHz) is about 4 times faster than the host CPU used here. Thus, the GPU used here is about five to ten times faster than the latest commodity CPU. Note that the GPU results include the communication and host code execution time. The host code, which calculates the grid weights and exchange-correlation potential on points, consumed about half the GPU time for $\lambda_{AO} = 10^{-6}$. The execution time of the original Gaussian03 program was close to the time for $\lambda_{AO} = 10^{-6}$ on the host computer. This is in accordance with the observation that $\lambda_{AO} = 10^{-6}$ is the best cutoff for the proposed algorithm without degradation of the energy. Note that the Gaussian03 program uses a more complex two-step procedure to select the significant AOs. In short, a GPU running the proposed algorithm accelerates the density functional calculation considerably.

However, the application of the GPU to quantum chemistry has a limitation. It is useful only for simple tasks. To minimize the communication time between the host computer and GPU, a series of tasks should be done on the GPU. However, the GPU's internal memory will soon run short storing the intermediate data. The evaluation of two-electron repulsion integrals between four contracted Gaussians is a fundamental process of ab initio calculation. However, it is difficult for the current GPU to execute efficient algorithms of the McMurchie-Davidson or Obara-Saika[15] type, because of the shortage of fast memory. Thus, we need a new algorithm to accelerate the hybrid density functional calculation. This is why Vogt et al. evaluated the electron repulsion integrals on a host computer and used a GPU for the postcalculation.[4] Other ab initio methodologies require more or less new algorithms. In particular, it would be hopeless to automatically convert the current ab initio program of the host computer to run on the GPU.

On the other hand, the issue of accuracy is much easier to manage in practice. We often have to solve an equation iteratively. A GPU would be suitable for such problems, since lower accuracy is allowed in the early stages of the iteration. An example is the diagonalization of the Fock matrix with the linear scaling algorithm.[19] This method minimizes the energy in terms of the purified 1-RDM iteratively. The major computational task is to calculate the gradient from the products of the Fock matrix and 1-RDM. A GPU can be used without problems in the early stages of this iteration. A study on accelerating the Fock matrix diagonalization is in progress, and the results will be published soon.

Acknowledgment. I am grateful to Prof. H. Nakatsuji at the Quantum Chemistry Research Institute, Prof. J. Makino at the National Astronomical Observatory of Japan, Prof. T. Ebisuzaki at RIKEN, and Dr. H. Matsubara at RIKEN for their encouragement. This research was supported in part by a grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Appendix

The pseudocodes of the GPU kernels for calculating the density and the exchange-correlation potential are presented here (see Chart 1) . Each kernel describes the task of a thread; 128 threads on each multiprocessor execute the same kernel for different data, i.e., different grid points or basis functions. The host computer is responsible for organizing the input data, including the grid points, basis function parameters, and RDM elements. The abbreviation "gmem" stands for the external graphic memory on the GPU.

References

(1) Hamada, T.; Iitaka, T. http://arxiv.org/abs/astro-ph/0703100 (accessed April 1, 2007). See also http://www.gpgpu.org/ (accessed May 13, 2008).

(2) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. *J. Comput. Chem.* **2007**, *28*, 2618–2640.

(3) Yasuda, K. *J. Comput. Chem.* **2008**, *29*, 334–342.

(4) Vogt, L.; Olivares-Amaya, R.; Kermes, S.; Shao, Y.; Amador-Bedolla, C.; Aspuru-Guzik, A. *J. Phys. Chem. A* **2008**, *112*, 2049–2057.

(5) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2008**, *4*, 222–231.

(6) Anderson, A. G.; Goddard, W. A., III; Schroeder, P. *Comput. Phys. Commun.* **2007**, *177*, 298–306.

(7) (a) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133−A1138. (b) Parr, R. G.; Yang, W. In *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989. (c) Gross, E. K. U.; Dreizler, R. M. In *Density Functional Theory*; Plenum: New York, 1995.

(8) *Gaussian 03, Revision B.01*; 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.: Pittsburgh, PA, 2003.

(9) (a) Perdew, J. P. In *Electronic Structure of Solids '91*; Akademie Verlag: Berlin, 1991; p 11. (b) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533–16539.

(10) (a) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657–2664. (b) Binkley, J. S.; Pople J. A.; Hehre W. J, *J. Am. Chem. Soc.* **1980**, *102*, 939–947. (c) Hehre; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261. (d) Dill, J. D.; Pople, J. A. *J. Chem. Phys.* **1975**, *62*, 2921–2923.

(11) Johnson, G. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612–5626.

(12) http://www.nvidia.com/object/cuda_home.html (accessed April 1, 2007).

(13) *IEEE standard for binary foating-point arithmetic*; ANSI/IEEE Standard, Std 754-1985, New York, 1985.

(14) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547–2553.

(15) Helgaker, T.; Jorgensen, P.; Olsen, J. In *Molecular Electronic-Structure Theory*; John Wiley: 2000; pp 336−426.

(16) (a) Knuth, D. In *The Art of Computer Programming, volume 2, Seminumerical Algorithms*; Addison Wesley: MA, 1998. (b) Dekker, T. J. *Numerische Mathematik* **1971**, *18*, 224–242. (c) Hida, Y.; Li, X.; Bailey, D. H. In *15th IEEE Symposium on Computer Arithmetic*; Vail, CO, 2001; p 155.

(17) https://www.scientificlinux.org/ (accessed April 1, 2007).

(18) http://www.hpc.co.jp/appli/solution/gaussian/gaussian_bench_test397.html (accessed April 1, 2007).

(19) (a) Li, X. P.; Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 10891–10894. (b) Daw, M. S. *Phys. Rev. B* **1993**, *47*, 10895–10898. (c) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.

CT8001046

JCTC Journal of Chemical Theory and Computation

Interfacing ab Initio Quantum Mechanical Method with Classical Drude Osillator Polarizable Model for Molecular Dynamics Simulation of Chemical Reactions

Zhenyu Lu and Yingkai Zhang*

Department of Chemistry, New York University, New York 10003

Received April 5, 2008

Abstract: In order to further improve the accuracy and applicability of combined quantum mechanical/molecular mechanical (QM/MM) methods, we have interfaced the ab initio QM method with the classical Drude oscillator polarizable MM force field (ai-QM/MM-Drude). Different coupling approaches have been employed and compared: 1. the conventional dual self-consistent-field (SCF) procedure; 2. the direct SCF scheme, in which QM densities and MM Drude positions are converged simultaneously; 3. the microiterative SCF scheme, in which the Drude positions of the polarizable model are fully converged during each self-consistent field (SCF) step of QM calculations; 4. the one-step-Drude-update scheme, in which the MM Drude positions are updated only once instead of fully converged during each molecular dynamics (MD) step. The last three coupling approaches are found to be efficient and can achieve the desired convergence in a similar number of QM SCF steps comparing with the corresponding QM method coupled to a nonpolarizable force field. The feasibility and applicability of the implemented ai-QM/MM-Drude approach have been demonstrated by carrying out Born−Oppenheimer molecular dynamics simulations with the umbrella sampling method to determine potentials of mean force for both the methyl transfer reaction of the methyl chlorine-chlorine ion system and the glycine intramolecular proton transfer reaction in aqueous solution. Our results indicate that the ai-QM/MM-Drude approach is very promising, which provides a better description of QM/MM interactions while achieving quite similar computational efficiency in comparison with the corresponding conventional ab initio QM/MM method.

I. Introduction

The combined quantum mechanical/molecular mechanical (QM/MM) approaches[1,2] have been widely used in modeling chemical reactions in complex systems, from the solid and surface catalysis to solution and enzyme reactions.[3–12] With the increase of computer power and the development of more efficient algorithms which make high level electronic structure calculations more affordable, there is a great deal of interest in developing and applying QM/MM approaches based on ab initio quantum mechanical methods to achieve better accuracy and wider applicability. Over the past few years, this field is expanding rapidly, and molecular dynamics simulations with ab initio QM/MM methods have become increasingly feasible.[13–19] Meanwhile, it has been recognized that in order for ab initio QM/MM approaches to become an equal partner of experimental methods, significant development efforts are still needed.

In most QM/MM methods and applications, the conventional nonpolarizable molecular mechanical force fields have been employed, and QM/MM electrostatic interactions are calculated through a Coulombic term in an effective Hamiltonian with MM atoms as fixed point charges. In such a formulation, the electronic-configuration of the QM subsystem changes in response to the presence of MM electrostatic environment, while MM charges are always kept fixed. Thus it only takes into account the polarization effect of the MM charges on the QM subsystem but does not include the polarization effect of the QM subsystem on MM atoms or

* Corresponding author e-mail: yingkai.zhang@nyu.edu.

10.1021/ct800116e CCC: $40.75
Published on Web 06/28/2008

the polarization interactions among MM atoms. It is clear that the origin of this limitation comes from conventional pairwise additive force fields which use fixed atomic charges to model electrostatics. An intuitive and well-known step to make progress is to couple QM methods with polarizable MM force fields to achieve a consistent treatment of QM and MM electrostatic polarization interactions.[20–41]

The available polarizable MM force fields can be mainly divided into three categories:[42,43] the induced point dipole model,[1,44–47] the fluctuating point charge model (also known as electronegativity equalization model),[48–51] and the classical Drude oscillator model[52–65] (also known as the shell model or charge-on-spring model). In the first two approaches, either atomic point dipoles or atomic charges are allowed to fluctuate in response to the environmental electric field changes. In the Drude oscillator model, an induced dipole is represented as a pair of point charges connected with a harmonic spring.

Among polarizable force fields, the induced dipole model was the first employed to interface with QM methods, which was presented when the QM/MM approach was introduced by Warshel and Levitt.[1] Later, the QM/induced-point-dipole method was further developed and implemented by a number of groups and had been employed to investigate the molecular properties in ground state and excited states and study the spectroscopy of organic molecules and biological systems.[20–32] Instead of using the induced point dipole model, Bryce et al. and Field et al. used the fluctuating charge models in their QM/MM-pol calculations.[33,34] Zhang et al. also made use of the fluctuating charge model to treat the polarizability of MM boundary atoms, and they demonstrated that including the mutual polarization of the QM and MM subsystems can yield more accurate results when modeling proton affinities.[37] Compared to the other two categories of polarizable models, the applications of the oscillator Drude model to the QM/MM calculations are seldom seen. Until very recently Geerke et al. combined semiempirical QM method with their charge-on-spring polarizable model (also known as the classical Drude oscillator model or shell model) to perform potential of mean force (PMF) simulations for a S_N2 reaction in the solvent dimethyl ether.[35] Besides the polarizable force field approaches, the importance of the mutual polarization between the QM and MM subsystems is also realized in QM/MM studies treating the environment with the reaction field approach.[38–41]

Most of the above QM/MM-pol calculations so far employed a semiempirical Hamiltonian and a dual SCF procedure: at each iteration cycle, the MM induced dipoles/fluctuating charges were optimized in the presence of a frozen QM wave function, followed by the regular QM self-consistent field (SCF) calculation in the presence of external charges/dipoles. Generally, depending on the convergence criterion, two to five iterative cycles are usually needed to achieve the self-consistencies of both QM wave function and MM induced dipoles/fluctuating charges.[20,23,35] Therefore, the dual SCF scheme makes the QM/MM-pol calculation significantly slower than the corresponding conventional QM/MM calculation. This is especially problematic for the ab initio QM/MM-pol approach, in which the QM SCF calculation is much more expensive than the one in semiempirical QM methods. To circumvent the dual SCF scheme and improve the computational efficiency, Dupuis et al. presented a direct SCF approach,[26] in which QM wave functions and MM induced dipoles are converged simultaneously instead of iteratively. With this direct SCF algorithm, they studied the structure and energy of the formaldehyde and water complex in the ground state and excited state with ai-QM/MM-pol calculations.[26]

From the above account, it is very clear that to couple QM methods with polarizable force fields is both desirable and feasible. However, the QM/MM-pol approaches have been rarely adopted in QM/MM studies of chemical reactions so far, even with the recent renewed enthusiasm in the development of polarizable force fields. Besides the availability of polarizable force field parameters, the lack of popularity of the QM/MM-pol approaches may also be due to the following two concerns: one is the computational cost of such calculations, and the other is the effect of such consistent treatment of polarization on the final results. It can be easily envisioned that the QM/MM-pol approaches would meet much less resistance if they can be demonstrated to have a similar efficiency while better accuracy than the corresponding conventional QM/MM methods. Thus in this paper, we have interfaced ab initio QM methods with the classical Drude oscillator polarizable MM force field (ai-QM/MM-Drude) for Born-Oppenheim molecular dynamics simulations of chemical reactions. In order to improve the efficiency and stability of such simulations, we have explored several schemes of optimizing/updating the Drude particles during MD simulations. The resulted ai-QM/MM-Drude methods have been tested on the water dimer and applied to calculate the potentials of mean force for both the methyl transfer reaction of the methyl chlorine-chlorine ion system and the glycine intramolecular proton transfer reaction in aqueous phase. The results and computational efficiency have been compared with the corresponding conventional ai-QM/MM methods. Our work indicates that the ai-QM/MM-Drude approach is very promising, which provides a more realistic description of QM/MM interactions while achieving quite similar computational efficiency in comparison with the corresponding conventional ai-QM/MM method. We also found that the inclusion of polarization effects can have a significant effect on the calculated free energy profile for the glycine intramolecular proton transfer reaction in aqueous solution.

The outline of the paper is as follows: In section II we give a brief introduction to the Drude oscillator model and the conventional QM/MM approach, followed by the description about the coupling between the QM Hamiltonian and the Drude oscillator model, and the one-step-Drude-update scheme for QM/MM-Drude MD simulations. In section III we present the computational details. Results and discussion are given in section IV.

II. Methodology

A. Classical Drude Oscillator Model. In the classical Drude oscillator model, the induced dipole is represented

by a pair of point charges separated by a variable distance. One point charge q_α is fixed to the charge center site while the other point charge (called Drude particle) $q_{\alpha'}$ is bounded to the charge center site via a harmonic spring of force constant $k_{\alpha'}$. The net charge and dipole moment on this charge center site are $q_\alpha+q_{\alpha'}$ and $\boldsymbol{\mu}_{\alpha'} = q_{\alpha'}\,\mathbf{d}_{\alpha'}$, where $\mathbf{d}_{\alpha'} = |\mathbf{r}_{\alpha'}\text{-}\mathbf{r}_\alpha|$. The total electrostatic energy of the system for a Drude oscillator system is

$$\begin{aligned} E_{ele} &= E_{static-ele} + E_{Drude-ele} + E_{self} \\ &= \sum_{\alpha}\sum_{\beta>\alpha}\frac{q_\alpha q_\beta}{r_{\alpha\beta}} + \left(\sum_{\alpha}\sum_{\beta'}\frac{q_\alpha q_{\beta'}}{r_{\alpha\beta'}} + \sum_{\alpha'}\sum_{\beta'>\alpha'}\frac{q_{\alpha'}q_{\beta'}}{r_{\alpha'\beta'}}\right) + \\ &\qquad \frac{1}{2}\sum_{\alpha'}k_{\alpha'}d_{\alpha'}^2 \end{aligned} \tag{2.1}$$

where the prime denotes the Drude particle site. Comparing the self-energy term in the above equation with the one from the induced dipole model

$$E_{self} = \frac{1}{2}\sum_{\alpha'}\frac{\mu_{\alpha'}^2}{\alpha_{a'}}$$

it leads to the following expression for the isotropic atomic polarizability

$$\alpha_{\alpha'} = \frac{q_{\alpha'}^2}{k_{\alpha'}} \tag{2.2}$$

B. QM/MM Approach. The total energy of a QM/MM system can be written as

$$E_{tot} = E_{qm} + E_{qm/mm} + E_{mm} \tag{2.3}$$

For the Hartree-Fock theory or the Kohn−Sham density functional theory, E_{qm} can be written as

$$E_{qm} = \sum_{\mu\nu}^{AO} D_{\mu\nu}H_{\mu\nu}^{core} + \frac{1}{2}\sum_{\mu\nu\lambda\sigma}^{AO} D_{\mu\nu}D_{\lambda\sigma}(\mu\nu|\lambda\sigma) + E^{XC}[\rho] \tag{2.4}$$

where μ,υ denote the atomic basis set, $D_{\mu\upsilon}$ is the density matrix element, and $E^{XC}[\rho]$ is the exchange-correlation functional of electron density whose form depends on the theory used. $E_{qm/mm}$ is the coupling term between the QM and MM subsystem and can be decomposed into

$$E_{qm/mm} = E_{qm/mm}^{ele} + E_{qm/mm}^{vdw} + E_{qm/mm}^{MM-bonded} \tag{2.5}$$

where $Eqm/mm^{MM\text{-}bonded}$ refers to the bond, angle, and dihedral energy terms at the QM/MM interface. In the actual implementation, the QM/MM electrostatic coupling enters into the QM SCF calculation by adding the following one-electron core Hamiltonian into the Fock matrix $F_{\mu\upsilon}$

$$H_{\mu\upsilon}^{core,qm/mm} = \langle\mu| - \sum_{i}^{electron}\sum_{\alpha}^{MM}\frac{q_\alpha}{r_{i\alpha}}|\upsilon\rangle \tag{2.6}$$

where q_α is the MM atomic charge, and i is the index for electron. Then

$$E_{qm/mm}^{ele} = \sum_{\mu,\upsilon}^{AO} D_{\mu\upsilon}H_{\mu\upsilon}^{core,qm/mm} + \sum_{A}^{QM}\sum_{\alpha}^{MM}\frac{Q_A q_\alpha}{r_{A\alpha}} \tag{2.7}$$

where Q_A is the QM nuclei charge. Now combining eqs 2.4 and 2.7, the final electronic energy can be obtained after determining $D_{\mu\upsilon}$ with the SCF approach according to the variational principle.

C. QM/MM-Drude Oscillator Model. In the QM/MM-Drude approach, the total energy of the system can be written as

$$E_{tot} = E_{qm} + E_{qm/mm} + E_{mm} + E_{Drude-ele} + E_{self} \tag{2.8}$$

where $E_{Drude\text{-}ele}$ is the electrostatic energy term involving the Drude particles

$$\begin{aligned} E_{Drude-ele} &= \sum_{\alpha}^{MM}\sum_{\beta'}^{MM'}\frac{q_\alpha q_{\beta'}}{r_{\alpha\beta'}} + \sum_{\alpha'}^{MM'}\sum_{\beta'>\alpha'}^{MM'}\frac{q_{\alpha'}q_{\beta'}}{r_{\alpha'\beta'}} + \\ &\quad \sum_{A}^{QM}\sum_{\alpha}^{MM'}\frac{Q_A q_{\alpha'}}{r_{A\alpha'}} + \sum_{\mu,\upsilon}^{AO} D_{\mu\upsilon}\langle\mu| - \sum_{i}^{electron}\sum_{\alpha'}^{MM'}\frac{q_{\alpha'}}{r_{i\alpha'}}|\upsilon\rangle \\ &= \sum_{\alpha'}^{MM'} q_{\alpha'}\left(\varphi_{\alpha'}^{MM} + \frac{1}{2}\varphi_{\alpha'}^{MM'} + \varphi_{\alpha'}^{Nuc} + \varphi_{\alpha'}^{electron}\right) \end{aligned} \tag{2.9}$$

and

$$E_{self} = \frac{1}{2}\sum_{\alpha'}^{MM'} k_{\alpha'}d_{\alpha'}^2 \tag{2.10}$$

Here $\varphi_{\alpha'}^{MM}$, $\varphi_{\alpha'}^{MM'}$, $\varphi_{\alpha'}^{Nuc}$, $\varphi\alpha'^{electron}$ are the electrostatic potentials at the Drude site α' generated by the classical point charges (whose positions are immobile during the energy evaluation), the charges of Drude particles, the QM nuclei charge, and QM wave function, respectively. Again the prime indicates the Drude particle site.

It is clear from eqs 2.4, 2.7, 2.9, and 2.10 that now the total energy of the system should be minimized with respect to both $D_{\mu\upsilon}$ and the Drude particle position $\mathbf{r}_{\alpha'}$. Typically, a dual SCF coupling procedure can be employed:

(1) With a frozen QM density, according to eqs 2.9 and 2.10 the Drude particle positions are updated to satisfy

$$\begin{aligned} \nabla_{\mathbf{r}_{\alpha'}}\left(E_{Drude-ele} + E_{self}\right) &= -\sum_{\alpha'}^{MM'} q_{\alpha'}\left(\mathbf{E}_{\alpha'}^{MM} + \frac{1}{2}\mathbf{E}_{\alpha'}^{MM'} + \mathbf{E}_{\alpha'}^{Nuc} + \right. \\ &\quad \left.\mathbf{E}_{\alpha'}^{electron}\right) + k_{\alpha'}\mathbf{d}_{\alpha'} = 0 \end{aligned} \tag{2.11}$$

or rearranged as

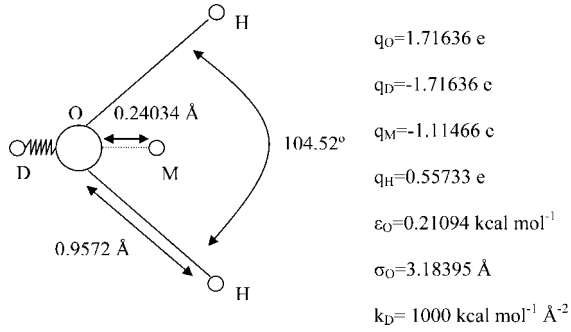
$$\mathbf{d}_{\alpha'} = \frac{1}{k_{\alpha'}}\sum_{\alpha'}^{MM'} q_{\alpha'}\left(\mathbf{E}_{\alpha'}^{MM} + \frac{1}{2}\mathbf{E}_{\alpha'}^{MM'} + \mathbf{E}_{\alpha'}^{Nuc} + \mathbf{E}_{\alpha'}^{electron}\right) \tag{2.12}$$

Either a SCF approach based on eq 2.12 (since the electric field $\mathbf{E}_{\alpha'}$ depends on $\mathbf{d}_{\alpha'}$) or a minimization method based on eq 2.11 can be applied to obtain $\mathbf{d}_{\alpha'}$ or equivalently $\mathbf{r}_{\alpha'}$.

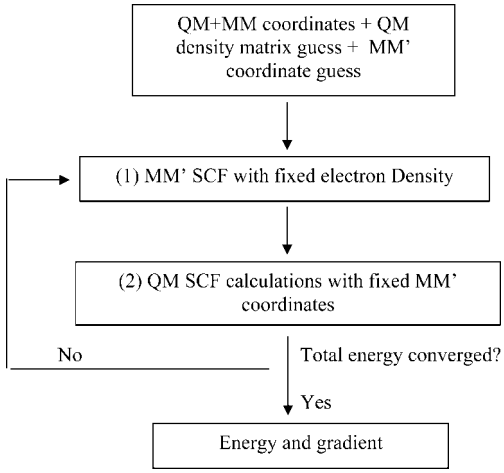
(2) With the updated Drude particle positions determined from step (1), regular QM SCF calculations as implemented in QM software packages are performed to obtain a converged $D_{\mu\upsilon}$. If the total energy is converged, exit; otherwise, go back to step (1).

For clarity, the above dual SCF coupling procedure is illustrated in Scheme 2, which has been employed in most QM/MM-pol methods.

Scheme 1. Parameters of the SWM4-NDP Water Model[64,65]



Scheme 2. Dual SCF Scheme for QM/MM-Drude Calculations



D. Direct SCF and Microiterative SCF Coupling Schemes. Although the dual SCF coupling is very straightforward to implement, it increases QM SCF steps and makes QM/MM-pol calculations significantly slower than the corresponding conventional QM/MM calculations. The alternative is the direct SCF coupling scheme presented by Dupuis et al.,[26] in which step (1) and step (2) are combined so that the QM wave function and the MM induced dipoles can be converged simultaneously. In our current work, we extend the direct SCF algorithm to treat the coupling between the Drude oscillators and the QM wave function, and the details are described as follows:

(A) At each QM SCF step, the current density matrix is used to update the Drude particle positions only once according to eq 2.12.

(B) The one-electron integral

$$\langle \mu | - \sum_{i}^{electron} \sum_{\alpha'}^{MM'} \frac{q_{\alpha'}}{r_{i\alpha'}} | \upsilon \rangle$$

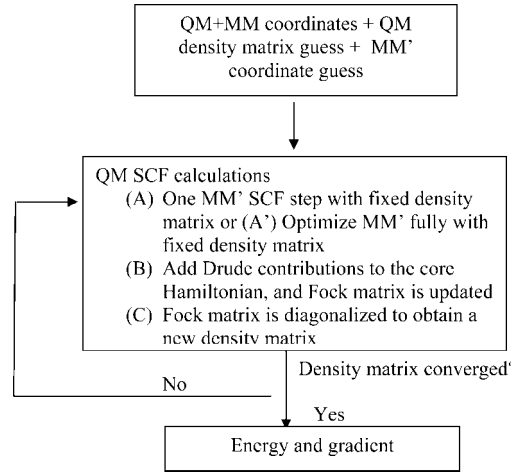
is calculated and added to the core Hamiltonian;

(C) The updated Fock matrix is diagonalized to obtain a new density matrix. If the density matrix and the total energy are converged, exit; otherwise start another QM SCF step.

If we replace the step (A) with a fully optimization step (A′):

(A′) At each QM SCF step, the current density matrix is used to fully optimize the Drude particle positions according to eq 2.12.

Scheme 3. Direct and Microiterative SCF Schemes for QM/MM-Drude Calculations



We obtain a microiterative SCF coupling scheme. From the above description, we can see that step (1) in the dual SCF algorithm corresponds to step (A′), a microiteration step as part of the QM SCF step. In the direct SCF and microiterative coupling schemes, due to the re-evaluations of the one-electron integral

$$\langle \mu | - \sum_{i}^{electron} \sum_{\alpha'}^{MM'} \frac{q_{\alpha'}}{r_{i\alpha'}} | \upsilon \rangle$$

at each QM SCF step, the core Hamiltonian keeps changing until convergence. For clarity, the direct SCF and microiterative SCF coupling schemes are illustrated in Scheme 3.

Compared to the conventional QM/MM approach, the increase of computational time in the QM/MM-Drude calculation using the direct SCF algorithm or the microiterative SCF coupling scheme may also come from the following two aspects:

(1) Computing $\mathbf{E}_{\alpha'}$ according to eq 2.12 for updating the Drude particle positions at step (A) or (A′) and evaluating

$$\langle \mu | - \sum_{i}^{electron} \sum_{\alpha'}^{MM'} \frac{q_{\alpha'}}{r_{i\alpha'}} | \upsilon \rangle$$

at step (B).

(2) The possible increase of QM SCF steps in comparison to the conventional QM/MM calculation.

For the second concern, as demonstrated later from our tests, there is almost no increase of QM SCF steps in our ai-QM/MM-Drude MD simulations. The cost due to (1) needs some comments. Although the two-electron coulomb integral and exchange-correlation integral calculations are computationally much more expensive than the one-electron integral and MM calculations, the cost of computing $\mathbf{E}_{\alpha'}$ and one-electron integrals repetitively involving the Drude particles is still not ignorable, especially when the size of MM-Drude system is large. If $\mathbf{d}_{\alpha'}$ remains small enough (which means the current Drude oscillator model is a good approximation to the induced point dipole model), then $\mathbf{E}_{\alpha'}$ can be well approximated by $\mathbf{E}_{\alpha}$, the electric field at the fixed charge center site. Unlike $\mathbf{E}_{\alpha'}$ which is dependent on the position of the mobile Drude particles, $\mathbf{E}_{\alpha}$ needs to be

computed only once at each QM SCF step. The work from Thompson[21] indicates that for $d_{\alpha'} < 0.1$ Å, the error in energy caused by such an approximation is small. This approximation was taken by Geerke et al. in their semiempirical QM/MM-Drude MD simulations.[35] However, it has been very recently found that the resulted error can be quite significant unless the charges of Drude particles are very large.[66] Here we did not take such an approximation and evaluated $\mathbf{E}_{\alpha'}^{electron}$ whenever the Drude particles move. Therefore, the microiterative SCF coupling scheme is more computationally demanding than the direct SCF scheme, because of the repetitive $\mathbf{E}_{\alpha'}$ calculations at each QM step in the former approach.

E. One-Step-Drude-Update Scheme. In MD simulations with polarizable force fields, instead of obtaining the converged $\mathbf{d}_{\alpha'}$ with either an optimizer or a SCF approach, there are two more efficient ways for updating the Drude particle positions at each MD step. One is called the extended Lagrangian method,[67] in which the Drude particles are assigned with a fictitious mass and treated as dynamic variables and their positions are propagated according to the extended Lagrangian function. By coupling the motions of the Drude particles with a low-temperature thermostat, the system can remain close to the SCF regime.[68] Another method to avoid the SCF calculations in MD polarizable force field simulations is called the always stable predictor-corrector (ASPC) method.[69,70] In this method, the induced dipoles (or $\mathbf{d}_{\alpha'}$ for Drude oscillator model) at time $t+h$ is

$$\boldsymbol{\mu}(t+h) = \omega M(\boldsymbol{\mu}^p(t+h)) + (1-\omega)\boldsymbol{\mu}^p(t+h) \tag{2.13}$$

where h denotes the time step, ω is the relaxation parameter to ensure the stability, which prevents the error in $\boldsymbol{\mu}(t+h)$ accumulating along the trajectory, and M represents the right-hand side (RHS) equation for $\boldsymbol{\mu}$ in the SCF procedure. From eq 2.12 we can see that the RHS equation of $\mathbf{d}_{\alpha'}$ in the Drude oscillator model is just

$$M \equiv \frac{1}{k_{\alpha'}} \sum_{\alpha'}^{MM'} q_{\alpha'} \left(\mathbf{E}_{\alpha'}^{MM} + \frac{1}{2}\mathbf{E}_{\alpha'}^{MM'} + \mathbf{E}_{\alpha'}^{Nuc} + \mathbf{E}_{\alpha'}^{electron} \right) \tag{2.14}$$

In eq 2.13, $\boldsymbol{\mu}^p(t+h)$ is the guess of the induced dipole based on the historical information from previous MD steps. While the simplest form is just a linear extrapolation, $\boldsymbol{\mu}^p(t+h) = 2\boldsymbol{\mu}(t)\text{-}\boldsymbol{\mu}(t\text{-}h)$, Kolafa gave more elaborate forms to improve the accuracy and time reversibility.[69,70]

Recently Niklasson et al. proposed a lossless time-reversible ab initio QM MD scheme,[71,72] which allows stable Hatree-Fock MD simulations with only one single QM SCF cycle per time step. In our current work, we extend their method to update the positions of Drude particles in a similar manner as they update the density matrix along the MD simulations. Specifically

$$\mathbf{d}_{\alpha'}^p(t+h) = 2\mathbf{d}_{\alpha'}(t) - \mathbf{d}_{\alpha'}^p(t-h) \tag{2.15}$$

$$\mathbf{d}_{\alpha'}(t+h) = M(\mathbf{d}_{\alpha'}^p(t+h)) \tag{2.16}$$

where M is defined in eq 2.14. In eq 2.15 a simple linear extrapolation form is used, and the time reversibility is reserved by using $\mathbf{d}_{\alpha'}^p(t\text{-}h)$ instead of $\mathbf{d}_{\alpha'}(t\text{-}h)$. As tested, the resulted lossless time-reversible MD with one-step-Drude-update scheme allows a stable MM Drude oscillator MD simulation by updating the positions of Drude particles only once at each time step. The details of integrating this lossless time-reversible MD scheme into QM/MM-Drude MD simulations are described in Scheme 4.

III. Computational Details

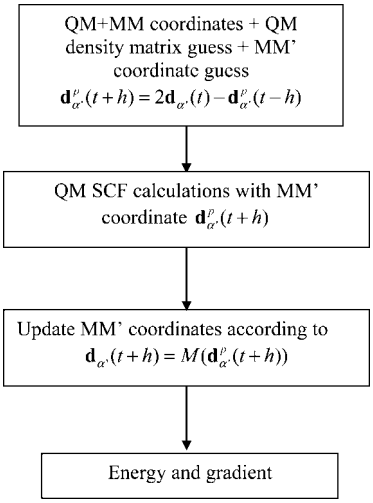
In the QM/MM-Drude oscillator approach, for the water solvent, we employed a SWM4-NDP model, a Drude oscillator model from Lamoureux et al.,[64,65] which is calibrated to reproduce important bulk properties of the water at room temperature and pressure. The parameters of the SWM4-NDP water model are shown in Scheme 1. In conventional QM/MM calculations, we use the TIP3P water model.[73] The QM/MM and QM/MM-Drude schemes were implemented by modifying Gaussian03[74] and the TINKER program.[75] The correctness of the implementation has been carefully checked by comparing the analytical and numerical gradients and the energy conservation from short NVE MD runs.

For the PMF calculation of glycine intramolecular proton transfer reaction, the solutes were solvated with a 20 Å sphere of waters. Only the waters within 15 Å of the sphere center were allowed to move during the MD simulations. For the methyl-transfer reaction, the solute was solvated with a 18 Å sphere of waters, and only 13 Å of the sphere center were allowed to move during MD simulations. The bonds in water molecules were constrained using the RATTLE algorithm.[76,77] A time step of 1 fs was employed for the MD simulations, and the mass of deuterium was used for the hydrogen atoms in the glycine molecule. No cutoff was used for the nonbonded interactions. The velocity verlet integrator implemented in the TINKER program was used, and the temperatures of the systems were controlled at 300 K with the Berendsen velocity scaling method.[78] The PMF calculations were performed with the umbrella sampling and the Weighted Histogram Analysis method.[79,80]

For the methyl transfer reaction of the methyl chlorine-chlorine ion system, the reaction coordinate (rc) is chosen as $rc = d_{c\text{-}cl1}\text{-}d_{c\text{-}cl2}$, and 22 windows centering from reaction coordinate of −3.8 to 0.0 Å were used. The symmetry was used to obtain a full PMF curve. For each window, 10 ps equilibration was performed, followed by a 20 ps data collection. HF/6−31G(d) was employed for the QM calculations. The convergence criterion in the microiterative SCF algorithm was set to rms gradient 0.001 kcal mol^{-1} Å^{-2} per Drude.

For the glycine intramolecular proton transfer reaction, the reaction coordinate is defined as $rc = d_{H10\text{-}N1}\text{-}d_{H10\text{-}O5}$ (see Figure 1), and 17 windows were employed. For each window, 12 ps equilibration were performed, followed by 24 ps data collection. B3LYP/6−31G(d) was employed for the QM calculations. The convergence criterion in the microiterative SCF algorithm was set to rms gradient 0.01 kcal mol^{-1} Å^{-2} per Drude. The Lennard-Jones parameters from the Amber94 force field[81] were used for the QM subsystems in both ai-QM/MM-Drude and ai-QM/MM

Scheme 4. One-Step-Drude-Update Scheme for QM/MM-Drude MD Simulations



calculations. It should be noted that ideally different QM van der Waals parameters should be used to further improve the description of QM/MM interactions. Some work was already done in parametrizing the QM van der Waals parameters for the QM/MM models,[82–84] but the influence of the QM van der Waals parameters on the ai-QM/MM-Drude calculations has not been addressed. The work along this direction is currently in progress and will be presented in future publications.

IV. Results and Discussion

A. Comparison between the SWM4-NDP Water and the QM Water. As illustrated in Scheme 1, the SWM4-NDP water model[64,65] uses the point charges for electrostatic interactions and the 12−6 Lennard-Jones potentials for the repulsive and dispersion interactions. The molecular polarizability is modeled by a Drude particle attached to the oxygen atom. According to eq 2.2, the molecular polarizability given by this model is 0.97825 Å^3, much smaller than the experimental value 1.44 Å^3. The reduced molecular polarizability was found to be essential in reproducing the liquid properties,[64] which may be due to the simplicity of this model.

Before applying this water model in our ai-QM/MM-Drude simulations, it will be informative to learn how accurate this simple polarizable water model is in modeling the electrostatic interactions. To avoid the complexity caused by choosing Lennard-Jones parameters, we studied the interactions between a water and a point charge. As shown in the insets of Figure 2, we moved either a +1e or −1e point charge in four different directions and calculated the interaction energies along the charge-oxygen distance for the SWM4-NDP water, the TIP3P water, and the QM water in the MP2/aug-cc-pvtz description. From Figure 2 we can see that in general the SWM4-NDP water model yields a better agreement with the QM description than the TIP3P water, although the electrostatic interactions are underestimated in the SWM4-NDP water, which is not a surprise considering that the molecular polarizability given by SWM4-NDP is

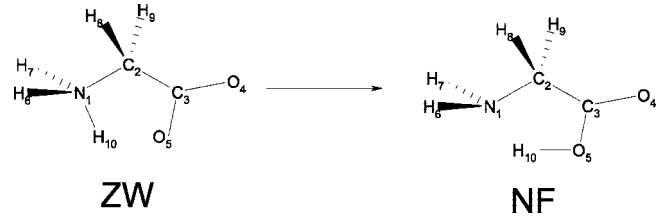


Figure 1. Glycine intramolecular proton transfer reaction.

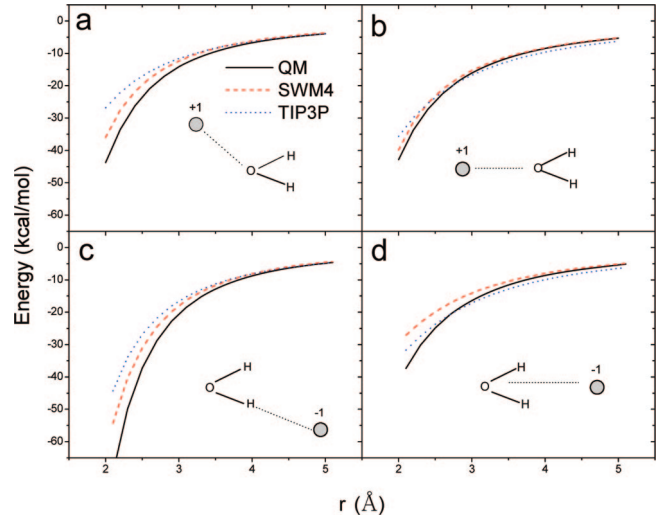


Figure 2. Electrostatic interaction energies between a water and a point charge as a function of charge-oxygen distance. The insets illustrate the approaching directions of the point charge (+1 or −1): (a) along the oxygen electron lone pair direction; (b) along the C_2 axis of the symmetry; (c) along the O−H bond direction; and (d) along the C_2 axis of the symmetry from the side of the hydrogen atoms.

significantly lower than the values from experiments or the MP2/aug-cc-pvtz calculation.

Another test we performed is to calculate the water dimer interaction energies. The QM calculations were done at the B3LYP/6−31G(d,p) level. Two cases were considered for the QM/MM-Drude and QM/MM calculations. One corresponds to the QM water as the hydrogen bond donor, while in the other case the QM water is the hydrogen bond acceptor. For the QM/MM-Drude calculations, the Lennard-Jones parameters of the QM water are taken from those of the SWM4-NDP water model, while in the QM/MM calculations, the Lennard-Jones parameters of the QM water are taken from those of the TIP3P water model. As illustrated in Figure 3, in contrary to those from the QM/MM calculations, the QM/MM-Drude calculations for the two cases considered yield the consistent binding energies and binding distances and have close agreement with those from pure QM calculations, which suggest that the SWM4-NDP water has a better performance in terms of "mimicking" the QM water than the TIP3P water does. These results are encouraging, which indicate that the ai-QM/MM-Drude method can provide a better description of QM/MM interactions than the corresponding conventional ai-QM/MM method.

B. Computational Efficiency of ai-QM/MM-Drude MD Simulations. To evaluate the feasibility of ai-QM/MM-Drude calculations, we performed short-time MD simulations for several systems with both conventional ai-QM/MM and

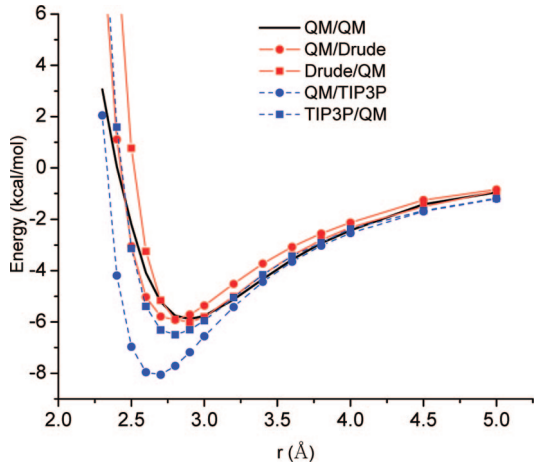


Figure 3. Water dimer interaction energies as a function of the O−O distance. QM-QM denotes QM treatment for the whole water dimer; QM-SWM4 (TIP3P) denotes QM for the acceptor and SWM4-NDP (TIP3P) for the donor; and SWM4 (TIP3P)-QM denotes QM for the donor and SWM4-NDP (TIP3P) for the acceptor.

ai-QM/MM-Drude simulations. The computational efficiency of the dual SCF, direct SCF, and microiterative SCF schemes as well as the one-step-Drude-update scheme have been tested. The results of these short MD simulations (0.1 ps ∼ 1.0 ps) are summarized in Table 1. First we observe that in comparison with conventional QM/MM calculations, except the dual SCF scheme, coupling the QM subsystem with the MM-Drude subsystem will not increase the QM SCF steps for all the other three Drude optimization/updating schemes. Second, among the different Drude optimization/updating schemes, the one-step-Drude-update scheme is the most efficient one. This is fully expected because comparing with other schemes, there is no increase in the QM SCF steps and no repetitive one-electron integrals calculation involved in this scheme. Third, the increase in the size of the QM subsystem leads to the decrease of the time ratio between ai-QM/MM MD and ai-QM/MM-Drude MD, which becomes more close to the unity. We can see that for a medium QM subsystem of ∼300 basis functions, the increase of computational time due to employing the Drude oscillator polarizable force field becomes insignificant for the direct SCF, microiterative SCF, and one-step-Drude-update schemes.

C. Potential of Mean Force Calculations of the Chemical Reactions in Solution. The chemical reactions often involve the charge transfer steps, which cause significant changes of the electrostatic properties around the reaction center. The polarizable waters should be able to adapt to the change of electronic configurations of the reaction center by adjusting its dipole moments, which in turn has an impact on the charge distributions of the reaction center. Therefore, the polarizable water model may have a strong influence on the calculated energetics of chemical reactions. In this work, we have investigated two chemical reactions in solution, one is a S_N2 reaction, the methyl transfer of the methyl-chlorine-chlorine ion system, and the other is the glycine intramolecular proton transfer reaction (see Figure 1).

The potential of mean force curves for the S_N2 reaction is shown in Figure 4. We obtained a barrier of 26.2 kcal/mol from the HF(6−31G(d))/TIP3P simulation and 27.5 kcal/mol from the HF(6−31G(d))/SWM4-NDP simulation. There have been extensive studies on the self-exchange of the $Cl^- + CH_3Cl$ reaction by different groups,[85–89] and the reported free energy barriers are in the range of 26−27 kcal/mol. Our results are in good agreement with those from the literature. The inclusion of the explicit water polarizability increases the barrier by ∼1.3 kcal/mol (or ∼5%).

The intramolecular proton transfer reaction of glycine is illustrated in Figure 1. While the neutral form (NF) of the glycine is energetically more stable than the zwitterion (ZW) form in the gas phase by more than 20 kcal/mol (see Figure 5), the ZW is the predominant form in water, which indicates that the water solvent plays a critical role in stabilizing the ZW form. The glycine intramolecular proton transfer reaction has also been widely studied through many different theoretical methods, including the polarizable continuum model (PCM),[90] the water cluster plus the PCM modeling,[91–93] the EVB description,[94,95] the QM/MM treatment,[96–98] and the CPMD simulation.[99] While the experimentally measured free energy difference of the form (ZW) and the neutral form (NF) of glycine in aqueous solution is about 7.27 kcal/mol,[100] the reported theoretical values of the free energy difference between the ZW conformer and the NF conformer were in a wide range, from about 4.8 to 11.2 kcal/mol. In this work, we used B3LYP/6−31G(d) for the QM calculations and the TIP3P and SWM4-NDP water models. As shown in Figure 5, by comparing with the results from MP2/aug-cc-pvtz, we find that B3LYP/6−31G(d) is fairly accurate in describing this intramolecule proton reaction in the gas phase. Figure 6 demonstrates that the employment of the polarizable force field has a strong effect on the resulted free energy reaction profiles in the aqueous solution. We obtained a 7.4 kcal/mol transition state barrier and a free energy difference of 4.4 kcal/mol between the ZW and the NF from the B3LYP(6−31G)(d))/TIP3P simulation. The corresponding B3LYP(6−31G(d))/SWM4-NDP simulation gives a barrier of 8.9 kcal/mol and a free energy difference of 6.4 kcal/mol. By collecting the data from the reactant region (ZW) defined as -1.5 Å $< rc < -0.8$ Å and the product region (NF) defined as 0.6 Å $< rc <$1.0 Å, we find the difference of the averaged glycine-water interaction energy between the ZW and NF is 73 kcal/mol from the QM/MM simulation and 83 kcal/mol from the QM/MM-Drude simulation. Apparently the polarizable water model gives a relatively stronger glycine-water interaction for the ZW. As shown in Figure 7, there is no significant difference of the radial distribution functions between the TIP3P and SWM4-NDP water model. However, as illustrated in Figure 8, we observe that the averaged SWM4-NDP water dipoles within the first solvation shell of the N1 atom of glycine in the ZW form are significantly larger than the dipole of TIP3P water (2.35 Deybe). Our results here suggest that the polarizable water model can strongly affect the resulted free energy reaction profiles by influencing the QM/MM interactions without changing the solvation structure.

Table 1. Comparison of SCF Steps and Time Ratios between ai-QM/MM and ai-QM/MM-Drude with Different Schemes of Optimizing/Updating the Positions of Drude Particles during MD Simulations[a]

			ai-QM/MM-Drude			
		ai-QM/MM	microiterative SCF	direct SCF	one-step-Drude-update	dual SCF
$Cl^- + CH_3Cl$ (HF 59 basis sets)	SCF steps	6.2	6.1	6.1	6.1	10.4
	time ratio	1.0	2.4	2.0	1.6	3.4
glycine (B3LYP 85 basis sets)	SCF steps	7.0	7.0	7.0	7.0	11.5
	time ratio	1.0	1.7	1.5	1.25	2.0
adenosine (HF 311 basis sets)	SCF steps	7.0	7.0	7.0	7.0	10.7
	time ratio	1.0	1.13	1.09	1.05	1.44

[a] For calculating the time ratio, we have employed the computational cost of ai-QM/MM calculations as the reference.

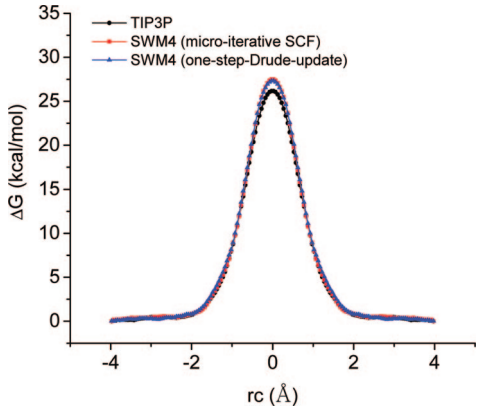


Figure 4. Potential of mean force for the $Cl^- + CH_3Cl$ reaction in TIP3P and in SWM4-NDP obtained using the microiterative SCF scheme and the one-step-Drude-update scheme.

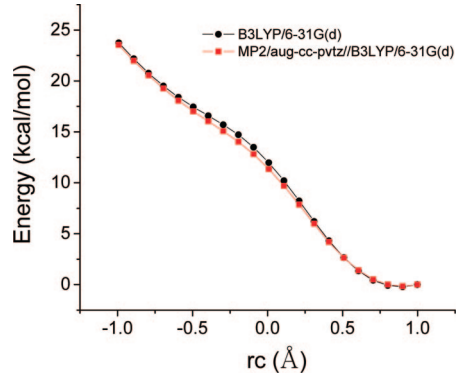


Figure 5. Energies of the glycine intramolecular proton transfer reaction in gas phase.

All the calculations described above have employed the microiterative SCF scheme so that the Drude positions are fully converged at each MD step. Such simulations are more expensive than the one-step-Drude-update scheme as presented in section II.E. Since the Drude positions are not fully converged in the one-step-Drude-update scheme, one concern is that the Drude particles may exert systematic drag forces on the physical atoms along the MD simulations, which affects the resulted PMF. As a test, we employed the one-step-Drude-update scheme in our ai-QM/MM-Drude MD simulations for the PMF calculations of the two chemical reactions described above. To make sure the Drude particles stay close to the SCF regime, we switched to the microiterative SCF scheme every 100 MD steps. From Figures 4 and 6, we can see that the curves obtained with different

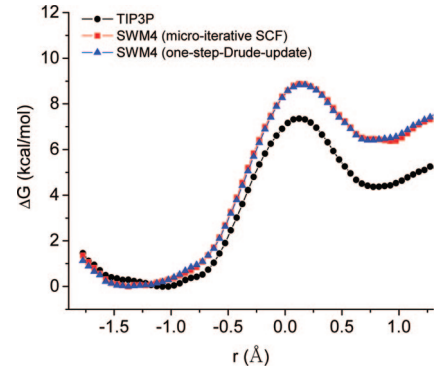


Figure 6. Potential of mean force for the glycine intramolecular proton transfer reaction in TIP3P and in SWM4-NDP obtained using the microiterative SCF scheme and the one-step-Drude-update scheme.

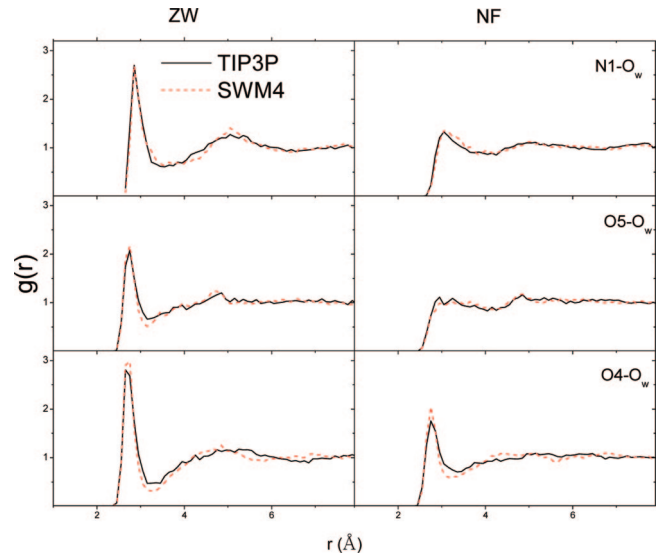


Figure 7. Radial distribution functions $g_{N1\text{-}Ow}$ (top), $g_{O5\text{-}Ow}$ (middle), and $g_{O4\text{-}Ow}$ (bottom) of the glycine in the ZW form (left panel) and the NF form (right panel) solvated in TIP3P and SWM4-NDP waters.

schemes overlap very well with each other, which suggests the promise of the one-step-Drude-update scheme in the QM/MM-Drude MD simulations.

V. Conclusion

In this work, we have presented a detailed description of the methodologies of coupling the ab inito QM methods with the classical Drude oscillator model and applied the ai-QM/MM-Drude approach to the PMF calculations of two

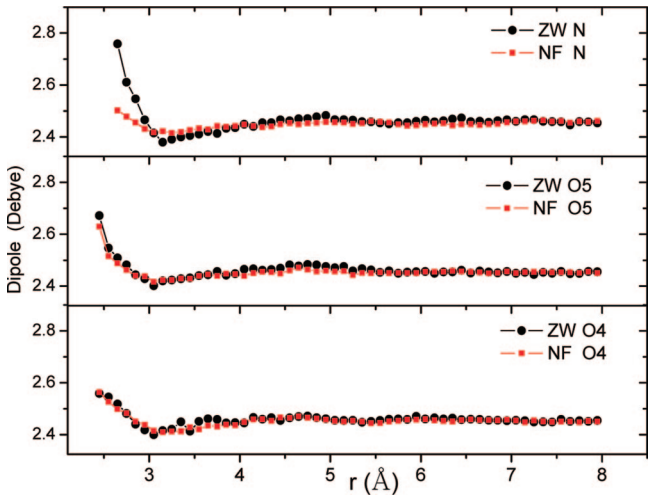


Figure 8. Averaged water dipoles in the ZW form and the NF form of glycine along $r_{N1\text{-}Ow}$, $r_{O5\text{-}Ow}$, and $r_{O4\text{-}Ow}$.

chemical reactions in solution. Besides the dual SCF, direct SCF, and microiterative SCF schemes, we have presented a one-step-Drude-update scheme, in which the Drude positions of MM subsystem are updated only once instead of fully converged during each molecular dynamics simulation step. The resulted ai-QM/MM-Drude MD simulations are found to be highly efficient and yield chemical reaction free energy profiles in quantitative agreement with the corresponding fully converged-Drude-update simulations. In comparison with the corresponding ai-QM/MM calculations, the computational cost overhead for the ai-QM/MM-Drude calculations with efficient implementation schemes is rather insignificant when the QM subsystem consists of tens of atoms, which is typical for most ai-QM/MM applications. The feasibility and applicability of the implemented ai-QM/MM-Drude approach have been demonstrated by performing the potentials of mean force calculations for both the methyl transfer reaction of the methyl chlorine-chlorine ion system and the glycine intramolecular proton transfer reaction in aqueous solution. Compared with the results from ai-QM/MM, the relative effects on the free energy profiles by switching from TIP3P to a polarizable water model for the MM environment are found to be significant for the glycine intramolecular transfer reaction. With the continuing development of polarizable MM force fields and the further improvement in describing QM/MM interactions, the ai-QM/MM-Drude approach should become a more robust approach than the conventional QM/MM approach in modeling chemical reactions in solutions and biological systems.

Acknowledgment. We thank Dr. Shenglong Wang for helpful discussion. This work has been supported by the National Institutes of Health (R01-GM079223) and the National Science Foundation (CHE-CAREER-0448156). We thank NYU-ITS and NCSA for providing computational resources and support.

References

(1) 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.

(2) 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.

(3) Gao, J. L.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467–505.

(4) Hammes-schiffer, S. Quantum-classical simulation methods for hydrogen transfer in enzymes: a case study of dihydrofolate reductase. *Curr. Opin. Struct. Biol.* **2004**, *14*, 192–201.

(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) Friesner, R. A.; Guallar, V. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389–427.

(7) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H. B.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G. H.; Xu, D. G.; 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.

(8) Zhang, Y. Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theor. Chem. Acc.* **2006**, *116*, 43–50.

(9) Senn, H. M.; Thiel, W. QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182–187.

(10) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here. *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(11) Shurki, A.; Warshel, A. Structure/function correlations of proteins using mm, qm/mm, and related approaches: methods, concepts, pitfalls, and current progress. *Adv. Protein Chem.* **2003**, *66*, 249–313.

(12) Yang, W.; Hu, H. Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanical/molecular mechanics methods. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.

(13) Singh, U. C.; Kollman, 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.

(14) Stanton, R. V.; Hartsough, D. S.; Merz, K. M. Calculation of Solvation Free Energies Using a Density Functional/Molecular Dynamics Coupled Potential. *J. Phys. Chem.* **1993**, *97*, 11868–11870.

(15) Yarne, D. A.; Tuckerman, M. E.; Martyna, G. J. A dual length scale method for plane-wave-based, simulation studies of chemical systems modeled using mixed ab initio/empirical force field descriptions. *J. Chem. Phys.* **2001**, *115*, 3531–3539.

(16) Carloni, P.; Rothlisberger, U.; Parrinello, M. The role and perspective of a initio molecular dynamics in the study of biological systems. *Acc. Chem. Res.* **2002**, *35*, 455–464.

(17) Rega, N.; Iyengar, S. S.; Voth, G. A.; Schlegel, H. B.; Vreven, T.; Frisch, M. J. Hybrid ab-initio/empirical molecular dynamics: combining the oniom scheme with the atom-

centered density matrix propagation (admp) approach. *J. Phys. Chem. B* **2004**, *108*, 4210–4220.

(18) Hu, P.; Wang, S.; Zhang, Y. How do SET-domain protein lysine methyltransferases achieve the methylation state specificity? Revisited by ab initio QM/MM molecular dynamics simulations. *J. Am. Chem. Soc.* **2008**, *130*, 3806–3813.

(19) Wang, S.; Hu, P.; Zhang, Y. Ab initio quantum mechanical/molecular mechanical molecular dynamics simulation of enzyme catalysis: The case of histone lysine methyltransferase SET7/9. *J. Phys. Chem. B* **2007**, *111*, 3758–3764.

(20) Thompson, M. A.; Schenter, G. K. Excited states of the bacteriochlorophyll b dimmer of Rhodopseudomonas viridis: A QM/MM study of the photosynthetic reaction center that includes MM polarization. *J. Phys. Chem.* **1995**, *99*, 6374–6386.

(21) Thompson, M. A. Qm/mmpol: a consistent model for solute/solvent polarization. application to the aqueous solvation and spectroscopy of formaldehyde, acetaldehyde, and acetone. *J. Phys. Chem.* **1996**, *100*, 14492–14507.

(22) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(23) Gao, J. L. Energy components of aqueous solution: insight from hybrid qm/mm simulations using a polarizable solvent model. *J. Comput. Chem.* **1997**, *18*, 1061–1071.

(24) Aida, M.; Yamataka, H.; Dupuis, M. Critical assessment of the hybrid QM/MM-pol-vib approach: Small water clusters using polarizable flexible water potentials. *Int. J. Quantum Chem.* **2000**, *77*, 199–210.

(25) Houjou, H.; Inoue, Y.; Sakurai, M. Study of the opsin shift of bacteriorhodopsin: insight from qm/mm calculations with electronic polarization effects of the protein environment. *J. Phys. Chem. B* **2001**, *105*, 867–879.

(26) Dupuis, M.; Aida, M.; Kawashima, Y.; Hirao, K. A polarizable mixed hamiltonian model of electronic structure for micro-solvated excited states. i. energy and gradients formulation and application to formaldehyde ($^1A_2$). *J. Chem. Phys.* **2002**, *117*, 1242–1255.

(27) Dupuis, M.; Kawashima, Y.; Hirao, K. A polarizable mixed Hamiltonian model of electronic structure for solvated excited states. II. Application to the blue shift of the H2CO $^1(\pi^*\leftarrow n)$ excitation in water. *J. Chem. Phys.* **2002**, *117*, 1256–1268.

(28) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. Coupled cluster/molecular mechanics method: Implementation and application to liquid water. *J. Phys. Chem. A* **2003**, *107*, 2578–2588.

(29) Jensen, L.; Duijnen, P. T.v.; Snijders, J. G. A discrete solvent reaction field model within density functional theory. *J. Chem. Phys.* **2003**, *118*, 514–521.

(30) Illingworth, C. J. R.; Gooding, S. R.; Winn, P. J.; Jones, G. A.; Ferenczy, G. G.; Reynolds, C. A. Classical polarization in hybrid QM/MM methods. *J. Phys. Chem. A* **2006**, *110*, 6487–6497.

(31) Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J. Density functional self-consistent quantum mechanics/molecular mechanics theory for linear and nonlinear molecular properties: Applications to solvated water and formaldehyde. *J. Chem. Phys.* **2007**, *126* (15), 154112–154118.

(32) Lin, Y. L.; Gao, J. L. Solvatochromic shifts of the $n\rightarrow\pi^*$ transition of acetone from steam vapor to ambient aqueous solution: A combined configuration interaction QM/MM simulation study incorporating solvent polarization. *J. Chem. Theory Comput.* **2007**, *3*, 1484–1493.

(33) Bryce, R. A.; Buesnel, R.; Hillier, I. H.; Burton, N. A. A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges. *Chem. Phys. Lett.* **1997**, *279*, 367–371.

(34) Field, M. J. Hybrid quantum mechanical molecular mechanical fluctuating charge models for condensed phase simulations. *Mol. Phys.* **1997**, *91*, 835–845.

(35) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. Combined QM/MM molecular dynamics study on a condensed-phase S_N2 reaction at nitrogen: The effect of explicitly including solvent polarization. *J. Chem. Theory Comput.* **2007**, *3*, 1499–1509.

(36) Washel, A.; Kato, M.; Pisliakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.

(37) Zhang, Y.; Lin, H.; Truhlar, D. G. Self-consistent polarization of the boundary in the redistributed charge and dipole scheme for combined quantum-mechanical and molecular-mechanical calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1378–1398.

(38) Hannachi, Y.; Angyan, J. G. The role of induction forces in infrared matrix shifts - quantum chemical calculations with reaction field model hamiltonian. *THEOCHEM - J. Mol. Struct.* **1991**, *78*, 97–110.

(39) Jansen, G.; Colonna, F.; Angyan, J. G. Mixed quantum-classical calculations on the water molecule in liquid phase: Influence of a polarizable environment on electronic properties. *Int. J. Quantum Chem.* **1996**, *58*, 251–265.

(40) Thole, B. T.; Vanduijnen, P. T. On the quantum-mechanical treatment of solvent effects. *Theor. Chim. Acta* **1980**, *55* (4), 307–318.

(41) van Duijnen, P. T.; Grozema, F.; Swart, M. Some applications of the direct reaction field approach. *THEOCHEM - J. Mol. Struct.* **1999**, *464*, 191–198.

(42) Rick, S. W.; Stuart, S. J. Potentials and algorithms for incorporating polarizability in computer simulations. In *Review in Computational Chemistry*; VCH: New York, 2002; Vol. *18*, pp 89−146.

(43) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(44) Cieplak, P.; Caldwell, J.; Kollman, P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and n-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.* **2001**, *22*, 1048–1057.

(45) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.* **2002**, *23*, 1515–1531.

(46) Ren, P. Y.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(47) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an accurate and robust polarizable molecular

mechanics force field from ab initio quantum chemistry. *J. Phys. Chem. A* **2004**, *108*, 621–627.

(48) Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular-dynamics simulations. *J. Phys. Chem.* **1991**, *95*, 3358–3363.

(49) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force-fields - application to liquid water. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(50) York, D. M.; Yang, W. T. A chemical potential equalization method for molecular simulations. *J. Chem. Phys.* **1996**, *104*, 159–172.

(51) Patel, S.; Brooks, C. L. Charmm fluctuating charge force field for proteins: i parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25*, 1–15.

(52) Drude, P. *The Theory of Optics*; Longmans, Green, and Co.: New York, 1902.

(53) Born, M.; Huang, K. *Dynamic Theory of Crystal Lattices*; Oxford University Press: Oxford, U.K., 1954.

(54) Straatsma, T. P.; McCammon, J. A. Molecular Dynamics Simulations with Interaction Potentials Including Polarization. Development of a Noniterative Method and Appplication to Water. *Mol. Simul.* **1990**, *5*, 181–192.

(55) van Maaren, P. J.; van der Spoel, D. Molecular dynamics simulations of water with novel shell-model potentials. *J. Phys. Chem. B* **2001**, *105*, 2618–2626.

(56) Yu, H. B.; Hansson, T.; van Gunsteren, W. F. Development of a simple, self-consistent polarizable model for liquid water. *J. Chem. Phys.* **2003**, *118*, 221–234.

(57) Lamoureux, G.; Roux, B. Modeling induced polarization with classical drude oscillators: theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

(58) Yu, H. B.; van Gunsteren, W. F. Charge-on-spring polarizable water models revisited: From water clusters to liquid water to ice. *J. Chem. Phys.* **2004**, *121*, 9549–9564.

(59) Yu, H. B.; van Gunsteren, W. F. Accounting for polarization in molecular simulation. *Comput. Phys. Commun.* **2005**, *172*, 69–85.

(60) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.

(61) Vorobyov, I. V.; Anisimov, V. M.; MacKerell, A. D. Polarizable empirical force field for alkanes based on the classical drude oscillator model. *J. Phys. Chem. B* **2005**, *109*, 18988–18999.

(62) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.* **2006**, *2*, 1587–1597.

(63) Geerke, D. P.; van Gunsteren, W. F. Calculation of the free energy of polarization: Quantifying the effect of explicitly treating electronic polarization on the transferability of force-field parameters. *J. Phys. Chem. B* **2007**, *111*, 6425–6436.

(64) Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119*, 5185–5197.

(65) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245–249.

(66) Geerke, D. P.; van Gunsteren, W. On the Calculation of Atomic Forces in Classical Simulation Using the Charge-on-Spring Method To Explicitly Treat Electronic Polarization. *J. Chem. Theory Comput.* **2007**, *3*, 2128–2137.

(67) Sprik, M.; Klein, M. L. A polarizable model for water using distributed charge sites. *J. Chem. Phys.* **1988**, *89*, 7556–7560.

(68) Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

(69) Kolafa, J. Numerical integration of equations of motion with a self-consistent field given by an implicit equation. *Mol. Simul.* **1996**, *18*, 193–212.

(70) Kolafa, J. Time-reversible always stable predictor-corrector method for molecular dynamics of polarizable molecules. *J. Comput. Chem.* **2004**, *25*, 335–342.

(71) Niklasson, A. M. N.; Tymczak, C. J.; Challacombe, M. Time-reversible Born-Oppenheimer molecular dynamics. *Phys. Rev. Lett.* **2006**, 97.

(72) Niklasson, A. M. N.; Tymczak, C. J.; Challacombe, M. Time-reversible ab initio molecular dynamics. *J. Chem. Phys.* **2007**, 126.

(73) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(74) 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 03,* Revision D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(75) Ponder, J. W. TINKER, Software Tools for Molecular Design, Version 4.2. The most updated version for the TINKER program can be obtained from J. W. Ponder's World Wide Web at http://dasher.wustl.edu/tinker, June 2004.

(76) 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.

(77) Andersen, H. C. RATTLE - A velocity version of the shake algorithm for molecular-dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

(78) Berendsen, H. J. C.; Postma, J. P. M.; Gunstereno, W. F. V.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 684–3690.

(79) Ferrenberg, A. M.; Swendsen, R. H. Optimized monte-carlo data-analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.

(80) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. the method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(81) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(82) Freindorf, M.; Gao, J. L. Optimization of the lennard-jones parameters for a combined ab initio quantum mechanical and molecular mechanical potential using the 3−21g basis set. *J. Comput. Chem.* **1996**, *17*, 386–395.

(83) Riccardi, D.; Li, G. H.; Cui, Q. Importance of van der waals interactions in qm/mm simulations. *J. Phys. Chem. B* **2004**, *108*, 6467–6478.

(84) Freindorf, M.; Shao, Y. H.; Furlani, T. R.; Kong, J. Lennard-Jones parameters for the combined QM/MM method using the B3LYP/6−31+G*/AMBER potential. *J. Comput. Chem.* **2005**, *26* (12), 1270–1278.

(85) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. Theoretical examination of S_N2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.* **1985**, *107*, 154–162.

(86) Hwang, J. K.; King, G.; Creighton, S.; Warshel, A. Simulation of free-energy relationships and dynamics of sn2 reactions in aqueous-solution. *J. Am. Chem. Soc.* **1988**, *110*, 5297–5311.

(87) Jorgensen, W. L. Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.

(88) Bash, P. A.; Field, M. J.; Karplus, M. Free-energy perturbation method for chemical-reactions in the condensed phase - a dynamical-approach based on a combined quantum and molecular mechanics potential. *J. Am. Chem. Soc.* **1987**, *109*, 8092–8094.

(89) Mo, Y. R.; Gao, J. L. Ab initio QM/MM simulations with a molecular orbital-valence bond (MOVB) method: Application to an S_N2 reaction in water. *J. Comput. Chem.* **2000**, *21*, 1458–1469.

(90) Tunon, I.; Silla, E.; Ruiz-Lopez, M. F. On the tautomerization process of glycine in aqueous solution. *Chem. Phys. Lett.* **2000**, *321*, 433–437.

(91) Fernandez-Ramos, A.; Smedarchina, Z.; Siebrand, W.; Zgierski, M. Z. A direct-dynamics study of the zwitterion-to-neutral interconversion of glycine in aqueous solution. *J. Chem. Phys.* **2000**, *113*, 9714–9721.

(92) Bandyopadhyay, P.; Gordon, M. S.; Mennucci, B.; Tomasi, J. An integrated effective fragment-polarizable continuum approach to solvation: Theory and application to glycine. *J. Chem. Phys.* **2002**, *116*, 5023–5032.

(93) Aikens, C. M.; Gordon, M. S. Incremental solvation of nonionized and zwitterionic glycine. *J. Am. Chem. Soc.* **2006**, *128*, 12835–12850.

(94) 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.

(95) 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.

(96) Tunon, I.; Silla, E.; Millot, C.; Martins-Costa, M. T. C.; Ruiz-Lopez, M. F. Intramolecular proton transfer of glycine in aqueous solution using quantum mechanics-molecular mechanics simulations. *J. Phys. Chem. A* **1998**, *102*, 8673–8678.

(97) Shoeib, T.; Ruggiero, G. D.; Siu, K. W. M.; Hopkinson, A. C.; Williams, I. H. A hybrid quantum mechanical molecular mechanical method: Application to hydration free energy calculations. *J. Chem. Phys.* **2002**, *117*, 2762–2770.

(98) Takahashi, H.; Kawashima, Y.; Nitta, T.; Matubayasi, N. A novel quantum mechanical/molecular mechanical approach to the free energy calculation for isomerization of glycine in aqueous solution. *J. Chem. Phys.* **2005**, 123.

(99) Leung, K.; Rempe, S. B. Ab initio molecular dynamics study of glycine intramolecular proton transfer in water. *J. Chem. Phys.* **2005**, 122.

(100) Wada, G.; Tamura, E.; Okina, M.; Nakamura, M. On the ratio of zwitterion form to uncharged form of glycine at equilibrium in various aqueous-media. *Bull. Chem. Soc. Jpn.* **1982**, *55*, 3064–3067.

CT800116E

JCTC Journal of Chemical Theory and Computation

Hybrid Meta-Generalized Gradient Functional Modeling of Boron−Nitrogen Coordinate Covalent Bonds

Joshua A. Plumley and Jeffrey D. Evanseck*

Center for Computational Sciences and the Department of Chemistry and Biochemistry, Duquesne University, 600 Forbes Avenue, Pittsburgh, Pennsylvania 15282-1530

Received May 5, 2008

Abstract: Truhlar's new generation of hybrid meta-generalized gradient functionals has been evaluated in modeling the binding enthalpies of substituted B−N coordinate covalent bonds. The short-range exchange correlation (XC) energy of coordinate covalent bonding coupled with the medium-range XC energy of noncovalent interactions results in a particularly difficult case for density functional theory (DFT). In this study, M06, M06−2X, M05, M05−2X, MPWB1K, and MPW1B95 with the 6−311++G(3df,2p) basis set have been used to evaluate four methylated ammonia trimethylboranes, $(CH_3)_3B-N(CH_3)_nH_{3-n}$ ($n = 0$ to 3), along with $H_3B\text{-}NH_3$. The predicted binding enthalpies from the new functionals have been compared to experiment as well as previous DFT (B3LYP, MPW1K) and ab initio (HF, MP2, QCISD, and QCISD(T)) results. Previously, only MP2, QCISD, and QCISD(T) were found to model the experimental energetic trend accurately. The mean absolute deviation (MAD) from experimental binding enthalpies for M06−2X and M05−2X is 0.3 and 1.6 kcal/mol, respectively. M06−2X yields a lower MAD than more expensive ab initio methods (MP2 = 1.9 kcal/mol and QCISD = 2.3 kcal/mol) and a comparable MAD to QCISD(T) (MAD = 0.4 kcal/mol). M06−2X is shown to provide a balanced account of the short- and medium-range XC energies necessary to describe the binding enthalpy of coordinate covalent bonds accurately in sterically congested molecular systems.

Introduction

Kohn−Sham density functional theory (DFT)[1] has become a well-established tool in the computational chemistry community. Despite broad applicability across many areas of chemistry, DFT computations have resulted in a number of limitations involving unreliable barrier heights, noncovalent interactions, and aspects of transition metal chemistry, as recently reviewed by Zhao and Truhlar[2] and Sousa et al.[3] It is known that DFT accuracy issues originate from the approximation of the exchange-correlation (XC) functional within the Kohn−Sham formalism.[1,3] Current DFT implementations, such as the local spin density approximation (LSDA),[1,4] generalized gradient approximation (GGA),[5−10] meta-GGA (M-GGA),[11,12] hybrid-GGA (H-GGA),[5,9,11,13,14] and hybrid meta-GGA (HM-GGA),[15−17] treat the XC functional differently. LSDA assumes that the XC energy at any point in space depends only on the spin density at that specific spatial region. GGA improves upon LSDA by considering the gradient of the density along with the spin density. Further improvements of GGA include the spin-dependent electronic kinetic energy (M-GGA) and a certain percentage of Hartree−Fock exchange (H-GGA). Finally, the HM-GGA incorporates both approximations from M-GGA and H-GGA.

A known case reflecting a challenge for DFT computations[18] is presented by the experimental binding enthalpies of four methylated ammonia borane trimethylboranes, $(CH_3)_3B-N(CH_3)_nH_{3-n}$ ($n = 0$ to 3) and $H_3B\text{-}NH_3$.[19] The B−N coordinate covalent bond strength has been observed by experiment to increase for each methyl group added to the nitrogen atom ($n = 0$ to 2) until the third methyl group ($n = 3$), in which the strength decreases to that of the one

* Corresponding author e-mail:evanseck@duq.edu.

10.1021/ct800210e CCC: $40.75
Published on Web 07/19/2008

methyl case.[19] The experimental uncertainty was determined to be less than or equal to 0.3 kcal/mol.[19] Although an experimental binding enthalpy for ammonia borane is not known, estimates from experimental data[20,21] and ab initio methods[18,22,23] report that the ammonia borane B−N coordinate covalent bond is stronger than the methyl substituted ammonia trimethylboranes considered here. The relative experimental binding enthalpies across the five adducts is difficult to model by computational methods accurately, due to the critical mix of short- and medium-range XC energy necessary to describe the coordinate covalent bond and the nonbond interactions with increasing alkyl substitution.

Recently, we have reported the binding enthalpies for $(CH_3)_3B-N(CH_3)_nH_{3-n}$ ($n = 0$ to 3) and H_3B-NH_3 employing HF, B3LYP, MPW1K, MP2, QCISD, and QCISD(T) with the 6−311++G(3df,2p) basis set.[18] Only MP2, QCISD, and QCISD(T) were found to model the experimental trend in binding enthalpies accurately. Others have also reported deficiencies with commonly employed functionals when modeling coordinate covalent bonds.[22,23] For example, B3LYP binding energies and structures resulted with increasing error, as methyl groups were systematically added to the donor and acceptor atoms of ammonia borane.[22] Similarly, B3LYP computational results were unable to predict the energetics of the B−N coordinate covalent bond within borate systems.[23] However, MP2 predicted trends in binding enthalpies are consistent with experiment.[22] In a systematic analysis of ammonia borane and $(CH_3)_3B-N(CH_3)_nH_{3-n}$, quantitative accuracy (MAD = 0.4 kcal/mol) of binding enthalpies was determined with QCISD(T)/6−311++G(3df,2p) single point energy evaluations on MP2/6−311++G(3df,2p) optimized structures. (Reference 18 concentrated on only the four alkyl substituted adducts. The MAD computed in that paper omitted ammonia borane, and resulted in a QCISD(T) MAD of 0.5 and MP2 MAD of 2.2 kcal/mol. Statistical analysis of binding enthalpies including ammonia borane is given in the present work.) The electronic energy was corrected for BSSE, basis set limit, and appropriate thermal factors.[18] However, this strategy is resource intensive and limited to small systems. MP2/6−311++G(3df,2p) delivers semiquantitative results (MAD = 1.9 kcal/mol), maintaining the qualitative trend, when corrected as given above. Although MP2 is not as resource intensive as QCISD(T), MP2 remains an impractical and inefficient choice for large chemical systems. It would be beneficial to identify improved DFT functionals, so larger chemical assemblies possessing coordinate covalent chemical bonding may be investigated with structural and thermodynamic accuracy.

Truhlar and co-workers have developed a new generation of DFT methods that have shown significant accuracy improvements when considering thermochemistry, kinetics, noncovalent interactions, excited states, and treatment of transition metals.[2] Of the many functionals parameterized and designed by Truhlar and co-workers, the present work considers the HM-GGA functionals, M06−2X, M06, M05−2X, M05, MPW1B95, and MPWB1K. Before the existence of the Minnesota functionals (M06−2X, M06, M05−2X, and M05), MPWB1K and MPW1B95 were among the best for analyzing noncovalent interactions dominated by medium-range XC energy.[16,24] M05−2X and M05 were initially recommended when a combination of nonmetallic thermochemistry, kinetics, and noncovalent interactions were investigated,[15,25] while M05 was suggested for exploring organometallic and inorganometallic thermo chemistry.[15,25] Subsequently, M06−2X and M06 have been recommended for a wide range of chemical phenomena.[17] A recent review by Truhlar assessed the accuracy of many functionals against 496 data values within 32 databases including thermochemistry, barrier heights, noncovalent interactions, electronic spectroscopy, and structural data.[2] Finally, M06−2X, M05−2X, and M06 were suggested for systems where main-group thermochemistry, kinetics, and noncovalent interactions are important.[2] M06 was recommended specifically for transition metal thermochemistry involving both reactive organic and transition metal bonds.[2]

Proper description of short-range XC energy of coordinate covalent bonding coupled with the medium-range XC energy of noncovalent substituent interactions is important in many fields of chemistry. Since the Lewis acid adducts explored within the present work possess short- and medium-range XC, the M06, M06−2X, M05, M05−2X, MPW1B95, and MPWB1K functionals have been employed to explore the thermochemistry and electronics within the B−N coordinate covalent bond and alkyl substituted cases. The $(CH_3)_3B-N(CH_3)_nH_{3-n}$ complexes and ammonia borane have been chosen for analysis due to the reported difficulty in modeling the energetic character of the corresponding B−N coordinate covalent bonds, where only computationally intense ab initio methods (MP2, QCISD, and QCISD(T)) have been able to model the experimental trend of binding enthalpies accurately.[18] Besides interest in the fundamental treatment of coordinate covalent bonds,[20,26] the new generation of density functionals may provide improved structural and energetic description of coordinate covalent bonds that occur in larger biochemical systems[27,28] and new materials.[29,30]

Computational Details

All electronic structure calculations were carried out with G03[31] and NWChem 5.1[32] using the computational resources at the Center for Computational Sciences at Duquesne University. NWChem 5.1 was utilized for M06 and M06−2X computations. Full electronic structure optimizations on H_3B-NH_3 and $(CH_3)_3B-N(CH_3)_nH_{3-n}$ ($n = 0$ to 3) were performed with HM-GGA DFT, specifically MPW1B95,[16] MPWB1K,[16] M05−2X,[15] M05,[15,25] M06−2X,[17] and M06.[17] The design and parameterization of these functionals have been discussed in great detail within the corresponding references and recently reviewed.[2] The computational protocol within the present analysis is the same as reported previously,[18] except for the implementation of NWChem and the consideration of Radom and co-worker's updated scaling factors. Specifically, it is known that Gaussian03 and NWChem 5.1 utilize 5 *d* and 7 *d* (pure) and 6 *d* and 10 *f* (Cartesian) angular momentum basis functions as the default, respectively. The SPHERICAL keyword has been implemented within NWChem to utilize 5 *d* and 7 *f* functions. In addition, Radom and co-workers recently reported new scaling factors for B3LYP/

Table 1. Predicted and Experimental Binding Enthalpies, ΔH_T (kcal/mol)[a]

	BH_3-NH_3	$B(CH_3)_3$-NH_3	$B(CH_3)_3$-NH_2CH_3	$B(CH_3)_3$-$NH(CH_3)_2$	$B(CH_3)_3$-$N(CH_3)_3$	MAD
HF[18]	−16.5	−2.0	−3.5	−1.2	3.6	15.2
M05	−23.2	−6.8	−9.0	−7.3	−2.4	9.4
B3LYP[18]	−24.7	−6.8	−8.8	−7.0	−2.4	9.2
MPW1K[18]	−29.2	−11.9	−14.2	−12.9	−9.1	4.4
MPW1B95	−30.0	−12.5	−15.4	−14.9	−12.5	3.1
QCISD/MP2[18]	**−26.1**	**−12.6**	−13.8	**−16.8**	**−14.7**	**2.3**
M06	**−25.3**	**−10.3**	**−14.0**	**−14.5**	**−12.5**	**3.9**
MPWB1K	−30.3	−13.4	−16.4	−16.1	−13.9	2.3
MP2[18]	**−28.0**	**−15.6**	**−20.1**	**−21.4**	**−20.2**	**1.9**
M05−2X	**−25.9**	**−12.2**	**−16.3**	**−17.3**	**−16.1**	**1.6**
QCISD(T)/MP2[18]	**−27.5**	**−14.4**	**−18.6**	**−19.5**	**−17.8**	**0.4**
M06−2X	**−27.8**	**−14.2**	**−18.4**	**−19.2**	**−17.8**	**0.3**
experiment[19]	-27.5 ± 0.5[b]	-13.8 ± 0.3	-17.6 ± 0.2	-19.3 ± 0.3	-17.6 ± 0.2	

[a] BSSE corrections and convergence corrections of 1.5, 1.3, and 1.6 kcal/mol for MP2, QCISD, and QCISD(T), respectively, have been applied according to ref 18. The temperatures are 298 for BH_3-NH_3 and 373 for $(CH_3)_3B-N(CH_3)_nH_{3-n}$; $n = 0$ to 3. The 6−311++G(3df,2p) basis set has been employed with all chemical methods. Chemical methods in boldface indicate that the experimental trend was reproduced. [b] Best estimate predicted by QCISD(T)/aug-cc-pV5Z//MP2/6−311++G(3df,2p), see ref 18.

Table 2. Basis Set Superposition Error (kcal/mol)[a]

	BH_3-NH_3	$B(CH_3)_3$-NH_3	$B(CH_3)_3$-NH_2CH_3	$B(CH_3)_3$-$NH(CH_3)_2$	$B(CH_3)_3$-$N(CH_3)_3$	average
HF[18]	0.2	0.3	0.3	0.3	0.4	0.3
M05	0.2	0.4	0.3	0.4	0.4	0.4
B3LYP[18]	0.2	0.3	0.3	0.4	0.4	0.3
MPW1K[18]	0.2	0.3	0.4	0.4	0.5	0.4
M06	0.3	0.4	0.5	0.6	0.7	0.5
MPW1B95	0.2	0.3	0.3	0.4	0.5	0.3
MPWB1K	0.2	0.3	0.3	0.3	0.4	0.3
M05−2X	0.2	0.3	0.4	0.5	0.6	0.4
M06−2X	0.2	0.3	0.5	0.5	0.7	0.4

[a] BSSE computed by subtracting the ΔH BSSE corrected value from the ΔH uncorrected value. The 6−311++G(3df,2p) basis set has been utilized.

6−31G(d) predicted enthalpy corrections.[33] However, the new enthalpy corrections corresponding to $T = 298$ and 373 K are 1.0004 and 0.9953 (interpolated), yielding negligible differences between predicted binding enthalpies using the newer and older scaling factors. Consequently, the older scaling factors have been used for consistency and comparison with previous HF, B3LYP, MPW1K, MP2, QCISD, and QCISD(T) results.[18]

Results and Discussion

Binding Enthalpies of B−N Coordinate Covalent Bonds. The strength of the coordinate covalent bond has been assessed by predicting the binding enthalpy of each adduct utilizing M06, M06−2X, M05−2X, M05, MPWB1K, and MPW1B95. The results are compared to experimental binding enthalpies reported by Brown and co-workers[19] and previous theoretical values.[18] Previous analysis of HF, B3LYP, MPW1K, MP2, QCISD, and QCISD(T) determined that only ab initio methods were able to replicate the experimental binding enthalpy trend, as shown by Table 1.[18] M06−2X, M06, and M05−2X are the only DFT methods found to replicate the B−N coordinate covalent bond strengthening upon methyl substitution on the donor side ($n = 0$ to 2) with subsequent weakening as the last methyl group is added ($n = 3$). All methods predict a stronger ammonia borane B−N coordinate covalent bond as compared to the trimethylboranes.

M06 yields the highest MAD (3.9 kcal/mol) out of the functionals able to reproduce the experimental trend and consistently underestimates the strength of the B−N coordinate covalent bond. M05−2X yields a comparable MAD to MP2 at a much lower computational cost. Formally, MP2 scales as N^5, while DFT scales as N^3 (N = number of basis functions).[3] M05−2X and MP2 MADs are 1.6 and 1.9 kcal/mol, respectively; however, MP2 tends to overestimate the strength of the coordinate covalent bond, while M05−2X underestimates it. M06−2X is comparable to QCISD(T), yielding a MAD of 0.3 kcal/mol, 0.1 kcal/mol lower than that of QCISD(T)'s MAD. Thus, M06−2X is a practical and efficient functional suitable for modeling the energetic character of the B−N coordinate covalent bond. In contrast of resources, QCISD(T) scales as N^7 compared to DFT methods which grow as N^3.[3] Furthermore, M06−2X does not require a convergence correction for the basis set limit, a requirement for QCISD(T) to realign predicted results with experiment, due to the slow convergence of predicted post SCF results on larger molecular systems.[18]

Basis Set Superposition Error for DFT. As reported for ammonia borane and methyl substituted ammonia boranes, post SCF methods result in more BSSE than HF and DFT methods, and convergence to the complete basis set limit is much slower.[18] When post SCF methods are corrected for BSSE, the complete basis set limit was not reached with the 6−311++G(3df,2p) basis set. For practicality, residual convergence corrections were added to yield the appropriate binding enthalpies determined using the 6−311++G(3df,2p) basis set. However, HF and DFT methods possessed less

than 0.5 kcal/mol of BSSE, and binding enthalpies were considered converged.

The amount of BSSE for all functionals and HF is displayed in Table 2. M06, M06−2X, M05, M05−2X, MPWB1K, and MPW1B95 have comparable amounts of BSSE when employed with the 6−311++G(3df,2p) basis set to that of the other functionals and HF.

Since HF, B3LYP, and MPW1K yield a converged binding enthalpy at this basis set, it is reasonable to assume that the other functionals, M06, M06−2X, M05, M05−2X, MPWB1K, and MPW1B95, behave similarly. A M06−2X/aug-cc-pVQZ optimization of ammonia borane supports this assumption, where the B−N coordinate covalent bond lengths and BSSE corrected binding enthalpies differ by 0.0008 Å and 0.2 kcal/mol, respectively, as compared to that computed by M06-2X/6-311++G(3df,2p). Consequently, M06−2X is used with the 6−311++G(3df,2p) basis set to deliver a value near convergence with a minimum amount of BSSE without jeopardizing efficiency and practicality.

M06−2X Origin of Success. M06−2X, M06, and M05−2X are the only DFT methods found in this study that reproduce the experimental trend in binding enthalpies. Furthermore, only M06−2X and M05−2X yield comparable MADs to QCISD(T) and MP2, respectively. The Minnesota functionals were designed with a balance of kinetic energy density between the exchange and correlation functionals. In contrast, MPWB1K and MPW1B95 incorporate kinetic energy density within the correlation functional only, while B3LYP and MPW1K consider it in neither the exchange nor the correlation functionals. However, this cannot be the only reason for the relative quantitative success of M06−2X and M05−2X, since M05 also incorporates kinetic energy density in both the exchange and correlation functionals and is unable to model the experimental trend. The two main differences between the successful functionals (M06−2X and M05−2X) and M05 involves nearly double the amount of Hartree−Fock exchange (M06−2X and M05−2X) and a different formulation of the functional form (M06−2X). The linear combination of the M05 class functional and VSXC[12,34] (yielding the M06 style) allows both M06 and M06−2X to model the experimental binding enthalpy *trend* accurately; however, doubling the Hartree−Fock exchange allows a more *quantitative* assessment of the B−N coordinate covalent bond by M06−2X compared to M06. The empirical nature of the functional forms associated with M06−2X, M06, M05−2X, and M05, possessing 33, 36, 23, and 23 optimized parameters, respectively, may contribute to the high accuracy as well. However, M05 possesses the same number of optimized parameters as M05−2X and is unable to account for the experimental trend. In contrast, M05−2X is able to reproduce the experimental trend by just doubling the amount of Hartree−Fock exchange but with a higher MAD compared to M06−2X. Consequently, the balance of kinetic energy density in both the exchange and correlation functionals, the higher amount of Hartree−Fock exchange, and the functional forms of the exchange and correlation functionals within the M06 class of functionals are critical in order to model a balance of short- and medium-range XC interactions that exist within the B−N coordinate covalent bond and steric interactions, respectively.

Conclusions

The inherent difficulty of modeling the B−N coordinate covalent bond strength within ammonia borane and four methyl substituted trimethylboranes using Truhlar's hybrid meta-generalized gradient functionals (M06−2X, M06, M05−2X, M05, MPWB1K, MPW1B95) has been investigated. Previously, only ab initio methods (MP2, QCISD, and QCISD(T)) have been able to model the experimental trend accurately, with MP2 and QCISD(T) yielding lower MADs than QCISD. Of the new functionals explored, M06−2X, M06, and M05−2X are able to reproduce the experimental trend with MADs of 0.3, 3.9, and 1.6 kcal/mol, respectively. M05−2X and M06−2X yield comparable MADs to that of MP2 (MAD = 1.9 kcal/mol) and QCISD(T) (MAD = 0.4 kcal/mol), respectively. M06−2X is able to model the B−N coordinate covalent bond efficiently and accurately without the residual convergence correction and intense computational resources necessary for MP2, QCISD, and QCISD(T). M06−2X incorporates the proper balance of short-range XC energy necessary to model the electronics within the B−N coordinate covalent chemical bond as well as the medium-range XC energy required by sterically congested environments near the coordinate covalent bond. M06−2X computations result in superior performance, due to the balance of kinetic energy density in the exchange and correlation functionals, the higher amount of Hartree−Fock exchange, and the functional forms of the exchange and correlation functionals. M06−2X/6−311++G(3df,2p) is able to model the electronics of the B−N chemical bond and is a more practical and efficient choice for investigating chemical systems possessing B−N coordinate covalent bonds, where quantitative trends in energetics are necessary.

Acknowledgment. This work was funded in part by the NSF and DoD (CHE-0723109, CHE-0649182, CHE-0321147, CHE-0354052, AAB/PSC CHE-030008P), Department of Education (P116Z040100 and P116Z050331), and the Silicon Graphics Inc., James River Technical, Inc., and Gaussian Corporations. The authors acknowledge Dr. Doug Fox (Gaussian) for his expertise, knowledge, and discussions crucial to this study.

Supporting Information Available: Full Gaussian and NWChem citation, all optimized structures, frequencies, thermodynamic data, interpolation of scaling factor at 373 K, and BSSE uncorrected binding enthalpies. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Sham, J.; Kohn, W. *Phys. Rev.* **1965**, *140*, 1133–1138.

(2) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157–167.

(3) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439–10452.

(4) Painter, G. S. *J. Phys. Chem.* **1986**, *90*, 5530–5535.

(5) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098–3100.

(6) Perdew, J. P.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800–8802.

(7) Becke, A. D. *J. Chem. Phys.* **1986**, *84*, 4524–4529.

(8) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(9) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, *37*, 785–789.

(10) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053–1062.

(11) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040–1046.

(12) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400–410.

(13) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715–2719.

(14) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811–4815.

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

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

(17) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(18) Plumley, J. A.; Evanseck, J. D. *J. Phys. Chem. A* **2007**, *111*, 13472–13483.

(19) Brown, H. C.; Bartholomay, H., Jr.; Taylor, M. D. *J. Am. Chem. Soc.* **1944**, *66*, 435–442.

(20) Haaland, A. *Angew. Chem.* **1989**, *101*, 1017–1032.

(21) Gurvich, L. V.; Veyts, I. V.; Alcock, C. B. *Thermodynamic Properties of Individual Substances*, 4th ed.; CRC Press: Boca Raton, 1994; Vol. 3, pp 106−107.

(22) Gilbert, T. M. *J. Phys. Chem. A* **2004**, *108*, 2550–2554.

(23) LeTourneau, H. A.; Birsch, R. E.; Korbeck, G.; Radkiewicz-Poutsma, J. L. *J. Phys. Chem. A* **2005**, *109*, 12014–12019.

(24) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415–432.

(25) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103/161101–161103/161104.

(26) Gutmann, V. *The Donor-Acceptor Approach to Molecular Interactions*, 1st ed.; Plenum Press: New York, 1978; p 279.

(27) Burnham, B. S. *Curr. Med. Chem* **2005**, *12*, 1995–2010.

(28) Li, P.; Sergueeva, Z. A.; Dobrikov, M.; Shaw, B. R. *Chem. Rev.* **2007**, *107*, 4746–4796.

(29) Haubner, R.; Wilhelm, M.; Weissenbacher, R.; Lux, B. *Struct. Bonding (Berlin)* **2002**, *102*, 1–45.

(30) Gu, Y.; Zheng, M.; Liu, Y.; Xu, Z. *J. Am. Ceram. Soc.* **2007**, *90*, 1589–1591.

(31) Frisch, M. J., et. al. *Gaussian 03, Revision E.01*; Gaussian, Inc.: Wallington CT, 2004.

(32) Bylaska, E. J., et. al. NWChem *A Computational Chemistry Package for Parallel Computers, 5.1*; Pacific Northwest National Laboratory: Richland, WA, 2007.

(33) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.

(34) Van Voorhis, T.; Scuseria, G. E. *Mol. Phys.* **1997**, *92*, 601–608.

CT800210E

JCTC Journal of Chemical Theory and Computation

Assessment of a Middle-Range Hybrid Functional

Thomas M. Henderson,† Artur F. Izmaylov,† Gustavo E. Scuseria,*,† and Andreas Savin‡

Department of Chemistry, Rice University, 6100 Main Street, Houston, Texas 77005-1892, and Laboratoire de Chimie Théorique, CNRS, Université Pierre et Marie Curie, 4 Place Jussieu, F-75252 Paris, France

Received May 6, 2008

Abstract: While hybrid functionals are largely responsible for the utility of modern Kohn−Sham density functional theory, they are not without their weaknesses. In the solid state, the slow decay of their nonlocal Hartree−Fock-type exchange makes hybrids computationally demanding and can introduce unphysical effects. Both problems can be remedied by a screened hybrid which uses exact exchange only at short-range. Many molecular properties, in contrast, benefit from the inclusion of long-range exact exchange. Recently, the authors reconciled these two seemingly contradictory requirements by introducing the HISS functional [*J. Chem. Phys.* 2007, *127*, 221103], which uses exact exchange only in the *middle* range. In this paper, we expand upon our previous work, benchmarking the performance of the HISS functional for several simple properties and applying it to the dissociation of homonuclear diatomic cations and to the polarizability of linear H_2 chains to determine the importance of middle-range exact exchange for these systems, which are expected to be sensitive to the asymptotic exchange potential.

1. Introduction

Over the past several years, the Kohn−Sham (KS) construction in density functional theory[1,2] (DFT) has become the dominant technique for predictions of the electronic structure of molecules and solids.[3] This success is due to the method's combination of reasonable accuracy and low computational cost. The accuracy of a given KS calculation is largely controlled by the choice of the exchange-correlation functional, $E_{xc}[n]$, which is known to be a functional of the density but whose precise form is unknown except in certain limiting cases.

The simplest class of exchange-correlation functionals are termed semilocal because they approximate the exchange-correlation energy density at a point **r** as a function of the density at **r** and, possibly, its derivatives (which serve to incorporate information also in the neighborhood of the point **r**). Typical examples include the local density approximation (LDA), generalized gradient approximations (GGAs) such as the GGA of Perdew, Burke, and Ernzerhof (PBE),[4] and meta-GGAs such as the functional of Tao, Perdew, Staroverov, and Scuseria (TPSS).[5] Semilocal functionals are often derived from first principles without resort to empirical parameters, but unfortunately, they perform relatively poorly for many properties of interest. Therefore, it is common to use hybrid functionals, which mix a fraction of exact nonlocal Hartree−Fock-type exchange with the semilocal exchange provided by conventional functionals.[6−10] Hybrid functionals tend to be significantly more accurate than their parent semilocal functionals, albeit at the cost of adding an extra parameter (the fraction of Hartree−Fock-type exchange) which is usually determined empirically.

But hybrids, also, suffer from several deficiencies. In extended systems, nonlocal Hartree−Fock-type exchange is expensive to compute since the lattice summations required converge very slowly in semiconductors and metals due to the slow spatial decay of the one-particle density matrix.[11] On the other hand, in finite systems the exchange potential of a semilocal functional decays too rapidly, leading to errors in properties which sample density tails, and hybrids only partially remedy this defect.

Both of these problems can be alleviated by the use of range-separated hybrids,[12−15] which split the Coulomb

* Corresponding author. E-mail: guscus@rice.edu.
† Rice University.
‡ Université Pierre et Marie Curie.

10.1021/ct800149y CCC: $40.75
Published on Web 08/12/2008

operator into short-range (SR) and long-range (LR) pieces, typically as

$$\frac{1}{r} = \underbrace{\frac{\mathrm{erfc}(\omega r)}{r}}_{\mathrm{SR}} + \underbrace{\frac{\mathrm{erf}(\omega r)}{r}}_{\mathrm{LR}}, \quad (1)$$

and then mix in different fractions of nonlocal exchange for the different regions. The screened hybrid functional of Heyd, Scuseria, and Ernzerhof,[16,18] (HSE) uses a fraction of exact exchange in the short-range but none in the long-range; in doing so, it improves the convergence of the lattice sums required to evaluate the nonlocal exchange potential, thereby improving computational efficiency for solids and large molecules.[19] The long-range-corrected LC-ωPBE[20] hybrid of Vydrov and Scuseria uses no exact exchange at short-range, but treats the long-range exclusively with Hartree−Fock-type exchange, thus yielding the right answer in density tails and improving accuracy for a host of properties.

Recently, we introduced the HISS functional[21] which uses exact exchange in the *middle* range (MR) only, with the intent of providing a functional with the computational advantages of HSE but with much of the accuracy of LC-ωPBE. Preliminary results suggest that the functional referred to as HISS-B in ref 21 and called simply HISS here provides a sort of "best of both worlds" approach in that it yields uniformly accurate results for total atomic energies, heats of formation, reaction barrier heights, and band gaps of solids. In contrast, while both HSE and LC-ωPBE perform well for atomic energies and for heats of formation, HSE performs poorly for reaction barriers and LC-ωPBE performs quite badly for band gaps.

The present work aims to systematically investigate the performance of the HISS functional for a wide variety of properties, providing benchmark results so that we can further assess what the functional does right and why it does so. In section 2, we discuss the HISS functional, after which we elaborate on its performance for several properties in section 3. We conclude with section 4.

2. HISS Functional

The HISS functional is a multi-range hybrid which partitions the Coulomb operator into three pieces rather than the two given in eq 1. This is accomplished by writing

$$\frac{1}{r} = \underbrace{\frac{\mathrm{erfc}(\omega_{\mathrm{SR}} r)}{r}}_{\mathrm{SR}} + \underbrace{\frac{\mathrm{erf}(\omega_{\mathrm{LR}} r)}{r}}_{\mathrm{LR}} + \underbrace{\frac{\mathrm{erfc}(\omega_{\mathrm{LR}} r) - \mathrm{erfc}(\omega_{\mathrm{SR}} r)}{r}}_{\mathrm{MR}} \quad (2)$$

where ω_{SR} and ω_{LR} are the two parameters defining the ranges and here take numerical values of $0.84{a_0}^{-1}$ and $0.20{a_0}^{-1}$, respectively. Given this partition, the exchange-correlation energy within HISS is written as

$$E_{\mathrm{xc}}^{\mathrm{HISS}} = E_{\mathrm{x}}^{\mathrm{SR\text{-}PBE}} + \frac{2}{5}E_{\mathrm{x}}^{\mathrm{MR\text{-}PBE}} + \frac{3}{5}E_{\mathrm{x}}^{\mathrm{MR\text{-}HF}} + E_{\mathrm{x}}^{\mathrm{LR\text{-}PBE}} + E_{\mathrm{c}}^{\mathrm{PBE}} \quad (3)$$

where HF refers to Hartree−Fock-type exchange, PBE for exchange refers to the exchange functional built from the PBE model exchange hole,[22] and PBE for correlation refers to the original functional of Perdew, Burke, and Ernzerhof. The various parameters (ω_{SR}, ω_{LR}, and the fraction of middle-range exact exchange) were obtained by fitting to the (small) AE6 set of atomization energies[23] and the BH6 set of barrier heights[23] while insisting that performance for band gaps in solids be reasonable. Note that, as with most hybrids, we use a generalized Kohn−Sham scheme[24] in which we employ the nonlocal exchange potential.

Table 1. Mean Errors and Mean Absolute Errors per Electron (mH) in Total Atomic Energies for the First Eighteen Atoms from a Variety of Functionals Based on the PBE GGA[a]

functional	ME	MAE	max +	max −
HISS	4.50	4.84	7.69 (Be)	−3.11 (H)
HSE	6.06	6.21	8.07 (Ar)	−1.33 (H)
LC-ωPBE	4.37	5.05	7.45 (Ar)	−6.09 (H)
PBE	8.55	8.55	11.60 (Ar)	
PBEh	6.98	7.10	9.15 (Ar)	−1.15 (H)

[a] Also shown are maximum positive and negative deviations and the atoms to which they correspond.

3. Benchmarks

With the HISS functional defined, we are ready to begin its assessment. Our intent in this work is to more exhaustively examine the effects of middle-range exact exchange than was done in ref 21.

Throughout, we will compare with other functionals in the "PBE family", including PBE itself, the PBEh global hybrid with 25% exact exchange, as well as the screened HSE06 hybrid and the long-range-corrected LC-ωPBE hybrid. All calculations are performed self-consistently using a development version of the GAUSSIAN program.[25] Except where mentioned otherwise, we use the 6-311++G(3df,3pd) basis set.

3.1. Previous Results. For the sake of completeness, we begin by expanding upon the results of ref 21, which studied atomic total energies, thermochemistry, reaction barrier heights, and band gaps in solids. We will discuss each of these in turn.

3.1.1. Atomic Energies. One important requirement of any successful functional is the ability to properly describe atomic total energies. Atomic energies are particularly important in describing heats of formation, atomization energies, cohesive energies, and so on.[26] To test our results for total atomic energies, we consider the first eighteen atoms (H−Ar). In Table 1, we show mean error (ME) and mean absolute error (MAE) per electron for a variety of functionals; reference data is taken from the work of Chakravorty et al.[27] Errors are defined as the result from the functional minus the reference result; positive errors thus indicate that the functional gives a result too low in absolute value. We clearly see that HISS, while not parametrized to atomic energies, nevertheless reproduces them very well, with accuracy comparable to that of LC-ωPBE and significantly better than that of PBE or PBEh. These results may partially explain the success of HISS for heats of formation, which we address next.

3.1.2. Heats of Formation. Performance for equilibrium thermochemistry has long been one of the main requirements

Table 2. Mean Errors and Mean Absolute Errors (kcal/mol) in the G2 and G3 Sets of Heats of Formation from a Variety of Functionals Based on the PBE GGA[a]

	G2				G3			
functional	ME	MAE	max +	max −	ME	MAE	max +	max −
HISS	3.12	4.39	26.1 (O_3)	−9.4 (C_2F_4)	2.46	4.34	26.1 (O_3)	−10.6 ($C_{10}H_8$)
HSE	−0.73	3.88	22.1 (SiF_4)	−15.3 (C_5H_5N)	−2.10	4.86	22.1 (SiF_4)	−28.0 ($C_{10}H_8$)
LC-ωPBE	−0.36	3.73	16.3 (P_2)	−20.6 (C_2F_4)	−0.93	4.25	16.3 (P_2)	−21.4 (C_2F_6)
PBE	−16.07	16.87	10.8 (Si_2H_6)	−50.5 (C_2F_4)	−21.69	22.22	10.8 (Si_2H_6)	−79.7 ($C_{10}H_8$)
PBEh	−2.42	4.87	21.3 (SiF_4)	−19.8 (C_5H_5N)	−4.72	6.66	21.3 (SiF_4)	−35.6 ($C_{10}H_8$)

[a] Also shown are maximum positive and negative deviations and the systems to which they correspond.

Table 3. Mean Errors and Mean Absolute Errors (kcal/mol) in the HTBH38 and NHTBH38 Sets of Reaction Barrier Heights from a Variety of Functionals Based on the PBE GGA

			non-hydrogen-transfer reactions of the NHTBH38 set							
	HTBH38 hydrogen transfer (38)		heavy-atom transfer (12)		nucleophilic substitution (16)		unimolecular and association (10)		full NHTBH38 (38)	
functional	ME	MAE	ME	MAE	ME	MAE	ME	MAE	ME	MAE
HISS	−1.3	1.7	−2.1	2.4	0.6	0.8	1.0	2.7	−0.2	1.8
HSE	−4.6	4.6	−7.4	7.4	−2.3	2.4	−0.9	2.2	−3.5	3.9
LC-ωPBE	−0.5	1.3	−0.6	1.9	2.8	2.8	1.4	2.3	1.4	2.4
PBE	−9.7	9.7	−15.3	15.3	−6.8	6.8	−3.1	3.5	−8.5	8.6
PBEh	−4.6	4.6	−7.0	7.0	−1.7	1.9	−0.8	2.3	−3.1	3.6

for methods in quantum chemistry. To assess this performance, we calculate heats of formation ($\Delta_f H_{298^\circ}$) in two standard benchmark test sets, namely the G2/97 set of 148 molecules,[28] and the G3/99 set of 223 molecules;[29] G3/99 is a superset of G2/97 which primarily adds larger organic molecules.

In accordance with refs 28 and 30, we use equilibrium geometries and zero-point energies tabulated at the B3LYP/6-31G(2df,p) level of theory, with a frequency scale factor of 0.9854. We calculate the total energies of the atoms and molecules in the test set with our chosen functional and basis set. Enthalpies of formation are then obtained by calculating atomization energies, adding tabulated thermal corrections and zero-point energies, and using the experimental enthalpies of formation of free atoms. Errors in $\Delta_f H_{298^\circ}$ thus reflect errors in atomization energies. While this procedure does not produce a truly consistent atomization energy, since it does not use the molecular geometry predicted by the HISS functional, this is standard practice in benchmarking calculations (so as to separate errors in geometry from errors in energetics). Here and in what follows, we define errors as theory minus experiment.

Table 2 shows that HISS is as accurate as the best functionals we consider (and in ref 20, it was shown that LC-ωPBE delivers thermochemical performance competitive with heavily parametrized functionals). In particular, HISS clearly performs better than PBEh, and much better than PBE. Both of the latter are markedly less accurate for the G3 set than they are for the G2 set, which implies size-dependent errors for these functionals which are reduced significantly in HSE, LC-ωPBE, and HISS. We should point out that while the other functionals all tend to underestimate enthalpies of formation (i.e., the mean error is negative, albeit only slightly so in the case of LC-ωPBE), HISS tends to overestimate them instead. Calculations on the AE6 set of atomization energies show that using the HISS geometry instead of that specified by the benchmarking set decreases the MAE by 0.15 kcal/mol.

3.1.3. Reaction Barrier Heights. Semilocal functionals generally underestimate reaction barriers, and often dramatically so. Transition states are not infrequently predicted to be lower in energy than the reactants or products. This error is connected to one-electron self-interaction error, which is particularly large for systems with stretched bonds which allow electrons to delocalize readily. Many semilocal functionals, in fact, deliver reasonable reaction barriers when the orbitals are taken from a self-interaction free method.[31]

A second, more consistent, route toward predicting reaction barrier heights with DFT is to use hybrid functionals with a large (40−60%) fraction of nonlocal exchange. But while these functionals can accurately predict barrier heights, they perform poorly for equilibrium thermochemistry. Relatively few functionals do well for both, and most of those are heavily parametrized. Long-range-corrected hybrids, however, are also accurate for both reaction barriers and equilibrium thermochemistry. Since HISS contains a significant fraction of exact exchange over the length scales relevant in transition states, one might expect HISS to be another functional that gives accurate barrier height as well as thermochemistry.

To assess performance for reaction barriers, we consider the HTBH38/04 set of forward and reverse barrier heights for 19 hydrogen transfer reactions,[32] and the NHTBH38/04 set of forward and reverse barrier heights for 19 non-hydrogen-transfer reactions.[33] Geometries of all species as well as the best estimates for the experimental barrier heights are taken from ref 33.

As can be seen from Table 3, only HISS and LC-ωPBE describe hydrogen transfer or heavy atoms transfer barriers with satisfactory accuracy, while for these reactions PBEh and HSE are inadequate and PBE is clearly unacceptable. For nucleophilic substitution reactions, PBEh performs quite

well and HSE is almost as good as LC-ωPBE but HISS gives excellent results. Note that LC-ωPBE uniformly overestimates the barriers to nucleophilic substitution, while PBE underestimates them; apparently, LC-ωPBE includes too much exact exchange for this type of reaction, and HISS, which generally uses less exact exchange than LC-ωPBE, happens to give the right answer. All of the functionals perform similarly for unimolecular and association reactions. To examine the effects of geometry on the reaction barriers, we optimize the geometries of the reactants, products, and transition states in the BH6 set using HISS, which decreases the MAE by 0.09 kcal/mol. Generally, LC-ωPBE outperforms HISS for reaction barriers, but HISS performs quite adequately for all categories and, in particular, yields better answers than does HSE [Note that there is an error in Table 1 of ref 21; the number given here for the mean absolute error in the HSE calculations of barrier heights in the NHTBH38 set is correct]; as we shall see in the following section, the opposite conclusions can be draw for band gaps.

3.1.4. Band Gaps in Solids. Finally, we consider a small set of band gap calculations to assess the performance of the HISS functional in solids, examining Si, C, SiC, BN, and BP. As seen from Table 4, both PBEh and LC-ωPBE consistently overestimate band gaps, and in the case of LC-ωPBE, disastrously so. While the magnitude of the overestimation is somewhat in question due to the limited data available, the fact that these functionals do generally overestimate the band gap is certainly true, especially given that Hartree−Fock theory does the same.

While both PBEh and LC-ωPBE perform poorly for band gaps, HISS is somewhat better behaved. It still overestimates band gaps consistently, indicating that the amount of exact exchange included is (on average) more than is required for reproducing experimental band gaps. In this respect, the HSE functional has close to the optimal amount of exact exchange. The existence and, more interestingly, approximate universality of this optimal amount of exact exchange between two limits (PBE and PBEh) has been rationalized recently by examining excitation energies from time-dependent density functional calculations with periodic boundary conditions.[34]

We have also calculated the band gap in MgO and NaCl, which we have excluded from the statistics above. For wide band gap insulators, the optical and fundamental gaps differ significantly. We have reasons to believe that band gaps obtained as band energy differences in band structure calculations with screened functionals tend to approximate optical instead of fundamental gaps.[37] Thus, we have excluded these systems from the statistics in Table 4, which is then limited to semiconductors. For the two ionic solids we have considered, HISS is in error by about 0.9 eV, again comparable to HSE (which underestimates the gaps by about 1 eV). Note, however, that in contrast to HSE, the HISS functional does not consistently underestimate the band gaps in these two systems, but rather underestimates the gap in NaCl and overestimates it in MgO. Interestingly, the PBEh global hybrid performs better than either HISS or HSE for these two systems, yielding band gaps in error by 0.6 eV.

Table 4. Mean Errors and Mean Absolute Errors (eV) in the Band Gaps for Si, C, SiC, BN, and BP from a Variety of Functionals Based on the PBE GGA[a]

functional	ME	MAE	max
HISS	0.38	0.38	0.63 (C)
HSE	−0.09	0.09	−0.28 (BP)
LC-ωPBE[b]	3.81	3.81	4.11 (CSi)
PBE	−1.10	1.10	−1.70 (BN)
PBEh	0.57	0.57	0.73 (Si)

[a] All basis sets and geometries are given in ref 35. For all functionals, the geometries have been optimized at the HSE/m-6311G** level of theory.[36] Also shown are maximum absolute deviation and the system to which it corresponds. [b] Excluding BN and C, calculations on which we were unable to converge.

Table 5. Mean Errors and Mean Absolute Errors in IPs and EAs (eV) and in PAs (kcal/mol) from a Variety of Functionals Based on the PBE GGA[a]

functional	ME	MAE	max +	max −
		IP (eV)		
HISS	0.006	0.188	2.01 (CN)	−0.49 (B_2F_4)
HSE	−0.068	0.201	1.60 (CN)	−0.68 (B_2F_4)
LC-ωPBE	0.073	0.192	1.97 (CN)	−0.46 (B_2F_4)
PBE	−0.105	0.235	1.11 (CN)	−1.01 (BF_3)
PBEh	−0.064	0.199	1.61 (CN)	−0.67 (B_2F_4)
		EA (eV)		
HISS	−0.047	0.211	1.34 (C_2)	−0.42 (CH_3)
HSE	−0.025	0.165	1.09 (C_2)	−0.37 (HO_2)
LC-ωPBE	0.019	0.177	1.41 (C_2)	−0.31 (CH_3)
PBE	0.061	0.118	0.78 (C_2)	−0.29 (NO_2)
PBEh	−0.027	0.165	1.09 (C_2)	−0.39 (HO_2)
		PA (kcal/mol)		
HISS	1.07	1.51	4.6 (C_2H_2)	−1.5 (SiH_4)
HSE	0.09	1.11	3.6 (C_2H_2)	−1.5 (SiH_4)
LC-ωPBE	0.86	1.42	4.7 (C_2H_2)	−1.8 (SiH_4)
PBE	−0.82	1.60	2.4 (C_2H_2)	−3.6 (PH_3)
PBEh	0.18	1.14	3.9 (C_2H_2)	−1.7 (SiH_4)

[a] Also shown are maximum positive and negative deviations and the systems to which they correspond.

3.2. Ionization Potentials, Electron Affinities, and Proton Affinities. Thus far, we have considered performance only for charge-neutral species. We can also consider positively or negatively charged systems, however, which play important roles in chemistry. Further, they can pose interesting challenges for density functionals, since electron attachment energies tend to sample the long-range part of the density which is poorly described by semilocal functionals.

To assess the quality of the HISS functional for charged species, we examine the G2 ion test set,[38] excluding N_2^+ and H_2S^+ which do not converge at the GGA level.[38] The set thus includes 86 ionization potentials (IPs) and 58 electron affinities (EAs). We also consider the eight proton affinities (PAs) in the G3/99 test set. Results are reported in Table 5 and are calculated as the difference between the self-consistent energies of the neutral and charged species (i.e., these are "ΔSCF" calculations).

In general, HISS performs quite well for IPs, yielding results essentially equivalent to those from HSE, PBEh, or LC-ωPBE, while having a very low mean signed error. For EAs, however, HISS underperforms significantly. Indeed, the performance for electron affinities is almost opposite what one might expect, since PBE is quite accurate while

Table 6. Mean Errors and Mean Absolute Errors in Equilibrium Bond Lengths (Å) from a Variety of Functionals Based on the PBE GGA[a]

functional	ME	MAE	max +	max −
HISS	−0.0100	0.0146	0.073 (Li_2)	−0.070 (F_2^+)
HSE	−0.0007	0.0090	0.057 (Li_2)	−0.050 (F_2^+)
LC-ωPBE	−0.0087	0.0139	0.044 (Li_2)	−0.055 (P_4)
PBE	0.0156	0.0160	0.055 (Li_2)	−0.009 (F_2^+)
PBEh	−0.0013	0.0091	0.055 (Li_2)	−0.052 (F_2^+)

[a] Also shown are maximum positive and negative deviations and the systems to which they correspond.

functionals with better long-range exchange potentials have much higher error. For proton affinities, HISS improves only slightly over PBE.

3.3. Bond Lengths and Vibrational Frequencies. Aside from energetic quantities, we wish to predict structural information and other properties as well. We test the performance of HISS for molecular geometries by considering a test set of equilibrium bond lengths (r_e) compiled in ref 39. Since Be_2 is bound primarily by van der Waals forces, we have excluded it, and the remaining 95 covalently bonded molecules form the T-95R set of bond lengths defined by Vydrov and Scuseria.[20] Most molecules are diatomic, though several polyatomic molecules with sufficiently high symmetry that the geometry is determined entirely by one bond length are also included. We point out that although for each molecule in this test set we need optimize only a single geometrical parameter, we use analytic energy gradients to do so, and we are not limited to predicting geometries of simple systems.

As seen in Table 6, HISS is comparable in accuracy to LC-ωPBE and slightly outperforms PBE, though without the systematic bias toward stretched bonds that the latter functional exhibits. Both PBEh and HSE are significantly more accurate.

Once equilibrium molecular geometries are obtained—thus providing some assessment of the accuracy of the first derivatives of the potential energy surface with respect to nuclear displacements—it is natural to calculate vibrational frequencies, thereby assessing the second derivatives as well. To this end, we consider the T-82F test set of 82 diatomic molecules,[40] whose experimental harmonic vibrational frequencies have been well studied; once again, we exclude Be_2, leaving us with 81 diatomics. Considering the excellent performance of PBEh and HSE for molecular geometries, we might expect them to be similarly accurate for vibrational frequencies, while HISS and LC-ωPBE might be expected to suffer somewhat from their inclusion of a large portion of exact exchange (since Hartree−Fock is known to strongly overestimate vibrational frequencies). As seen in Table 7, this is indeed the case. While PBE tends to underestimate vibrational frequencies, hybrid functionals do the opposite, and as the amount of exact exchange included gets larger, so too do errors in vibrational frequencies. This points toward inadequacies in the PBE correlation functional as the likely source of error, and we suggest that, with an improved correlation functional, HISS and LC-ωPBE are quite likely to deliver improved accuracy for these properties.

Table 7. Mean Errors and Mean Absolute Errors (cm^{-1}) for Harmonic Vibrational Frequencies from a Variety of Functionals Based on the PBE GGA[a]

functional	ME	MAE	max +	max −
HISS	72.3	74.6	328.2 (O_2^+)	−20.7 (AlF)
HSE	31.9	42.4	234.0 (O_2^+)	−39.1 (AlH)
LC-ωPBE	55.1	62.3	237.4 (F_2^+)	−87.1 (HF^+)
PBE	−33.1	41.5	75.9 (F_2^+)	−175.3 (HF^+)
PBEh	34.7	43.8	236.3 (O_2^+)	−36.2 (AlH)

[a] .Also shown are maximum positive and negative deviations and the systems to which they correspond.

Table 8. Mean Absolute Errors (eV) for 19 Valence and 20 Rydberg Excitation Energies for CO, N_2, and H_2CO Molecules from a Variety of Functionals Based on the PBE GGA[a]

functional	valence	Rydberg
HISS	0.45	0.18
HSE	0.40	0.79
LC-ωPBE	0.33	0.15
PBE	0.43	1.74
PBEh	0.39	0.85

[a] The augmented Sadlej pVTZ basis is used.[39]

3.4. Electronic Excitation Energies. While the first several benchmarks have focused on equilibrium properties of molecules in their ground state, these can be predicted accurately by many semilocal functionals or global hybrids as well, and for these properties, the complexity of range-separated hybrids are not essential. More interesting is performance for properties sensitive to the exchange potential, and especially for those properties sensitive to the long-range exchange potential for which many functionals perform rather poorly.

We begin by considering electronic excitation spectra of three small molecules: CO, N_2, and H_2CO. Excitation spectra of these systems have been studied before, and in Table 8, we investigate the same 19 valence and 20 Rydberg excitations that have been assessed in ref 42. All functionals give roughly similar performance for valence excitations, with errors on the order of 0.4 eV. This changes drastically for the Rydberg excitations, which are more sensitive to the long-range exchange-correlation potential and its derivative. Thus, it is interesting to note that although the exchange potential from HISS inherits the exponential decay of PBE rather than the $1/r$ decay of LC-ωPBE, HISS nevertheless performs quite well for these Rydberg excitations. We speculate that these Rydberg excitations, which represent only the lower-lying members of the Rydberg series, are not high enough in energy to sample the truly long-range part of the potential. As we go further up in the Rydberg series, however, note that deficiencies in the basis set generally dominate other considerations.

3.5. Rydberg Excitation in the Hydrogen Atom. Let us return to the issue of Rydberg states in greater detail. Rydberg states are characterized by their progression toward an ionization threshold, with the n^{th} Rydberg excitation energy given by

$$\omega_n = \mathrm{IP} - \frac{1}{2(n-\delta)^2} \tag{4}$$

Table 9. TDDFT Excitation Energies in the Hydrogen Atom from a Variety of Functionals Based on the PBE GGA Compared to the Exact Result[a]

n	exact	HF	LC-ωPBE	HISS	HSE	PBEh	PBE
2	10.2043	10.2043	9.1435	9.2510	8.5344	8.4411	7.5806
3	12.0940	12.0939	10.7855	9.4311	8.7252	8.9546	7.5976
4	12.7553	12.7553	11.3885	9.4394	8.7336	9.0478	7.6075
5	13.0615	13.0615	11.6742	9.4558	8.7500	9.0824	7.6263
6	13.2278	13.2278	11.8315	9.4847	8.7791	9.1001	7.6589
7	13.3280	13.3280	11.9273	9.5347	8.8295	9.1234	7.7144
8	13.3931	13.3931	11.9898	9.6206	8.9161	9.1753	7.8086
9	13.4377	13.4377	12.0330	9.7684	9.0657	9.2826	7.9691
10	13.4696	13.4700	12.0639	10.0236	9.3242	9.4917	8.2541
IP[b]	13.6057	13.6057	12.1972	9.4285	8.7226	9.1302	7.5944
IP[c]	13.6057	13.6057	13.7765	13.6953	13.6467	13.6417	13.6054

[a] Also included are Hartree−Fock and TDHF, errors in which are purely due to the basis set. The last two rows contain the ionization potential as calculated in two different ways. All results are in electronvolts, using a conversion factor of $1E_h = 27.2114$ eV. [b] Ionization potential obtained as $-\varepsilon_{HOMO}$. [c] Ionization potential obtained via ΔSCF between the neutral and the cation.

where δ is the quantum defect (and is zero in the hydrogen atom). We can thus, in principle, extract ionization potentials in at least three ways. First, we can directly evaluate the difference between the self-consistent energies of the charged and neutral species in a ΔSCF calculation. Second, we can use the fact that, in the exact theory, the ionization potential is given by the opposite of the highest occupied orbital energy:[1]

$$\mathrm{IP} = -\varepsilon_{\mathrm{HOMO}} \tag{5}$$

Third, we can evaluate the limit of the Rydberg excitation spectra. All three methods should in principle give the same result but in practice give different answers due to deficiencies in the functionals and basis. Additionally, the quality of the Rydberg excitation spectra is generally rendered suspect for semilocal functionals since most calculations neglect both the current-dependence and the frequency-dependence of the exchange-correlation kernel.

To examine the relation between the IP as calculated in these approaches, we wish to consider the hydrogen atom, in which there is no question that all excitations are Rydberg in character. Further, the ΔSCF IP is simply the ground-state energy, so there is no question of error cancelation in evaluating it. Self-interaction error is a particularly large problem for one-electron systems, and *all* errors are in some way or another manifestations of it. We use an even-tempered expansion of 48 Gaussian functions with exponents between 10^6 and 10^{-4}, inclusive. Since all basis functions are spherical we use a one-point angular grid, and we use a 999-point radial grid. Results are given in Table 9. As discussed above, we make the adiabatic approximation for the exchange-correlation kernel and neglect any current dependence.

Let us begin with Hartree−Fock, which is of course exact within the limitations imposed by our finite basis set. The IP as calculated in a Δ-SCF fashion is identical to that given by the occupied orbital energy, and within our basis, we describe the first nine Rydberg states ($n = 2$ through $n = 10$) or so almost exactly (error in the excitation energy for $n = 10$ is only 0.4 meV). The Rydberg excitation energies approach the IP, and at $n = 8$ are still about 0.2 eV away, exactly as they should be.

Turning to LC-ωPBE, we find a slightly different story. The IP as calculated by ΔSCF is 13.8 eV, with the error due primarily to the PBE correlation energy. But the IP as calculated by $-\varepsilon$ is only 12.2 eV; the difference reflects self-interaction error. Past the first few excitations, there is an almost constant error of 1.4 eV in the excitation energies; the LC-ωPBE Rydberg series is converging (as it should) to $-\varepsilon$. In other words, while having the right asymptotic potential is necessary, it is insufficient to guarantee that Rydberg excitations are properly described, and in long-range-corrected functionals, we *also* require that the orbital energy is correct.

Considering next PBE, the story changes again. The IP as calculated by ΔSCF is 13.6 V, indicating minimal self-interaction error in the energy. On the other hand, the IP as calculated by $-\varepsilon$ is only 7.6 eV, indicating a very large self-interaction error in the potential. The excitation energies calculated by TDDFT do not even form an obvious Rydberg series, and only one of them is below $-\varepsilon$. Several other excitations lie closely spaced but just above $-\varepsilon$. The fact that these states do not lie below $-\varepsilon$ may be due to inadequacies of our basis set.

The same qualitative story is true in HISS and HSE, which far from the nucleus resemble PBE. The PBEh global hybrid has several excitations below $-\varepsilon$, but it is not clear that they form a Rydberg series per se.

3.6. Dissociation of Homonuclear Diatomic Cations. The dissociation of homonuclear diatomic cations (X_2^+) has proven to be a challenge for density functional methods. Asymptotically, these systems can dissociate symmetrically to two fractionally charged species ($2X^{+1/2}$) or asymmetrically to one neutral and one charged component ($X + X^+$). In the exact theory, these limits must be degenerate. However, this is rarely the prediction of single determinant methods. Hartree−Fock theory predicts the symmetric dissociation limit to lie above the asymmetric limit—it frequently prefers to break symmetry and localize electrons. Semilocal functionals predict the opposite, and the potential energy curve goes through a maximum before turning over into a $1/(4R)$ decay far away as the two species interact Coulombically. Global hybrids like PBEh behave similarly, though the turnover is not as pronounced. The LC-ωPBE functional is somewhat better behaved—the potential energy curve does not have a transition state as in PBE, PBEh, and

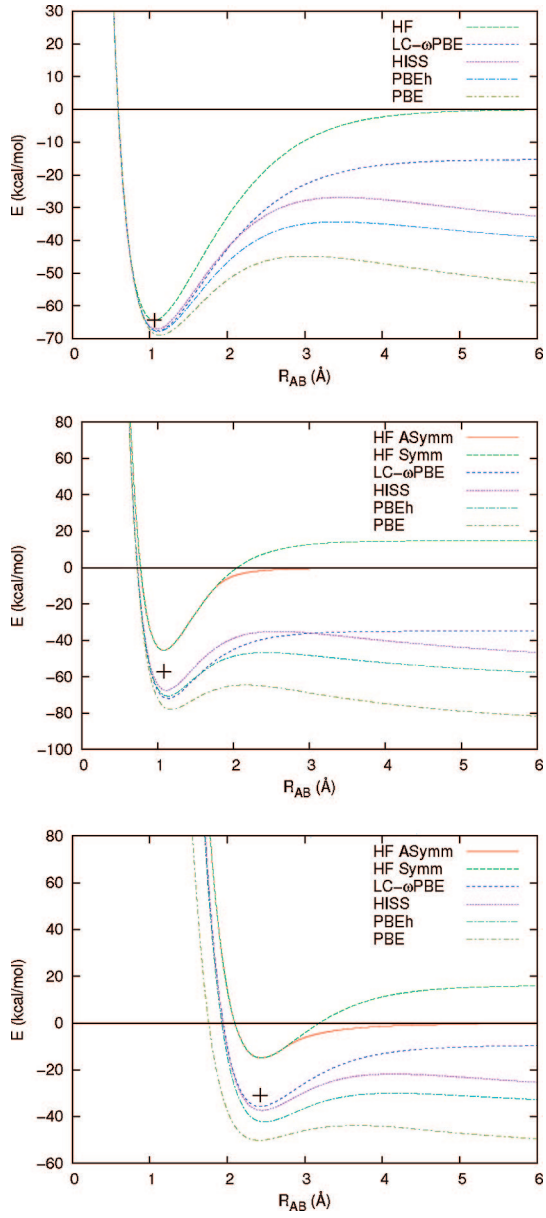


Figure 1. Dissociation of several homonuclear diatomic cations with HF theory and with several functionals based on the PBE GGA. The experimental equilibrium bond length and dissociation energy is marked by the symbol "+". (top panel) H_2^+ (experiment, ref 41). (middle panel) He_2^+ (experiment, ref 42). (bottom panel) Ar_2^+ (experiment, ref 43).

HSE, but it predicts the symmetric dissociation limit to be lower in energy than the asymmetric limit.

We therefore examine the performance of the HISS functional for the dissociation of H_2^+, He_2^+, and Ar_2^+, with the potential energy curves given in Figure 1. The qualitative behavior of HISS is similar to that of PBEh and PBE, with an artificial transition state in the potential energy curve. However, HISS tends to push this point off to slightly larger internuclear separation. In other words, eventually the lack of long-range exact exchange causes HISS to break down and gives the type of result we expect to see for a semilocal functional, but this is delayed to longer bond lengths. Note that these calculations used Dunning's aug-cc-pVQZ basis set.

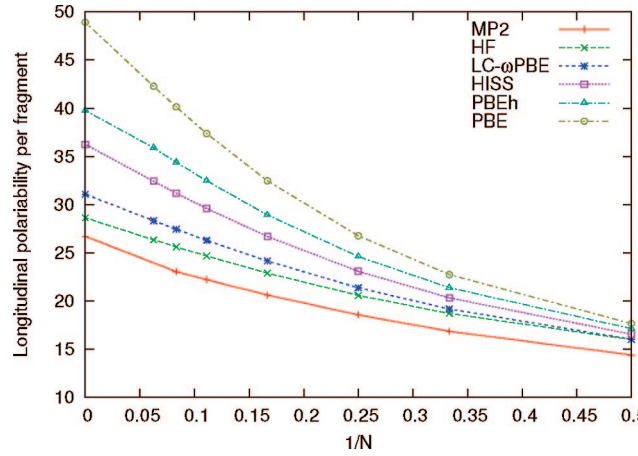


Figure 2. Longitudinal polarizability per H_2 fragment (a_0^3) with HF, MP2, and several functionals based on the PBE GGA, plotted against one over the number of H_2 fragments. The PBC limit is thus $1/N \rightarrow 0$.

3.7. Polarizabilities in H_2 Chains. One well-known failure of semilocal functionals is in describing the polarizability of extended systems. Essentially, by artificially favoring electron delocalization due to self-interaction error, semilocal functionals tend to overestimate polarizabilities, particularly for large systems. This failure is largely corrected by long-range corrected exchange functionals.[46,47] We therefore wish to test the performance of HISS for this case, to determine whether the long-range-correction is critical or whether polarizabilities, like reaction barriers, are more sensitive to the presence of middle range exact exchange.

Our test systems constitute one-dimensional chains of hydrogen molecules aligned colinearly with interatomic and intermolecular distances of 2 and 3 au, respectively. As shown in ref 48, accurately predicting the polarizability of the H_2 linear chain is particularly difficult for semilocal functionals. Figure 2 presents the longitudinal polarizability per H_2 fragment for oligomers of various lengths and the periodic chain. All periodic estimates except that for MP2 have been obtained by using our analytic polarizability code with periodic boundary conditions.[47] For the periodic MP2 polarizabilities, we used finite electric field calculations[50] of MP2 unit cell energies.[51] All calculations used the 6-311G** basis set.

The polarizabilities obtained correlate with the amount of exact exchange in a straightforward way, with HISS yielding results intermediate between LC-ωPBE and PBEh, in agreement with the qualitative results for homonuclear diatomic cations. Apparently, the amount of exact exchange is more important for this property than the range in which it is included. This is in agreement with self-interaction-corrected calculations of the same system,[52] which show that as the self-interaction error decreases, so does the error in the polarizability.

4. Conclusions

Range-separated exchange functionals have much to offer over conventional hybrids, allowing for simple functionals which can do many things right. A screened hybrid which

includes exact exchange only at short-range can provide accuracy essentially equivalent to that of the parent hybrid while increasing computational efficiency in solids and large molecules. A long-range-corrected hybrid which includes exact exchange only at long-range can simultaneously describe both thermochemistry and kinetics, as well as a wide variety of problems sensitive to the asymptotic form of the exchange potential.

By compromising between these two approaches, the HISS functional can accurately and efficiently predict atomization energies, barrier heights, and band gaps in solids. For properties sensitive to the longer-range exchange potential, HISS yields results, as one might expect, intermediate between those from the PBEh global hybrid and the LC-ωPBE long-range-corrected hybrid. In the cases of the dissociation curves of homonuclear diatomic cations and the polarizabilities of H_2 chains, for which both semilocal and global hybrids fail, HISS is reasonably accurate and begins to fail only as the system gets large or spreads far apart. The same general observation holds for electronically excited states: HISS describes low-lying Rydberg states very well, presumably because these states are not too diffuse, and it is only as the states get higher in energy and more spread out that HISS will break down even if the basis set is otherwise capable of describing these excitations.

Of course, middle-range exact exchange is not a panacea, and for some properties, HISS performs no better than, or sometimes worse than, a global hybrid. As a rough rule of thumb, HISS underperforms in many of the same situations as does LC-ωPBE. This need not be too surprising, as both functionals use a large amount of exact exchange but use the standard PBE correlation functional. With their large fraction of exact exchange, neither HISS nor LC-ωPBE can be expected to benefit from the full error cancelation in semilocal functionals. By refining the correlation functional used with range-separated hybrids, performance for problematic cases should be improved further. Local range-separated exchange hybrids, which use make the fraction of exact exchange at each range a function of space, are also a route toward improved functionals.[53−56]

Acknowledgment. A.F.I. would like to thank Vincenzo Russo for help with the TDDFT assessment. This work was supported by the National Science Foundation (CHE-0457030), the Department of Energy (DE-FG02−04ER15523), the Welch Foundation (C-0036), and ARO-MURI DAAD-19-03-1-0169.

Supporting Information Available: Details of the basis sets used for our calculations in extended systems. 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.

(2) Dreizler, R. M.; Gross, E. K. U. *Density Functional Theory*; Plenum Press: New York, 1995.

(3) 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*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, The Netherlands, 2005; pp 669−724.

(4) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.

(5) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(6) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(7) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(8) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(9) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029.

(10) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

(11) Monkhorst, H. J. *Phys. Rev. B* **1979**, *20*, 1504.

(12) Savin, A. On degeneracy, near-degenaracy and density functional theory. In *Recent Developments and Applications of Modern Density Functional Theory*; Seminario, J. M., Ed.; Elsevier: Amsterdam, The Netherlands, 1996; p *327.*

(13) Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. *Chem. Phys. Lett.* **1997**, *275*, 151.

(14) Gerber, I. C.; Angyan, J. G. *Chem. Phys. Lett.* **2005**, *415*, 100.

(15) Gerber, I. C.; Angyan, J. G.; Marsman, M.; Kresse, G. *J. Chem. Phys.* **2007**, *127*, 054101.

(16) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2003**, *118*, 8207.

(17) Heyd, J.; Scuseria, G. E. *J. Chem. Phys.* **2004**, *120*, 7274.

(18) Heyd, J.; Scuseria, G. E. *J. Chem. Phys.* **2004**, *121*, 1187.

(19) Izmaylov, A. F.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2006**, *125*, 104103.

(20) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 234109.

(21) Henderson, T. M.; Izmaylov, A. F.; Scuseria, G. E.; Savin, A. *J. Chem. Phys.* **2007**, *127*, 221103.

(22) Ernzerhof, M.; Perdew, J. P. *J. Chem. Phys.* **1998**, *109*, 3313.

(23) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 8996.

(24) Seidl, A.; Görling, A.; Vogl, P.; Majewski, J. A.; Levy, M. *Phys. Rev. B* **1996**, *53*, 3764.

(25) Frisch, M. J. ; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A. ; Cheeseman, J. R. ; Montgomery, J. A., Jr.; Vreven, T.; Scalmani, G.; Kudin, K. N.; Iyengar, S. S. ; Tomasi, J. ; Barone, V. ; Mennucci, B.; Cossi, M.; 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.; Li, X.; Hratchian, H. P.; Peralta, J. E. ; Izmaylov, A. F.; Brothers, E.; Staroverov, V.; Kobayashi, R. ; Normand, J. ; Burant, J. C.; Millam, J. M.; Klene, M. J.; Knox, E.; 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.; Chen, W.; Wong, M. W.; Pople, J. A.

Gaussian Development Version, revision E.05; Gaussian, Inc.: Wallingford CT, 2006.

(26) Brothers, E. N.; Scuseria, G. E. *J. Chem. Theory Comput.* **2006**, *4*, 1045.

(27) Chakravorty, S. J.; Gwaltney, S. R.; Davidson, E. R.; Parpia, F. A.; Fischer, C. F. *Phys. Rev. A* **1993**, *47*, 3649.

(28) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063.

(29) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374.

(30) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 108.

(31) Janesko, B. G.; Scuseria, G. E. *J. Chem. Phys.*, in press.

(32) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2004**, *7*, 43.

(33) Zhao, Y.; González-García, N.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 4942.

(34) Brothers, E. N.; Izmaylov, A. F.; Normand, J. O.; Barone, V.; Scuseria, G. E. *J. Chem. Phys.* **2008**, *129*, 011102.

(35) See the Supporting Information.

(36) Heyd, J.; Peralta, J. E.; Scuseria, G. E.; Martin, R. L. *J. Chem. Phys.* **2005**, *123*, 174101.

(37) Izmaylov, A. F.; Scuseria, G. E. Why are TD-DFT excitations in solids equal to band structure energy gaps for semilocal functionals and how does nonlocal Hartree-Fock type exchange introduce excitonic effects? *J. Chem. Phys.* **2008**, *129*, 034101.

(38) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42.

(39) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2004**, *121*, 11507.

(40) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129.

(41) Sadlej, A. J. *Theor. Chim. Acta* **1991**, *79*, 123.

(42) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126.

(43) Schaad, L. J.; Hicks, W. V. *J. Chem. Phys.* **1970**, *53*, 851.

(44) Xie, J.; Poirier, B.; Gellene, G. I. *J. Chem. Phys.* **2005**, *122*, 184310.

(45) Wüest, A.; Merkt, F. *J. Chem. Phys.* **2004**, *210*, 638.

(46) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 3540.

(47) Sekino, H.; Maeda, Y.; Kamiya, M. *Mol. Phys.* **2005**, *103*, 2183.

(48) van Faassen, M.; de Boeij, P. L.; van Leeuwen, R.; Berger, J. A.; Snijders, J. G. *Phys. Rev. Lett.* **2002**, *88*, 186401.

(49) Izmaylov, A. F.; Brothers, E. N.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 224105.

(50) Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2000**, *113*, 7779.

(51) Ayala, P. Y.; Kudin, K. N.; Scuseria, G. *J. Chem. Phys.* **2001**, *115*, 9698.

(52) Ruzsinksky, A.; Perdew, J. P.; Csonka, G. I.; Scuseria, G. E.; Vydrov, O. A. *Phys. Rev. A*, in press.

(53) Jaramillo, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2003**, *118*, 1068.

(54) Janesko, B. G.; Scuseria, G. E. *J. Chem. Phys.* **2007**, *127*, 164117.

(55) Arbuznikov, A. V.; Kaupp, M. *J. Chem. Phys.* **2007**, *127*, 194102.

(56) Janesko, B. G.; Scuseria, G. E. *J. Chem. Phys.* **2008**, *128*, 084111.

CT800149Y

JCTC Journal of Chemical Theory and Computation

Photoisomerization Reactions of Cyclopropene and 1,3,3-Trimethylcyclopropene: A Theoretical Study

Ming-Der Su*

Department of Applied Chemistry, National Chiayi University, Chiayi 60004, Taiwan

Received March 6, 2008

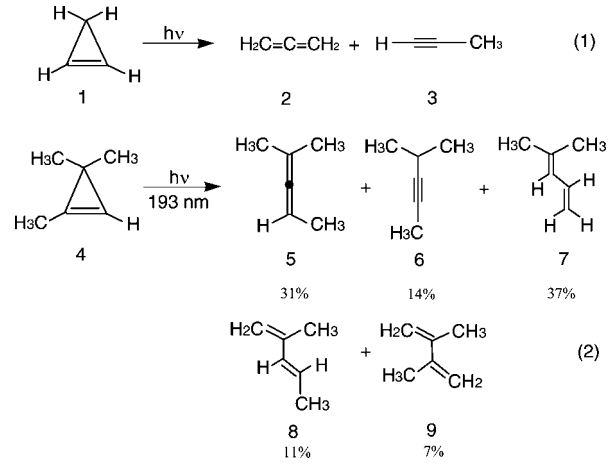
Abstract: The mechanisms of the photochemical isomerization reactions were investigated theoretically using two model systems, cyclopropene and 1,3,3-trimethylcyclopropene with the CASSCF/6−311G(d) (six-electron/six-orbital active space) and MP2-CAS/6−311G(d,p)//CASSCF/6−311G(d) methods. The structures of the conical intersections, which play a decisive role in such photorearrangements, were obtained. The intermediates and transition structures of the ground states were also calculated to assist in providing a qualitative explanation of the reaction pathways. Our model investigations suggest that the preferred reaction route for both cyclopropene and 1,3,3-trimethylcyclopropene is as follows: reactant → Franck−Condon region → local minimum → transition state → conical intersection → local intermediate → transition state → photoproduct. The theoretical findings suggest that the conical intersection mechanism found in this work gives a good explanation and supports the experimental observations. We also investigated the thermal (dark) reaction mechanisms for the hydrogen migration reactions. Again, all the relative yields of final products predicted based on the present work are in good agreement with the available experimental findings.

I. Introduction

Experimentalists and theoreticians alike have been long fascinated by the inclusion of double bonds in small carbocyclic rings.[1] For instance, the chemical and physical properties of the ground- and excited-state potential energy surfaces which link the cyclopropene (C_3H_4) hydrocarbons have been the subject of considerable experimental and theoretical investigations over many years.[2] Nevertheless, the photochemistry of cyclopropenic compounds has drawn surprisingly little attention.[3] The first report of cyclopropene photolysis appeared in the work of Chapman.[4] He reported that photolysis of cyclopropene itself (**1**) in an argon matrix at 8 K (eq 1 in Scheme 1) yields allene (**2**) and propyne (**3**). However, the relative yields of the two photoproducts have never been reported. On the other hand, thermolysis of cyclopropene at 190−240 °C yields **2** and **3** in yields of 1−2% and ca. 98%, respectively.[5]

Also, other intriguing photochemical results have been found in an unsymmetrical, alkyl-substituted cyclopropene (1,3,3-trimethylcyclopropene, **4**). See eq 2 in Scheme 1. This study represents the first investigation of the photochemistry of monocyclic, alkyl-substituted cyclopropene derivatives in solution. Photolysis of **4** in hydrocarbon solution with far-uv light results in rearrangement to the five primary products (**5**−**9**) shown in eq 2.[3,6] Similar product yields have been observed in the thermolysis of **4**. Gas-phase thermolysis of

Scheme 1



* Corresponding author e-mail: midesu@mail.ncyu.edu.tw.

10.1021/ct800078j CCC: $40.75
Published on Web 06/21/2008

4 has been reported to produce **8** and **9** in yields of 71 and 21%, respectively; no other products were reported.[7] Leigh and Fahie considered that most of the photolysis products are formally derived from the vinylcarbene formed by cleavage of the more highly substituted cyclopropene σ-bond in the excited singlet state. Moreover, two-bond cleavage of excited cyclopropenes to yield carbenes and alkynes is unimportant in solution, even for cyclopropenes with high excitation energies.[6] However, to the best of our knowledge, until now no theoretical investigations have been devoted to the study of such photochemical isomerization reactions.

Since the photochemical rearrangements of cyclopropene and its related compound mentioned above are both unusual and useful, we were curious about exactly how they occur and wanted detailed mechanistic knowledge in order to exercise greater control over them. Although these experimental results help in understanding the potential energy surfaces of excited states in cyclopropene systems, they are at present not capable of providing complete mechanistic detail. In fact, a detailed understanding of the photochemical reactions of cyclopropene as well as its derivatives is of interest not only for the advancement of basic science but also for further precise control of the total reaction process.

In principle, most photochemical reactions of organic molecules start on an excited electronic potential surface but cross over to a lower surface somewhere along the reaction pathway.[8,9] They finally reach the ground-state surface by a sequence of radiationless transitions (i.e., conical intersections) and move on the ground-state surface toward the product.[8] This mechanism is not controlled by the avoided surface crossing and the resulting energy gap between ground and excited states but rather by the presence of minima and transition states on the ground and excited states themselves. Furthermore, the existence of a conical intersection region provides access to a number of ground-state pathways that can lead to different photoproducts.[10] Indeed, interest in the area of photochemistry over the past decade has remained high as a result of a number of efforts focusing on excited-state processes that are promoted by conical intersections. This has already been both experimentally and theoretically proved to be a general feature of the excited states relevant to photochemical reactions.[11]

In order to understand the reaction mechanism of the photoreactions of cyclopropenes,[8–11] we have undertaken an investigation of the potential energy excited-state surfaces of C_3H_4 (**1**) and C_6H_{10} (**4**) in their singlet states. It is the aim of this paper to generate the essential parts of the potential surfaces by quantum chemical calculations and to describe the consequences for the reaction mechanisms from such an explicit calculation of the reaction pathways. It will be shown below that the conical intersection[10,11] plays a crucial role in the photochemistry of cyclopropenes systems.

II. Methology

All the geometries were fully optimized without imposing any symmetry constraints, although in some instances the resulting structures showed various elements of symmetry. The *ab initio* molecular orbital calculations were performed using the Gaussian 03 software package.[12]

In the investigation of the photochemical reaction pathways, the stationary point structures on the S_0 and S_1 surfaces were optimized at the CASSCF (the complete-active-space SCF) level of calculations using the standard 6−311G(d) basis set.[13] The active space for describing the photoisomerizations of **1** and **4** comprises six electrons in six orbitals, i.e., two p-π orbitals plus two σ(C−C) and two σ*(C−C) orbitals. This is referred to as CASSCF(6,6). In some hydrogen shift molecules, their active space is chosen as two p-π orbitals plus two σ (i.e., C−C and C−H) and two σ* (i.e., C−C and C−H) orbitals. The state-averaged CASSCF(6,6) method was used to determine geometry on the intersection space. The optimization of conical intersections was achieved in the (f - 2)-dimensional intersection space using the method of Bearpark et al.[14] implemented in the Gaussian 03 program.

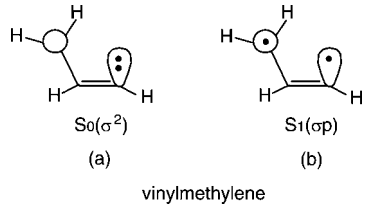
Every stationary point was characterized by its harmonic frequencies computed analytically at the CASSCF level. The harmonic vibrational frequencies of all the stationary points were computed analytically to characterize them as minima (all frequencies are real) or transition states (only one imaginary frequency). The optimization was determined when the maximum force and its root-mean-square (rms) were less than 0.00045 and 0.00005 hartree/bohr, respectively. For each transition state, an IRC (CASSCF) calculation[15] has been performed as well. Accordingly, the products and reagents connected by the respective transition states have been unambiguously localized. The IRC results are given in the Supporting Information. Localization of the minima, transition states, conical intersection minima, and mapping of the intrinsic reaction coordinates has been performed in mass-weighted Cartesian coordinates; therefore, the results are independent of any specific choice of internal variables.

To correct the energetics for dynamic electron correlation, we have used the multireference Møller−Plesset algorithm[16] as implemented in the program package GAUSSIAN 03. Unless otherwise noted, the relative energies given in the text are those determined at the MP2-CAS-(6,6)/6−311G(d,p) level using the CAS(6,6)/6−311G(d) (hereafter designed MP2-CAS and CASSCF, respectively) geometry.

III. General Consideration

Even though the cyclopropene and cyclopropene derivative photoisomerization reported experimentally[3,6,7] show wide variance in their reaction types, it is possible to construct a certain consistency in these reactions, which at least serves as a basis for discussion. In this section, we describe possible excited-state reaction paths that lead to the conical intersections. It has been shown that thermolysis of cyclopropenes results in the formation of products consistent with initial ring opening to yield vinylcarbene intermediates.[3,17] Also, it has been found that the same situation occurs during the photolysis of cyclopropenes.[6,17] Basically, thermochemical cyclopropene ring opening produces the S_0 state (σ^2), whereas photochemical ring opening is considered to involve the S_1 state (σp; planar vinyl diradical).[3,17] See Scheme 2. Moreover, according to previous theoretical calculations, it was noted that the lowest excited triplet state of cyclopropene

Scheme 2



is unreactive toward ring opening and generally undergoes dimerization.[3,18] We thus do not consider the excited triplet state any further.

Again, in the case of eq 2, Leigh and Fahie proposed a mechanism involving the vinylcarbene intermediate to explain the formation of a variety of photoproducts due to the photolysis of **4**.[3,6] As demonstrated in Scheme 3, they concluded that 70−90% of the observable products are derived from the two possible vinylcarbene intermediates (**10** and **11**) formed from **4**.[3,6] Namely, they suggested that the photoproducts **5**−**8** should result from the vinylcarbene **10** formed by cleavage of the more highly substituted cyclopropene single bond in **4**, whereas only diene **9** derives from the vinylcarbene **11** obtained by cleavage of the less substituted cyclopropene single bond. Further supporting evidence comes from the fact that, according to the bond order analysis based on the CAS(6,6)/6−311G(d) level of theory, it was found that the C−C bond order increases in the order C_1-C_3 (1.042) < C_2-C_3 (1.082) < C_1-C_2 (1.175) at the S_1 excited state of 1,3,3-trimethylcyclopropene (**4**). Since it is well-known that absorption of light leads to cleavage of the weakest of the chemical bonds in the ring,c one may predict that bond cleavage should decrease in the order $C_1-C_3 > C_2-C_3 > C_1-C_2$. That is to say, the most substituted (C_1-C_3) 1,3,3-trimethylcyclopropene (**4**) single bond should be more easily broken than the less substituted (C_2-C_3) **4** single bond. From the above discussion, it is clear that vinylcarbene intermediates should play a central role in the photochemical mechanisms of cyclopropene and its derivatives. Nevertheless, attempts to chemically trap such intermediates using methanol or alkene have all been unsuccessful.[6]

We shall use the above ideas to help locate the "funnel" from the excited-state surface to the ground-state surface that corresponds to a conical intersection in the following section.

IV. Results and Discussion

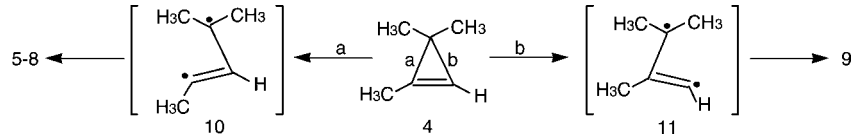
(1) Cyclopropene (C_3H_4). Let us first consider the photoisomerizations of cyclopropene (**1**). For an understanding of its both ground- and excited-state mechanism, it is best to start the discussion with the reaction profiles as summarized in Figure 1, which also contains the relative energies of all the critical points with respect to the energy of the reactant **1**. Some selected geometrical parameters optimized for the stationary points and conical intersections are collected in Figure 2. The energies relative to the reactant molecule (**1**) are also listed in Table 1. Cartesian coordinates and energetics calculated for the various points at the CASSCF/6−311G(d) level are available as Supporting Information.

Our theoretical investigations are based on the mechanism indicated in Figure 1. In the first step the reactant (cyclopropene, **1**) is promoted to its excited singlet state by a vertical excitation as shown in the left-hand of Figure 1. After the vertical excitation process, the molecule is situated on the excited singlet surface but still possesses the S_0 (ground-state) geometry (**FC-1**).[15] The vertical excitation energy ($S_0 \rightarrow S_1(S_0$ geom)) was calculated to be 169 and 200 kcal/mol at the MP2-CAS and CASSCF levels of theory. As there are no relevant experimental and theoretical data on the cyclopropene system, the above result is a prediction.

From the point reached by the vertical excitation $S_1(S_0$ geom), the molecule relaxes to a local minimum **Int-1** ($S_1(S_1$ geom)),[19] which is predicted to be 60 kcal/mol lower in energy than **FC-1** ($S_1(S_0$ geom)). The optimized geometrical parameters of **Int-1** ($S_1(S_1$ geom)) are given in Figure 2. Comparing the **Int-1** geometry (Figure 2) with that of its corresponding ground-state minimum **1** (Figure 2), it is readily seen that the former has one significantly longer C−C bond (1.827 Å) and two somewhat shorter C−C bonds (1.472 and 1.449 Å) than its closed shell singlet state. As a result, the three-membered ring in the intermediate **Int-1** is to some extent broken. From this local minimum, **Int-1**, a transition state search for the ring-opening based on the model of the **Int-1** conformation was investigated. As can be seen in Figures 1 and 2, the cyclopropene ring opening process proceeds via a transition state **TS-1** (and **TS-1′**). Vibrational frequency calculations show that **TS-1** (and **TS-1′**) is a real transition state with one imaginary frequency (1047*i* and 927*i* cm^{-1}) on the singlet potential energy surface. Our MP2-CAS calculation predicts that this route possesses a small barrier, of about 12 kcal/mol. Due to the large excess energy of about 60 kcal/mol resulting from the relaxation from **FC-1** ($S_1(S_0$ geom)) to **Int-1** ($S_1(S_1$ geom)), the barrier can easily be surmounted.

From the **TS-1** point cyclopropene reaches an S_1/S_0 **CI** where the photoexcited system decays nonradiatively to S_0. Specifically, the photochemically active relaxation path, starting from the **TS-1** excited-state of cyclopropene, leads to either S_1/S_0 **CI-c-1** (path 1) or **CI-t-1** (path 2). As already mentioned in the previous section, one may foresee that the vinylcarbene-like species (**CI-c-1** and **CI-t-1**) should play a key role in the photorearrangement reactions of **1**. That is, along the C_1C_2 bending reaction path, two conical intersections, **CI-c-1** and **CI-t-1**, are obtained at the $\angle C_1C_2C_3$ bond angle of 123° and 125°, respectively. Their optimized structures are given in the left- and right-hand side of Figure 2 along with several key structural parameters. Moreover, the energy difference between these two conical intersections is only 1.1 kcal/mol, the former being lower than the latter, at the MP2-CAS level of theory. In Figure 2, we also give the directions of the derivative coupling and gradient difference vectors for both S_1/S_0 **CI-c-1** and **CI-t-1** conical intersections. As a result, for instance, funneling through the S_1/S_0 **CI-c-1** conical intersection leads to two different reaction pathways on the ground-state surface via either the derivative coupling vector or the gradient difference vector.[10] Any linear combination of these vectors causes the degeneracy to be lifted, and therefore these vectors give an

Scheme 3



indication of possible reaction pathways available on the ground-state surface after decay. According to the results demonstrated in Figure 2, examination of these two vectors in **CI-c-1** provides important information about the photoisomerization process of **1**: the major contribution to the derivative coupling vector involves C−C bond stretching that gives the planar **Int-c-2** intermediate on the S_0 surface. However, the gradient difference vector gives an asymmetric CCC bending motion that leads to a vibrationally hot **1**-S_0 species. Similarly, this phenomenon can also be found in path 2. That is, as shown in Figure 2, the derivative coupling vector of **CI-t-1** results in the planar **Int-t-2** intermediate, while its gradient difference vector gives a vibrationally hot **1**-S_0 species.

From these local minima (**Int-c-2** and **Int-t-2**), a hydrogen migration must take place via transition states (**TS-c-2** and **TS-t-2**) to give the final products **2** and **3**, respectively. Our theoretical findings indicate that both hydrogen migration pathways possess their own barrier heights, computed to be

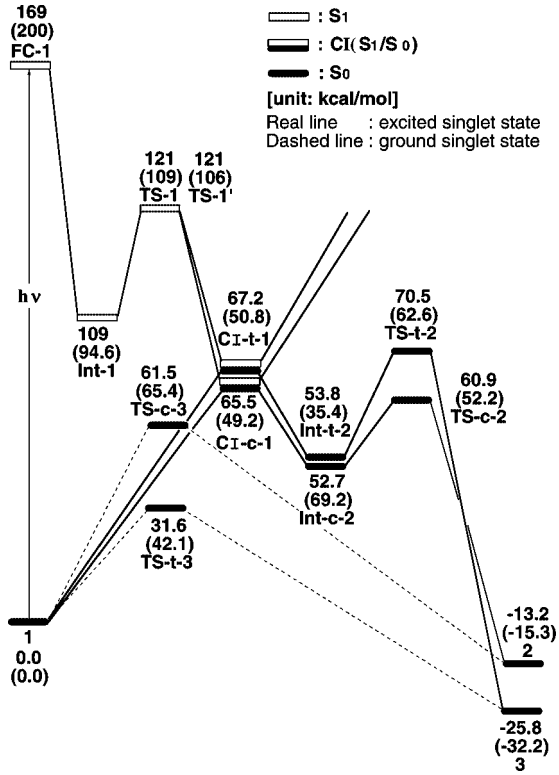


Figure 1. Energy profiles for the photoisomerization modes of cyclopropene (**1**). The abbreviations FC and CI stand for Frank-Condon and conical intersection, respectively. The relative energies were obtained at the MP2-CAS-(6,6)/6−311G(d,p)//CAS(6,6)/6−311G(d) and CAS(6,6)/6−311G(d) (in parentheses) levels of theory. All energies (in kcal/mol) are given with respect to the reactant (**1**). The CASSCF optimized structures of the crucial points, see Figures 2 and 3. For more information see the text.

17 kcal/mol for the formation of allene **2** (path 2) and 8.2 kcal/mol for the production of propyne **3** (path 1). Additionally, both photoproducts **2** and **3** are thermodynamically stable by 13 and 26 kcal/mol compared with reactant **1**, respectively. Thus, our computational results suggest that the mechanisms for path 1 and path 2 should proceed as follows:

$$\text{Path 1: } \mathbf{1}(S_0) + h\nu \rightarrow \mathbf{FC-1} \rightarrow \mathbf{Int-1} \rightarrow \mathbf{TS-1} \rightarrow \mathbf{CI-t-1} \rightarrow \mathbf{Int-t-2} \rightarrow \mathbf{TS-t-2} \rightarrow \mathbf{3} \quad (1)$$

$$\text{Path 2: } \mathbf{1}(S_0) + h\nu \rightarrow \mathbf{FC-1} \rightarrow \mathbf{Int-1} \rightarrow \mathbf{TS-1} \rightarrow \mathbf{CI-c-1} \rightarrow \mathbf{Int-c-2} \rightarrow \mathbf{TS-c-2} \rightarrow \mathbf{3} \quad (2)$$

Furthermore, from the computational data discussed above, one can readily see that the barrier to path 1 is larger than that for path 2, by 8.5 kcal/mol in energy. This finding suggests that path 2 should be more favorable than path 1 from a kinetic viewpoint. Based on the present theoretical investigations as demonstrated in Figure 1 and Table 1, we thus predict that the photoproduct (allene, **2**) produced by path 2 should be in a larger quantum yield than the photoproduct (propyne, **3**) produced by path 1 (see below).

On the other hand, we have also examined the dark (thermal) reaction on the ground-state potential energy surface. Although photoexcitation raises cyclopropene into an excited electronic state, the products of the photochemical process are controlled by the ground-state (thermal) potential surface.[7] Two reaction pathways (path 3 and path 4) were thus explored in this work. The search for transition states on the S_0 surface near the structures of **CI-c-1** and **CI-t-1** gives **TS-c-3** and **TS-t-3** for the products allene **2** (path 4) and propyne **3** (path 3), respectively. The optimized geometrical parameters of these transition states are collected in Figure 3. In addition, the MP2-CAS computational results indicate that the energy of **TS-c-3** relative to the ground-state minimum (**1**) is 62 kcal/mol and lower than the S_1/S_0 **CI-c-1** by only 4.0 kcal/mol. Nevertheless, the energy of the **TS-t-3** connecting **1** and **3** on the S_0 surface lies 32 kcal/mol above the energy of reactant **1**. This strongly suggests that in the dark reaction path 3 (formation of propyne **3**) is more favorable than path 4 (formation of allene **2**). Thus, our theoretical calculations demonstrate that both thermal reactions, path 3 and path 4, can be represented as follows:

$$\text{Path 3: } \mathbf{1} \rightarrow \mathbf{TS-t-3} \rightarrow \mathbf{3}$$

$$\text{Path 4: } \mathbf{1} \rightarrow \mathbf{TS-c-3} \rightarrow \mathbf{2}$$

Based on the above discussion, we can then propose a picture of the photochemistry of **1**, which is already shown schematically in Figure 1. Comparing the two reaction pathways (path 1 and path 2), our *ab initio* CASSCF and MP2-CAS calculations show that path 2 is preferred over path 1. That is, an efficient photoisomerization occurs when the photoexcited reactant **FC-1** evolves along a small barrier excited-state pathway, decays at a conical intersection point

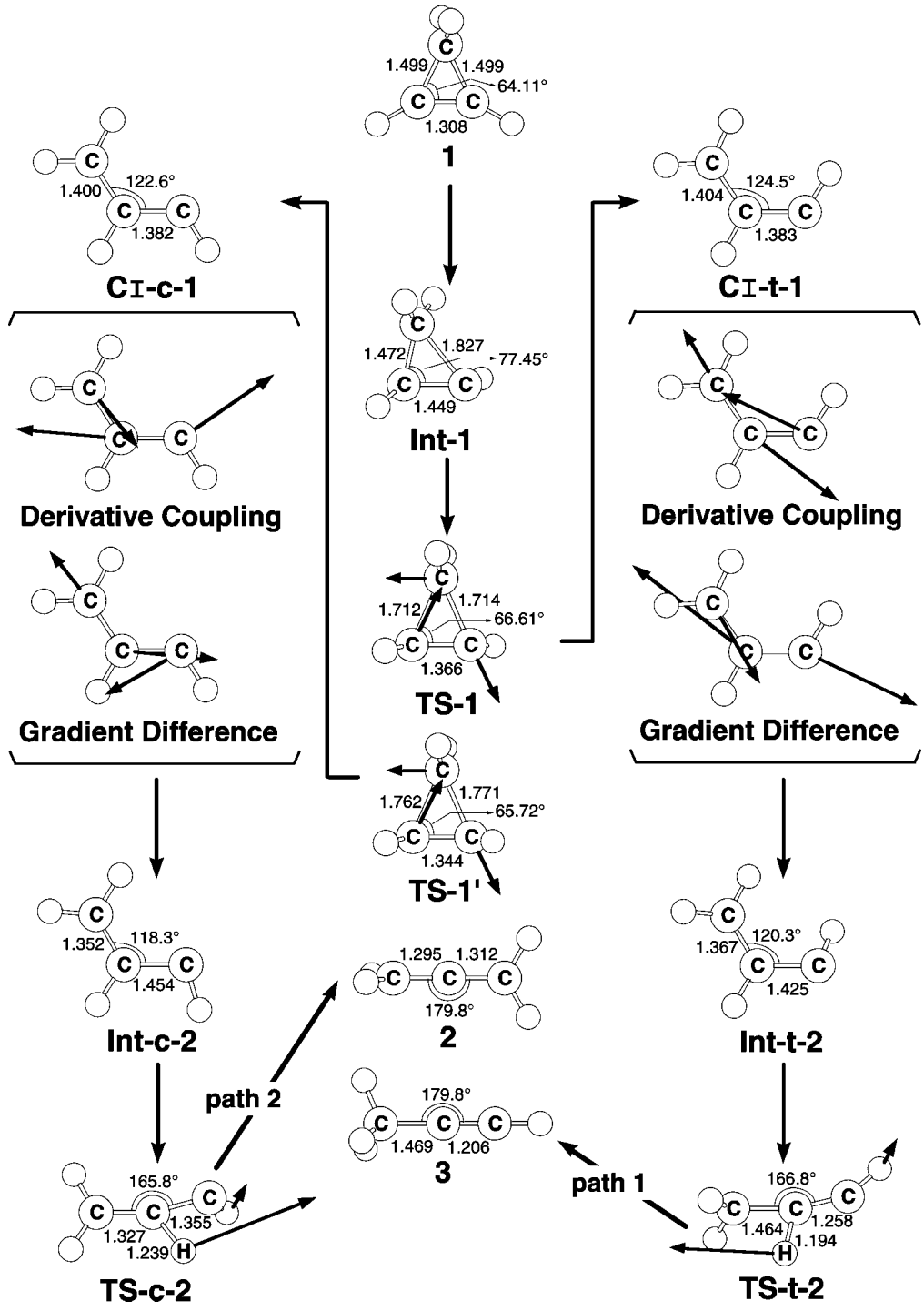


Figure 2. The CAS(6,6)/6−311G(d) geometries (in Å and deg) for path 1 and path 2 of cyclopropene (**1**), conical intersection (**CI**), intermediate (**Int**), transition state (**TS**), and isomer products. The derivative coupling and gradient difference vectors—those which lift the degeneracy—computed with CASSCF at the conical intersections **CI-c-1** and **CI-t-1**. The corresponding CASSCF vectors are shown inset. Also, the heavy arrows indicate the main atomic motions in the transition state eigenvector. For more information see the Supporting Information.

(S_1/S_0 **CI-c-1** or **CI-t-1**),[15] and finally relaxes to the ground-state between reactant (**1**) and photoproduct allene (**2**) or propyne (**3**), respectively. Also, our theoretical investigations strongly predict that the quantum yield of allene **2** should be greater than that of propyne **3** from a kinetic viewpoint. As mentioned earlier, neither experimental nor theoretical work on such a molecule has been reported so far.

However, in the dark (thermal) reaction the above situation is completely inverted. In the competition between the formation of allene **2** (path 4) and propyne **3** (path 3), the former has the larger energy requirement and the lower exothermicity. As a result, this makes path 3 the most energetically favorable pathway for thermal cyclopropene isomerization. In principle, these two reactions are concerted C−C bond cleavage accompanied by 1,2-H shift. Accordingly, one can foresee that in the dark reaction the yield of propyne **3** should be larger than that of allene **2.** This conclusion is in good agreement with experimental observations.[5]

Table 1. Energies (in kcal/mol) of the Critical Points Located along the Pathways 1-4 at the MP2-CAS(6,6)/6-311G(d,p)//CAS(6,6)/6-311G(d) and CAS(6,6)/6-311G(d) (in Parentheses) Levels of Theory[a]

structure	state	ΔErel[b]
cyclopropene **(1)**	S_0	0.0 (0.0)
FC-1	S_1	168.9 (199.6)
Int-1	S_1	108.6 (94.57)
TS-1	S_1	120.8 (109.2)
CI-t-1	S_1/S_0	67.23 (50.83)
TS-1′	S_1	120.5 (106.3)
CI-c-1	S_1/S_0	65.47 (49.20)
Int-t-2	S_0	53.78 (35.44)
Int-c-2	S_0	52.71 (69.23)
TS-t-2	S_0	70.46 (62.64)
TS-c-2	S_0	60.89 (52.20)
TS-t-3	S_0	31.61 (42,12)
TS-c-3	S_0	61.54 (65.42)
alene **(2)**	S_0	−13.15 (−15.32)
propyne **(3)**	S_0	−25.84 (−32.18)

[a] See Figures 2 and 3. [b] Energy relative to cyclopropene.

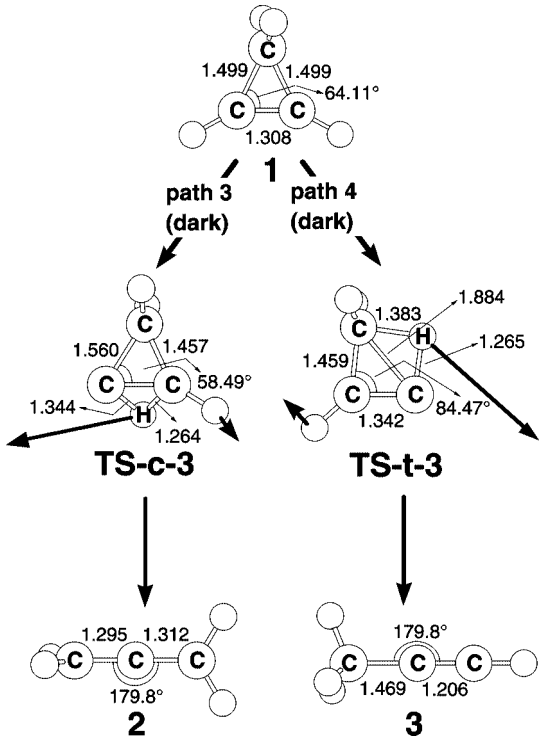


Figure 3. The CAS(6,6)/6−311G(d) geometries (in Å and deg) for path 3 and path 4 of cyclopropene (**1**), transition state (**TS**), and isomer products in the dark (thermal) reactions. The heavy arrows indicate the main atomic motions in the transition state eigenvector. For more information see the Supporting Information.

Furthermore, as already shown in Figures 1 and 2, our theoretical computations strongly demonstrate that both photolysis and thermolysis of cyclopropene (**1**) can be formally attributed to the initial formation of the vinylcarbene intermediate. This is consistent with what we predicted in the previous section.

(2) 1,3,3-Trimethylcyclopropene (C_6H_{10}). Next, we consider the photochemical rearrangement reactions of 1,3,3-trimethylcyclopropene (**4**). As stated earlier, there are two kinds of reaction pathways for the photoisomerization

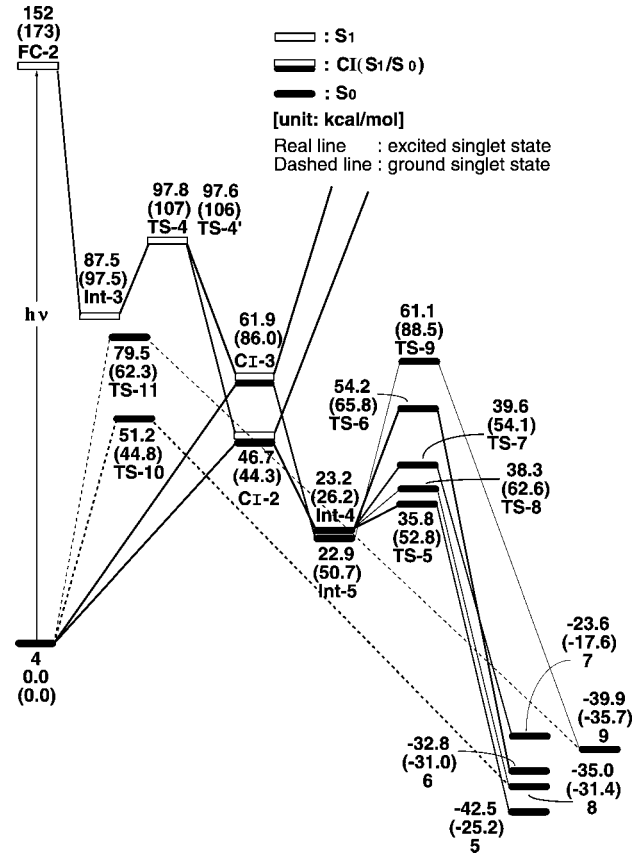


Figure 4. Energy profiles for the photoisomerization modes of 1,3,3-trimethylcyclopropene (**4**). The abbreviations **FC** and **CI** stand for Frank-Condon and conical intersection, respectively. The relative energies were obtained at the MP2-CAS(6,6)/6−311G(d,p)//CAS(6,6)/6−311G(d) and CAS(6,6)/6−311G(d) (in parentheses) levels of theory. All energies (in kcal/mol) are given with respect to the reactant (**4**). The CASSCF optimized structures of the crucial points, see Figures 5 and 6. For more information see the text.

reactions of **4**, i.e., path 5 and path 6, which may lead to the final photoproducts as given in Scheme 1. Figure 4 shows the reaction profiles computed for eq 2 and contains the relative energies of the various points with respect to the energy of the reactant **4**. Selected geometrical values and the relative energies based on the CASSCF and MP2-CAS calculations for all the stationary points of **4** are reported in Figures 5 and 6 and Table 2, respectively. Cartesian coordinates are given in the Supporting Information.

The vertical excitation energy (**FC-2**) is calculated to lie 173 kcal/mol above the ground-state surface at the CAS(6,6)/6−311G(d) optimized reactant geometry of **4**. This value drops to 152 kcal/mol after correction using MP2-CAS calculations, which is surprisingly close to the experimental absorption band (193 nm = 148 kcal/mol).[6] It is thus believed that the present model compound with the current method employed in this study should provide reliable information for the discussion of the singlet photochemical reaction mechanisms of **4**.

Once $S_0 \rightarrow S_1$ vertical excitation occurs, the system will relax from the **FC-2** point ($S_1(S_0$ geom)) to the S_1 minimum **Int-3** ($S_1(S_1$ geom)), the latter being lower in energy by 65 kcal/mol than the former. The optimized geometrical pa-

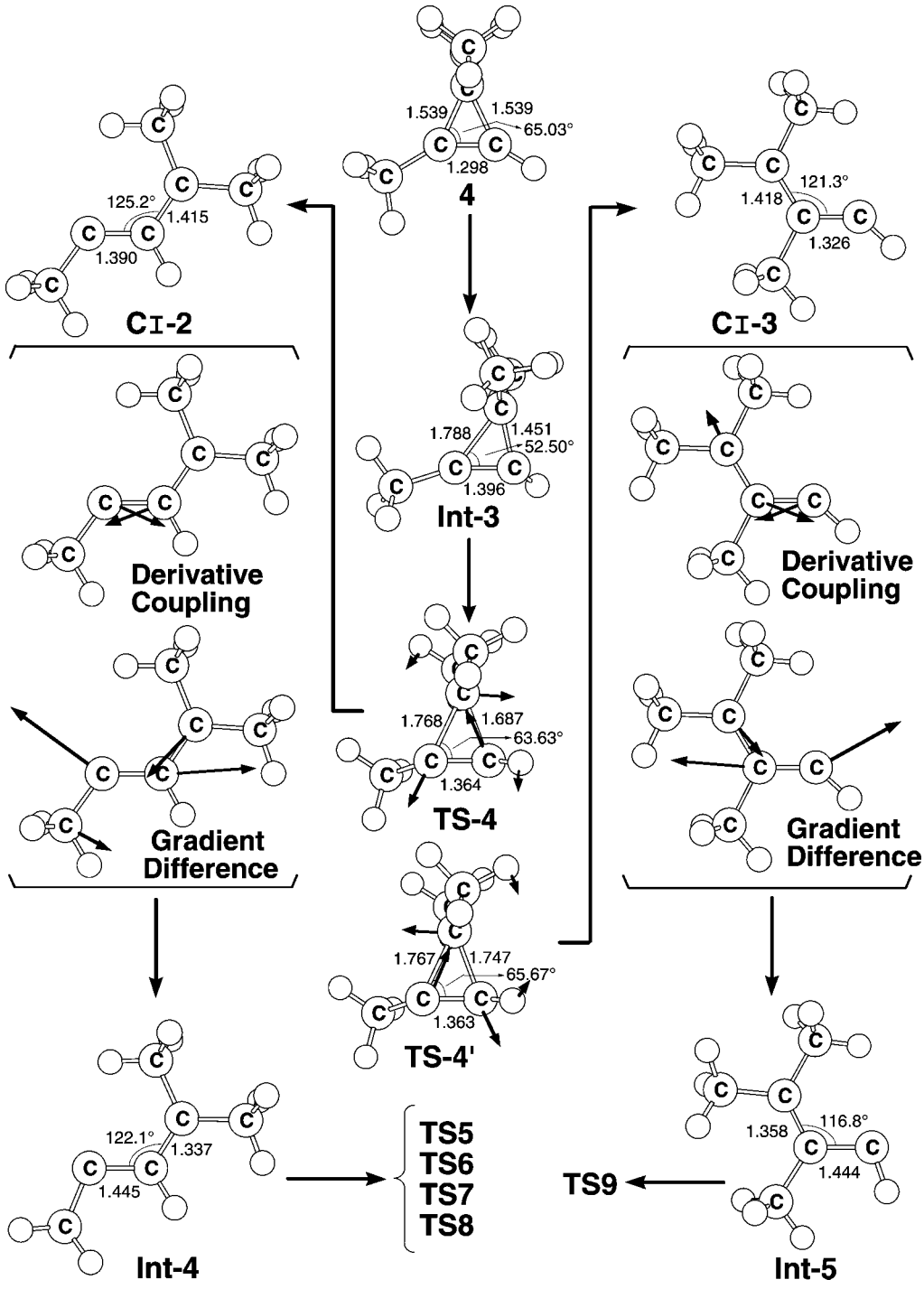


Figure 5. The CAS(6,6)/6−311G(d,p) geometries (in Å and deg) for path 5 and path 6 of 1,3,3-trimethylcyclopropene (**4**), conical intersection (**CI**), intermediate (**Int**), and transition state (**TS**). The derivative coupling and gradient difference vectors—those which lift the degeneracy— computed with CASSCF at the conical intersections **CI-3**. The corresponding CASSCF vectors are shown inset. Also, the heavy arrows indicate the main atomic motions in the transition state eigenvector. For more information see the Supporting Information.

rameters of **Int-3** can be found in Figure 5. As with the case of cyclopropene (**1**), our CASSCF results indicate that the **Int-3** structure has two short C−C bonds (1.420 and 1.428 Å) and one longer bond one (1.872 Å) compared with its closed shell singlet state. Again, this strongly implies that one C−C bond of the triangular ring (**Int-3**) is broken during the photochemical isomerization of **4**.

After the local minimum **Int-3**, we located two transition states **TS-4** and **TS-4′**, which are based on the model of reactant **4**. The optimized transition states structure (**TS-4** and **TS-4′**) along with the calculated transition vectors at the CASSCF level are given in Figure 5. The arrows indicate the direction in which the three carbon atoms in the 3-membered ring vibrate in the normal coordinate corresponding to the imaginary frequency (750*i* and 726*i* cm^{-1}). As seen in Figure 4 and Table 2, it is apparent that the transition structure (**TS-4**) for the ring opening process is lower in energy, by 54 kcal/mol, than the corresponding

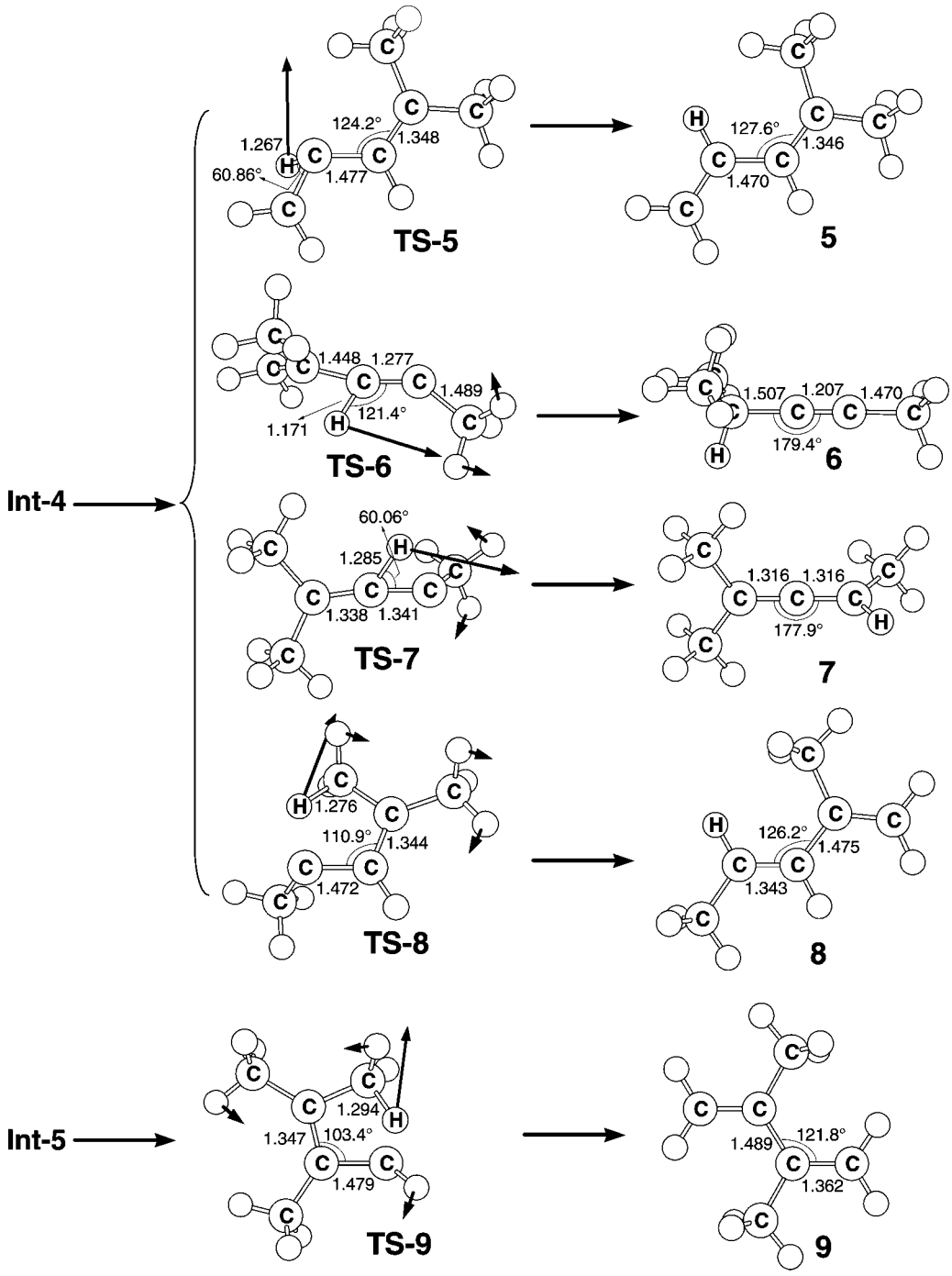


Figure 6. The CAS(6,6)/6−311G(d) geometries (in Å and deg) for path 7 of 1,3,3-trimethylcyclopropene (**4**), conical intersection (**CI**), and isomer product. The derivative coupling and gradient difference vectors—those which lift the degeneracy—computed with CASSCF at the conical intersections **CI-4**. The corresponding CASSCF vectors are shown inset. For more information see the Supporting Information.

FC-2 point but higher than the local intermediate **Int-3** by 10 kcal/mol. Accordingly, owing to the large excess energy (65 kcal/mol) obtained from the decay of **FC-2** to **Int-3**, it is expected that this relaxation energy is sufficient to drive the effective photoisomerization reactions (paths 5 and 6) for **4** (vide infra).

Furthermore, through the transition state **TS-4** (and **TS-4′**), the lowest energy point of the intersection seam of the S_0 and S_1 state was located for the ring opening process. This was identified as **CI-2** as presented in Figures 4 and 5. As suggested earlier, this can be equated to the more substituted vinylcarbene formed by cleavage of the C_1-C_2 cyclopropenyl bond. Then, funneling through the S_0/S_1 **CI-2** conical intersection can lead to two different reaction pathways on the ground-state surface via either the derivative coupling vector or the gradient difference vector.[10] The derivative coupling vector for **CI-2** corresponds to a C−C vibrating motion, which leads to a vibrationally hot species at the S_0 configuration. On the other hand, the gradient difference vector corresponds to a C−C bond stretching motion, which results in another planar intermediate **Int-4**. It should be noted here that although the energy is minimized,

Table 2. Energies (in kcal/mol) of the Critical Points Located along the Pathways 5-10 at the MP2-CAS(6,6)/6-311G(d,p)//CAS(6,6)/6-311G(d) and CAS(6,6)/6-311G(d) (in Parentheses) Levels of Theory[a]

structure	state	ΔErel[b]
1,3,3-trimethylcyclopropene **(4)**	S_0	0.0 (0.0)
FC-2	S_1	151.8 (172.9)
Int-3	S_1	87.51 (97.52)
TS-4	S_1	97.79 (107.1)
CI-2	S_1/S_0	46.67 (44.26)
Int-4	S_0	23.24 (26.19)
TS-5	S_0	35.75 (52.82)
5	S_0	−42.53 (−25.24)
TS-6	S_0	54.21 (65.82)
6	S_0	−32.82 (−31.00)
TS-7	S_0	39.63 (54.12)
7	S_0	−23.55 (−17.65)
TS-8	S_0	38.30 (62.64)
8	S_0	−34.96 (−31.43)
TS-4′	S_1	97.64 (106.3)
CI-3	S_1/S_0	61.93 (85.95)
Int-5	S_0	22.86 (50.74)
TS-9	S_0	61.14 (88.46)
9	S_0	−39.89 (−35.69)
TS-10	S_0	51.22 (49.47)
TS-11	S_0	62.41 (79.48)

[a] See Figures 5 and 6. [b] Energy relative to 1,3,3-trimethylcyclopropene (**4**).

the structure **CI-2** shown in Figures 4 and 5 is just a conical intersection rather than a true minimum,[10] because its energy gradient does not go to zero. Moreover the MP2-CAS computational results indicate that the energy of **CI-2** is 46 kcal/mol higher in energy than the ground-state minimum (**4**) and 105 kcal/mol lower than **FC-2**. The MP2-CAS results also demonstrate that the energy of **Int-4** is 129 kcal/mol lower than **FC-2** but 23 kcal/mol higher than that of the reactant **4**.

Again, after the local minimum **Int-4** on the S_0 potential energy surface is arrived at, there are four possible reaction pathways (**TS-5**, **TS-6**, **TS-7**, and **TS-8**) from which one may obtain the final photoproducts **5**, **6**, **7**,and **8**, respectively. Our MP2-CAS computations indicate that the barrier heights for these transition states from intermediate **Int-4** increase in the order **TS-5** (13 kcal/mol) < **TS-8** (15 kcal/mol) < **TS-7** (16 kcal/mol) < **TS-6** (31 kcal/mol). That is to say, our theoretical investigations are in reasonable agreement with the photoproduct distributions obtained experimentally.[6] Consequently, the calculations suggest that the reaction mechanisms for paths 5−8 should proceed as follows:

$$\text{Path 5: } \mathbf{4(S_0)} + h\nu \rightarrow \mathbf{FC-2} \rightarrow \mathbf{Int-3} \rightarrow \mathbf{TS-4} \rightarrow \mathbf{CI-2} \rightarrow \mathbf{Int-4} \rightarrow \mathbf{TS-5} \rightarrow \mathbf{5}$$

$$\text{Path 6: } \mathbf{4(S_0)} + h\nu \rightarrow \mathbf{FC-2} \rightarrow \mathbf{Int-3} \rightarrow \mathbf{TS-4} \rightarrow \mathbf{CI-2} \rightarrow \mathbf{Int-4} \rightarrow \mathbf{TS-6} \rightarrow \mathbf{6}$$

$$\text{Path 7: } \mathbf{4(S_0)} + h\nu \rightarrow \mathbf{FC-2} \rightarrow \mathbf{Int-3} \rightarrow \mathbf{TS-4} \rightarrow \mathbf{CI-2} \rightarrow \mathbf{Int-4} \rightarrow \mathbf{TS-7} \rightarrow \mathbf{7}$$

$$\text{Path 8: } \mathbf{4(S_0)} + h\nu \rightarrow \mathbf{FC-2} \rightarrow \mathbf{Int-3} \rightarrow \mathbf{TS-4} \rightarrow \mathbf{CI-2} \rightarrow \mathbf{Int-4} \rightarrow \mathbf{TS-8} \rightarrow \mathbf{8}$$

We have also explored the mechanism of path 9, which contains another conical intersection point (S_1/S_0 **CI-3**). As before, the structure of **CI-3** can be equated to the vinylcarbene derived from the cleavage of the less substituted cyclopropene σ-bond. The derivative coupling and gradient difference vectors obtained at the conical intersection are given in Figure 5. Our MP2-CAS results suggest that S_1/S_0 **CI-3** is lower in energy than **FC-2** by 129 kcal/mol but higher than the corresponding reactant **4** by 23 kcal/mol. The existence of a low-lying conical intersection provides a highly effective radiationless decay channel.[10] Besides this, the computations predict that the photochemical rearrangement reaction of path 9 should be the highest barrier process. That is, starting from the **FC-2** point, 1,3,3-trimethylcyclopropene (**4**) enters an extremely efficient decay channel, S_1/S_0 **CI-3**. After decay at this conical intersection point, this molecule will proceed via **Int-5** and **TS-9** points to reach the photoisomer **9**. Accordingly, this work suggests that the reaction mechanism for path 9 should be as follows:

$$\text{Path 9: } \mathbf{4(S_0)} + h\nu \rightarrow \mathbf{FC-2} \rightarrow \mathbf{Int-3} \rightarrow \mathbf{TS-4'} \rightarrow \mathbf{CI-2} \rightarrow \mathbf{Int-5} \rightarrow \mathbf{TS-9} \rightarrow \mathbf{9}$$

In addition, as shown in Figure 4, our computational results indicate that the two conical intersection points (i.e., **CI-2** and **CI-3**) are 47 and 62 kcal/mol lower in energy than **FC-2**, respectively. Competition between these reaction paths is presumably governed by the relative energies of the conical intersections involved. In consequence, the reaction paths through the **CI-2** point should be much more favorable through the **CI-3** point. Beside this, as discussed above, the reaction route through the **CI-2** point can lead to products **5**-**8** (path 5−8), while the reaction path through the **CI-3** point can only result in product **9** (path 9). Although these five routes (from path 5 to path 9) are competing with each other, only the last one is energetically unfeasible from a kinetic viewpoint. We would therefore expect a smaller quantum yield of **9** than of **5**-**8**. The energetic arguments are in qualitative agreement with the experimentally observed relative product distributions of **5**, **6**, **7**, **8**, and **9**, as already shown in eq 2.[6]

Before proceeding further, it should be pointed out that the first conical intersection (**CI-2**) is related to the more substituted vinylcarbene formed by cleavage of the cyclic (C_1-C_2) cyclopropenyl bond and producing the photoproducts **5**-**8**. Similarly, **CI-3** can be rationalized as arising from the less substituted vinylcarbene formed by cleavage of the cyclic C_1-C_3 bond and yielding photoproduct **9**. That is, the product distribution from photolysis is normally characteristic of reactivity associated with vinylcarbene intermediates, formed by cleavage of the C_1-C_3 or C_2-C_3 cyclopropene bonds in the first excited singlet state. As a result, our theoretical findings are in accordance with the predictions based on Fahie and Leigh's proposal.[6]

As in the case of cyclopropene (**1**), we also investigated the ground-state (thermal) potential surfaces of **4**, which are given in Figures 4 and 7. The search for transition states on the S_0 surface near the structures of S_1/S_0 **CI-2** gives **TS-10** and **TS-11**, which connect **4** and **8** and **4** and **9**, respectively. Our model calculations demonstrate that these transition states all involve a hydrogen migration with the calculated imaginary frequencies 148*i* and 152*i* cm^{-1}, respectively. In

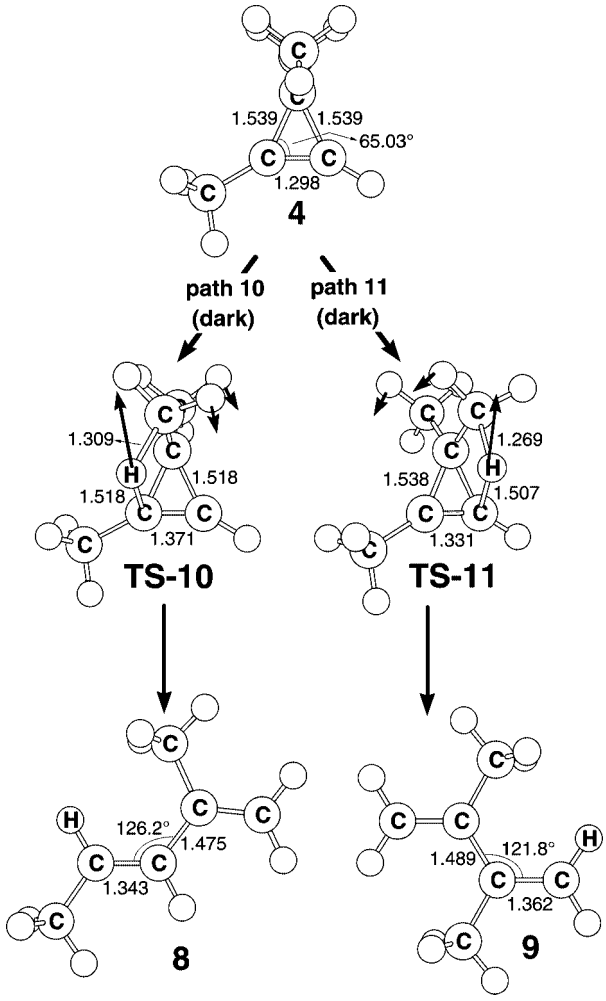


Figure 7. The CAS(6,6)/6−311G(d) geometries (in Å and deg) for path 8, path 9, and path 10 of spiro[2,4]hept-1-ene (**4**), transition state (**TS**), and isomer products in the dark (thermal) reactions. The heavy arrows indicate the main atomic motions in the transition state eigenvector. For more information see the Supporting Information.

consequence, our theoretical investigations suggest that the reaction mechanisms for paths 10 and 11 should proceed as follows:

$$\text{Path 10: } \mathbf{4} \rightarrow \mathbf{TS-10} \rightarrow \mathbf{8}$$

$$\text{Path 11: } \mathbf{4} \rightarrow \mathbf{TS-11} \rightarrow \mathbf{9}$$

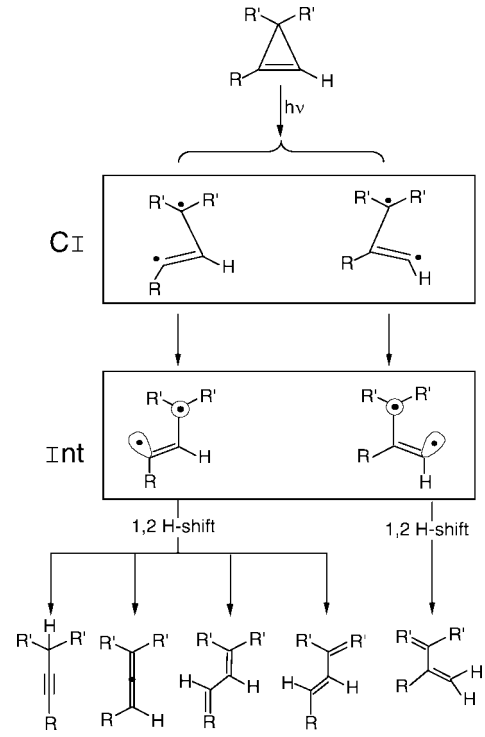
Additionally, our theoretical work indicates that the barrier heights for these two reaction pathways are 51 and 80 kcal/mol, respectively. This strongly implies that path 10 should be the preferred route for the thermal reactions of **4**. Thus product **8** should have the largest yield. Again, our theoretical findings are in good agreement with the available experimental work.[7]

V. Conclusion

Photochemical mechanisms of cyclopropene **1** (eq 1) and 1,3,3-trimethylcyclopropene **4** (eq 2) have been investigated in the present work. Taking both systems studied in this paper together, one can draw the following conclusions:

(1) We summarize the results of the experimental and theoretical investigations of the photochemistry of cyclopropenes disscussed in this work in Scheme 4. In this scheme, we show the role of vinylcarbene intermediates in the photochemical and thermal reactions of cyclopropene and its derivatives. That is, our computational results are consistent with the intermediacy of a vinylidene species formed by ring opening. According to our theoretical findings, the vinylcarbene is formed in the S_1 (σp) state from the first excited singlet state of cyclopropene. It can then decay to photoproducts on the S_0 surface by 1,2-H migration. Our theoretical observations for the cyclopropene hydrocarbons are in good accordance with Leigh's experimental findings.[3]

Scheme 4



(2) Both cyclopropene and its derivative have a similar photochemical as well as thermal potential energy profile. Moreover, our calculations indicate that the differences between their energy barriers are not so large that small steric effects on the transition states would control the reaction mechanism as in other analogous molecular systems.

(3) From the present results, we can elaborate on the standard model of the photochemistry of cyclopropene and of its derivatives. It is found that knowledge of the conical intersection of the cyclopropene species is of great importance in understanding its reaction mechanism since it can affect the driving force for photochemistry. That is to say, the conical intersections can efficiently funnel molecules from the $^1(\pi\pi^*)$-state to the ground-state surface. Also, these funnels, which are easily accessed once the $^1(\pi\pi^*)$ surface is populated, determine the reaction path taken on the ground-state surface.[10] Accordingly, these findings, based on the conical intersection viewpoint, have helped us to better understand the photochemical reactions and to support the experimental observations.[3,6,7]

The photochemistry of cyclopropenes is seen to be intriguing in the varied types of photochemical reaction

encountered. In particular, the subtle variation provides both a mechanistic challenge and a promise of reward. It is hoped that the present work can stimulate further research into this subject.

Acknowledgment. The author is grateful to the National Center for High-Performance Computing of Taiwan for generous amounts of computing time and the National Science Council of Taiwan for the financial support. The author also wishes to thank Professor Michael A. Robb, Dr. Michael J. Bearpark (University of London, U.K.), and Professor Massimo Olivucci (Universita degli Studi di Siena, Italy) for their encouragement and support. Special thanks are also due to Reviewer 1 and Reviewer 2 for very helpful suggestions and comments.

Supporting Information Available: Details for the optimized atomic coordinates and energies of compounds studied in this work (PDF format). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) For instance, see: (a) Greenberg, A.; Liebman, J. F. *Strained Organic Molecules*; Academic Press: New York, 1978. (b) Kobrich, G. *Angew. Chem., Int. Ed. Engl.* **1973**, *12*, 464. (c) De. Mayo, P. In *Rearrangements in Ground and Excited States*; Academic Press: New York, 1980, Vol. 3, p 501.

(2) For instance, see: (a) Padwa, A. *Org. Photochem.* **1979**, *4*, 261. (b) Padwa, A. *Acc. Chem. Res.* **1979**, *12*, 310. (c) Steinmetz, M. G.; Srinivasan, R.; Leigh, W. J. *Rev. Chem. Intermed.* **1984**, *5*, 57, and references cited therein.

(3) Leigh, W. J. *Chem. Rev.* **1993**, *93*, 487, and references cited therein.

(4) Chapman, O. L. *Pure Appl. Chem.* **1974**, *6*, 511.

(5) (a) Lifshitz, M.; Frenklach, M. J. *Phys. Chem.* **1975**, *79*, 1148. (b) Walsh, R. J. *Chem. Soc., Faraday Trans.* **1 1976**, *72*, 2137. (c) Bailey, I. M.; Walsh, R. J. *Chem. Soc., Faraday Trans.* **1 1978**, *74*, 1146.

(6) Fahie, B. J.; Leigh, W. J. *Can. J. Chem.* **1989**, *67*, 1859, and references cited therein.

(7) Srinivasan, R. J. *Chem. Soc., Chem. Commun.* **1971**, 1041.

(8) We will use the conical intersections to rationalize the mechanisms of the photoreactions of cyclopropenes. For details about conical intersections, see: Steinmetz, M. G.; Yen, Y.-P.; Poch, G. K. *J. Chem. Soc., Chem. Commun.* **1983**, 1504. (b) Reference 10. (c) Reference 11.

(9) It has been found that, according to many investigations of the reaction path of different organic reactions, in general, these processes are nonadiabatic; i.e., the path begins on an electronic excited state of the reactant species and ends on the ground-state energy surface where the product is formed. The central feature of such processes is that at some point along the excited-state reaction coordinate the system enters an efficiently decay channel, which takes the form of a conical intersection between the ground- and excited-state potential energy surfaces. Also see ref. (10).

(10) (a) Bernardi, F.; Olivucci, M.; Robb, M. A. *Isr. J. Chem.* **1993**, 265. (b) Klessinger, M. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 549. (c) Bernardi, F.; Olivucci, M.; Robb, M. A. *Chem. Soc. Rev.* **1996**, 321. (d) Bernardi, F.; Olivucci, M.; Robb, M. A. J. *Photochem. Photobiol. A: Chem.* **1997**, *105*, 365. (e) Klessinger, M. *Pure Appl. Chem.* **1997**, *69*, 773. (f) Klessinger, M.; MichlJ. In *Excited States and Photochemistry of Organic Molecules*;VCH Publishers: New York, 1995.

(11) (a) Olivucci, M.; Ragazos, I. N.; Bernardi, F.; Robb, M. A. *J. Am. Chem. Soc.* **1993**, *115*, 3710. (b) Bernardi, F.; Olivucci, M.; Ragazos, I. N.; Robb, M. A. *J. Am. Chem. Soc.* **1992**, *114*, 8211.

(12) 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, Liashenko, G.; Piskorz, A.; Komaromi, P. 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, Inc.: Pittsburgh, PA, 2003.

(13) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(14) Bearpark, M. J.; Robb, M. A.; Schlegel, H. B. *Chem. Phys. Lett.* **1994**, *223*, 269.

(15) (a) Gonzale, C.; Schlegel, H. B. J. *Chem. Phys.* **1989**, *90*, 2154. (b) Gonzale, C.; Schlegel, H. B. J. *Phys. Chem.* **1990**, *94*, 5523.

(16) McDouall, J. J. W.; Peasley, K.; Robb, M. A. *Chem. Phys. Lett.* **1988**, *148*, 183.

(17) (a) Steinmetz, M. G.; Srinivasan, R.; Leigh, W. J. *Rev. Chem. Intermed.* **1984**, *5*, 57, and references cited therein. (b) Padwa, A. *Acc. Chem. Res.* **1979**, *12*, 310. (c) Padwa, A. *Org. Photochem.* **1979**, *4*, 261.

(18) In addition to vinylcarbene formation from the first excited state, it was mentioned that other decay pathways (such as side-chain cleavage leading to free radicals) can intervene if the system bears appropriate substituents. See: (a) Reference 3. (b) Padwa, A.; Blacklock, T. J.; Cordova, D. M.; Loza, R. J. *Am. Chem. Soc.* **1980**, *102*, 5648.

(19) As pointed out by one referee, the geometry of intermediate **Int-1** is very interesting from the point of view of the chemical bond. In this three-membered ring species, two C—C bonds are tightly bonding (1.472 and 1.449 Å) and one C—C bond is nearly broken (1.827 Å). However, as discussed latter, this nearly broken C—C bond will be slightly closed together (about 1.714 or 1.771 Å) in the next excited transition state (**TS-1** or **TS-1′**) steps.

CT800078J

JCTC Journal of Chemical Theory and Computation

Computational Study of Tungsten(II)-Catalyzed Rearrangements of Norbornadiene

Allan L. L. East,* Greg M. Berner, Adam D. Morcom, and Lynn Mihichuk*

Department of Chemistry and Biochemistry, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

Received March 12, 2008

Abstract: Rearrangements of norbornadiene (NBD, C_7H_8) to various alkylidenes, via a hypothetical 7-coordinate tungsten(II) complex $W(CO)_3I_2$(NBD), were studied using density-functional theory computations. An extensive search for intermediates and transition states of rearrangement was made. The theoretical method (basis sets and level of DFT) used was justified by new benchmark studies which compare optimized structural parameters to those from crystal structures of several different tungsten complexes. Transition-metal-catalyzed rearrangements of NBD are not as well-known as those of norbornene and are considerably more complicated than had been thought. This work predicts a large variety of intermediates which may be feasible targets for experimental synthesis. All the rearrangement paths to alkylidenes found here feature high activation energies of over 45 kcal mol^{-1}, implying that self-initiation for the ring-opening metathesis polymerization of NBD via tungsten(II) complexes must occur via an alternative mechanism.

Introduction

Norbornadiene (NBD, a C_7H_8 isomer) can undergo ring-opening metathesis polymerization (ROMP) via homogeneous catalysis by Group 6 transition-metal complexes without the need of a preformed metal carbene.[1–5] The propagation (metathesis) step is shown in Figure 1, with a tungsten catalyst WL_x, and proceeds via a metallacyclobutane structure as originally postulated by Hérisson and Chauvin.[6]

An outstanding mystery is the nature of the NBD rearrangement on the "unprepared" (i.e., nonmetal-carbene) catalyst to generate the required initial carbene (alkylidene). In the simpler case of the ROMP of norbornene (NBE) on similar catalysts, the activation step is believed[5,7] to be a net 1,2-H-shift along the metal-coordinated C═C double bond, based on experimental evidence.[8] However, NBD offers other possibilities due to its extra degree of unsaturation, and experimental evidence is even less clear-cut.[3] The only computational quantum-chemistry applications of the NBD rearrangement mechanism to date are the exploratory density-functional-theory (DFT) ones of Handzlik, Szymańska-Buzar, and co-workers.[9,10] Their most recent paper concludes that there are many possible rearrangements, that an acceptable mechanism remains elusive, and that more study is required.[10] Our study, begun before ref 10 was in print, fortuitously anticipated this call.

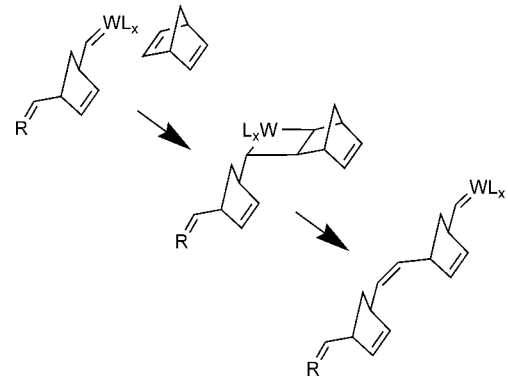


Figure 1. Propagation step in the ROMP of norbornadiene.

We set out to use DFT to thoroughly explore single-NBD rearrangements on an L_5W(NBD) complex, i.e. hypothesis category I of the four plausible ones presented in Figure 2. Our system of interest is $W(CO)I_2(\eta^2$-dppm$)(\eta^4$-NBD) (dppm

* Corresponding author e-mail: (A.L.L.E) allan.east@uregina.ca, (L.M.) lynn.mihichuk@uregina.ca.

10.1021/ct800088e CCC: $40.75
Published on Web 07/02/2008

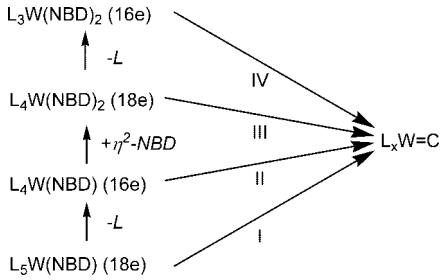


Figure 2. Four hypothesis categories for initial carbene generation, based on starting complex. Each category contains several hypothetical mechanisms to be explored; the present work studied category I.

$= Ph_2PCH_2PPh_2$), a known ROMP system,[11] but a complex far too large for thorough computational transition-state exploration for C_7H_8 rearrangement. Instead, we performed the transition-calculations with the analogous stereoisomer of the model complex $W(CO)_3I_2(\eta^4\text{-NBD})$. This "small-model approximation" is well justified for three reasons: (i) the qualitative mechanism for initial carbene generation should be fairly model-independent, since several Group 6 M-NBD complexes are capable of self-initiation,[1,2,10] (ii) our results contain a subset that qualitatively match the limited results of ref 10 for $Mo(CO)_3Cl(SnCl_3)(NBD)$ and $Mo(CH_3CN)(CO)_2Cl(SnCl_3)(NBD)$,[10] and (iii) our limited collection of computed intermediates for the dppm-containing system has been found using the smaller $W(CO)_3I_2(\eta^4\text{-NBD})$ system.

Our data are laid out as follows. First are data from several collected basis sets and levels of theory, which were benchmarked by using them to optimize the geometries of complexes for which there exists experimentally measured geometrical data: $W(CO)_6$,[12,13] *trans*-$W(CO)_4(\eta^2\text{-}C_2H_4)_2$,[14] $W(CO)_4(\eta^4\text{-NBD})$,[15] and the seven-coordinate tungsten(II) structure $W(CO)_3I_2(NCMe)_2$.[16] Second are data for NBD rearrangement in the model system $W(CO)_3I_2(\eta^4\text{-NBD})$: geometry-optimized energies and pathways for a large variety of transition states and intermediates that connect this reactant to various metal carbene intermediates, using the most successful theoretical method from the benchmarking study. Third is a brief data comparison to calculated data of other Group 6 M-NBD complexes, supporting our hypothesis that the DFT-approximated mechanism appears to be largely independent of the complex used.

Theoretical Methods

All calculations were performed using the Gaussian 03 software package.[17] Several levels of theory and basis sets were used in the preliminary benchmarking study. The theoretical method denoted BP86/RDZP:ITZ2DF:6−31G(d), as explained below, was chosen and used for the ensuing reaction-path study, and all energies are reported without zero-point corrections.

Levels of theory included the three density functional theory (DFT) approximations B3LYP,[18,19] BLYP,[19,20] and BP86[20,21] as well as the Møller−Plesset second-order perturbation theory (MP2),[22] and the expensive but very accurate coupled-cluster approximation CCSD(T).[23]

The basis set 6−31G(d)[17] was employed for all atoms except the large W and I atoms. Basis sets for W atoms all employed the Hay-Wadt small-core relativistic effective core potential,[24] which replaces all W core electrons up to 4*s*4*p*4*d*4*f*; hence, basis functions are needed for the 5*s* and 5*p* "core" as well as the 6*s*, 5*d*, and 6*p* valence orbitals. We tested the well-known Los Alamos LANL2MB[24] and LANL2DZ[17] basis sets as well as new even-tempered[25] sets developed by us at the University of Regina (RDZ, RDZP, RTZ, and RTZP, see the Appendix). The new sets were developed principally to add polarization functions but also to demonstrate that simple even-tempered basis sets work just as well as optimized ones for obtaining molecular geometries. For the I atom, three basis sets were used: (i) the LANL2DZ[17] basis set, with its complete-core relativistic effective core potential,[26] and two modifications proposed by Glukhovtsev et al.[27] which East and co-workers have used before, [28] (ii) LANL2DZ with added *d*- and *f*-orbital polarization functions (LANL2DZDF), and (iii) a triple-ζ uncontraction of the *s* and *p* sets with added 2*d* and 1*f* polarization sets (ITZ2DF).

Transition-state calculations were optimized for $W(CO)_3I_2NBD$ using the eigenvector−following transition-state algorithm (TS,EF).[29] However, since the TS,EF algorithm in G03 allows for simultaneous optimization of no more than 50 of the 66 geometry variables, several others were held fixed during the optimizations: carbonyl-group coordinates and/or H-atom coordinates of rigid H atoms in the NBD group. Each optimized transition state was subjected to a vibrational frequency analysis, to ensure that the magnitudes of the positive and one imaginary frequency were greater than the residual noise (the six "zero frequencies" for translations and rotations from normal mode diagonalization). Each one was also verified by two energy-minimization runs, begun from geometries displaced from the transition state (0.05 Å or 5−10°), which more definitively identifies the two intermediates that it connects. This precaution was essential, for we routinely discovered extra and unexpected intermediates in this manner.

Results and Discussion

Tungsten Basis Set Benchmarking. Four new and two traditional basis sets for tungsten orbitals were tested for their ability to reproduce experimental gas-phase[12] or crystal-phase[13] bond lengths of $W(CO)_6$ (Table 1).

Note from the experimental values that crystal-packing effects[30] reduce R_{WC} and R_{CO} by 0.033 and 0.003 Å, respectively, from their gas-phase values. Computed results for an isolated molecule should be properly compared with gas-phase results, and this comparison for values of R_{WC} reveals that the Regina double-ζ basis sets strike the best balance with B3LYP, but the LANL2DZ basis set performs better when paired with BP86 or the ab initio non-DFT method MP2. Note that a different choice, MP2/RDZP, performs best if one is instead interested in a level of theory that gives the best agreement with crystal-phase bond lengths.

The computed carbonyl R_{CO} values are hardly affected by the basis set choice for tungsten. These R_{CO} values are all

Table 1. Computed Bond Lengths (Å) and Percent Errors for $W(CO)_6$

level of theory	W basis set	W-CO (Å)	% error vs crystal	% error vs gas	C−O (Å)	% error vs crystal	% error vs gas
B3LYP	RDZ	2.060	1.7	0.1	1.151	0.5	0.3
	RDZP	2.058	1.6	0.0	1.151	0.5	0.3
	RTZ	2.069	2.2	0.5	1.152	0.6	0.3
	RTZP	2.068	2.1	0.5	1.152	0.6	0.3
	LANL2MB	2.071	2.3	0.6	1.150	0.4	0.2
	LANL2DZ	2.068	2.1	0.5	1.150	0.4	0.2
BLYP	RDZP	2.067	2.1	0.4	1.165	1.7	1.5
	RTZP	2.077	2.6	0.9	1.166	1.8	1.6
	LANL2DZ	2.076	2.5	0.9	1.165	1.7	1.5
BP86	RDZP	2.052	1.3	−0.3	1.164	1.7	1.4
	RTZP	2.062	1.8	0.2	1.165	1.7	1.5
	LANL2DZ	2.061	1.8	0.1	1.164	1.7	1.4
MP2	RDZP	2.033	0.4	−1.2	1.167	1.9	1.7
	RTZP	2.041	0.8	−0.8	1.168	2.0	1.7
	LANL2DZ	2.053	1.4	−0.2	1.166	1.8	1.6
CCSD(T)	RDZP	2.058	1.6	0.0	1.159	1.2	1.0
experiment[12]	(gas)	2.058			1.148		
experiment[13]	(crystal)	2.025(8)			1.145(15)		

systematically high, due both to the small basis set used for C and O (6−31G(d)) and the moderate level of theory. We also performed a high-level CCSD(T) optimization, which showed improvement in R_{CO} and the adequacy of the new RDZP basis set for tungsten.

Since we also desired a tungsten basis set appropriate for η^2 interactions of C═C double bonds with W, we further tested the RDZP, RTZP, and LANL2DZ sets on the R_{WC} distances to alkene carbons in *trans*-$W(CO)_4(\eta^2\text{-}C_2H_4)_2$ and $W(CO)_4(\eta^4$-NBD) (Figure 3).[14,15] The RDZP basis set outperformed the others with DFT methods, producing errors of less than 1.2% when used with either B3LYP or BP86.

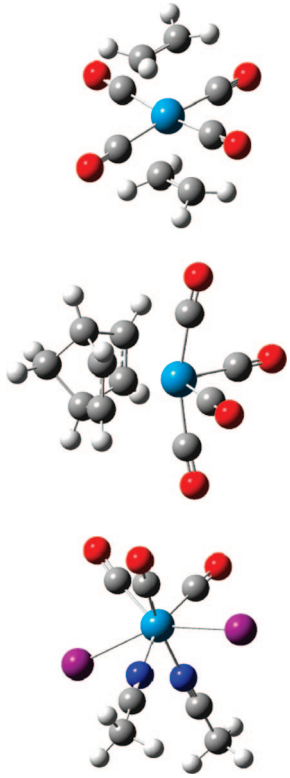


Figure 3. Structures of *trans*-$W(CO)_4(\eta^2\text{-}C_2H_4)_2$, $W(CO)_4(\eta^4$-NBD), and $W(CO)_3I_2(NCMe)_2$, optimized via BP86/RDZP:ITZ2DF:6−31G(d).

Iodine Basis Set Benchmarking. In order to test potential iodine basis sets in conjunction with the already established W basis sets and to extend our tests to a 7-coordinate complex, optimizations of the complex $W(CO)_3I_2(NCMe)_2$ (see Figure 3) were performed. The optimizations were conducted using the B3LYP, BP86, and MP2 levels of theory with the RDZP, RTZP, and LANL2DZ mixed basis sets for W each combined with the ITZ2DF, LANL2DZ, and LANL2DZDF basis sets for I. Computed bond lengths to W and percent error data for these computations are presented in Table 2. The experimental data[16] for $W(CO)_3I_2(NCMe)_2$ show two degenerate R_{WC} values ("long") and two degenerate R_{WN} values; hence, for comparisons, we averaged the near-degenerate computed values in these cases.

In comparing the results from optimizations using the LANL2DZ and LANL2DZDF basis sets for I it was found that the added *d* and *f* polarization functions significantly improved the accuracy. As a result it was decided to try yet another basis set for W, by adding *f*-polarization to LANL2DZ (LANL2DZF). These results are also shown in Table 2, and although the polarization functions did provide an improvement over the optimizations performed with LANL2DZ, it did not lead to the optimal method. The optimal method, which kept all R_{WX} percent errors (relative to experimental crystal-structure values) at or below 2.4%, was BP86/RDZP:ITZ2DF:6−31G(d), and this was the level of theory chosen for the rearrangement study.

$W(CO)_3I_2(\eta^4$-NBD) and Intermediates. We found over 40 different minimum-energy structures and speculate that over 100 probably exist. Broadly, these structures could be grouped in three classes: (i) ordinary π-complexes of stable alkenes, (ii) alkylidenes (carbenes), and (iii) C−H and C−C insertion (oxidative addition) compounds. Figure 4 provides summary sketches of the structures obtained, while full images (top and side views) are provided in the Supporting Information. Computed relative energies (no zero-point corrections) appear in Table 3.

The alkenes we considered were the NBD reactant, in η^4 (**1**) and η^2 (**2**) forms, and two vinylcyclopentadienes (**3**, **4**) which are closely related to the diene alkylidenes (**9**, **10**).

Table 2. Computed Bond Lengths (Å) and Percent Errors for $W(CO)_3I_2(NCMe)_2$ Using Various Methods and Heavy-Atom Basis Sets

W basis set	I basis set	W−I short	W−I long	W−CO short	W−CO long	W−N	max % error
			B3LYP				
RDZP	ITZ2DF	2.865	2.941	1.990	1.993	2.176	3.3
	LANL2DZDF	2.865	2.941	1.988	1.992	2.177	3.3
	LANL2DZ	2.916	2.962	1.982	2.002	2.173	4.3
RTZP	ITZ2DF	2.865	2.947	2.001	2.005	2.190	3.6
	LANL2DZDF[a]	2.864	2.946	2.001	2.005	2.191	3.5
	LANL2DZ[a]	2.888	2.978	2.000	2.005	2.188	4.6
LANL2DZ	ITZ2DF	2.882	2.963	2.003	2.007	2.197	4.1
	LANL2DZDF	2.887	2.969	2.003	2.006	2.197	4.4
	LANL2DZ	2.926	3.015	2.003	2.007	2.193	5.9
LANL2DZF	LANL2DZDF[a]	2.875	2.949	1.993	1.999	2.200	3.6
			BP86				
RDZP	ITZ2DF	2.845	2.915	1.986	1.989	2.152	2.4
	LANL2DZDF[a]	2.845	2.917	1.985	1.987	2.152	2.5
	LANL2DZ	2.877	2.956	1.983	1.986	2.150	3.9
RTZP	ITZ2DF	2.844	2.923	1.998	2.000	2.166	2.7
	LANL2DZDF	2.845	2.923	1.997	1.999	2.167	2.7
	LANL2DZ[a]	2.870	2.955	1.996	1.999	2.165	3.8
LANL2DZ	ITZ2DF[a]	2.860	2.941	1.999	2.000	2.173	3.3
	LANL2DZDF	2.866	2.946	1.999	2.000	2.173	3.5
	LANL2DZ	2.906	2.992	1.997	2.000	2.170	5.1
LANL2DZF	ITZ2DF	2.850	2.923	1.989	1.992	2.176	2.7
	LANL2DZDF	2.856	2.922	1.991	1.994	2.175	2.7
			MP2				
RDZP	ITZ2DF	2.777	2.821	1.960	1.965	2.146	3.3
	LANL2DZDF	2.769	2.813	1.957	1.963	2.148	3.4
	LANL2DZ	2.812	2.845	1.949	1.958	2.146	3.7
RTZP	ITZ2DF	2.773	2.824	1.969	1.974	2.154	2.9
	LANL2DZ	2.790	2.830	1.961	1.968	2.152	3.2
LANL2DZ	ITZ2DF	2.821	2.863	1.971	1.979	2.175	2.7
	LANL2DZDF	2.824	2.861	1.967	1.976	2.171	2.8
	LANL2DZ	2.880	2.910	1.962	1.971	2.163	3.1
experiment[16]	(crystal)	2.797	2.846	1.99	2.03	2.17	

[a] These computations were assessed converged; automatic convergence failed due to continually large displacements of nearly free methyl rotation modes.

The regular (**2a**) and inverted (**2b,c**) η^2-NBD complexes are 17 and 12−13 kcal mol^{-1} higher in energy than the η^4-NBD reference point **1**, at the BP86 level of DFT used here.

The alkylidene structures could be further subdivided into tricyclic no-ene (**5**), bicyclic monoene (**6**−**8**), and monocyclic diene (**9**, **10**) versions. The lowest-energy versions were **5** and **6**, being only 1 and 5−6 kcal mol^{-1} above **1**. The monocyclic diene alkylidenes **9** and **10** exhibited both ring-η^2 backbiting (**9a**, **9d, 9e, 10a**, **10b, 10e**) and nonbackbiting (**9b**, **9c**, **10c**, **10d, 10f**) forms, with the ring-η^2 interaction providing 3−6 kcal/mol of extra stabilization. One of our attempts at an alkylidene resulted in carbene insertion into a nearby carbonyl group, resulting in an η^2-ketene (**11**).

The C−H bond insertions resulted in various positions for the H atom, depending on the nearby ligands: a simple W−H bond (**12**), μ-bridging I and I (**13**), or μ-bridging W and C (**14**). In one case, the insertion resulted in new C−C bond formation between C_7H_7 and a carbonyl group in an acyl fashion (**15**).

The C−C bond insertions result in an exotic mix of structures having two connections from C_7H_8 to W, and the lowest-energy ones were quadricyclane-like dialkyl intermediates (**16a,b**) lying 11−13 kcal mol^{-1} above **1**, and three π-allyl (η^3-allyl) intermediates (**20a, 21b, 22a**) lying 9−12 kcal mol^{-1} above **1**.

Experimentally, none of these types of W-containing intermediates have been isolated. Tungsten(II)-alkylidene structures **5**−**10** and the C−C-insertion intermediates **16**−**23** would be highly reactive in the presence of excess hydrocarbons, and the C−H-insertion intermediates **12**−**15** would be highly reactive because they are high in energy. Ruthenium versions of the vinylcyclopentadiene π-complexes **3** and **4** have been observed as results of Ru-NBD rearrangements.[31] The π-allyl structures **20**−**23** might be experimentally isolable, since their energies are moderate, and organometallic complexes with π-allyl C_3H_5 ligands have been prepared.[32]

Rearrangements via C−H Bond Insertions. Results for C−H insertion routes appear in Figure 5. Routes to alkylidenes tended to involve H-atom transfer to the catalyst, followed by its back-transfer to a different carbon atom on the C_7H_7 moiety to produce bicyclic or tricyclic carbenes. However, Figure 5 shows that we never actually found a 2-step mechanism from the η^4-NBD structure **1**. Two-step mechanisms from η^2-NBD structures were found. The lowest-barrier paths to alkylidenes found were one-step ones: one from **1** to **6b** and one from **2b** to **6b**. All of the paths to alkylidenes here require rather high activation barriers of over 45 kcal mol^{-1}.

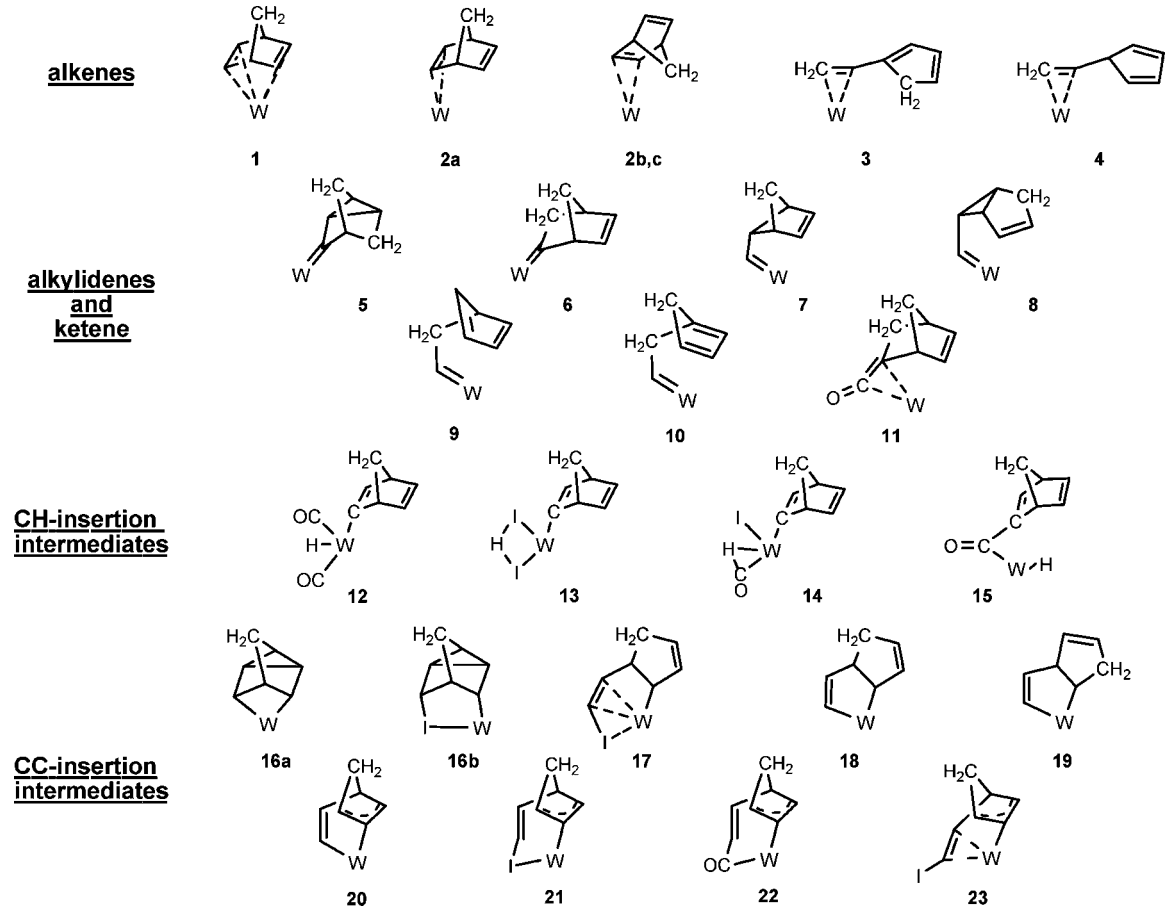


Figure 4. Structures found as intermediates (local minima) in this work.

Table 3. Computed Relative Energies (kcal mol^{-1}) for Intermediates of $WI_2(CO)_3(C_7H_8)$

structure	E	structure	E	structure	E	structure	E
1	0.0	**8b**	19.0	**10f**	11.7	**19**	43.4
2a	17.1	**9a**	15.7	**11**	22.6	**20a**	8.9
2b	12.2	**9b**	22.1	**12a**	34.8	**20b**	23.8
2c	13.4	**9c**	27.3	**12b**	34.1	**21a**	15.0
3	−2.6	**9d**	7.7	**13**	42.4	**21b**	12.0
4	11.8	**9e**	7.1	**14**	44.3	**22a**	10.2
5	1.1	**10a**	38.7	**15**	40.6	**22b**	15.8
6a	5.9	**10b**	14.9	**16a**	13.1	**23**	26.0
6b	5.4	**10c**	18.1	**16b**	11.5		
7	34.3	**10d**	19.2	**17**	33.8		
8a	16.5	**10e**	5.5	**18**	32.3		

The route we accidentally found to the ketene **11** was a surprise, as it came in one step from NBD without first forming an alkylidene. The barrier for this direct step is quite high, however, at 71 kcal mol^{-1}.

Rearrangements via C−C Bond Insertions. Results for C−C insertion routes appear in Figure 6. Four of the six routes to alkylidenes involved intermediates with σ-vinyl + π-allyl interactions to the catalyst (**20**−**22**). The six C−C insertion routes resulted in either monocyclic carbenes **9** and **10** or the peculiar bicyclic carbene **8a** with a 3-membered ring; this is unlike the four C−H insertion routes which led to different bi- and tricyclic alkylidenes.

The routes via π-allyl structures **20a** and **20b** were clean 2-step processes in which the intermediate featured both vinyl and π-allyl interactions with the tungsten atom. The vinyl arm of C_7H_8 was positioned between two like ligands in these cases: two carbonyl groups in **20a** and two iodine atoms in **20b**. When we tried putting this vinyl arm between unlike groups, the mechanism became complex: it began with a non-π-allyl structure **17** and then branched to three different structures (**21a**, **18,** or **23**), two of which we successfully connected to alkylidenes.

We would like to point out the similarities between resonant η^3 π-allyl structures (**20**−**23**) and localized η^1 non-π-allyl ones (**17**−**19**). The localization of the resonance in **21** can lead to **17** without any H shifts, and localization of the resonance in **20** can similarly lead to either **18** or **19**. However, we see indications that such interconversions, while geometrically simple, may have sizable energy barriers: we cite the example of **17** to **21a** and the absence of **18** or **19** when the mechanism passes through **20a** or **20b**.

The η^3 π-allyl intermediates led directly to alkylidenes with a backbiting η^2 interaction with W, the structures on the left side of the gray area in Figure 6. The route to **10** via 1,2-H-shift has an insurmountable barrier of 90 kcal/mol, and hence structure **9** (via 1,3-H-shift) is favored over **10**. We also pursued the additional step of removal of the backbiting interactions in alkylidenes 8−10 (right side of gray area). We found that this could be done in two ways: the cyclopentadiene ring could either flip or shear away from the W atom. From structure **10b**, the flip barrier to **10d** was 3 kcal mol^{-1} lower than the shear barrier to **10c**. We only pursued shear steps for the other cases: **8a** to **8b** and **9a** to **9b**.

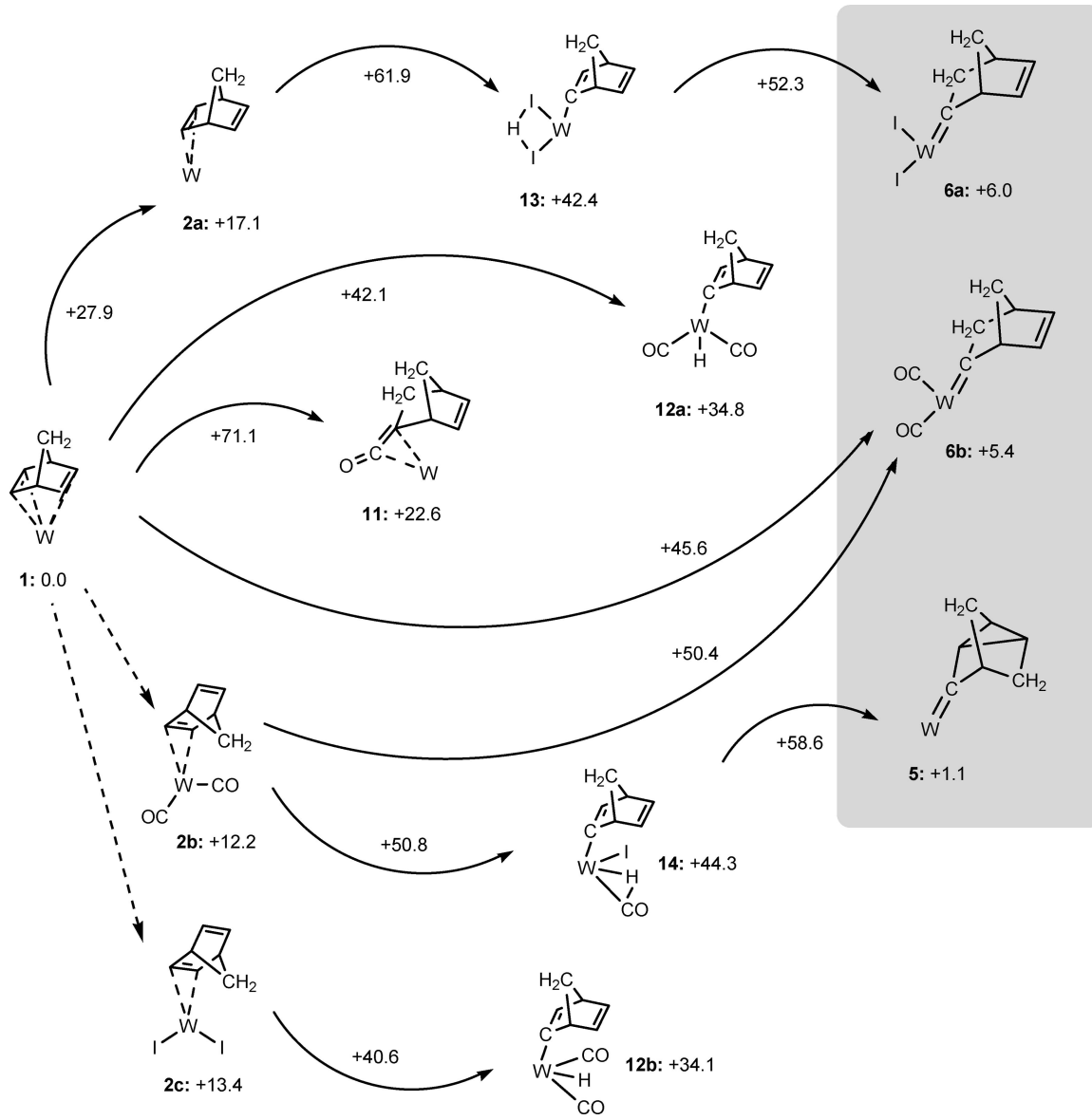


Figure 5. Reaction pathways via C−H insertion as found computationally. Alkylidene structures are in the gray area on the right. All energies are in kcal mol^{-1}; the ones alongside reaction arrows represent transition-state energies relative to **1** (i.e., not individual-step barrier heights).

Table 4. Overall Computed Energy Barriers (kcal mol^{-1}) for Selected Rearrangements of NBD to Alkylidenes via Group 6 Complexes

molecule	level of theory	**1** to **6**	**1** to **2?** to **6**	**1** to **16** to **8**	**1** to **17?** to **18** to **10**
$W(CO)_3I_2$(NBD), this work	BP86/mixed	45.6	50.4	53.5	61.9
$Mo(CO)_2$(MeCN)Cl($SnCl_3$)(NBD)[10]	B3LYP/LANL2DZ	49.2	52.3	47.6	59.5

All of our C−C insertion paths to alkylidenes also require rather high activation barriers, of over 50 kcal mol^{-1}.

Comparison to Results Using Other Group 6 M-NBD Complexes. Handzlik et al. recently published a limited computational study of NBD rearrangement, using B3LYP/LANL2DZ as the computational method and either $Mo(CO)_2$(MeCN)Cl($SnCl_3$)(NBD) or $Mo(CO)_3$Cl($SnCl_3$)-(NBD) as catalysts.[10] Of the single-NBD pathways they located, they present three (with 5 transition states) and discuss 3 others (with perhaps 4 transition states). Of these 6 paths, 4−6 appear to be among our 10 paths, hence strongly supporting our hypothesis that that mechanistic similarities should exist within a family of W(II) or Mo(II)-based catalysts.

The 4 paths of Handzlik et al. that are common to ours have similar large overall barrier heights, as shown in Table 4. Their single-NBD study was certainly not as thorough as ours, as we found three times as many transition states, and our intermediates **2**−**5**, **7**, **9**, **11**−**15**, **17**, and **19**−**23** are all new, with several being likely to exist with their Mo-based systems.

We have managed to obtain a BP86-optimized structure of the larger complex $W(CO)I_2(\eta^2$-dppm$)(\eta^4$-NBD) (dppm = $Ph_2PCH_2PPh_2$) as well as a few intermediate structures we

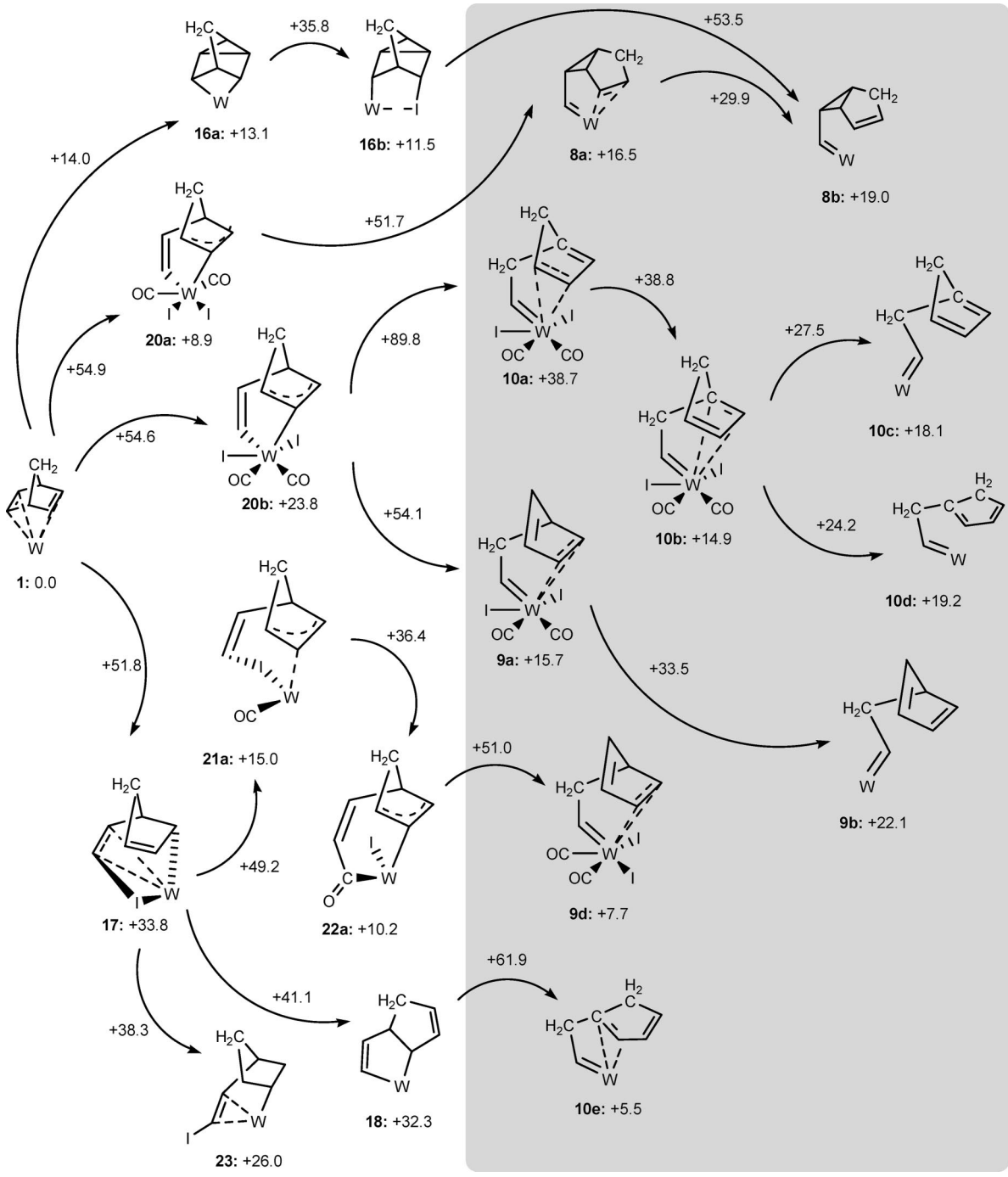


Figure 6. Reaction pathways via CC insertion as found computationally. Alkylidene structures are in the gray area on the right. All energies are in kcal mol^{-1}; the ones alongside reaction arrows represent transition-state energies relative to **1** (i.e., not individual-step barrier heights).

Table 5. Relative BP86 Energies (kcal mol^{-1}) of Intermediates of Two Tungsten(II) Systems

molecule	1	6a	34	8a	9b	10c	12a	16a
$W(CO)_3I_2(C_7H_8)$	0	6	34	17	22	18	35	13
$W(CO)I_2(\eta^2$-dppm$)(C_7H_8)$	0	14	40	29	19	11	46	24

obtained with the smaller $W(CO)_3I_2(\eta^4$-NBD) system. The same basis sets were used, except that STO-3G was used for the nonphosphorus atoms of dppm. No new intermediates were found with the larger system. Table 5 compares their relative energies. There is a general shift of 6−12 kcal mol^{-1}, likely due to steric effects of the bulky dppm ligand; the exceptions are in structures **9** and **10**, where the C_7H_8 moiety is more flexible and appears to benefit from π-π T-shaped stacking of its dangling cyclopentadiene ring with a phenyl group of dppm.

Conclusions

We make four main conclusions. (i) The BP86 DFT level of theory, with appropriate atomic basis sets, can reproduce crystalline organometallic R_{WX} bond lengths to within 2.4%. (ii) A large variety of intermediates should be considered as candidates in the rearrangement of a single norbornadiene with a tungsten(II) catalyst, and we hope that our Figure 4 will serve as a useful catalog for future studies. (iii) Generally, carbenes

Chart 1. W Atom, RDZP Set

```
W 0
S       3       1.00    0.000000000000
3.2400000000D+00        -0.5D+00
1.0800000000D+00        +0.9D+00
0.3600000000D+00        +0.5D+00
S       4       1.00    0.000000000000
3.2400000000D+00        +0.2D+00
1.0800000000D+00        -0.5D+00
0.3600000000D+00        -0.4D+00
0.1200000000D+00        +0.9D+00
S       1       1.00    0.000000000000
0.0400000000D+00        +1.0D+00
P       3       1.00    0.000000000000
3.2400000000D+00        -0.2D+00
1.0800000000D+00        +0.8D+00
0.3600000000D+00        +0.4D+00
P       2       1.00    0.000000000000
1.0800000000D+00        -0.3D+00
0.1200000000D+00        +0.3D+00
P       1       1.00    0.000000000000
0.0400000000D+00        +1.0D+00
D       2       1.00    0.000000000000
1.0800000000D+00        +0.3D+00
0.3600000000D+00        +0.5D+00
D       1       1.00    0.000000000000
0.1200000000D+00        +1.0D+00
F       1       1.00    0.000000000000
0.3600000000D+00        +1.0D+00
```

Chart 2. W Atom, RTZP Set

```
W 0
S       3       1.00    0.000000000000
3.2400000000D+00        -0.5D+00
1.0800000000D+00        +0.9D+00
0.3600000000D+00        +0.5D+00
S       3       1.00    0.000000000000
3.2400000000D+00        +0.2D+00
1.0800000000D+00        -0.5D+00
0.3600000000D+00        -0.4D+00
S       1       1.00    0.000000000000
0.1200000000D+00        +1.0D+00
S       1       1.00    0.000000000000
0.0400000000D+00        +1.0D+00
P       3       1.00    0.000000000000
3.2400000000D+00        -0.2D+00
1.0800000000D+00        +0.8D+00
0.3600000000D+00        +0.4D+00
P       1       1.00    0.000000000000
1.0800000000D+00        +1.0D+00
P       1       1.00    0.000000000000
0.1200000000D+00        +1.0D+00
P       1       1.00    0.000000000000
0.0400000000D+00        +1.0D+00
D       1       1.00    0.000000000000
1.0800000000D+00        +1.0D+00
D       1       1.00    0.000000000000
0.3600000000D+00        +1.0D+00
D       1       1.00    0.000000000000
0.1200000000D+00        +1.0D+00
F       1       1.00    0.000000000000
0.3600000000D+00        +1.0D+00
```

obtained via C−C bond insertion of NBD have (with one exception) a single cyclopentadiene ring, while carbenes obtained by C−H bond insertion of NBD remain bicyclic or become tricyclic. (iv) None of the rearrangement pathways found here had an activation barrier lower than 45 kcal mol^{-1}, suggesting that the ROMP of norbornadiene may *not* start with any of the rearrangements studied here.

On point (iv) we add the following remarks. Since our study dealt only with hypothesis category I (Figure 2), the other categories should be investigated. The exploratory calculation of a category III pathway by Handzlik et al.[10] produced an overall barrier of 35 kcal mol^{-1}, which is still high, but it is intriguingly lower than those of all the single-NBD rearrangements studied here. The ultimate goal of understanding the activation steps of ROMP will continue to benefit from additional careful computational research.

Acknowledgment. We thank the Natural Science and Engineering Research Council and the Canada Foundation for Innovation for funding and the Laboratory for Computational Discovery (Regina) for supercomputer maintenance. Dr. Szymańska-Buzar is thanked for alerting us to her latest manuscript while we were finishing our paper.

Appendix

The following Gaussian03-ready input files are for new sets of basis functions to represent the 5*s*, 6*s*, 5*p*, 6*p*, and 5*d* orbitals of W atom. The input files present columns of coefficients c_{ij} and exponents α_{ij} that define a contracted basis function Φ_i as a sum of primitive basis functions

$$\Phi_{ilm}(r,\theta,\varphi) = \sum_j c_{ij} e^{-\alpha_{ij} r^2} Y_l^m(\theta,\varphi)$$

where $Y_l^m(\theta,\phi)$ indicates a spherical harmonic function.

The first input file is for the RDZP set (see Chart 1): one contracted *s* function for the 5*s* orbital, a double-ζ pair of *s* functions for the 6*s* orbital, one contracted *p* function for each 5*p* orbital, a double-ζ pair for each 6*p* orbital, and a double-ζ pair for each 5*d* orbital. The final two lines represent the uncontracted set of polarization *f* functions, used in RDZP but removed for RDZ. The second input file is for the RTZP set (see Chart 2), with triple-ζ sets for the 6*s*, 6*p*, and 5*d* orbitals; again, the final uncontracted set of polarization *f* functions is used for RTZP but removed for RTZ.

These basis sets are based on a 5531 set of *spdf* primitive basis functions with very simple even-tempered exponents $\alpha_j = \alpha_{mid}\beta^{\pm j}$; our choice of $\alpha_{mid} = 0.36$ and $\beta = 3$ were made to span roughly the same exponent range as that of the LANL2DZ primitives. The coefficients c_{ij} in the contractions were chosen to mimic the Kohn−Sham atomic orbital coefficients observed in a B3LYP calculation for the ^{7}S state of W atom.

Supporting Information Available: Side and top views of each optimized intermediate of Table 3 and tables of Cartesian coordinates of all 30 transition states. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Amass, A. J.; McGourtey, T. A.; Tuck, C. N. *Eur. Polym. J.* **1976**, *12*, 93.

(2) Sen, A.; Thomas, R. R. *Organometallics* **1982**, *1*, 1251.

(3) (a) Szymańska-Buzar, T.; Głowiak, T.; Czeluśniak, I. *J. Organomet. Chem.* **2001**, *640*, 72. (b) Szymańska-Buzar, T.; Głowiak, T.; Czeluśniak, I. *Polyhedron* **2002**, *21*, 2505.

(4) Baker, P. K.; Drew, M. G. B.; Meehan, M. M.; Müller, J. *Z. Anorg. Allg. Chem.* **2002**, *628*, 1727.

(5) Bencze, L.; Bíró, N.; Szabó-Ravasz, B.; Mihichuk, L. *Can. J. Chem.* **2004**, *82*, 499.

(6) Hérisson, J.-L.; Chauvin, Y. *Makromol. Chem.* **1971**, *141*, 161.

(7) Górski, M.; Szymańska-Buzar, T. *J. Mol. Catal. A* **2006**, *257*, 41.

(8) (a) Bencze, L.; Kraut-Vass, A.; Prókai, L. *J. Chem. Soc., Chem. Commun.* **1985**, 911. (b) Veige, A. S.; Wolczanski, P. T.; Lobkovsky, E. B. *Angew. Chem., Int. Ed.* **2001**, *40*, 3629.

(9) Handzlik, J.; Górski, M.; Szymańska-Buzar, T. *J. Mol. Struct. THEOCHEM* **2005**, *718*, 191.

(10) Handzlik, J.; Stosur, M.; Kochel, A.; Szymańska-Buzar, T. *Inorg. Chim. Acta* **2008**, *502*, 41.

(11) Tosh, E. K. M.Sc. Thesis, University of Regina, Regina, Canada, 2006.

(12) Arnesen, S. V.; Seip, H. M. *Acta Chem. Scand.* **1966**, *20*, 2711.

(13) Heinemann, F.; Schmidt, H.; Peters, K.; Thiery, D. *Z. Kristallogr.* **1992**, *198*, 123.

(14) Szymańska-Buzar, T.; Kern, K.; Downs, A. J.; Greene, T. M.; Morris, L. J.; Parsons, S. *New J. Chem.* **1999**, *23*, 407.

(15) Grevels, F.-W.; Jacke, J.; Betz, P.; Krüger, C.; Tsay, Y.-H. *Organometallics* **1989**, *8*, 293.

(16) Drew, M. G. B.; Baker, P. K.; Armstrong, E. M.; Fraser, S. G.; Muldoon, D. J.; Lavery, A. J.; Shawcross, A. *Polyhedron* **1995**, *14*, 617.

(17) *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.; 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, Inc.: Wallingford, CT, 2004.

(18) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(19) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(20) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(21) (a) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406.

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

(23) Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 1697.

(24) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.

(25) (a) Bardo, R. D.; Ruedenberg, K. *J. Chem. Phys.* **1974**, *60*, 918. (b) Bardo, R. D.; Ruedenberg, K. *J. Chem. Phys.* **1974**, *60*, 932.

(26) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284.

(27) Glukhotsev, M. N.; Pross, A.; McGrath, M. P.; Radom, L. *J. Chem. Phys.* **1995**, *103*, 1878.

(28) Hepperle, S. S.; Li, Q.; East, A. L. L. *J. Phys. Chem. A* **2005**, *109*, 10975.

(29) Baker, J. *J. Comput. Chem.* **1986**, *7*, 385.

(30) Szilagyi, R. K.; Frenking, G. *Organometallics* **1997**, *16*, 4807.

(31) (a) Suzuki, H.; Kakigano, T.; Fukui, H.; Tanaka, M.; Morooka, Y. *J. Organomet. Chem.* **1994**, *473*, 295. (b) Jia, G.; Ng, W. S.; Lau, C. P. *Organometallics* **1998**, *17*, 4538.

(32) (a) McClellan, W. R.; Hoehn, H. H.; Cripps, H. N.; Muetterties, E. L.; Howk, B. W. *J. Am. Chem. Soc.* **1961**, *83*, 1601. (b) Shigeyoshi, S.; Takeuchi, K.; Sugimoto, M. *Organometallics* **1997**, *16*, 2995.

CT800088E

JCTC Journal of Chemical Theory and Computation

Theoretical Investigation of the Mechanism of Acid-Catalyzed Oxygenation of a Pd(II)-Hydride To Produce a Pd(II)-Hydroperoxide

Sugata Chowdhury, Ivan Rivalta, Nino Russo,* and Emilia Sicilia

Dipartimento di Chimica and Centro di Calcolo ad Alte Prestazioni per Elaborazioni Parallele e Distribuite-Centro d'Eccellenza MURST, Università della Calabria, I-87030 Arcavacata di Rende, Italy

Received April 29, 2008; Revised Manuscript Received June 6, 2008; Accepted June 9, 2008

Abstract: Density Functional Theory (DFT) has been applied to a comprehensive mechanistic study of the conversion reaction of the Pd(II)-hydride complex, $(IMe)_2(RCO_2)PdH$ (R=CH_3, Ph, and *p*-$O_2NC_6H_4$), to the corresponding Pd(II)-hydroperoxide in the presence of molecular oxygen. The calculations have evaluated the two mechanistic proposed alternatives, that are both considered viable on the basis of current data, of slow RCO_2H reductive elimination followed by oxygenation (*Path A*) and direct O_2 insertion (*Path B*). Results suggest that the mechanism of direct insertion of molecular oxygen into the Pd−H bond of the initial complex is energetically preferred. The activation energy relative to the rate-determining step of *Path A*, indeed, is calculated to be lower than the activation energy of the rate determining step of the alternative *Path B*, whatever ligand (CH_3CO_2, Ph, CO_2, *p*-$O_2NC_6H_4CO_2$) is coordinated to the Pd center. The calculated free activation energy of the rate-determining hydrogen abstraction step ($\Delta G^* =$ 24.8 kcal/mol) in the case of the oxygenation reaction of the benzoate-ligated Pd(II)-hydride complex is in very good agreement with the experimentally determined value of 24.4 kcal/mol. In addition, according to the experimentally detected enhancement of the reaction rate due to the presence of a nitro group on the benzoate ligand, our calculations show that the transition state for the hydrogen atom abstraction by molecular oxygen along the pathway for the oxygenation reaction of $(IMe)_2$(*p*-$O_2NC_6H_4CO_2$)PdH lies lower in energy with respect to the analogous transition state calculated for R=Ph.

1. Introduction

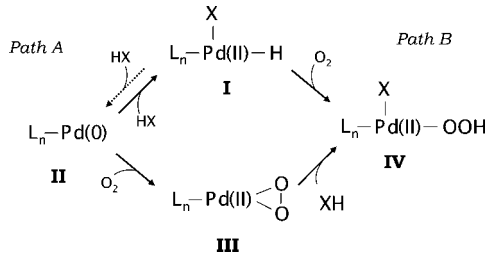
The use of molecular oxygen for the oxidative functionalization of organic molecules is a highly attractive option because O_2 is a readily available and nontoxic reagent. Additionally, a common byproduct of aerobic oxidations is the innocuous substance water. However, uncatalyzed chemical reactions between molecular oxygen and organic substrates generally result in complete combustion of the starting materials, and, consequently, the synthetic advantages these reactions possess cannot be exploited unless catalysts are used. The homogeneously Pd-catalyzed oxidations have emerged as a particularly promising reaction type for selective and efficient aerobic oxidation[1–7] even if the development of new and more efficient catalytic strategies has been hampered by an inadequate understanding of how O_2 interacts with the Pd center.[7] Two are the pathways proposed to explain how the oxygen activation in palladium-catalyzed oxidations proceeds (Scheme 1).

One pathway (*Path A*) is the formation of a Pd(0) species **II** through the reductive elimination of a HX species from the Pd(II)-hydride complex followed by the oxygenation of the Pd(0) center to form a η^2-peroxo Pd(II) species **III**. Upon protonation of the peroxopalladium(II) intermediate a pal-

* Corresponding author fax: +39-0984-492044; e-mail: nrusso@unical.it.

10.1021/ct8001442 CCC: $40.75
Published on Web 07/16/2008

Scheme 1



ladium hydroperoxide species **IV** is formed.[8–10] The second mechanistic possibility[11–13] (*Path B*) involves the direct insertion of molecular oxygen into a palladium(II)-hydride intermediate **I** to form the same palladium hydroperoxide species **IV** (Scheme 1) and, therefore, avoiding to proceed through Pd(0). The former mechanism has substantial experimental support even if the reaction between Pd(0) and triplet oxygen is spin forbidden.[8–10] Landis, Stahl, and co-workers have investigated this aspect performing a DFT theoretical study that showed the exothermic formation of the singlet Pd(II)-peroxido species requiring spin crossing between triplet and singlet surfaces mediated by spin−orbit coupling.[14]

The first observation of what can be assumed to be a direct insertion of molecular oxygen into a Pd(II)-hydride bond to form a hydroperoxopalladium complex has been reported by Goldberg, Kemp, and co-workers for a Pd complex that, due to the nature of the ligand, cannot undergo reductive elimination.[15] The authors report that benzene solutions of ($^{t\mathrm{Bu}}$PCP)PdH ($^{t\mathrm{Bu}}$PCP=1,3-$(CH_2P^tBu_2)_2C_6H_3]^-$) exposed to O_2 cleanly react to yield ($^{t\mathrm{Bu}}$PCP)PdOOH, which is relatively stable at ambient temperature as a solid and structurally characterizable. A DFT theoretical investigation, carried out by Goddard and co-workers, of the involved mechanism has recently appeared in literature that demonstrates how an insertion mechanism is possible and plausible and supports the hypothesis that the hydrogen atom abstraction is the key step of the reaction.[16] Namely, hydrogen atom is abstracted from the Pd center by molecular oxygen to form a HOO fragment that interacts only weakly with the Pd(I) center. Rotation of HO_2 moiety enables formation of a Pd−O bond to give the final Pd(II)-hydroperoxo product. A minimum energy crossing point (MECP) between triplet and singlet surfaces has been located at the exit channel of the reaction, whereas the mechanism for the rearrangement of the HOO fragment to give the hydroperoxide product has not been individuated. In the same work is discussed the difference with respect to a previous study of the same authors that underlined the subordination of the O_2 insertion process to the presence of a H-bond acceptor *cis* to the hydride.[17]

Stahl and co-workers have investigated the feasibility of the Pd(0) direct oxygenation pathway by subjecting [Pd(0)$(IMes)_2$] (IMes=Mesityl) to one equivalent of a carboxylic acid (acetic, benzoic, and *p*-nitrobenzoic) to produce the *trans*-[$(IMes)_2(RCO_2)$PdH] hydride, which subsequently exposed to O_2 yielded the corresponding hydroperoxide complex.[18] Since the experimental findings could support both mechanisms, that is direct insertion of O_2 into the Pd(II)-H bond and oxygenation of Pd(0), both pathways have been considered viable by the authors. Indeed, four possible alternative pathways have been computationally examined by Popp and Stahl to probe the reaction mechanism of the formal insertion of molecular oxygen into the Pd−H bond of *trans*-[PdH(OAc)$(IMes)_2$].[19] The main conclusion of this investigation is that the energetically preferred pathways, exhibiting very similar kinetic barriers, are those corresponding to (a) hydrogen atom abstraction from the Pd−H bond by molecular oxygen and (b) reductive elimination AcOH followed by oxygenation of Pd(0) and protonolysis of the formed η^2-peroxo-Pd(II) species. Also in this last case the proposed steps of the hydrogen abstraction mechanism are as follows: formation of a HOO moiety as a consequence of the hydrogen atom abstraction by molecular oxygen and formation of the final hydroperoxo product through rearrangement of the HO_2 fragment. In analogy with Goddard's analysis[16] the MECP between singlet and triplet PESs has been located at the exit channel of the reaction, but not one of the surmised mechanisms for the rearrangement of the HO_2 fragment has been confirmed.

In the framework of a more extended project aiming to unravel the mechanistic details of the selective oxidation of organic molecules by molecular oxygen in the presence of metal catalysts we have theoretically investigated the direct O_2 insertion mechanism for both ($^{t\mathrm{Bu}}$PCP)PdH and [$(IMes)_2$(AcO)PdH] hydrides.[20,21] With respect to the preceding computational analyses carried out on the same subject[16,19] the new results are as follows: the recognition of a different pathway for the process and the determination of the rearrangement mechanism of the OOH fragment as the last step of the overall hydride oxygenation reaction. In the present work we have examined the possible alternative mechanisms of the oxygenation process of the Pd-hydride complex *trans*-[Pd(H)$(O_2CR)(IMes)_2$] with R=CH_3, Ph, *p*-$O_2NC_6H_4$ to yield the corresponding hydroperoxide, taking into account also the influence on the oxygenation pathways of the employed carboxylic acid. Kinetic measurements,[18] indeed, have shown that the reaction rate exhibits a clear first-order dependence on the concentration of the Pd-hydride, and the presence of an electron-withdrawing nitro group on the benzoate ligand significantly enhances the reaction rate (when R=*p*-$O_2NC_6H_4$ the oxygenation is about eight times faster than when R=Ph). The ligand dependence could suggest that carboxylate dissociation is the rate determining step of the process and, as a consequence, supports the reaction mechanism involving reductive elimination of the carboxylic acid. Since many questions concerning the mechanistic details of the Pd-catalyzed oxidation reactions are still open,[22] the results of our computations can contribute to the determination of the true pathway for the interaction between molecular oxygen and Pd(II)-H complexes.

2. Computational Details

Geometry optimizations as well as frequency calculations for all the reactants, intermediates, products, and transition states have been performed at the Density Functional level of theory, employing the Becke's three-parameter hybrid functional[23] combined with the Lee, Yang, and Parr (LYP)[24] correlation functional, denoted as B3LYP, as implemented

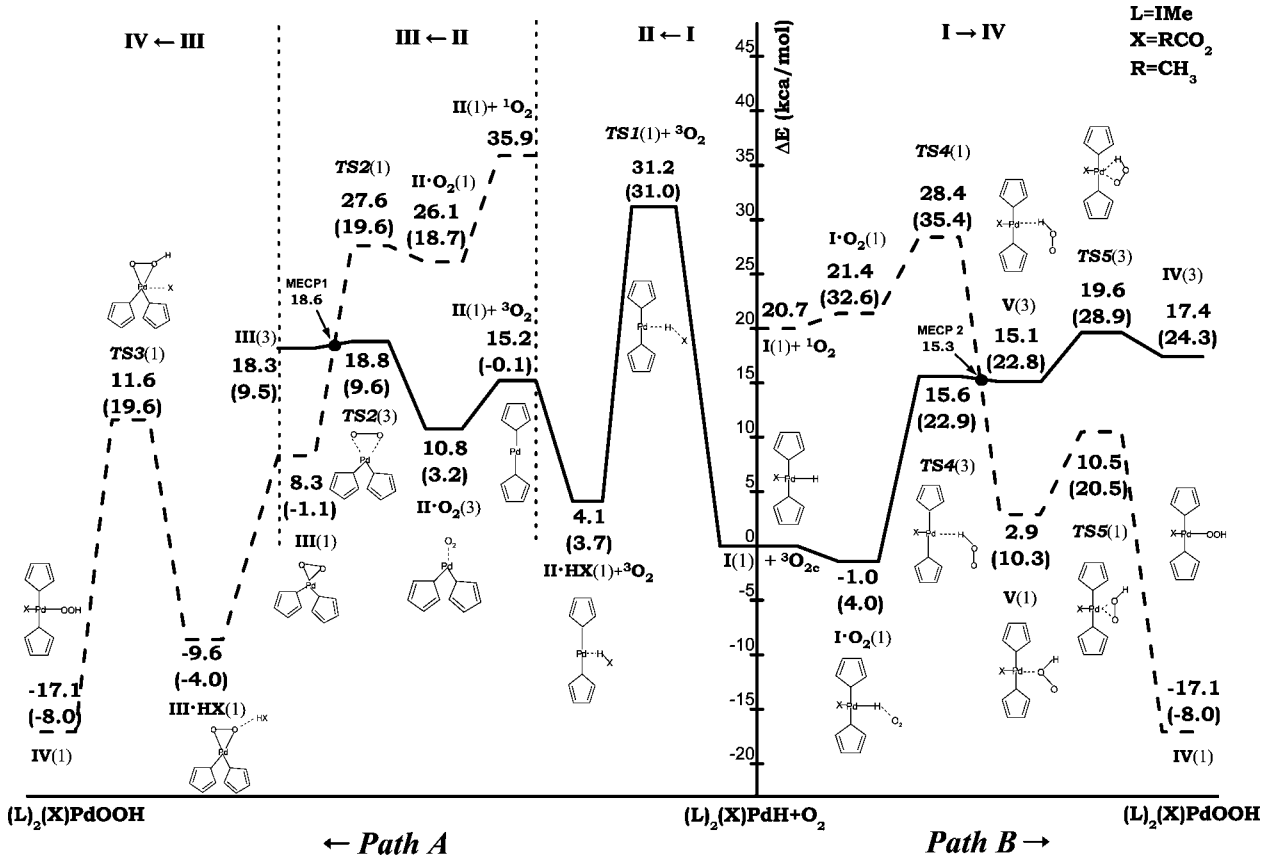


Figure 1. Calculated B3LYP PESs for the oxygenation reaction of $(IMe)_2(CH_3CO_2)Pd(II)$-H to give $(IMe)_2(CH_3CO_2)Pd(II)$-OOH. Gibbs free energies changes at 298.15 K in benzene are also reported in parentheses. Energies are in kcal/mol and relative to ground-state reactants. On the left the pathway that involves reductive elimination of CH_3COOH (*Path A*) and on the right the pathway for the direct O_2 insertion (*Path B*) are reported.

in Gaussian03 code.[25] For Pd the relativistic compact Stuttgart/Dresden effective core potential[26] has been used in conjunction with its split valence basis set. The 6−311G* basis sets of Pople and co-workers have been employed for the rest of the atoms. For each optimized stationary point vibrational analysis has been performed to determine its character (minimum or saddle point), and zero-point vibrational energy (ZPVE) corrections were included in all relative energies (ΔE). For transition states it was carefully checked that the vibrational mode associated with the imaginary frequency corresponded to the correct movement of involved atoms. Furthermore, the intrinsic reaction coordinate (IRC)[27,28] method has been used to assess that the localized TSs correctly connect to the corresponding minima along the imaginary mode of vibration.

To speed up calculations the 2,4,6-trimethylphenyl groups of the IMes ligands have been substituted with methyl ones to obtain a palladium hydride complex that, although less sterically hindered, can give analogous results.[19] For this reason, from now on we will indicate the 1,3-di(methyl)imidazoline-2-ylidene model of the ligand as IMe, and the starting points for our study are the *trans*[(RCO_2)(1,3-di(methyl)imidazoline-2-ylidene)$_2$PdH] species (R=CH_3, Ph, *p*-$O_2NC_6H_4$), also abbreviated as *trans*[$(RCO_2)(IMe)_2$PdH].

Both triplet and singlet reaction paths have been examined and for all the studied species $<S^2>$ values have been checked to assess whether spin contamination can influence the quality of the results. For triplet state structures no significant contamination was found by unrestricted calculations. Unrestricted calculations, instead, revealed in some cases triplet spin contamination corresponding to $<S^2>$ values close to 1.0. For molecular oxygen, due to the contamination of the singlet wave function with the triplet state, a highly stable singlet $^1\Delta_g$ state has been obtained corresponding to an excitation energy of 10.5 kcal/mol, compared with the experimental value of 22.5 kcal/mol.[29] Adopting the method proposed by Ovchinnikov and Labanowski[30] for correcting the mixed spin energies and removing the foreign spin components, a triplet-singlet energy gap of 20.7 kcal/mol was obtained, in very good agreement with the experimental value. The same scheme was adopted to correct the triplet contaminated energies of some structures along the singlet path, as will be underlined in the next paragraphs.

Solvent effects, both electrostatic and nonelectrostatic, have been treated implicitly making use of the Tomasi's integral equation formalism for the polarizable continuum model (IEF-PCM)[31,32] as implemented in Gaussian03. All stationary points structures obtained from vacuum calculations were reoptimized in implicit benzene with the above IEF-PCM method.

The solvation Gibbs free energies have been evaluated using the well-known thermodynamic cycle,[33] where the reaction Gibbs free energy in solution, ΔG_{sol}, is calculated for each process as the sum of two contributions: a gas-

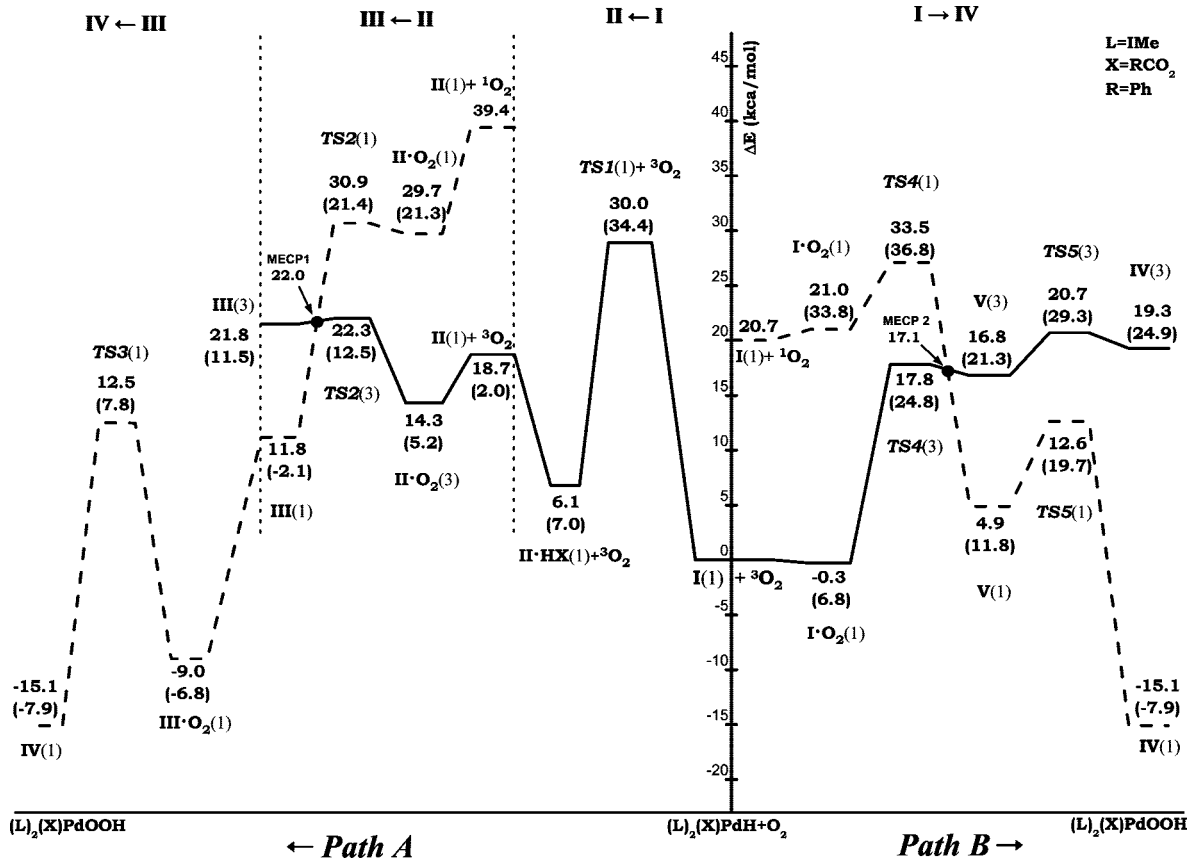


Figure 2. Calculated B3LYP PESs for the oxygenation reaction of $(IMe)_2(PhCO_2)Pd(II)$-H to give $(IMe)_2(PhCO_2)Pd(II)$-OOH. Gibbs free energies changes at 298.15 K in benzene are also reported in parentheses. Energies are in kcal/mol and relative to the ground-state reactants. On the left the pathway that involves reductive elimination of PhCOOH (*Path A*) and on the left the pathway for the direct O_2 insertion (*Path B*) are reported.

phase reaction free energy, ΔG_{gas}, and a solvation reaction free energy term calculated with the continuum approach, ΔG_{solv}:

$$\Delta G_{sol} = \Delta G_{gas} + \Delta G_{solv} \quad (1)$$

The gas-phase reaction free energy is the sum of two parts: electronic plus nuclear repulsion energy (ΔE_{ele}) and thermal contribution including zero-point energy ($\Delta G_{gas} = \Delta H + T\Delta S$). The last term, $T\Delta S$, that is the thermal correction, is evaluated using the calculated quantum mechanical vibrational frequencies.

To locate the minimum energy crossing points (MECP) between the triplet surfaces of reactants and the singlet ones of the products the methodology introduced by Harvey and co-workers has been used.[34]

3. Results and Discussion

Calculated B3LYP PESs for the oxygenation process of $(RCO_2)(IMe)_2PdH$ hydrides are drawn in Figures 1–3 for R=CH_3, Ph, *p*-$O_2NC_6H_4$, respectively. Both singlet and triplet PESs have been computed. Relative energies are calculated with respect to the ground-state reactants asymptote (**I**(1)+$^3O_2$). The energy profile for the steps of the pathway that, starting from the hydride species **I**, involves reductive elimination of RCOOH (R=CH_3, Ph, *p*-$O_2NC_6H_4$) to yield Pd(0) complex (**II**), oxygenation of **II** to yield the peroxo complex **III**, and subsequent protonation to form the hydroperoxide complex **IV** are shown on the left side of each figure (*Path A*). The calculated PESs for the direct insertion of O_2 into the Pd−H bond of the hydride species to obtain the same hydroperoxides are drawn on the right side of each figure (*Path B*). Relative free energies calculated in benzene solvent are also reported in Figures 1–3.

The ground-state optimized structure of the reference hydride species chosen as the starting point for our study is shown in Figure 4 along with the optimized structure of the final hydroperoxide complex in the case of R=Ph. Ground-state geometrical structures of all stationary points intercepted along the pathway for the $(PhCO_2)(IMe)_2PdH$ hydride oxygenation through reductive elimination, Pd(0) oxygenation, and protonolysis of peroxo complex are reported in Figure 5, whereas ground-state optimized structures of the sequence of minima and transition states involved by the direct O_2 insertion pathway are shown in Figure 6. In Figures 5 and 6 are sketched also the structures of the calculated MECPs. Labels employed in Figures 4–6 to individuate geometrical parameters have to be used to read information reported in Table 1 for *Path A* and Table 2 for *Path B*, which collect selected bond lengths and angles along with dihedral angles of stationary points and MECPs for R=CH_3, Ph, and *p*-$O_2NC_6H_4$.

In the next two paragraphs we are going to illustrate the outcomes of our calculations for both *Path A* and *Path B* in the case of R=CH_3. Subsequently, the results of our DFT

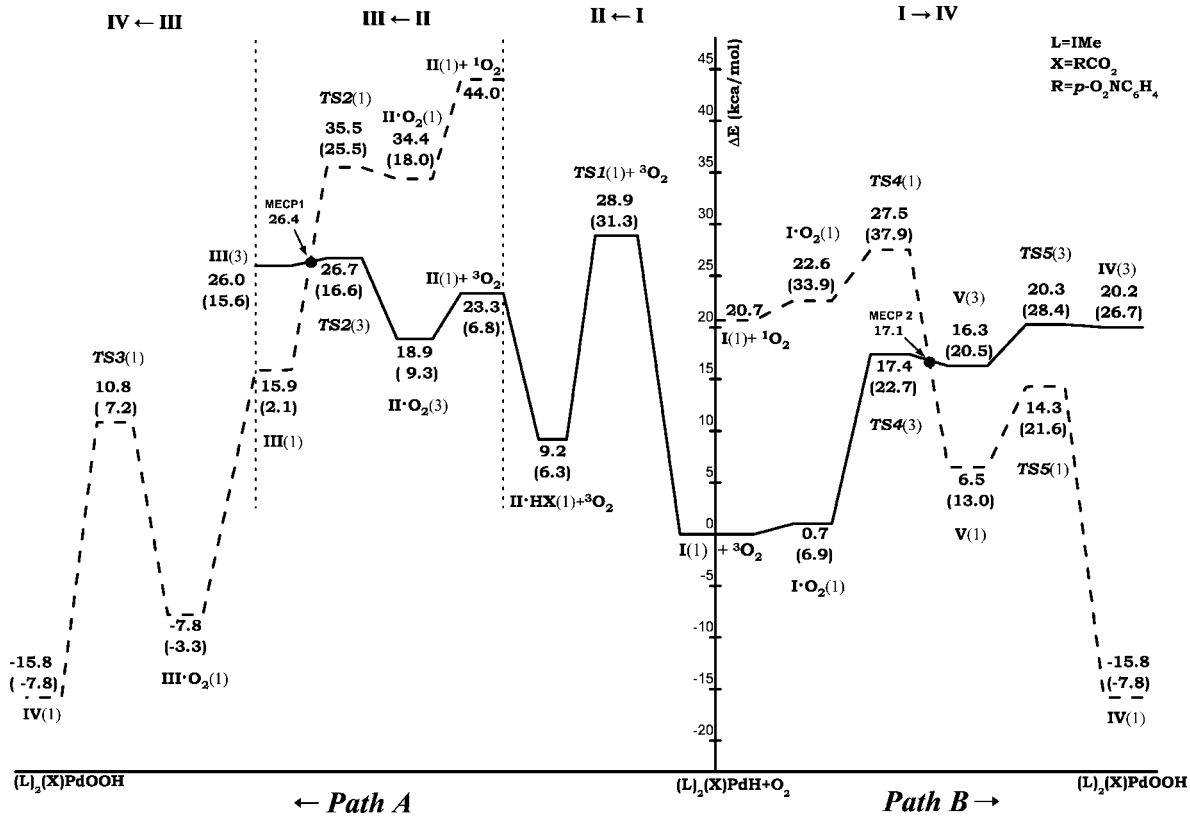


Figure 3. Calculated B3LYP PESs for the oxygenation reaction of $(IMe)_2(p\text{-}O_2NC_6H_5CO_2)Pd(II)$-H to give $(IMe)_2(p\text{-}O_2NC_6H_5CO_2)Pd(II)$-OOH. Gibbs free energies changes at 298.15 K in benzene are also reported in parentheses. Energies are in kcal/mol and relative to the ground-state reactants. On the left the pathway that involves reductive elimination of $p\text{-}O_2NC_6H_5COOH$ (*Path A*) and on the right the pathway for the direct O_2 insertion (*Path B*) are reported.

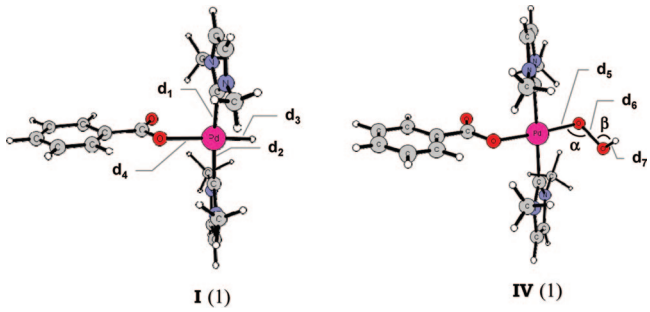


Figure 4. Ground-state optimized structures of the Pd(II)-H and Pd(II)-OOH complexes in the case of R=Ph. Labels employed to individuate geometrical parameters have to be used to read information reported in Tables 1 and 2.

analysis for R=Ph, $p\text{-}O_2NC_6H_4$ will be pointed out for comparison, and general conclusions on the viability of the alternative proposed mechanisms will be drawn.

3.1. Mechanism Proceeding via the Pd(0) Intermediate. The first step of the mechanism that involves formation of the Pd(0) species proceeds by reductive elimination of CH_3COOH from the initial Pd-hydride complex (Step **I** → **II**). As shown in Figure 1 the transition state **TS1***(1)*, which the reaction evolves through is very high in energy, with a calculated relative energy of about 31 kcal/mol both in gas phase and in solution. An inspection of Figure 1 clearly shows that this is the rate-determining step of the overall transformation. The calculated imaginary frequency is 485*i* cm^{-1} and is mainly associated with the displacement of the hydrogen atom from palladium to carbon atom. In the next minimum intercepted along the PES, **II·HX**(1), the carboxylic acid molecule still interacts with the Pd(0) formed species, and the process results to be endothermic by 4.7 kcal/mol (4.1 kcal/mol in gas phase). Formation of final products, upon reductive elimination of CH_3COOH from the initial Pd−H complex, is practically thermoneutral in solvent and endothermic by 15.2 kcal/mol in gas phase. The reaction between the formed Pd(0) complex with triplet molecular oxygen (Step **II**→**III**) requires that both singlet and triplet multiplicities are examined and leads initially to the formation of an η^1-O_2 adduct, **II·**O_2, in which the IMe ligands retain their *trans* configuration. A stable closed-shell singlet complex, **III**, is formed by surpassing a low energy barrier (6.4 kcal/mol in solvent and 10 kcal/mol in gas phase) corresponding to the transition state, **TS2**(3), for the IMe ligands *trans* to *cis* isomerization. The characteristics of this adduct **III** (O−O distance of 1.4 Å and O−O stretching frequency of 978 cm^{-1}) allow us to design it as a η^2-peroxide.[35] Due to the multiplicity change, in this region of the PES the system must undergo a spin inversion that takes place after passage of the **TS2** transition state. In the framework of the Two-State Reactivity paradigm[36] such a kind of spin crossover is not considered a rate-limiting factor as it involves species that are formed with excess energy, and even electronically excited states become accessible. Following the procedure outlined before, the MECP between the triplet and singlet surfaces has been individuated, and the structure at this point, **MECP1**, exhibits a geometry very

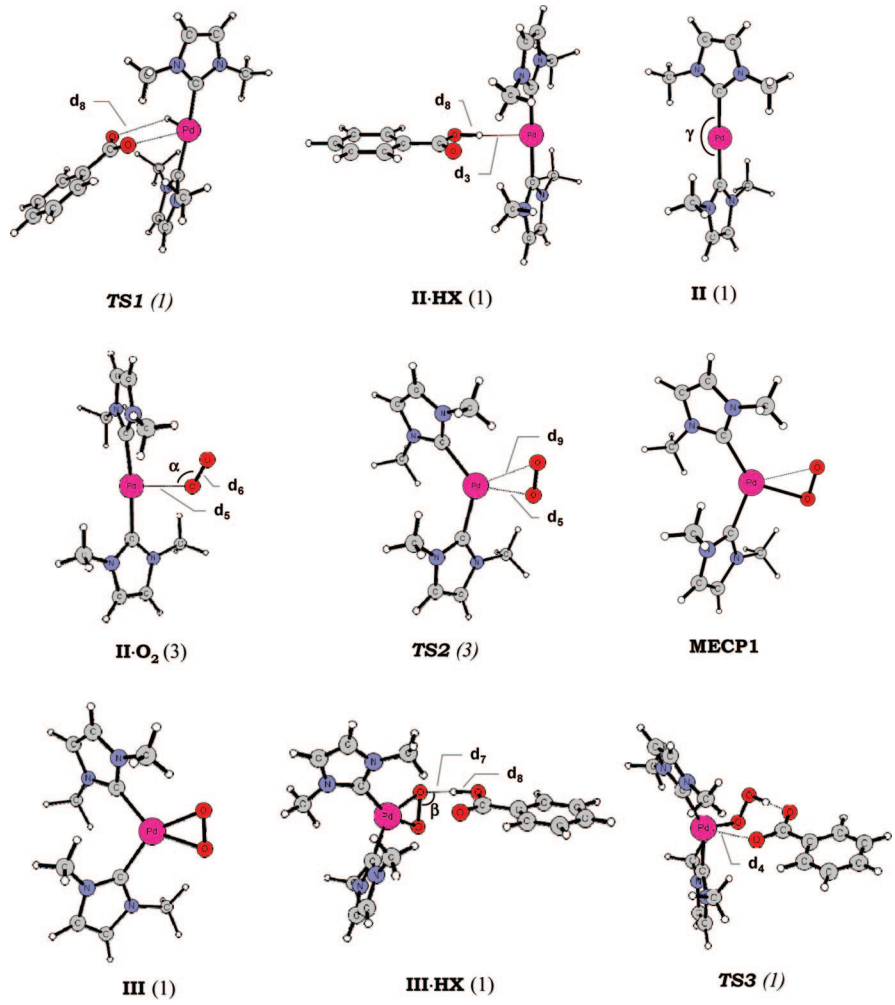


Figure 5. Ground-state optimized structures of stationary points and MECP intercepted along *Path A* in the case of R=Ph. Labels employed to individuate geometrical parameters have to be used to read information reported in Tables 1 and 2.

similar to **TS2**(3). Crossing point relative energy calculated in gas phase is reported in Figure 1. The third and final step (Step **III**→**IV**) of the examined pathway involves protonolysis of the Pd−O bond of the η^2-peroxo complex **III**. Computational analysis shows that initially CH_3COOH weakly interacts with **III** to give a hydrogen-bonded complex, **III**·HX, stabilized by 4 kcal/mol in solvent (9.6 kcal/mol in gas phase) with respect to ground-state reactants asymptote. Formation of the final hydroperoxide complex **IV** takes place overcoming an energy barrier of 23.6 and 21.2 kcal/mol in solvent and in gas phase, respectively, corresponding to the **TS3**(1) transition state. The normal mode associated with the imaginary frequency, calculated to be 161*i* cm^{-1}, corresponds to the concerted *cis* to *trans* isomerization of the IMe ligands, proton transfer from acetic acid to one of the oxygen atoms of the peroxide, and coordination of the formed acetate to the Pd center. As a result, the final singlet hydroperoxide product (CH_3CO_2)-$(IMe)_2$Pd-OOH in its *trans* configuration is obtained, and the overall process is calculated to be exothermic by 8.0 kcal/mol in benzene solvent and 17.1 kcal/mol in gas phase.

3.2. Mechanism of Direct Dioxygen Insertion into the Pd−H Bond of *trans*-[$(CH_3COO)(IMe)_2$Pd(H)]. Along the triplet pathway (see *Path B* Figure 1) the interaction of oxygen with the *trans*[$(IMe)_2(CH_3COO)$Pd−H] complex leads to the formation of a weakly bound van der Waals complex, **I·O$_2$**(3), that is stabilized by about 1 kcal/mol in solution. The fully optimized structure of the corresponding complex in a singlet state is substantially different as O_2 coordinates directly to the Pd(II) center. The formed adduct, **I·O$_2$**(1), is not stable and lies about 21 kcal/mol above the entrance channel of separated reactants in gas phase and 32.6 kcal/mol above, when solvation effects are included. The reported energy for this complex has been properly corrected, following the mentioned scheme,[30] since unrestricted calculations on this adduct showed a significant amount of triplet spin contamination.

The next step of the reaction, that results to be the rate-determining one, involves the abstraction of the hydrogen atom from the palladium center by O_2. This step requires an activation energy of 15.6 kcal/mol in solution and 22.9 kcal/mol in gas phase for the formation of the **TS4***(3)* transition state. In going from the first adduct to **TS4**(3) the Pd−H bond distance stretches from 1.565 to 1.763 Å, and at the same time the O−O bond length increases from 1.205 to 1.278 Å. The O−H distance is 1.237 Å indicating that the bond between oxygen and hydrogen is forming. The imaginary frequency for **TS4***(3)* transition state is calculated to be 1516*i* cm^{-1} and clearly corresponds to the movement of the hydrogen atom detaching from Pd and bonding to O

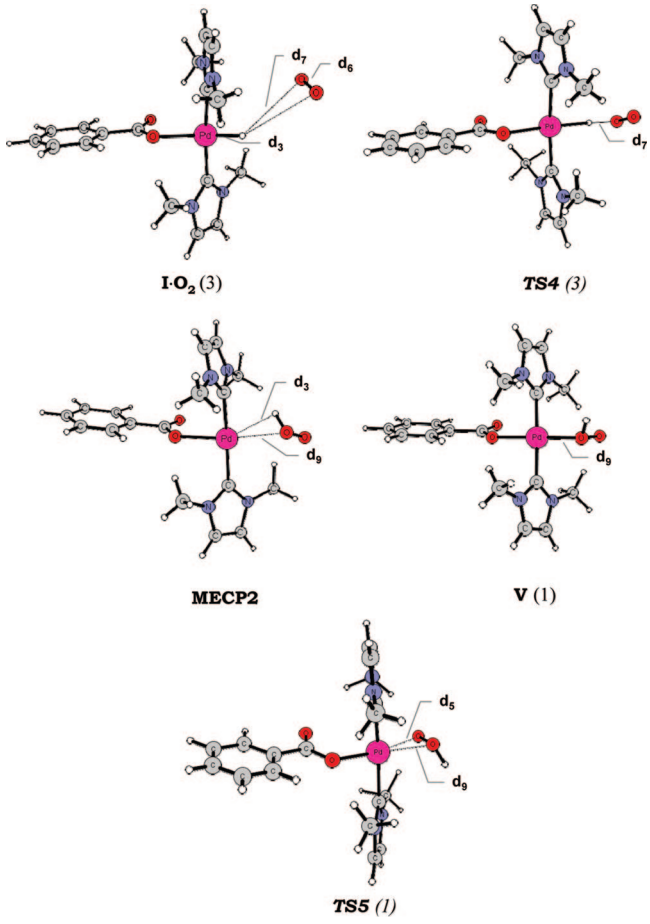


Figure 6. Ground-state optimized structures of stationary points and MECP intercepted along *Path B* in the case of R=Ph. Labels employed to individuate geometrical parameters have to be used to read information reported in Table 2.

atom. As a result, an intermediate is formed in which a peroxyl HOO radical is coordinated to the T-shaped Pd(I) complex through a weak interaction between hydrogen and Pd atoms. Indeed, this intermediate is only slightly more stable than **TS4***(3)* both in gas phase and solution. Along the singlet surface, instead, once molecular oxygen approaches the Pd center, at a distance of 1.472 Å between the closest oxygen and hydrogen, a transition state, **TS4***(1)*, has been intercepted and characterized. Also for this singlet transition state, which is less stable than the separate reactants by 35.4 kcal/mol in solution (28.4 kcal/mol in gas phase), the energy has been corrected for spin contamination. From **TS4***(1)* the reaction proceeds to yield a stable singlet intermediate, **V**(1), whose formation is endothermic by 2.9 kcal/mol in gas phase and 10.3 kcal/mol in benzene. The main feature of this intermediate is that the hydrogen atom is bonded (H−O=0.963 Å) to the proximal oxygen. Until now the existence of this kind of intermediates has not been evidenced by previous theoretical investigations.[16,19] IRC calculations confirmed that the **TS4***(1)* transition state is connected to both reactant **I·O$_2$**(3) and product **IV**(1). The analysis of the structures along the reaction coordinate from the transition state to the **IV**(1) intermediate shows that the hydrogen atom approaches the proximal oxygen lengthening the O−O bond, whereas the distal oxygen atom rotates upward.

Since the singlet state of the **IV**(1) intermediate is more stable than the triplet one, in this region of the PES the system undergoes a spin inversion that takes place after passage of the **TS4***(3)* transition state. Also in this case the spin crossover occurs after formation of the transition state and, as a consequence, cannot be considered a rate-limiting factor.[35] Relative energy calculated in gas phase of the minimum energy crossing point, **MECP2**, is reported in Figure 1. The crossing occurs in the vicinity of the triplet transition state, and the **MECP2** structure lies very close in energy to the preceding transition state. It is worthwhile to underline that all the attempts to individuate a singlet intermediate with a structure analogous to that of the triplet complex were unsuccessful in spite of the numerous strategies used to find it out.

Formation of the final hydroperoxo palladium(II) complex, along the singlet pathway involves breaking the bond between the proximal oxygen and the Pd center that makes a new bond with the distal oxygen. Triplet hydroperoxide product formation involves, instead, rearrangement of the HOO fragment in such a way that terminal oxygen and hydrogen atoms move in opposite directions toward and away from the Pd center, respectively. Both the corresponding transition states, **TS5***(1)* and **TS5***(3)*, have been intercepted and confirmed by IRC analysis. The singlet **TS5***(1)* structure lies 10.5 kcal/mol, 20.5 kcal/mol when solvent effects are included, above the reactants dissociation limit and is characterized by an imaginary frequency of 339*i* cm^{-1}. Along the triplet PES the **TS5***(3)* structure is destabilized respect to reactants by 19.6 kcal/mol in gas phase and by 28.9 kcal/mol in benzene, and the corresponding imaginary frequency is calculated to be 102*i* cm^{-1}.

Compared to the exothermic formation of the hydroperoxide product in its singlet multiplicity, **IV**(1), the overall conversion process of the palladium hydride to the Pd-OOH species, **IV**(3), along the spin conserving triplet surface is endothermic by 17.4 kcal/mol in gas phase and 24.3 kcal/mol in solvent.

3.3. Overall Mechanism for the RCO$_2$H Reductive Elimination Pathway with R=Ph, *p*-O$_2$NC$_6$H$_4$. As it appears, at a first glance, from the comparison between the left sides of Figures 1,2, and 3 the presence of a different ligand does not introduce any significant qualitative change in the calculated energy profiles. We briefly comment the quantitative description of the envisaged multistep mechanism focusing our attention on those changes that could have an influence on the determination of the most likely pathway among the two proposed ones.

The first step of the process, that is reductive elimination of the carboxylic acids, involves again formation of the weak hydrogen-bonding adduct **II·HX**(1) that occurs surmounting a high free energy barrier (see Figures 2 and 3), corresponding to the ***TS1****(1)* transition state, of 34.4 kcal/mol for R=Ph and of 31.3 kcal/mol for R=*p*-O$_2$NC$_6$H$_4$. These values are very similar to that computed for the examined acetic acid and confirm that this is the rate-determining step of the whole

Table 1. Selected Geometrical Parameters of Ground-State Reactants, Products, Intermediates, MECP, and Transition States Computed along *Path A*[a]

	I(1)(exp)	TS1(1)	II·HX(1)	II(1)	II·O_2 (3)	TS2(3)	MECP1	III(1)	III·HX(1)	TS3 (1)	IV(1)
d_1	**2.052**	**2.065**	**2.053**	2.047	2.074	2.103	2.099	2.061	**2.032**	**2.227**	**2.073**
	2.045(2.016)	*2.059*	*2.052*						*2.027*	*2.218*	*2.071*
	2.053	2.059	2.053						2.025	2.215	2.071
d_2	**2.050**	**2.058**	**2.053**	2.047	2.072	2.105	2.099	2.061	**2.061**	**2.001**	**2.058**
	2.051(2.022)	*2.058*	*2.052*						*2.061*	*2.006*	*2.061*
	2.053	2.057	2.053						2.059	2.005	2.064
d_3	**1.565**	**1.510**	**2.173**	-	-	-	-	-	-	-	-
	1.563(1.540)	*1.510*	*2.154*								
	1.559	1.510	2.102								
d_4	**2.185**	**2.932**	**3.840**	-	-	-	-	-	**3.383**	**2.903**	**2.097**
	2.187(2.134)	*3.039*	*3.819*						*3.379*	*3.145*	*2.106*
	2.195	3.116	3.816						3.399	3.238	2.113
d_5	-	-	-	-	2.392	2.256	2.279	2.036	**2.009**	**1.982**	**2.030**
									2.008	*1.980*	*2.027*
									2.007	1.979	2.016
d_6	-	-	-	-	1.259	1.284	1.305	1.400	**1.418**	**1.373**	**1.454**
									1.420	*1.371*	*1.452*
									1.422	1.374	0.967
d_7	-	-	-	-	-	-	-	-	**1.551**	**1.067**	**0.966**
									1.498	*1.048*	*0.966*
									1.424	1.035	1.452
d_8	-	**2.414**	**1.000**						**1.031**	**1.435**	-
	-	*2.517*	*1.014*	-	-	-	-	-	*1.046*	*1.479*	
		2.578	1.024						1.074	1.514	
d_9					3.218	2.572	2.312	2.036	**2.084**	**2.539**	-
	-	-	-	-					*2.089*	*2.574*	
									2.097	2.579	
α					120.5	88.8	74.8	69.9	**72.6**	**96.7**	**113.4**
	-	-	-	-					*72.9*	*98.7*	*113.5*
									76.2	99.0	101.3
β				-	-	-	-	-	**104.0**	**96.7**	**101.2**
	-	-	-						*97.9*	*103.7*	*101.3*
									102.6	103.5	113.7
γ	**175.9**	**175.1**	**178.5**	179.9	174.6	127.2	119.8	105.8	**100.8**	**140.5**	**177.6**
	175.3	*175.1*	*179.2*						*100.2*	*150.8*	*177.2*
	174.5	174.8	178.7						99.8	150.6	177.4
θ	**48.6**	**−71.1**	**0.0**	-	-	-	-	-	-	-	-
	10.9	*−67.4*	*0.0*								
	0.3	−69.7	0.0								

[a] Available experimental geometrical parameters for the benzoate analogue of the hydride complex are also reported in parentheses.Values are reported in **bold** for R=CH_3, in *italic* for R=Ph, and in regular for *p*-$O_2NC_6H_4CO_2$. Bond lengths are in Å and angles are in degrees. θ = N−C−C−N dihedral angle to individuate relative positions of IMe ligands.

oxygenation pathway. Formation of the Pd(0) complex is more endothermic for both R=Ph and R=*p*-$O_2NC_6H_4$ with respect to the elimination of acetic acid as well as, consequently, formation of the subsequent two minima and transition state since the $(IMe)_2Pd(0)$ oxygenation step does not depend on the employed carboxylic acid. Identical is also the structure of the minimum energy crossing point, **MECP1**, between the triplet and singlet surfaces.

The interaction of the carboxylic acid RCOOH with the η^2-peroxo complex **III**(1) is exothermic in solvent by −6.8 and −3.3 kcal/mol with respect to ground-state reactants asymptote for R=Ph and *p*-$O_2NC_6H_4$, respectively. The free energy barrier that is necessary to surmount to obtain the final hydroperoxide product **IV** is 14.6 kcal/mol for R=Ph and 10.5 kcal/mol for R=*p*-$O_2NC_6H_4$. Proton transfer from the carboxylic acid to one of the oxygen atoms of the peroxide, coordination of the formed anion to the Pd center, and *cis* to *trans* isomerization of the IMe ligands allow formation of the final singlet hydroperoxide complex (RCO_2)-$(IMe)_2$Pd-OOH in its *trans* configuration. The overall process is calculated to be exothermic by about 8 kcal/mol in benzene solvent for both R=Ph and *p*-$O_2NC_6H_4$, being that this value is very close to that calculated in the case of the reductive elimination of acetic acid.

3.4. Mechanism of Direct Dioxygen Insertion into the Pd−H Bond of *trans*-[RCOO(IMe)$_2$Pd(H)] with R=Ph and *p*-$O_2NC_6H_4$. As it was pointed out yet in the previous paragraph, no significant qualitative differences emerge between the calculated energy profiles when the right sides of Figures 1, 2, and 3 are compared. However, it is worth highlighting some quantitative differences. The interaction of triplet oxygen with the *trans*-$[(RCO_2)(IMe)_2PdH]$ complex leads to the formation of a weakly bound van der Waals complex, **II·HX**(1), that is endothermic in solvent by about 7 kcal/mol both for R=Ph and R=*p*-$O_2NC_6H_4$ and more endothermic than R=CH_3. Along the triplet surface the reaction evolves through the abstraction of the hydrogen atom from the palladium center by O_2. This step takes place overcoming a free energy barrier of 24.8 and 22.7 kcal/mol for R=Ph and R=*p*-$O_2NC_6H_4$, respectively, corresponding to the formation of the ***TS4(3)*** transition states. The imaginary frequencies that confirm the nature of these stationary points correspond to the shift of the hydrogen atom from Pd to one of the O atoms. The intermediate **V**(3) that is formed can be again classified as an adduct in which the peroxyl HOO radical is coordinated to the T-shaped Pd(I) complex through a weak interaction between hydrogen and Pd atoms. Indeed, this reaction intermediate lies just a few

Table 2. Selected Geometrical Parameters of Ground-State Intermediates, MECP, and Transition States Computed along *Path B*[a]

	I·O$_2$ (3)	**TS4(3)**	**MECP2**	**V(1)**	**TS5(1)**
	2.051	**2.063**	**2.093**	**2.075**	**2.074**
d$_1$	*2.052*	*2.064*	*2.091*	*2.073*	*2.073*
	2.052	2.066	2.092	2.074	2.072
	2.049	**2.063**	**2.082**	**2.066**	**2.042**
d$_2$	*2.049*	*2.062*	*2.106*	*2.064*	*2.042*
	2.051	2.062	2.108	2.066	2.044
	1.565	**1.762**	**2.452**		
d$_3$	*1.562*	*1.764*	*2.455*	-	-
	1.559	1.764	2.457		
	2.186	**2.220**	**2.226**	**2.070**	**2.066**
d$_4$	*2.191*	*2.218*	*2.263*	*2.076*	*2.069*
	2.199	2.227	2.284	2.082	2.075
	4.662	**3.679**	**3.449**	**2.999**	**2.478**
d$_5$	*4.573*	*3.649*	*3.444*	*3.000*	*2.476*
	4.548	3.646	3.441	2.990	2.468
	1.205	**1.277**	**1.352**	**1.481**	**1.517**
d$_6$	*1.205*	*1.277*	*1.351*	*1.479*	*1.518*
	3.466	1.235	1.348	0.964	0.965
	3.386	**1.237**	**0.984**	**0.963**	**0.965**
d$_7$	*3.488*	*1.239*	*0.983*	*0.963*	*0.965*
	1.205	1.276	0.983	1.478	1.519
			2.735	**2.086**	**2.258**
d$_9$	-	-	*2.737*	*2.088*	*2.251*
			2.731	2.083	2.243
			48.0	**39.7**	**63.5**
α	-	-	*47.7*	*39.7*	*63.4*
			47.8	39.8	63.3
		111.1	**106.7**	**101.8**	**101.8**
β	-	*111.2*	*106.5*	*101.7*	*101.7*
		111.2	106.7	101.8	101.7
	176.1	**174.4**	**177.7**	**177.2**	**176.9**
γ	*175.8*	*175.2*	*178.3*	*177.6*	*177*
	175.3	174.7	178.2	178.0	177.5
	47.8	**48.2**	**17.4**	**6.4**	**10.2**
θ	*49.2*	*8.1*	*26.9*	*7.1*	*9.7*
	49.7	2.1	27.4	6.7	8.1

[a] Values are reported in **bold** for R=CH_3, in *italic* for R=Ph, and in regular for *p*-$O_2NC_6H_4CO_2$. Bond lengths are in Å and angles are in degrees. θ = N−C−C−N dihedral angle to individuate relative positions of IMe ligands.

kcal/mol below the ***TS4(3)*** transition state for both $PhCO_2$ and *p*-$O_2NC_6H_4CO_2$ ligands.

Along the singlet surface, instead, overcoming the energy barrier for the intercepted transition state, **TS4***(1)*, allows the formation of a singlet intermediate, **V**(1), whose main feature is that the hydrogen atom is bonded to the proximal oxygen. Also in this case IRC calculations confirmed that the ***TS4(1)*** transition states are connected to both reactant **II·O$_2$**(1) and product **V**(1). Formation of the singlet intermediate **V**(1) is more favorable with respect to formation of the corresponding triplet intermediate by about 10 kcal/mol in solvent, as for R=CH_3. This means that the system must undergo a spin inversion that takes place after passage of the **TS4**(3) transition state. Geometrical parameters of the computed structures of the minimum energy crossing points, **MECP2**, between the triplet and singlet surfaces, are reported in Table 2, and their relative energies calculated in gas phase are reported in Figures 2 and 3 for R=Ph and R=*p*-$O_2NC_6H_4$, respectively.

Formation of the final palladium(II) hydroperoxo product takes place, along the singlet pathway, through the **TS5***(1)* transition state that lies 19.7 kcal/mol for R=Ph and 21.6 kcal/mol for R=*p*-$O_2NC_6H_4$ above the ground-state reactants asymptote. Rearrangement of the HOO fragment, instead, allows triplet hydroperoxide product formation through the corresponding less stable, for both benzoate and *p*-nitrobenzoate ligands, **TS5***(3)* transition states. Energetics of the hydroperoxide product reaction formation along both singlet and spin conserving triplet surfaces have been reported in the preceding paragraph.

3.5. Preferred Mechanism and Ligand Influence. The detailed and systematic analysis of the two viable palladium-hydride oxygenation pathways carried out in this work reveals that the free activation energies for the rate determining-step of *Path A* and *Path B* are different enough, in spite of the intrinsic uncertainties of DFT computations, to determine which is the preferred mechanism. When the involved carboxylic acid is the acetic acid, the calculated barrier for the CH_3COOH reductive elimination initial step (*Path A*) is 31.2 kcal/mol in gas phase and 31.0 kcal/mol when salvation effects are included. The rate-determining step barrier that is necessary to overcome, instead, for the abstraction of the hydrogen atom by molecular oxygen (*Path B*) is 15.6 and 22.9 kcal/mol in gas phase and in benzene, respectively. The two examined mechanisms exhibit the same difference in the calculated activation barriers relative to the rate-determining step even when the ligands coordinated to the Pd center of the initial Pd-hydride complex are $PhCO_2$ and *p*-$O_2NC_6H_4CO_2$. Indeed, free activation energies calculated in solvent are 34.4 and 24.8 kcal/mol for the reductive elimination and hydrogen atom abstraction steps, respectively when R=Ph. Analogous barriers are 31.3 and 22.7 kcal/mol when R=*p*-$O_2NC_6H_4$. Therefore, on the basis of our results the oxygenation reaction is more likely to evolve through direct insertion of molecular oxygen into the Pd−H bond of the initial hydride complex. This conclusion is reinforced by the favorable comparison between the experimental activation energy, determined to be 24.4(5) kcal/mol for the benzoate ligand,[18] and the corresponding calculated value of 24.8 kcal/mol. In addition, the experimentally detected enhancement of the oxygenation rate due to the presence of the nitro group on the benzoate ligand corresponds to the lower activation energy calculated by us for R=*p*-$O_2NC_6H_4$.

Conclusions

A detailed DFT study of the mechanism for the conversion of the Pd(II)-hydride complex, $(IMe)_2(RCO_2)PdH$ (R=CH_3, Ph, and *p*-$O_2NC_6H_4$), to the corresponding Pd(II)-hydroperoxide in the presence of molecular oxygen has been carried out with the aim to probe the alternative reaction pathways, both considered viable on the basis of current data, of direct O_2 insertion, and slow reductive elimination followed by oxygenation. The outcome of our computational analysis supports the mechanism of direct insertion of molecular oxygen into the Pd−H bond of the initial complex. The activation energy relative to the rate-determining step of the direct insertion pathway is calculated to be lower than the activation energy of the rate determining step of the alternative pathway that

involves the carboxylic acid reductive elimination, whatever ligand (CH_3CO_2, Ph, CO_2, p-$O_2NC_6H_4CO_2$) is coordinated to the Pd center. The calculated free activation energy of the rate-determining hydrogen abstraction step ($\Delta G^* = 24.8$ kcal/mol) in the case of the oxygenation reaction of the benzoate-ligated Pd(II)-hydride complex is in very good agreement with the value experimentally determined ($\Delta G^* = 24.4$ kcal/mol). In addition, according to the experimentally detected enhancement of the reaction rate due to the presence of the nitro group on the benzoate ligand, our calculations show that the transition state for the hydrogen atom abstraction by molecular oxygen lies lower in energy along the pathway for the oxygenation reaction of $(IMe)_2(p\text{-}O_2NC_6H_4CO_2)PdH$.

Acknowledgment. We gratefully acknowledge financial help from the Università della Calabria.

References

(1) Stahl, S. S. *Science* **2005**, *309*, 1824–1826.

(2) Stahl, S. S. *Angew. Chem., Int. Ed.* **2004**, *43*, 3400–3420.

(3) Sigman, M. S.; Schultz, M. J. *Org. Biomol. Chem.* **2004**, *2*, 2551–2554.

(4) Stoltz, B. M. *Chem. Lett.* **2004**, *33*, 362–367.

(5) Nishimura, T.; Uemura, S. *Synlett* **2004**, 201–216.

(6) Sheldon, R. A.; Arends, I. W. C. E.; ten Brink, G.-J.; Dijksman, A. *Acc. Chem. Res.* **2002**, *35*, 774–781.

(7) Gligorich, K. M.; Sigman, M. S. *Angew. Chem., Int. Ed.* **2006**, *45*, 6612–6615.

(8) Konnick, M. M.; Guzei, I. A.; Stahl, S. S. *J. Am. Chem. Soc.* **2004**, *126*, 10212–10213.

(9) Stahl, S. S.; Thorman, J. L.; Nelson, R. C.; Kozee, M. A. *J. Am. Chem. Soc.* **2001**, *123*, 7188–7189.

(10) Thiel, W. R. *Angew. Chem.* **1999**, *111*, 3349–3351. *Angew. Chem., Int. Ed.* **1999**, *38*, 3157−3158.

(11) Nishimura, T.; Onoue, T.; Ohe, K.; Uemura, S. *J. Org. Chem.* **1999**, *64*, 6750–6755.

(12) Hosokawa, T.; Murahashi, S.-I. *Acc. Chem. Res.* **1990**, *23*, 49–54.

(13) Muzart, J.; Pete, J. P. *J. Mol. Catal.* **1982**, *15*, 373–376.

(14) Landis, C. R.; Morales, C. M.; Stahl, S. S. *J. Am. Chem. Soc.* **2004**, *126*, 16302–16303.

(15) Denney, M. C.; Smythe, N. A.; Cetto, K. L.; Kemp, R. A.; Goldberg, K. I. *J. Am. Chem. Soc.* **2006**, *128*, 2508–2509.

(16) Keith, J. M.; Muller, R. P.; Kemp, R. A.; Goldberg, K. I.; Goddard, W. A., III; Oxgaard, J. *Inorg. Chem.* **2006**, *45*, 9631–9633.

(17) Keith, J. M.; Nielsen, R. J.; Oxgaard, J.; Goddard, W. A., III *J. Am. Chem. Soc.* **2005**, *127*, 13172–13179.

(18) Konnick, M. M.; Gandhi, B. A.; Guzei, I. A.; Stahl, S. S. *Angew. Chem., Int. Ed.* **2006**, *45*, 2904–2907.

(19) Popp, B. V.; Stahl, S. S. *J. Am. Chem. Soc.* **2007**, *129*, 4410–4422.

(20) Chowdhury, S.; Rivalta, I.; Russo, N.; Sicilia, E. *Chem. Phys. Lett.* **2007**, *443*, 183–189.

(21) Chowdhury, S.; Rivalta, I.; Russo, N.; Sicilia, E. *Chem. Phys. Lett.* **2008**, *456*, 41–46.

(22) Gligorich, K. M.; Sigman, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 6612–6615.

(23) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(24) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(25) 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.; 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.; Ochtersky, ; Ayala, P. Y.; Morokuma, ; Voth, ; Salvador, ; Dannenberg, ; Zakrzewski, ; Dapprich, ; Daniels, ; Strain, ; Farkas, ; Malick, ; Rabuck, ; Raghavachari, ; Foresman, ; Ortiz, ; Cui, ; Baboul, ; Clifford, ; Cioslowski, ; Stefanov, ; Liu, ; Liashenko, ; Piskorz, ; Komaromi, ; Martin, ; Fox, ; Keith, ; Al-Laham, ; Peng, ; Nanayakkara, ; Challacombe, ; Gill, ; Johnson, ; Chen, ; Wong, ; Gonzalez, ; Pople, *Gaussian03, reVision B.05*; Gaussian, Inc.: Wallingford, CT, 2004.

(26) Andrae, D.; Häussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123–141.

(27) Fukui, K. *J. Phys. Chem.* **1970**, *74*, 4161–4163.

(28) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154–2161.

(29) Weissbluth, M. *Atoms and Molecules*; Academic Press: New York, 1978; p 587.

(30) Ovchinnicov, A. A.; Labanowski, J. K. *Phys. Rev. A* **1996**, *53*, 3946–3952.

(31) Cancès, M.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(32) Mennucci, B.; Cancès, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506–10507.

(33) Kollaman, P. *Chem. Rev.* **1993**, *93*, 2395.

(34) Harvey, J. N.; Aschi, M.; Schwarz, H.; Koch, W. *Theor. Chem. Acc.* **1998**, *99*, 95–99.

(35) Cramer, C. J.; Tolman, W. B.; Theopold, K. H.; Rheingold, A. L. *PNAS* **2003**, *100*, 3635–3640.

(36) Schröder, D.; Shaik, S.; Schwarz, H. *Acc. Chem. Res.* **2000**, *33*, 139–145.

CT8001442

JCTC Journal of Chemical Theory and Computation

The "Hot-Solvent/Cold-Solute" Problem Revisited

M. Lingenheil, R. Denschlag, R. Reichold, and P. Tavan*

Lehrstuhl für Biomolekulare Optik, Ludwig-Maximillians-Universität München, Oettingestrasse 67, 80538 München, Germany

Received February 1, 2008

Abstract: The temperature steers the equilibrium and nonequilibrium conformational dynamics of macromolecules in solution. Therefore, corresponding molecular dynamics simulations require a strategy for temperature control which should guarantee that the experimental statistical ensemble is also sampled *in silico*. Several algorithms for temperature control have been proposed. All these thermostats interfere with the macromolecule's "natural" dynamics as given by the Newtonian mechanics. Furthermore, using a single thermostat for an inhomogeneous solute−solvent system can lead to stationary temperature gradients. To avoid this "hot solvent/cold solute" problem, two separate thermostats are frequently applied, one to the solute and one to the solvent. However, such a separate temperature control will perturb the dynamics of the macromolecule much more strongly than a global one and, therefore, can introduce large artifacts into its conformational dynamics. Based on the concept that an explicit solvent environment represents an ideal thermostat concerning the magnitude and time correlation of temperature fluctuations of the solute, we propose a temperature control strategy that, on the one hand, provides a homogeneous temperature distribution throughout the system together with the correct statistical ensemble for the solute molecule while, on the other hand, minimally perturbing its dynamics.

1. Introduction

Molecular dynamics (MD) simulations using molecular mechanics (MM) force fields have become a widespread tool to study the equilibrium conformational dynamics of proteins and peptides in solution,[1] including processes of folding and refolding.[2] More recently, also nonequilibrium processes have been simulated in which a protein or peptide is destabilized, for example by applying an external force mimicking the action of an atomic-force microscope,[3–5] by exerting internal mechanical strain,[6,7] by introducing point mutations into the protein sequence,[8,9] or simply by elevating the temperature.[9,10]

The behavior of proteins in solution is steered by the thermodynamic conditions, notably by the temperature. The native state is stable only within a certain temperature range; processes of hot and cold unfolding have been observed.[11] The temperature influences the stability and function of proteins not only directly by changing the relative importance of the entropy but also indirectly via certain temperature dependent solvent properties such as the dielectric constant[12] or the viscosity.[13] Therefore, if one wants to describe experiments on proteins by MD simulations, the temperature must be properly controlled.

Clearly, an adequate method for temperature control is not the only precondition if one aims at quantitative descriptions of experimental data. In this respect, the quality of the employed force field, the sufficiency of statistical sampling achieved by finite simulation times, and other technical issues are also questions of concern.[14] However, the temperature is of key importance because many experimental observables that can be compared with the information obtained from MD simulations sensitively depend on this parameter. Examples are the temperature factors in X-ray crystallography,[15] the proton exchange and spin relaxation rates in nuclear magnetic resonance spectroscopy [see ref 16 and references therein], and the fluorescence depolarization rates[17] as well as the thermodynamical measures of protein stability.[18]

* Corresponding author e-mail: tavan@physik.uni-muenchen.de.

10.1021/ct8000365 CCC: $40.75
Published on Web 07/08/2008

The requirement of proper control does not only apply to the temperature, i.e. the average kinetic energy of the system, but also to other ensemble properties (e.g., energy fluctuations) associated with experimental observables. Thus, in a broader sense, the problem of temperature control in MD simulations is also that of generating the correct statistical ensemble (usually canonical or isothermal−isobaric). The accurate generation of a specific statistical ensemble by means of a MD simulation is also relevant for the application of generalized ensemble techniques like replica exchange molecular dynamics[19–21] which has recently become very popular in order to enhance the sampling efficiency. These techniques rely on the assumption that the applied MD method samples the canonical ensemble at the respective temperature.

When simulating macromolecules in solution, the solvent environment, which is essential for the properties of the solute, can either be treated implicitly using continuum approximations or explicitly by including part of the solvent into the simulation system.[14] The following discussions exclusively deal with the latter case and are devoted to the task of controlling the temperature of a solute macromolecule in explicit solvent. This task can comprise additional challenges if nonequilibrium relaxation processes are studied. Here, frequently, energy is released in one part of the system and then dissipated into the rest of the simulation box, e.g. from a solute molecule to the surrounding solvent. Since the kinetics of energy relaxation and heat transport can influence the dynamical properties of the solute,[22] any applied temperature control method should make sure that the natural energy relaxation processes are unimpaired.

Generally, the ideal temperature control scheme for solute−solvent systems would be to simulate the complete simulation system microcanonically, i.e. at constant total energy in the *NVE* ensemble. One can show[23] that, in this situation, an arbitrary subset of degrees of freedom in thermal contact with the rest of the system (e.g., the solute's kinetic degrees of freedom) will sample the canonical ensemble if the energy fluctuations of the subsystem are insignificant compared with the total energy in the rest of the system. Furthermore, one can show that the subsystem will sample the isothermal−isobaric ensemble if also the subsystem's volume fluctuations are negligible compared with the volume of the rest of the system. Finally, one expects that all those configurational degrees of freedom of the solute which directly interact with the solvent system will sample the canonical or isothermal−isobaric ensemble, respectively, if additionaly the solute−solvent interaction energy is neglibible compared with the solvent-internal interaction energy. In MD simulations systems, the latter condition is fulfilled if the solvent atoms by far outnumber those of the solute. Concurrently, by using the *NVE* approach, the solute's Newtonian dynamics are left completely undisturbed. The *NVE* strategy has been recommended[24] for studies of protein folding kinetics and is occasionally applied[25,26] to eliminate a putative influence of thermostat algorithms on the simulated dynamics.

Unfortunately, the simple *NVE* strategy is not easily applied to extended MD simulations. Numerical inaccuracies associated with approximation schemes serving to speed up the computations generally lead to a violation of energy conservation. For example, heating may be caused by certain approximate treatments of long-range electrostatic interactions[27,28] or by integrating the equations of motion with multiple-time-step algorithms.[29] Cooling may occur, for instance, if constraining bond lengths or angles with a too loose tolerance[30] or if neighbor lists for the calculation of the van der Waals interactions are not updated frequently enough.[31] The defect of energy conservation could, in principle, be compensated by using an ergostat algorithm which would just scale the velocities of all atoms at every time step by an appropriate factor to keep the total energy exactly constant. However, the rates of algorithmic energy drift can vary among the constituents of an inhomogeneous simulation system leading to unphysical steady state temperature gradients,[32] a problem sometimes referred to as the "hot-solvent/cold-solute" problem.[33] For example, such a gradient can result from an approximate treatment of the electrostatic interactions, which may render a mildly polar solute less affected by algorithmic noise than a strongly polar aqueous solvent.[32,34,35]

Thus, specifically for equilibrium simulations of macromolecules in solution, the applied temperature control has to fulfill an important requirement: The temperature distribution has to be homogeneous throughout the inhomogeneous simulation system. As a strategy guaranteeing such a homogeneous temperature distribution it has been suggested to couple the subsystems independently to separate thermostats.[36] Further below we will check this strategy among others because it is the central aim of this work to determine an optimal strategy for generating a homogeneous temperature distribution in solute−solvent simulation systems.

From a general point of view, the appropriateness of a given temperature control method involves the following three aspects:

a) Thermodynamics: Does the method generate the expected thermodynamical ensemble in principle (i.e., with simulations of infinite length and in the absence of numerical errors)?

b) Ergodicity: Does the method generate the expected ensemble within the time typically covered by modern MD simulations?

c) Dynamics: Is the time dependence and spatial distribution of the thermostatic forces realistic? For a solute in solution, for example, one would prefer to have no such forces at all beyond the thermostatting Newtonian interactions with the solvent.

A number of different algorithms has been proposed as realizations of the required thermostats (for a review see ref 36). Each of these algorithms has its specific merits and drawbacks. A critical discussion of these issues is another goal of our study.

For example, the widely used Berendsen thermostat[37] (BT) has the advantage to couple only weakly to the dynamics of the controlled system (see the original paper ref 37 for this issue). On the other hand, it is clear from theoretical considerations that the BT does not create a canonical distribution of microstates,[40] i.e. it introduces artifacts of type

a). Furthermore, the BT violates energy equipartition by redistributing energy from high to low frequency modes, which leads to the so-called "flying-ice-cube effect".[38,39] It is unclear whether this effect is specific to the Berendsen method and closely related methods or it can occur with any thermostat belonging to the more general class of velocity rescaling algorithms.[38]

The more strongly coupling Nosé-Hoover thermostat[41,42] (NHT) is theoretically expected to generate the canonical distribution of microstates if certain conditions are obeyed thus conforming with the above question a).[43] However, within the time covered by a typical MD simulation, amplitudes of temperature fluctuations were observed which were by 1 order of magnitude larger than those expected for a canonical ensemble.[44] Several studies[42,44–47] have shown that Nosé-Hoover coupled systems do not necessarily acquire ergodicity in a reasonable time [cf. question b) above] if these systems are small, stiff, or at low temperatures. Additionally, by its very construction as a velocity rescaling algorithm, also the NHT could show the flying-ice-cube artifact (although we are not aware of any reports on a corresponding example).

As a reaction to these problems, modifications to both the Berendsen and Nosé-Hoover schemes have been proposed. The most frequently employed variant of the Nosé-Hoover thermostat is the so-called Nosé-Hoover chain,[48] which has been successfully tested by Cheng and Merz[33] as a remedy to the hot-solvent/cold-solute problem. No artifacts or deviations from the canonical ensemble have been reported so far. Only recently, Bussi et al.[49] suggested a modification of the Berendsen scheme in order to reliably generate a canonical distribution for systems that otherwise would sample the microcanonical ensemble. Both, the Nosé-Hoover chain and the modified Berendsen thermostat induce temperature fluctuations of the correct size by artificially scaling the atomic velocities. For systems, however, which anyway sample the canonical ensemble, such a thermostat introduces an unnecessary perturbation of the dynamics, i.e. artifacts of type c). The generic example for such a system is a solute molecule in a sufficiently large explicit solvent system, which, as discussed above, always samples a canonical ensemble although possibly at the wrong temperature because of algorithmic inaccuracies.

Concerning temperature control of macromolecules in solution, we want to show how one can (i) generate the appropriate ensemble for the solute molecule in adequate time, (ii) leave invariant the time scales of energy relaxation and of equilibrium fluctuations, and (iii) guarantee a homogeneous temperature distribution in equilibrium simulations with (iv) minimal perturbation of the solute's Newtonian dynamics.

For this purpose we will scrutinize in section 2 the existing temperature control scenarios for MD simulations of solvent−solute systems by partially recollecting and partially developing associated theoretical concepts. These considerations will lead to the definition of strategies for a minimally invasive control of a solute temperature. In section 3 we will sketch the methods which we employed in a series of quite extended test simulations on peptides in aqueous solution. As explained in section 4, these simulations were specifically designed to estimate the extent to which the theoretically expected effects of temperature control do actually modify the properties of a solute peptide. Section 5 discusses the results and suggests a practical procedure ensuring a minimally invasive temperature control.

2. Theory

Thermostats. The most widely used class of thermostat algorithms is based on the rescaling of atomic velocities. The equation of motion for an atom which belongs to a system under the rule of such a thermostat is

$$m_i\ddot{\mathbf{r}}_i(t) = \mathbf{F}_{i,\mathrm{ff}}(t) - m_i\gamma(t)\dot{\mathbf{r}}_i(t) \tag{1}$$

Here, the acceleration $\ddot{\mathbf{r}}_i(t)$ of atom i is caused not only by the forces $\mathbf{F}_{i,\mathrm{ff}}(t)$ derived from an MM force-field but also by a second term $\mathbf{F}_{i,\mathrm{therm}}(t) \equiv -m_i\gamma(t)\dot{\mathbf{r}}_i(t)$, which is proportional to the atom's velocity $\dot{\mathbf{r}}_i(t)$ and to a generally time dependent thermostat parameter $\gamma(t)$.

In the Berendsen scheme, $\gamma(t)$ is directly given in terms of the instantaneous kinetic temperature[50] $\boldsymbol{T}(t)$ by

$$\gamma(t) = \frac{1}{2\tau}\left[1 - \frac{T_0}{\boldsymbol{T}(t)}\right] \tag{2}$$

with τ denoting the coupling time and T_0 the target temperature. For the Nosé-Hoover[41,42] thermostat, $\gamma(t)$ is coupled on a time scale τ_{NHT} to $\boldsymbol{T}(t)$ by the differential equation

$$\frac{d\gamma}{dt} = \frac{1}{\tau_{\mathrm{NHT}}^2}\left[\frac{\boldsymbol{T}(t)}{T_0} - 1\right] \tag{3}$$

Perturbation of the Dynamics. Every thermostat which is described by eq 1 perturbs the Newtonian dynamics generated by the forces $\mathbf{F}_{i,\mathrm{ff}}(t)$ through the admixture of additional thermostatic forces $\mathbf{F}_{i,\mathrm{therm}}(t)$. For a solute−solvent system, these thermostatic forces introduce artifacts of type c) concerning the dynamics (cf. section 1). The resulting perturbation can be measured for a selected atom i by the quotient

$$\xi_i^2 \equiv \langle \mathbf{F}_{i,\mathrm{therm}}^2 \rangle_D / \langle \mathbf{F}_{i,\mathrm{ff}}^2 \rangle_D \tag{4}$$

where the brackets $\langle ... \rangle_D$ denote temporal averages over a simulation of a given duration D. The perturbation quotients (4) will depend on the system size and on the particular thermostat, i.e. on the form of $\gamma(t)$, as well as on the coupling time.

The perturbation quotients ξ_i are strictly local measures for the influence of a thermostat on a simulated dynamics. However, one may also consider the local perturbation inflicted on a certain group G of atoms within a simulation system, e.g. on the C_α-atoms of a solute peptide embedded in a solvent environment. Then the root mean quotient $\bar{\xi}_G \equiv \sqrt{\langle \xi_i^2 \rangle_G}$ over the ξ_i^2 belonging to G can be used to compare how the dynamics of a solute is perturbed in different solute−solvent systems.

Instead of calculating the averages $\langle \mathbf{F}_{i,\mathrm{therm}}^2 \rangle_D$ required for the evaluation of the ξ_i^2 directly from a simulation, one can also give a simple estimate for these average square forces.

Assuming a sufficiently large simulation system, the velocities of the individual atoms will negligibly contribute to the temperature $\boldsymbol{T}(t)$. Hence, the correlation of γ^2 and $\dot{\mathbf{r}}_i^2$ vanishes and one obtains

$$\langle \mathbf{F}_{i,\mathrm{therm}}^2 \rangle_D = \langle m_i^2 \gamma^2 \dot{\mathbf{r}}_i^2 \rangle_D = m_i^2 \langle \gamma^2 \rangle_D \langle \dot{\mathbf{r}}_i^2 \rangle_D \tag{5}$$

Assuming furthermore that the system is in equilibrium, that the atomic velocity distributions are undisturbed by the thermostat, and that the system is free of internal constraints (such as fixed bond lengths), the mean square velocity of atom i is expected to be $\langle \dot{\mathbf{r}}_i^2 \rangle_D \approx 3k_B\hat{T}/m_i$, where $\hat{T} = \langle \boldsymbol{T} \rangle_D$ is the average temperature determined from the simulation. Equation 5 then becomes

$$\langle \mathbf{F}_{i,\mathrm{therm}}^2 \rangle_D \approx 3m_i k_B \hat{T} \langle \gamma^2 \rangle_D \tag{6}$$

We will check this estimate by sample simulations and show that it already holds for relatively small systems.

Inserting the estimate 6 into eq 4, one can recognize that the perturbation quotients of a given system which is simulated with different thermostatic strategies solely differ with respect to $\langle \gamma^2 \rangle_D$. Thus, in this case, comparisons of the mean square scaling activities $\langle \gamma^2 \rangle_D$ suffice for the evaluation of different thermostatic strategies concerning the size of local perturbations of the dynamics. However, thermostats do not only cause local perturbations of the Newtonian dynamics but may also interfere with ensemble properties like, for example, size and time scales of the temperature fluctuations.

Temperature Fluctuations. In an MD simulation, the statistics of the temperature fluctuations provides a probe for artifacts of type a) and b) pertaining the generation of the desired ensemble (section 1). For a system in contact with a heat bath of temperature T_b, the distribution of microstates is either given by the canonical or by the isothermal−isobaric ensemble. However, with respect to the temperature fluctuations, both ensembles are equal. The associated probability density $\rho(\boldsymbol{T})$ for the instantaneous kinetic temperature is a χ^2-distribution[51]

$$\rho(\boldsymbol{T}) = \frac{(N_{\mathrm{DoF}}\boldsymbol{T}/2T)^{N_{\mathrm{DoF}}/2}}{\Gamma(N_{\mathrm{DoF}}/2)\boldsymbol{T}} \exp\left[-\frac{N_{\mathrm{DoF}}\boldsymbol{T}}{2T}\right] \tag{7}$$

where $T = T_b$ is the expectation value of $\boldsymbol{T}$, N_{DoF} is the number of degrees of freedom (DoF) of the system, and $\Gamma(...)$ denotes the Euler Γ-function. Consequently, the variance σ_T^2 of the temperature fluctuations is

$$\sigma_{\boldsymbol{T}}^2 = \frac{2T^2}{N_{\mathrm{DoF}}} \tag{8}$$

Under the influence of a thermostat, the statistics can deviate from what is expected for a canonical ensemble. This deviation constitutes a measure for the global influence of the thermostat and for how close a simulation is to sampling the canonical ensemble. In the limit $N_{\mathrm{DoF}} \rightarrow \infty$, eq 7 becomes a normal distribution.

The size of σ_T^2 together with the autocorrelation time[50] τ_c of the temperature fluctuations critically influences the accuracy with which the equilibrium temperature T is determined by a given simulation. The variance $\sigma_{\hat{T}}^2$ of the time averages $\hat{T}$ obtained from a set of equilibrium simulations with durations D can be estimated[50] to be

$$\sigma_{\hat{T}}^2 = 2\sigma_T^2 \frac{\tau_c}{D} \tag{9}$$

In order to judge whether a particular strategy is suited to correctly tune the temperature T, one has to perform a test simulation which is long enough to determine T with sufficient accuracy. For a small solute (large σ_T^2) with a correlation time τ_c longer than 10 ps, an accuracy of 1 K may require simulation times of up to 10 ns.

Power of a Thermostat. By means of the observables introduced above, one can judge to what extent a thermostat can perturb the dynamical and equilibrium properties of a solute in solute−solvent simulations. Such perturbations can, of course, be avoided by using no thermostat at all. However, as outlined in section 1, this approach is generally not feasible because algorithmic inaccuracies, which are inevitable in large scale simulations using efficient MD codes, represent heat drains or sources that have to be compensated.

To properly tune this compensation, we consider the work performed by the thermostatic forces $\mathbf{F}_{i,\,\mathrm{therm}}(t)$ on the atoms i for an ensemble of simulation systems with the temperature $T(t) = \langle \boldsymbol{T}(t) \rangle_{\mathrm{ens}}$. The ensemble average power exerted by the thermostat on a given atom i is

$$\beta_i(t) = \langle \mathbf{F}_{i,\mathrm{therm}}(t) \cdot \dot{\mathbf{r}}_i(t) \rangle_{\mathrm{ens}} \tag{10}$$

Using the definition of $\mathbf{F}_{i,\,\mathrm{therm}}$ [see eq 1] and the Berendsen expression 2 for γ leads to

$$\beta_i(t) = \frac{1}{\tau} \langle \varepsilon_{i,\mathrm{kin}}(t) [T_0/\boldsymbol{T}(t) - 1] \rangle_{\mathrm{ens}} \tag{11}$$

with the usual definition for the kinetic energy $\varepsilon_{i,\,\mathrm{kin}}(t)$ of atom i. Employing once more the assumption of a negligible correlation between the velocity [and, thus, the kinetic energy $\varepsilon_{i,\,\mathrm{kin}}(t)$] of a single atom and the kinetic temperature $\boldsymbol{T}(t)$ of the system, one obtains

$$\beta_i(t) = \frac{3k_B T_i(t)}{2\tau} [T_0/T(t) - 1] \tag{12}$$

where k_B is the Boltzmann constant, and $T_i(t) \equiv 2/3k_B \langle \varepsilon_{i,\,\mathrm{kin}}(t) \rangle_{\mathrm{ens}}$ is the ensemble average temperature of atom i. For equilibrated systems the ensemble averages employed in eq 12 can be replaced by temporal averages $\langle ... \rangle_D$. This allows to calculate for every subsystem κ from a simulation the (time) average thermostatic power

$$\hat{\beta}_\kappa = \frac{k_B \hat{T}_\kappa}{2\tau} [T_0/\hat{T} - 1] \tag{13}$$

per degree of freedom using the average temperature $\hat{T}_\kappa \equiv \langle \mathbf{T}_\kappa \rangle_D$ of the subsystem κ, the corresponding average $\hat{T} \equiv \langle \mathbf{T} \rangle_D$ of the temperature $\boldsymbol{T}(t)$ controlled by the BT, and the thermostat parameters T_0 and τ.

Further below we will use eq 13 to determine the thermostatic power exerted by a BT on a solute peptide from sample simulations. These data will be used to check the validity of a heat conduction model which we will now introduce to analyze the hot-solvent/cold-solute problem occasionally hampering MD simulations of inhomogeneous systems.[33]

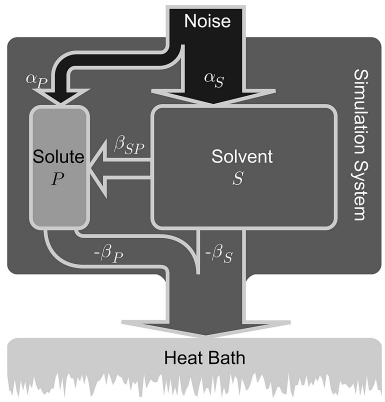


Figure 1. Heat flow model representing specifically the "hot-solvent/cold-solute" case for an inhomogeneous system consisting of two subsystems with different heating rates. The simulation system is coupled to a single thermostat, representing an external heat bath. Bright and dark colors code low and high temperatures, respectively. Heat flows driven by temperature gradients and heat sources are marked by arrows. A detailed discussion is given in the text.

Heat Flow Model. In simulations of solute−solvent systems, the algorithmic heat drains or sources may be inhomogeneously distributed and, thus, the temperature may likewise be inhomogeneous. According to requirement (iii) stated at the bottom of the Introduction, such inhomogeneous temperature distributions should be avoided. Figure 1 sketches a heat flow model from which one can derive strategies for the reliable control of the solute temperature. As drawn, the model refers to a particular strategy employing a single thermostat for the whole system. For further reference we denote this strategy by **G**.

The model depicted in Figure 1 consists of two subsystems $\kappa \in \{P, S\}$ with P denoting the solute and S denoting the solvent. The powers α_κ of algorithmic heating per DoF are assumed to be constant and homogeneous within the subsystems. Furthermore, the temperature is assumed to be homogeneous within each subsystem κ, i.e. for the atomic temperatures T_i we have $T_i = T_\kappa$ for all $i \in \kappa$. According to eq 12, the ensemble average work $\beta_i(t)$ exerted on atom i per unit time by the global thermostat then only depends on whether i is part of P or S, respectively. Thus, for the subsystems $\kappa \in \{P, S\}$ the respective thermostatic powers per DoF are given by

$$\beta_\kappa(t) = \frac{k_B T_\kappa(t)}{2\tau}[T_0/T(t) - 1] \tag{14}$$

If the local temperatures T_P and T_S differ, as is assumed in Figure 1, there will be a net heat flow

$$\beta_{SP}(t) = \frac{k_B[T_S(t) - T_P(t)]}{2\tau_{SP}} \tag{15}$$

between S and P, which we assume to depend linearly on the temperature difference. Here, the time constant τ_{SP} characterizes the thermal coupling of the subsystems.

The heat flowchart shown in Figure 1 immediately suggests stationarity conditions. In the steady state, the net heat flow must individually vanish for each of the two subsystems, i.e.

$$\alpha_P + \beta_{SP} + \beta_P = 0 \tag{16}$$

$$\alpha_S - \beta_{SP} + \beta_S = 0 \tag{17}$$

Now suppose for a moment that the temperature distribution is homogeneous throughout the system, i.e. $T_P(t) = T_S(t) = T(t)$. According to eq 14 the thermostatic powers $\beta_P(t)$ and $\beta_S(t)$ exerted on the subsystems are then equal, and, by eq 15, the heat flow $\beta_{SP}(t)$ between S and P vanishes. Equations 16 and 17 then immediately require as the stationarity condition that $\alpha_P = \alpha_S$, i.e. that the heat sources in the subsystems work at equal powers.

If this is not the case ($\alpha_P \neq \alpha_S$), the temperature cannot be homogeneously distributed in the stationary state, and, by eq 15, there will be a nonvanishing heat exchange $\beta_{SP} \neq 0$ between the subsystems. As a result, a steady state temperature difference is inevitable whenever, upon applying scenario **G**, a single global thermostat is used to thermostatize a system exhibiting inhomogeneities with respect to the rates α_κ of algorithmic heating. This is the origin of the hot-solvent/cold-solute problem as described e.g. in ref 34.

Separate Thermostats. To avoid temperature inhomogeneities, it has become a standard in simulations of macromolecules to couple *separate* thermostats to the subsystems.[36,47,52–57] We will denote this temperature control scenario by **P**. In the following discussion of scenario **P**, we will concentrate on the temperature control of the solute P, assuming that the temperature of the solvent S is reliably controlled. Such a reliable control can be achieved by a solvent thermostat combining a coupling time τ_S on the subpicosecond time scale (e.g., $\tau_S = 0.1$ ps) with a target temperature $T_{0,S}$ equal to the intended temperature. This choice of thermostat tuning actually is the standard (see e.g. refs 33, 36, 37, 47, 52–56), and, thus, we call it the classical setup.

For a scenario P, in which a separate BT is coupled to a (thermally homogeneous) solute P, the controlled temperature $T(t)$ is the solute temperature $T_P(t)$. Thus, we obtain from eq 14 the simplified expression

$$\beta_P(t) = \frac{k_B}{2\tau_P}[T_0 - T_P(t)] \tag{18}$$

for the power of the thermostat acting on P. With eqs 18 and 15, the solute's stationarity condition 16 may be rewritten as

$$\alpha_P + \frac{k_B(T_S - T_P)}{2\tau_{SP}} + \frac{k_B(T_{0,P} - T_P)}{2\tau_P} = 0 \tag{19}$$

where $T_{0,P}$ denotes the target temperature, and τ_P denotes the coupling time of the solute thermostat. The first term characterizes the algorithmic heating within P, the second term characterizes the heat flow between P and S, and the third term characterizes the power β_P of the thermostat separately coupled to P.

Equation 19 is the quintessence of our linear heat flow model and may be used to predict the effects of three different thermostatic strategies within scenario **P**. In all these strategies, S is coupled to a classical BT and P is decoupled from this thermostat. The three strategies are as follows:

P.1 The solute P is coupled to a classical thermostat. Here, the use of a correspondingly small coupling time $\tau_P \approx 0.1$

ps is the standard.[36,52–56,58] For such small τ_P, eq 19 is completely dominated by the thermostatic term. The reason for this dominance is that τ_P is by at least 1 order of magnitude smaller than the solute−solvent coupling time τ_{SP}, which is typically larger than 1 ps (see further below). Neglecting the heat flow contribution, the deviation $T_P - T_{0,P}$ from the target temperature is given by $2\alpha_P\tau_P/k_B$. For moderate algorithmic heating rates α_P, this deviation is expected to be small because of the short time scale τ_P.

P.2 No separate thermostat is coupled to the solute *P*, i.e. $\tau_P \rightarrow \infty$, and solely the thermostatted solvent *S* acts as a heat bath. We call this strategy "noninvasive" because it does neither alter the Newtonian dynamics nor the energy relaxation properties of *P*. The expected temperature difference $T_P - T_S = 2\alpha_P\tau_{SP}/k_B$ will be small if the local heating α_P is negligible on the time scale τ_{SP} of the thermal coupling between the subsystems.

P.3 The solute is coupled to a thermostat with a very large coupling time $\tau_P \gg \tau_{SP}$ to realize a "constant heat flow" (CHF) approach. As suggested by the heat balance eq 19, a homogeneous ($T_P = T_S$) and stationary temperature distribution only requires that the thermostatic power β_P cancels the power α_P of algorithmic heating, i.e., $\alpha_P = k_B(T_P - T_{0,P})/2\tau_P$. This condition can be satisfied for an arbitrarily large coupling time τ_P by a proper choice of the target temperature $T_{0,P}$. In the limit $\tau_P \rightarrow \infty$, the thermostat variable γ in eqs 1 and 2 becomes a constant γ_P, and the thermostat scheme may actually be descibed by this single parameter. At large τ_P, the thermostat works in a heating/cooling limit as a constant heat source/drain, and this activity solely serves to maintain the energy balance. Because of eq 2, the perturbation of the Newtonian dynamics [cf. eq 1] inflicted by such a CHF thermostat can be made very small. Therefore, we call the CHF approach to the solute's temperature control, which is applicable to non-negligible local heating rates α_P, "minimally invasive".

To set up a CHF simulation as required in strategy **P**.3, the a priori unknown power α_P of algorithmic heating has to be determined in order to specify the constant thermostat parameter γ_P, or, equivalently, the paramteres $T_{0,P}$ and τ_P if a traditional Berendsen thermostat is used in the heating/cooling limit. To this end, the solute temperature T_P has to be measured in two test simulations with different heating powers β_P of the thermostat. The two heat balance eqs (19) of these tests then constitute a system of linear equations which determines the unknown parameters β_{SP} and α_P. A detailed description of the setup protocol is given in Appendix A.

In the following we will examine the temperature control strategies **G** and **P**.1 −**P**.3 introduced above by test simulations. Based on these results, we formulate guidelines for a temperature control satisfying the four conditions summarized at the bottom of the Introduction.

3. Methods

MD Simulation Techniques. The software packages EGO-MMII[27] and GROMACS[59] were used in several series of MD simulations. Besides EGO we also applied GROMACS because it provides an NHT in addition to a BT, because it is computationally efficient for very small systems, and because it can provide data for a crosscheck of results. In EGO the electrostatic interactions are treated combining structure-adapted multipole expansions[60,61] with a moving-boundary reaction-field approach[62] and a multiple-time-step integration.[29,63] In the GROMACS simulations we used the PME method[28] with a 10 Å cutoff for the real space contribution, with a grid spacing of 0.5 Å, and with a sixth order interpolation of the charges to the grid. For both EGO and GROMACS simulations, the van der Waals interactions were truncated at 10 Å. If not stated otherwise, the simulations were carried out with explicit solvent using periodic boundary conditions and with a BT ($\tau_S = 0.1$ ps, $T_{0,S} = 300$ K) rapidly coupled to the solvent to guarantee that the solvent was closely kept at the target value. Bond lengths were constrained using the M-SHAKE algorithm[30] with relative tolerances of 10^{-4} when using GROMACS, which is the recommended default value, and 10^{-6} when using EGO, which is hard-coded in the source code in this case.

We applied different simulation protocols to vary the heating properties within the simulation systems. Here, the first parameter was the software used for simulation, which we denote by E for EGO or G for GROMACS. Since the M-SHAKE algorithm is known[30] to have a cooling effect, we varied the number of constraints by either constraining no bonds at all (N), only bonds involving a hydrogen atom (H), or all bonds (A). The last parameter which presumably influences the heating in the system is the length Δt of the basic integration time step, which we simply denote by its value in femtoseconds. Thus, a standard EGO simulation (constraints on bonds involving hydrogen atoms and $\Delta t = 1$ fs) would be denoted by E/H/1.

Model Systems. The first model system was a polyalanine octapeptide (8ALA) with charged termini described by the GROMOS96 force field[64] and embedded in a cubic box of 20 Å edge-length containing 236 simple point charge (SPC) water molecules.[65] The number of DoF for the peptide then is 153/143/103 for N-/H-/A-constraining, respectively. The starting conformation was always fully extended. The system was equilibrated for 300 ps during which solute and solvent were coupled to separate BTs ($\tau_P = \tau_S = 0.1_{\text{ps}}$, $T_{0,S} = T_{0,P} = 300$ K). The second model system was an alanine dipeptide (ALDI) described by the CHARMM22 force field[66] in a cubic box of 21.3 Å edge length containing 324 water molecules modeled by the transferable three-point intermolecular potential (TIP3P).[66,67] Here, the number of peptide DoF is 66/54/45 for N-/H-/A-constraining, respectively. The system was prepared as described for 8ALA, except that the equilibration time was only 100 ps.

MD Simulations. A first series of seven MD simulations of 8ALA in SPC water served to study the various situations encountered in the temperature control of inhomogeneous systems. Table 1 associates acronyms to these simulations and lists the employed parameters. In particular, in the last simulation G/A/2_**P**.3, the CHF approach was applied to the peptide. Using the data from the preceding simulation G/A/2_**P**.2 ($\tau_P = \infty$, $\hat{T}_P = 293.4$ K), in which only *S* was coupled to a classical BT and the data from an independent 10 ns test simulation with an additional CHF thermostat coupled

Table 1. Simulation Names and Associated Parameters in Series #1[a]

	protocol				thermostat parameters			
name	software	C	Δt/fs	D/ns	τ_{sys}/ps	τ_S/ps	τ_P/ps	$T_{0,P}$/K
E/H/1_**G**	EGO	H	1	20	0.1	—	—	—
E/H/2_**G**	EGO	H	2	20	0.1	—	—	—
E/H/2_***P***.2	EGO	H	2	20	—	0.1	—	—
G/H/2_***P***.2	GROMACS	H	2	20	—	0.1	—	—
G/A/2_***P***.2	GROMACS	A	2	20	—	0.1	—	—
G/A/2_***P***.1	GROMACS	A	2	20	—	0.1	0.1	300
G/A/2_***P***.3	GROMACS	A	2	20	—	0.1	500	2340

[a] The simulation names code the varied parameters and temperature control scenarios. C specifies the type of bond length constraints, Δt the size of the basic integration time step, and D the duration of the simulation. The parameters τ specify the coupling times of the BTs coupled to the whole system (sys), to the solvent (S), or to the solute (P). $T_{0,P}$ is the target temperature of a thermostat coupled to the solute. The solute peptide was 8ALA in SPC water. See the text for further information.

Table 2. Simulation Parameters in Series #2[a]

	protocol				thermostat parameters			
name	software	C	Δt/fs	D/ns	τ_{sys}/ps	τ_S/ps	τ_P/ps	$T_{0,P}$/K
CHF.0	GROMACS	A	2	200 × 2	—	0.1	—	—
CHF.1	GROMACS	A	2	200 × 2	—	0.1	500	2340
CHF.2	GROMACS	A	2	200 × 2	—	0.1	500	4800
CHF.3	GROMACS	A	2	200 × 2	—	0.1	500	7700
CHF.4	GROMACS	A	2	200 × 2	—	0.1	500	11100
CLS.1	GROMACS	A	2	200 × 2	—	0.1	0.1	300
CLS.2	GROMACS	A	2	200 × 2	—	0.1	0.1	340

[a] The model peptide was 8ALA in SPC water at $\hat{T}_S = 300$ K. Except for the simulation set CHF.0, in which only T_S was controlled, separate BTs were applied to S and P. See the caption to Table 1 for further information.

to the peptide ($\tau_P = 500$ ps, $T_{0,P} = 4800$ K, $\hat{T}_P = 307.9$ K), the unknown parameters in eq 19 were determined as described in the section 2. We found the values $\alpha_P = -2.04 k_B$ K/ps for the algorithmic heating rate and $\tau_{SP} = 1.61$ ps for the solute−solvent coupling time, which actually is in the picosecond time range as claimed further above. To realize a CHF thermostat maintaining the peptide at $T_P \approx 300.0$ K, these values were inserted into eq 19 yielding a "target temperature" $T_{0,P} = 2340$ K. If the assumptions underlying our heat-flow model are correct, this choice should compensate through $\beta_P = -\alpha_P$ the algorithmic energy drift in the G/A/2_**P**.3 simulation.

The setup of a second series of simulations was chosen such that the effects of the local temperature and of a thermostat on the dynamics of 8ALA can be studied. We performed seven sets of 200 simulations each. Every single simulation had a duration of 2 ns, amounting to 400 ns per set and a total of 2.8 μs of simulation time. The simulation parameters are summarized in Table 2. All simulations were performed with the G/A/2 protocol. In the first set (CHF.0), no thermostat was coupled to the peptide, while in the following four sets (CHF.1 to CHF.4) a BT targeting at increasingly large temperatures $T_{0,P}$ was coupled in an extremely slow fashion to the peptide. In the last two sets (CLS.1 and CLS.2), a separate classical BT was coupled to the peptide using either the same ($T_{0,P} = 300$ K) or a slightly higher ($T_{0,P} = 340$ K) target temperature as compared to $T_{0,S}$. The 200 initial conditions were obtained by taking snapshots every 20 ps from a 2 ns preparatory simulation at 300 K, with the peptide's C_α atoms harmonically coupled to their initial coordinates of an extended conformation.

To compare the different thermostatic strategies discussed in section 2, we determined the corresponding thermostatic forces (eqs 5 and 6) and perturbation ratios (eq 4) in a third series of relatively short 250 ps simulations. Simulations were performed for 8ALA with varying coupling strengths and BTs and NHTs, respectively. Additionally, we determined the thermostatic forces and the perturbation ratio also for ALDI and Berendsen coupling again varying the coupling strength. The simulation parameters of the third series are given in Table 3. As these simulations served to compare thermostatic and force-field forces, no bond lengths were constrained thus eliminating constraint forces.

Table 3. Simulation Parameters in Series #3[a]

	protocol				thermostat parameters		
peptide	software	C	Δt/fs	D/ns	τ_{sys}/ps	τ_P/ps	$T_{0,P}$/K
8ALA	GROMACS	N	1	0.25	NHT	0.064	300
8ALA	GROMACS	N	1	0.25	NHT	0.256	300
8ALA	GROMACS	N	1	0.25	NHT	1.024	300
8ALA	GROMACS	N	1	0.25	NHT	4.096	300
8ALA	GROMACS	N	1	0.25	NHT	16.384	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.001	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.004	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.016	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.064	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.256	300
8ALA/ALDI	EGO	N	1	0.25	BT	1.024	300
8ALA/ALDI	EGO	N	1	0.25	BT	4.096	300

[a] For nomenclature see the caption to Table 1. In all simulations the solvent was coupled with $\tau_S = 0.1$ ps to a Nosé-Hoover (NHT) or a Berendsen (BT) thermostat, respectively.

Finally, a fourth series of slightly more extended simulations (500 ps) was designed to examine how the solute's variance of temperature fluctuations (cf. the corresponding paragraph in section 2) is affected by the coupling times of a BT. We studied 8ALA and ALDI in water and in vacuum

Table 4. Simulation Parameters in Series #4[a]

system		protocol				thermostat parameters	
peptide	environment	software	C	Δt/fs	D/ns	τ_P/ps	$T_{0,P}$/K
8ALA	water/vac	EGO	A	2	0.5	0.001	300
8ALA	water/vac	EGO	A	2	0.5	0.004	300
8ALA	water/vac	EGO	A	2	0.5	0.016	300
8ALA	water/vac	EGO	A	2	0.5	0.064	300
8ALA	water/vac	EGO	A	2	0.5	0.256	300
8ALA	water/vac	EGO	A	2	0.5	1.024	300
8ALA	water/vac	EGO	A	2	0.5	4.096	300
ALDI	water/vac	EGO	H	2	0.5	0.001	300
ALDI	water/vac	EGO	H	2	0.5	0.004	300
ALDI	water/vac	EGO	H	2	0.5	0.016	300
ALDI	water/vac	EGO	H	2	0.5	0.064	300
ALDI	water/vac	EGO	H	2	0.5	0.256	300
ALDI	water/vac	EGO	H	2	0.5	1.024	300
ALDI	water/vac	EGO	H	2	0.5	4.096	300

[a] For nomenclature see the caption to Table 1. BTs were used for solvent and solute. The solvent was coupled with a coupling time of 0.1 ps.

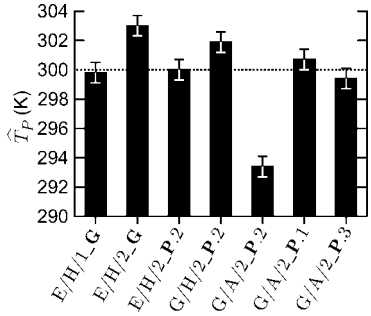


Figure 2. Average peptide temperature $\hat{T}_P$ observed in the first series of simulations on 8ALA in SPC water. The associated acronyms and parameters characterizing the members of the series are given in Table 1.

by E/H/2 simulations using the same set of coupling times as in series #3. Table 4 summarizes the simulations of the last series.

4. Results and Discussion

Temperature Control Scenarios. As outlined above, the series of equilibrium simulations on the model peptide 8ALA in explicit water as characterized by Table 1 served to exemplify the problems connected with the temperature control of inhomogeneous systems. Figure 2 shows the average peptide temperatures obtained in these sample simulations. Using eq 9, the remaining uncertainty of these average temperatures was estimated to $\sigma_{\hat{T}_P} < 0.7$ K. The solvent temperatures were 300.0 K where not mentioned explicitly.

In simulation E/H/1_**G**, we used the standard simulation protocol for EGO (see section 3 for details), which includes a classical BT coupled to the whole simulation system and, thus, represents an example for scenario **G** outlined in section 2. Neither the resulting temperatures of the peptide (cf. Figure 2) nor of the solvent showed any statistically significant deviations from the 300 K target value suggesting that in E/H/1_**G** the algorithmic noise was weak.

Figure 2 indicates that this behavior was lost in simulation E/H/2_**G**, in which the basic integration time step Δt was doubled to 2 fs. For our sample system, this doubling of Δt led to a 3.0 K increase of the peptide temperature, indicating that the modified simulation setup has caused certain algorithmic inaccuracies. When using EGO, the choice of a larger Δt is expected to reduce the accuracy of the integration algorithm because the employed highly efficient multiple-time-step algorithm does not exactly guarantee energy conservation and because the corresponding violation increases with the size of Δt (see refs 63 and 29 for a discussion). According to Figure 2, the combination of a global Berendsen thermostat with a reduced accuracy of integration in simulation E/H/2_**G** apparently led to a moderately elevated temperature for the peptide and to a slightly (0.3 K) cooler temperature for the larger solvent system. Nevertheless, the temperature of the total system was accurately kept at 300.0 K by the thermostat. Apart from changed signs (hot solute in cold solvent), this result is an example for the classical problem reported in the literature,[32,34,68] which can arise in scenario **G** from indiscriminately coupling a thermostat to all parts of an inhomogeneous system and which is described by the heat flow model sketched in Figure 1.

However, as demonstrated by the average peptide temperature displayed in Figure 2 for simulation E/H/2_**P**.2, this temperature control problem was eliminated by simply decoupling the peptide from the thermostat, i.e. by realizing scenario **P**.2. This observation suggests that in the E/H/2 simulations the solvent experiences a considerable cooling, whereas the level of algorithmic noise within the peptide is very low. According to our experience, such a decoupling of the solute is a proper solution for most temperature control problems which can occur in simulations of inhomogeneous systems using either EGO or GROMACS.

The fact that the application of scenario **P**.2 cannot always remove such problems is demonstrated by the results of simulation G/H/2_**P**.2, which was carried out with GROMACS using the same settings as in the EGO simulation E/H/2_**P**.2. According to Figure 2, in the G/H/2_**P**.2 simulation the peptide was by about 2 K too hot, indicating that the rate β_{SP} of heat transport from the peptide P into the solvent S was too slow to compensate the algorithmic heating $\alpha_P > 0$ of the solute occurring in this case.

It may be expected that introducing additional M-SHAKE constraints into the peptide system leads to a local cooling,[30] which might compensate the observed algorithmic heating of P. This is the reason why we carried out simulation G/A/2_**P**.2, which differs from G/H/2_**P**.2 only in the number of constraints (50 vs 10) within the peptide. In fact, Figure 2 displays for simulation G/A/2_**P**.2 a peptide temperature which is by 6.6 K cooler than that of the solvent, implying that the original heating has been overcompensated by the local cooling. A deviation of this size is unacceptable in simulations serving to probe the equilibrium properties of the solute. Thus, the simulation setup G/A/2 is a typical case in which one of the two remaining temperature control strategies **P**.1 and **P**.3 described in the section 2 should be applied.

Hence, in simulation G/A/2_**P**.1 we utilized a separate classical BT for temperature control of the peptide, while in simulation G/A/2_**P**.3 we applied a CHF thermostat. Figure 2 shows that in both cases there is no significant deviation of the observed peptide temperatures from the solvent temperature. Both methods are capable of correctly thermostatting the solute. For the CHF thermostat we conclude that the choice of parameters (cf. section 3) was correct and that the underlying heat flow model describes the situation in this case. This success has motivated us to further scrutinize the validity of this model.

Validity of the Heat Flow Model. The second quite extended series of simulations (see Table 2 for the parameters) can serve to assess the validity of eq 19, which expresses the contents of the model. With eq 18 the model 19 can be equivalently reformulated as

$$T_P = T_S + \frac{2\tau_{SP}}{k_B}(\beta_P + \alpha_P) \tag{20}$$

showing that the solute temperature T_P should depend linearly on the heating power β_P of the solute thermostat. To specify the unknown parameters α_P and τ_{SP} in eq 20, one needs measurements of T_S and T_P from two simulations employing different heating powers β_P.

Estimates $\hat{\beta}_P$ for the heating powers β_P can be determined from simulations by evaluating eq 13 specifically for the case of a solute thermostat, i.e. for $\kappa = P$, $T_0 = T_{0,P}$, $\tau = \tau_P$, and $\hat{T} = \hat{T}_P$. One obtains

$$\hat{\beta}_P = \frac{k_B}{2\tau_P}[T_{0,P} - \hat{T}_P] \tag{21}$$

which is, up to the use of different averages, identical to eq 18. Thus, at a constant coupling time τ_P, the heating power β_P is steered by the choice of the target temperature $T_{0,P}$ and measured through the average peptide temperature $\hat{T}_P$. Therefore, the linear relationship 20 between T_P and β_P can be checked by comparing with data points $(\hat{T}_P,\hat{\beta}_P)$ obtained from simulations employing different target temperatures $T_{0,P}$.

An inspection of the first five simulation sets in series #2 listed in Table 2 shows that this set qualifies both for the evaluation of the unknown parameters in eq 20 and for the check of this linear equation. In all simulation sets of series #2, the simulation protocol was G/A/2 just like in the simulation G/A/2.**P**.2 of the first series. However, the temperature control scenario **P**.2 (no separate thermostat for the peptide) was employed only in simulation CHF.0. In the remaining CHF simulations a BT was coupled to P using an extremely slow coupling time $\tau_P = 500$ ps combined with a large and increasing target temperature (cf. Table 2). According to eq 21 this choice leads to a heating power β_P of this thermostat, which increases from simulation CHF.0 ($\beta_P = 0$) to simulation CHF.4. Figure 3 shows the observed stationary peptide temperatures $\hat{T}_P$ as a function of the observed heating power $\hat{\beta}_P$. In the case of the simulation set CHF.0 (black dot) the result of simulation G/A/2_**P**.2 (cf. Figure 2) is closely recovered because the same temperature control setting **P**.2 was applied, i.e. $\hat{T}_P$ was by 6.5 K smaller than the solvent temperature of $\hat{T}_S = 300$ K. With nonzero and successively growing $\hat{\beta}_P$ the peptide temperature $\hat{T}_P$ is seen to increase.

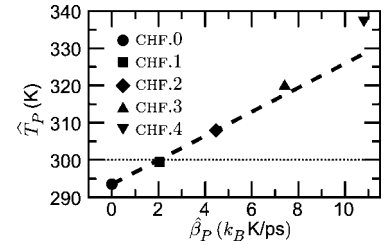


Figure 3. The average temperatures $\hat{T}_P$ of the peptide ALA8 (in SPC water at $\hat{T}_S = 300$ K) resulting from constant local heating with different powers β_P in the simulation sets CHF.0 to CHF.4 of series #2 (cf. Table 2). The prediction of linear heat flow model eq 20 is drawn as a dashed line, and the solvent temperature $\hat{T}_S$ is indicated as a dotted line (see the text for explanation).

The dashed line in Figure 3 expresses the linear relation 20 between β_P and T_P. The required parameters were determined as $\alpha_P = -2.02k_B$ K/ps and $\tau_{SP} = 1.60$ ps from the simulation sets CHF.0 and CHF.2. Therefore, the dashed line linearly interpolates between the data points $(\hat{\beta}_P,\hat{T}_P)$ of these two simulation sets. The above values of the parameters α_P and τ_{SP} closely agree with those calculated earlier (see Methods) for setting up the CHF thermostat used in simulation G/A/2.**P**.3. This result is expected because in both cases the parameters α_P and τ_{SP} were computed from simulations employing the same parameters.

In simulation set CHF.1, the peptide temperature was nearly identical to $\hat{T}_S$ with $\hat{T}_P = 299.5$ K (black square in Figure 3) because here the thermostat parameters were chosen equal to those of the simulation G/A/2.**P**.3 (series #1), which realizes the **P**.3 strategy. The temperature T_P predicted for CHF.1 by the dashed line deviates by only 0.5 K from the observed average. This deviation is probably significant because the temperature averages shown in the figure are extremely well converged ($\sigma_{\hat{T}_P} < 0.1$ K) due to the extended statistics. If a similar interpolation would be constructed using the data from the simulation sets CHF.3 or CHF.4 instead of CHF.2, the error in the prediction for CHF.1 would increase to 1.1 K or 2.2 K, respectively, with increasing violation of the approximate linear relation 20 between β_P and T_P. In the case of 8ALA in explicit water, the assumption of a linear thermal coupling between solvent

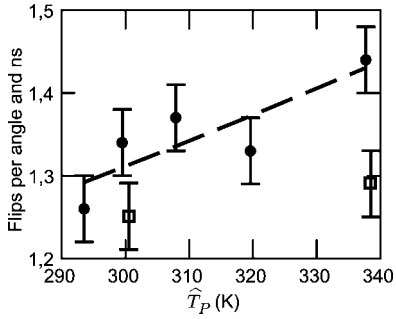


Figure 4. Temperature dependence of the peptide backbone dynamics of 8ALA. The graph shows the average number of transitions per angle and nanosecond of the ψ-dihedral angles between the α-type region [−60°,−30°] and the β-type region [95°, 145°] for the five CHF simulation sets (filled circles) and the two CLS sets (empty squares) over the observed average peptide temperature $\hat{T}_P$. The error bars give the range of plus/minus one standard deviation. Additionally, an Arrhenius[69] model (dashed line) fitted to the CHF data is plotted. The simulation parameters are summarized in Table 2.

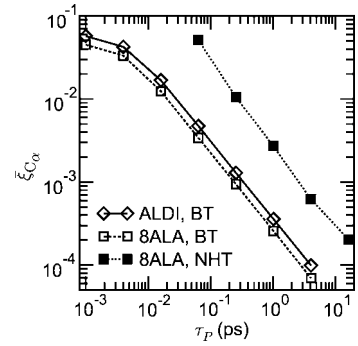


Figure 5. Root mean perturbation quotients $\bar{\xi}_{C\alpha}$ at the C_α atoms of 8ALA and ALDI evaluated from simulation series #3 for the NHTs and BTs, respectively, for different coupling times τ_P of the peptide thermostats.

and solute (eq 15), thus, obviously breaks down if $\hat{T}_P$ deviates by more than about 10 K from $\hat{T}_S$, which is probably also true for related simulation systems. In test simulations serving to set up a CHF thermostat through eq 19, the deviation $|\hat{T}_P - \hat{T}_S|$ should, thus, be smaller than about 10 K if one wants to guarantee an accurate tuning of T_P in applications of strategy **P**.3.

Backbone Dynamics. As we have seen further above, the use of an inappropriate strategy for temperature control can lead to peptide temperatures considerably deviating from that of the solvent. It seems likely that such a deviation can entail an altered conformational dynamics of the peptide. To check this expectation, we analyzed the second simulation series also in this respect. Due to the extremely slow thermostat coupling employed in CHF.0 to CHF.4, here, the dynamics should be exclusively affected by differences in the peptide temperatures.

Figure 4 shows how the kinetics of conformational transitions in 8ALA is modified by T_P in CHF.0−4 (black dots). This kinetics is measured by local flip rates of backbone torsional angles (see the figure caption). As expected, the flip rates increase with the temperature. A simple Arrhenius model[69] fitted to the CHF data is drawn as a dashed line. This model yields an energy barrier of $434k_B$ K for the backbone flips. This value is well in the range of typical barrier heights reported for biomolecules in the literature.[70]

Having estimated the influence of the temperature on the conformational dynamics of our sample peptide 8ALA in SPC water, it seems appropriate to check whether a separate classical BT (as frequently applied in strategy **P**.1) changes the dynamics. Here, particularly a slowing down seems possible because a rapidly coupled thermostat can interfere with long-lasting energy fluctuations within the peptide, which are caused by random in- and outflow of energy from the solvent. For the purpose of such a check, we carried out the simulation sets CLS.1−2 listed in Table 2, in which a classical BT separately coupled to P enforced temperatures $\hat{T}_P$ of about 300 K and 340 K, respectively.

Figure 4 compares the flip rates observed when using a classical Berendsen thermostat (open squares) with the data for the CHF thermostat (filled circles) and demonstrates that our expectation is actually met. Thus, if one wants to sample the equilibrium fluctuations of a peptide in solution by MD as rapidly as possible, or if one wants to gain access to the kinetics of nonequilibrium relaxation processes, the separate coupling of a classical BT to a small peptide seems counterproductive.

We interpret the above result by the following physical picture: A rapidly coupled BT likewise dampens fluctuations to higher and lower energies, thus leading to the correct average temperature. However, barrier crossings are enabled by rare accidental accumulations of a critical amount of energy in the respective collective coordinates. Particularly by dampening the higher energy fluctuations of the peptide, a classical BT makes such accumulations and, thus, barrier crossings less likely. Note that we have additionally checked the performance of a NHT in the same setting. We found no reduction of flip rates (data not shown) as could be expected for a thermostat maintaining the canonical energy fluctuations.

Local Perturbations of the Dynamics. The flips of backbone dihedral angles are collective movements and, therefore, are not directly related to the perturbation which a thermostat inflicts on the dynamics of individual atoms. To check the latter, we collected from simulation series #3 (cf. Table 3) all those forces acting on the C_α atoms of 8ALA which are required for the evaluation of the perturbation quotients (4). We carried out this data collection for BTs and NHTs with coupling times τ_P covering 4 orders of magnitude. In the case of the smaller ALDI model, we concentrated on the Berendsen approach.

Figure 5 shows the resulting perturbation ratios (4) evaluated using the approximate expression 6. As demonstrated by the squares marking the 8ALA results, the perturbations $\bar{\xi}_{C\alpha}$ are small for both thermostats and decrease over a wide range linearly with the inverse of τ_P. For the classical BT ($\tau_P = 0.1$ ps) the $\bar{\xi}_{C\alpha}$ are only about 0.5%. Furthermore, the smaller ALDI model exhibits slightly larger $\bar{\xi}_{C\alpha}$ (open diamonds) than 8ALA (open squares). However, this size-induced difference is much smaller than that between the NHTs and BTs. At a given τ_P, Nosé-Hoover coupling inflicts perturbations which are by 1 order of

magnitude larger than in the Berendsen case (cf. Figure 5). For a Berendsen coupling of maximal strength ($\tau_P = 0.001$ ps) the perturbation is comparable to that of a NHT with τ_P as large as 0.064 ps. Furthermore, for Nosé-Hoover coupling τ_P cannot be chosen larger than about 0.256 ps where $\bar{\xi}_{C\alpha}$ is about 1% and, thus, not particularly small. In the given case of 8ALA, one otherwise observes long-lasting and artificial temperature oscillations, i.e. the so-called Toda daemon[44] (data not shown).

One can compare the perturbations shown in Figure 5 to those which are inflicted by a CHF thermostat as employed in strategy **P**.3. In simulation G/A/2_**P**.3, the peptide 8ALA was kept at 300 K with a perturbation ratio of $\bar{\xi}_{C\alpha} \approx 10^{-4}$. As can be seen from Figure 5, this ratio corresponds to a Berendsen coupling time larger than 1 ps in the classical thermostat setup. However, a classical BT with $T_{0,P} = T_S$ and $\tau_p \geq 1$ ps cannot properly control the temperature because then τ_p is in the range of solvent−solute coupling time ($\tau_{SP} = 1.6$ ps), i.e. is too slow (cf. section 2). On the other hand, a more strongly coupled thermostat with $\tau_P = 0.1$ ps does the job, but then the perturbation is more than ten times stronger than for a CHF thermostat.

The above analysis was based on data for perturbation ratios derived through the approximate expression 6 and, therefore, depends on the validity of this equation. The first assumption made in the derivation (cf. section 2) of eq 6 was that the atomic velocities $\dot{\mathbf{r}}_i(t)$ and the thermostat variable $\gamma(t)$ are uncorrelated. We have checked this assumption for simulation series #3 by evaluating eq 5 with and without taking the correlation into account; the relative difference was less than 10^{-2} for both 8ALA and ALDI (data not shown).

The second assumption was that the individual atomic velocities $\dot{\mathbf{r}}_i(t)$ are drawn from an undisturbed Maxwell distribution and can be checked by comparing results of the exact expression 5 with results of the approximate expression 6. We evaluated these expressions for the trajectories of series #3 and determined the root-mean-square deviations. In the worst case of a BT at the maximum coupling strength ($\tau_P = 0.001$ ps), we found root-mean-square deviations amounting to 8.3% of the mean thermostatic force for 8ALA and to 14% for ALDI. In view of the moderate statistics provided by the 250 ps simulations employed in series #3, the estimate 6 is fairly reliable. Thus, eq 6 is adequate if one wants to estimate thermostatic forces.

Temperature Fluctuations. In our suggestion of the minimally invasive CHF thermostat characterizing strategy **P**.3 we were guided by the notion that a properly thermostatted explicit solvent system is a canonical heat bath for an uncontrolled solute. To check this assumption, we compare in Figure 6 the canonical χ^2-distribution (eq 7) for the instantaneous peptide temperature $\boldsymbol{T}_P(t)$ with results from simulation G/A/2_**P**.3. For the 103 degrees of freedom of 8ALA, the χ^2-distribution (solid line) resembles a Gaussian (dashed line), which is expected for very large systems. Remarkably, the MD results (circles) closely reproduce the slight asymmetry of the χ^2-distribution. This agreement strongly indicates that the peptide has sampled the canonical ensemble in the simulation G/A/2_**P**.3. We have verified this

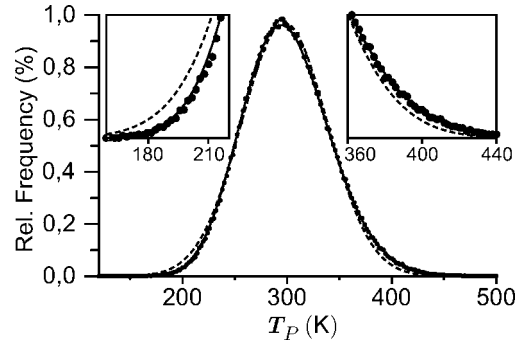


Figure 6. Distribution of the instantaneous temperature $\boldsymbol{T}_P(t)$ of 8ALA (in SPC water at 300 K) during the 20 ns MD simulation G/A/2_**P**.3 (dots). The dashed line is a Gaussian fit to the data. The canonical distribution (eq 7) is drawn as a solid line.

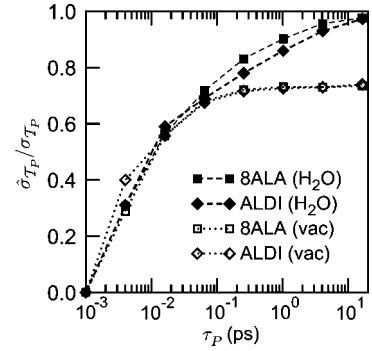


Figure 7. Ratio $\hat{\sigma}_{T_P}/\sigma_{T_P}$ of measured and canonical temperature fluctuations for various coupling times τ_P of a Berendsen solute thermostat. The model peptides are 8ALA (squares) and ALDI (diamonds). Simulations were performed in explicit water (H_2O, filled symbols) and vacuum (vac, empty symbols) for both peptides. Simulation parameters are given in Table 4.

result for a series of further CHF simulations. It did not change for larger solvent systems and was independent of the coupling time for the solvent thermostat provided that the solvent temperature remained well-tuned (data not shown).

To estimate how a classical BT separately coupled to a peptide (strategy **P**.1) affects its global statistical properties, we determined the temperature fluctuations of the peptides 8ALA and ALDI, respectively, as measured by the standard deviation $\hat{\sigma}_{T_P}$ in a fourth series of simulations (for details see Table 4). Figure 7 shows the ratio of $\hat{\sigma}_{T_P}$ and σ_{T_P}, which is the value theoretically expected for a canonical ensemble and is given by eq 8. For peptides in explicit solvent the figure shows that $\hat{\sigma}_{T_P}/\sigma_{T_P}$ is always smaller than one and approaches that limit for large τ_P. Thus, in the classical setting ($\tau_P = 10^{-2}$ ps) a BT strongly suppresses the canonical temperature fluctuations. These fluctuations successively become restored with increasing τ_P. The full range of canonical fluctuations is reached at coupling times $\tau_P > 10$ ps, i.e. at values exceeding the solvent-peptide heat coupling time τ_{SP} by a factor of 10. As a result, the separate BT is effectively disconnected from the peptide, the solute−solvent heat exchange term β_{SP} dominates the heat balance eq 19, and strategy **P**.1 reduces to the noninvasive strategy **P**.2.

Figure 7 not only reveals the general suppression of temperature fluctuations within a peptide by a classical BT but also demonstrates through a comparison with vacuum simulation data that these fluctuations are caused (i) by a fast exchange of kinetic and potential energy within a peptide and (ii) by a slower energy exchange with the solvent.

In vacuum simulations, the exchange of kinetic and potential energy within the peptide is the only cause of temperature fluctuations. As shown by the data, a rapidly coupled Berendsen thermostat ($\tau_P < 0.1$ ps) suppresses these microcanonical fluctuations in the same way as it suppresses the canonical temperature fluctuations of a solvated peptide. However, at slower coupling times τ_P the thermostat is seen to no longer affect the microcanonical fluctuations. The clear saturation of $\hat{\sigma}_{T_P}/\sigma_{T_P}$ at $\tau_P > 0.1$ ps demonstrates that the microcanonical fluctuations occur on time scales below 0.1 ps. In contrast, additional fluctuations of a solvated peptide are still suppressed by the thermostat with even slower coupling. Thus, as claimed above, they occur on longer time scales.

In order to retain the correct statistics for the solute, it is important to choose the coupling time τ_P for the thermostat longer than the typical time scale of the canonical fluctuations, which, in our case, is in the range of 10 ps, as can be seen from Figure 7. However, this time may even be longer for more weakly coupling solvents or larger solutes.

5. Conclusions

Every thermostat changes the dynamics of the controlled system to a larger or lesser extent. Measured on a microscopic scale, these changes are by about 1 order of magnitude smaller for BTs than for NHTs (cf. the data on the perturbation quotients displayed in Figure 5). On the other hand, NHTs, in contrast to BTs, guarantee the canonical ensemble. For instance, as shown by the results on the temperature fluctuations (Figure 7), BTs suppress all those canonical energy fluctuations which are slower than the time scale τ at which the BT is coupled to the system.

Whether such changes can modify the specific observables to be extracted from a simulation and to be compared with experimental data is *a priori* unclear in many cases. Even if one suspects that a given thermostat could possibly introduce an artifact into the computation of a certain observable, one may have to spend an enormous computational effort for a statistically clear proof. In fact, to prove a suspected dampening of peptide flip rates by a standard BT, we had to spend about 400 ns of simulation time on each of the data points to get the statistical certainty shown in Figure 4.

Especially if the popular strategy **P**.1 is applied to a solute−solvent system, the specific drawbacks of the various thermostat algorithms may directly affect the properties of the solute. The **P**.1 strategy with a BT is expected to cause artifacts of type a), i.e. artifacts resulting from an incorrect thermodynamical ensemble. In fact, as we have shown for a sample peptide, the dampening of the canonical energy fluctuations due to the BT can lead to reduced peptide flip rates. Furthermore, one expects that the combination of **P**.1 with the NHT will render the solute vulnerable to artifacts of type b), i.e. lacking ergodicity. Using the **P**.1 strategy with other thermostats which suffer neither from type a) nor type b) drawbacks (e.g., the Nosé-Hoover chain) still perturbs the dynamics much more strongly than necessary, i.e. such a strategy is prone to introduce artifacts of type c) (dynamics).

Given the need for some sort of temperature control in large scale MD simulations of complex systems, the optimal strategies to avoid artifacts of types a), b), and c) are **P**.2 or **P**.3, respectively. Here, the minimally invasive strategy **P**.3, which employs a constant heat flow to compensate the algorithmic heat production in the solute, has to be applied only if the noninvasive strategy **P**.2 turns out to be ineffective in a sufficiently extended test simulation. Strategy **P**.3 reduces the perturbation of the solute's dynamics to a minimum while keeping it nevertheless properly tempered. The precise protocol to set up a **P**.3 scheme is given in the Appendix.

The preservation of the canonical ensemble within the solute through strategies **P**.2 and **P**.3 (despite the use of a standard BT for the solvent which strongly perturbs the temperature fluctuations in this part of the system) is the most important result of this paper and proves our hypothesis that an explicitly simulated solvent of the correct temperature T_S represents the optimal thermostat for a solute. Admittedly, our quantitative analysis of the applicability of strategies **P**.2 and **P**.3 is restricted to relatively small peptides because an extended statistics is required for reliable results. Already for the small peptides with their short temperature autocorrelation times of 15 ps, it takes more than 10 ns to determine the average temperature with an accuracy of 1 K. For larger systems, the temperature autocorrelation times increase and so do the simulation times required for accurate temperature measurements. Too short simulations can easily lead to the false impression that the solute temperature sizably differs from the solvent temperature. To our experience, the noninvasive strategy **P**.2 can suffice for quite large solvent−solute systems. For instance, reinspecting a simulation[8] of the C-terminal domain of the human prion protein (residues 125−228), which employed a global thermostat coupling (strategy **G**), we found that the protein temperature deviated by more than 10 K from that of the solvent. Subsequent simulations of a slightly larger fragment (residues 114−228), which employed strategy **P**.2 but otherwise the same simulation setup, showed no significant temperature difference. In the few cases in which one observes a seemingly intolerable temperature difference between solute and solvent, one can still use the solvent as the heat bath by applying the minimally invasive strategy **P**.3 to keep the solute well tempered.

It should be noted that our heat flow model and the associated setup protocol for the constant heat flow strategy **P**.3 are restricted to two subsystems with homogeneous local algorithmic heating rates. For simulations of more complex systems such as protein-DNA assemblies in solution, for which one expects more than two different heating rates, a constant heat flow strategy can be analogously designed. However, it will become increasingly difficult to determine the local heating rates of the various subsystems which have to be compensated.

Acknowledgment. This work was supported by the Bavarian joint research project ForPrion (LMU02), by the VolkswagenStiftung (I/79 884), and by Deutsche Forschungsgemeinschaft (SFB533/C1, SFB749/C4).

Appendix: Setting up Strategies P.2 and P.3

Here, we give a detailed description of the steps needed in order to set up a simulation system containing a macromolecule P in thermal equilibrium with an explicit solvent enviroment S according to the strategies **P**.2 and **P**.3, respectively, using the standard Berendsen algorithm. After preparation (e.g., removal of close solvent−solute contacts by energy minimization), the following steps are necessary:

a) Heating phase: The subsystems are heated using two separate classical BTs (e.g., $\tau_S = \tau_P = 0.1$ ps) to the temperature T_{sim} desired in the production simulation. Depending on the initial deviations of the solute temperature T_P and solvent temperature T_S, it may take a simulation time of up to $30\tau_{S/P}$ for the respective subsystems to safely attune to T_{sim}.

b) Relaxation phase I: The solute is decoupled from its thermostat ($\tau_P = \infty$) and relaxes to its new steady state temperature $T_{P,1}$. The time constant for the relaxation to the steady state is the solvent−solute coupling time τ_{SP}. Since τ_{SP} is still unknown, an upper limit estimate (e.g., $\tau_{SP} \approx 20$ ps) should be used to determine the relaxation time $t_{\mathrm{relax}} \approx 10\tau_{SP}$.

c) Test simulation I: Here, the solute remains decoupled from its thermostat and the simulation serves to determine its average temperature $T_{P,1}$. If the deviation from equilibrium measured by $|T_{P,1} - T_{\mathrm{sim}}|$ is less than an acceptable tolerance ΔT_P, then the noninvasive strategy **P**.2 is applicable, and one may directly continue the simulation for data production f).

The necessary simulation time t_1 for the test depends on the tolerable uncertainty $\sigma^2_{\hat{T}_{P,1}}$ of the measured solute temperature $\hat{T}_{P,1}$, which forms an upper bound for the uncertainty $\sigma^2_{T_P}$ in the prediction of the production run temperature T_P. If ΔT_P is the accuracy required for the prediction, we should make sure that $\sigma^2_{\hat{T}_{P,1}} \leq \Delta T_P^2$. By eq 9 the simulation time then is $t_1 = 2\tau_c\sigma^2_{T_P}/\Delta T_P^2$, where τ_c is the temperature autocorrelation time of the solute, and σ_{T_P} is the standard deviation of its temperature fluctuations, which were observed during the test run. One typically obtains simulation times of several nanoseconds.

d) Relaxation phase II: The solute is coupled to a separate thermostat with a coupling time $\tau_P \geq 500$ ps intended for the **P**.3 production run. Using an estimate for τ_{SP} (e.g., 1 ps), a reasonable choice for the target temperature is given by $T_{0,P,2} = -\tau_P/\tau_{SP} \cdot |T_{P,1} - T_{\mathrm{sim}}|$ (leading to 2-fold overcompensation if τ_{SP} was exact). The duration of this relaxation phase is the same as in step b).

e) Test simulation II: The average temperature $T_{P,2}$ is determined. The simulation time t_2 should be equal to t_1 in step c).

f) Production simulation: If strategy **P**.2 turned out to be applicable in step c), the settings in this simulation are chosen identically (in fact, one may regard the test run as the initial part of the production simulation). Otherwise, the target temperature $T_{0,P}$ for a **P**.3 simulation is determined from the two test simulations by

$$T_{0,P} = T_{\mathrm{sim}} + \frac{T_{0,P,2} - T_{P,2}}{T_{P,2} - T_{P,1}}(T_{\mathrm{sim}} - T_{P,1}) \qquad (22)$$

References

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(2) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garca, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.

(3) Grubmüller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997–999.

(4) Rief, M.; Oesterhelt, F.; Heymann, B.; Gaub, H. E. *Science* **1997**, *275*, 1295–1297.

(5) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731–737.

(6) Kucera, K.; Lambry, J. C.; Martin, J. L.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5805–5807.

(7) Spörlein, S.; Carstens, H.; Satzger, H.; Renner, C.; Behrendt, R.; Moroder, L.; Tavan, P.; Zinth, W.; Wachtveitl, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7998–8002.

(8) Hirschberger, T.; Stork, M.; Schropp, B.; Winklhofer, K. F.; Tatzelt, J.; Tavan, P. *Biophys. J.* **2006**, *90*, 3908–3918.

(9) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863–867.

(10) Shea, J.-E.; Brooks, C. L. I. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499–535.

(11) Privalov, P. L. Physical basis of the stability of the folded conformations of Proteins. In *Protein Folding*; Creighton, T. E., Ed.; W. H. Freeman: San Francisco, 1992.

(12) Munishkina, L. A.; Phelan, C.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2003**, *42*, 2720–2730.

(13) Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M. *Nat. Struct. Biol.* **2000**, *7*, 34–38.

(14) Tavan, P.; Carstens, H.; Mathias, G. Molecular Dynamics Simulations of Proteins and Peptides: Problems, Achievements, and Perspectives. In *Protein Folding Handbook. Part I*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH: 2005.

(15) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature* **1979**, *280*, 558–563.

(16) Ishima, R.; Torchia, D. A. *Nat. Struct. Biol.* **2000**, *7*, 740–743.

(17) Ichiye, T.; Karplus, M. *Biochemistry* **1983**, *22*, 2884–2893.

(18) Lazaridis, T.; Karplus, M. *Biophys. Chem.* **2003**, *100*, 367–395.

(19) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(20) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

(21) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

(22) Carstens, H. Konformationsdynamik lichtschaltbarer Peptide: Molekulardynamiksimulationen und datengetriebene Modell-

bildung, Dissertation, Fakultät für Physik, Ludwig-Maximillians-Universität München, 2004.

(23) Koper, G. J. M.; Reiss, H. *J. Phys. Chem.* **1996**, *100*, 422–432.

(24) Swope, W. C.; Pitera, J. W.; Suits, F. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.

(25) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. *J. Phys. Chem. B* **2004**, *108*, 6582–6594.

(26) Nutt, D. R.; Smith, J. C. *J. Chem. Theory Comput.* **2007**, *3*, 1550–1560.

(27) Mathias, G.; Egwolf, B.; Nonella, M.; Tavan, P. *J. Chem. Phys.* **2003**, *118*, 10847–10860.

(28) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(29) Grubmüller, H.; Tavan, P. *J. Comput. Chem.* **1998**, *19*, 1534–1552.

(30) Kraeutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501–508.

(31) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.

(32) Oda, K.; Miyagawa, H.; Kitamura, K. *Mol. Simul.* **1996**, *16*, 167–177.

(33) Cheng, A.; Merz, K. M. *J. Phys. Chem.* **1996**, *100*, 1927–1937.

(34) Guenot, J.; Kollman, P. A. *Protein Sci.* **1992**, *1*, 1185–1205.

(35) Arnold, G. E.; Ornstein, R. L. *Proteins* **1994**, *18*, 19–33.

(36) Hünenberger, P. H. *Adv. Polym. Sci.* **2005**, *173*, 105–149.

(37) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(38) Harvey, S. C.; Tan, R. K.-Z.; Cheatham, T. E. *J. Comput. Chem.* **1998**, *19*, 726–740.

(39) Chui, S.-W.; Clark, M.; Subramaniam, S.; Jakobsson, E. *J. Comput. Chem.* **2000**, *21*, 121–131.

(40) Morishita, T. *J. Chem. Phys.* **2000**, *113*, 2976–2982.

(41) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(42) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(43) Tuckerman, M. E.; Liu, Y.; Ciccotti, G.; Martyna, G. J. *J. Chem. Phys.* **2001**, *115*, 1678–1702.

(44) Holian, B. L.; Voter, A. F.; Ravelo, R. *Phys. Rev. E* **1995**, *52*, 2338–2347.

(45) Posch, H. A.; Hoover, W. G.; Vesely, F. J. *Phys. Rev. A* **1986**, *33*, 4253–4265.

(46) D'Alessandro, M.; Tenenbaum, A.; Amadei, A. *J. Phys. Chem. B* **2002**, *106*, 5050–5057.

(47) Tobias, D. J.; Martyna, G. J.; Klein, M. L. *J. Phys. Chem.* **1993**, *97*, 12959–12966.

(48) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(49) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101(1−7)

(50) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: Oxford, 1987.

(51) Stange, K. *Angewandte Statistik*; Springer: Heidelberg, 1970.

(52) Weber, W.; Hünenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668–3675.

(53) Daura, X.; Gademann, K.; Schäfer, H.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 2393–2404.

(54) Villareal, M. A.; Montich, G. G. *J. Biomol. Struct. Dyn.* **2005**, *23*, 135–142.

(55) van den Bosch, M.; Swart, M.; Snijders, J. G.; Berendsen, H. J.; Mark, A. E.; Oostenbrink, C.; van Gunsteren, W. F.; Canters, G. W. *ChemBioChem* **2005**, *6*, 738–746.

(56) Monticelli, L.; Simoes, C.; Belvisi, L.; Colombo, G. *J. Phys.: Condens. Matter* **2006**, *18*, S329–S345.

(57) Fox, T.; Kollman, P. A. *Proteins* **1996**, *25*, 315–334.

(58) Kong, Y.; Ma, J.; Karplus, M.; Lipscomb, W. N. *J. Mol. Biol.* **2006**, *356*, 237–247.

(59) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.

(60) Niedermeier, C.; Tavan, P. *J. Chem. Phys.* **1994**, *101*, 734–748.

(61) Niedermeier, C.; Tavan, P. *Mol. Simul.* **1996**, *17*, 57–66.

(62) Mathias, G.; Tavan, P. *J. Chem. Phys.* **2004**, *120*, 4393–4403.

(63) Eichinger, M.; Grubmüller, H.; Heller, H.; Tavan, P. *J. Comput. Chem.* **1997**, *18*, 1729–1749.

(64) 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*, 3596–3607.

(65) 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.; D. Reidel Publishing Company: Dordrecht, 1981.

(66) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(67) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(68) Guenot, J.; Kollman, P. A. *J. Comput. Chem.* **1993**, *14*, 295–311.

(69) Gardiner, C. W. *Handbook of Stochastic Methods*, 2nd ed.; Springer: Berlin, 1985.

(70) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200–1202.

CT8000365

JCTC Journal of Chemical Theory and Computation

Re-Evaluation of the Reported Experimental Values of the Heat of Vaporization of N-Methylacetamide

Alexander D. MacKerell, Jr.,* Ji Hyun Shim, and Victor M. Anisimov

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201

Received March 26, 2008

Abstract: The accuracy of empirical force fields is inherently related to the quality of the target data used for optimization of the model. With the heat of vaporization (ΔH_{vap}) of N-methylacetamide (NMA), a range of values has been reported as target data for optimization of the nonbond parameters associated with the peptide bond in proteins. In the present work, the original experimental data and Antoine constants used for the determination of the ΔH_{vap} of NMA are reanalyzed. Based on this analysis, the wide range of ΔH_{vap} values reported in the literature is shown to be due to incorrect reporting of the temperatures at which the original values were extracted and limitations in the quality of experimental vapor pressure-temperature data over a wide range of temperatures. Taking these problems into account, a consistent ΔH_{vap} value is extracted from three studies for which experimental data are available. This analysis suggests that the most reliable value for ΔH_{vap} is 13.0 $\pm$ 0.1 at 410 K for use in force field optimization studies. The present results also indicate that similar analyses, including analysis of Antoine constants alone, may be of utility when reported ΔH_{vap} values are not consistent for a given neat liquid.

Introduction

Structure−function studies of peptides and proteins, including protein folding studies,[1] consume a significant volume of intellectual and financial resources. Of the many approaches used to study such systems empirical force field based calculations represent an effective and ever increasingly used methodology from which atomic details of structure−function relationships may be obtained.[2,3] In addition, force fields have the potential to be predictive allowing, for example, the impact of chemical modification of inhibitors on protein binding or the impact of mutations on protein activity to be made. Of the protein based predictions, accurate prediction of protein structure based on sequence alone,[4] the so-called protein folding problem, represents a grail of empirical force fields and, accordingly, a significant number of studies have and continue to be performed toward solving this problem.

Central to the success of empirical force field based methods is the accuracy of the force fields themselves. Simply put, the force field must correctly represent the change in energy of the system as a function of conformation and environment in order to effectively reproduce the experimental regimen.[5] Accordingly, significant effort has been and is still being made toward the optimization of empirical force fields for proteins as well as other biomolecules. While much of this effort is based on the reproduction of high-level quantum mechanical data, the most important data with respect to the condensed phase are experimental data, including thermodynamic data on small compounds representative of chemical moieties in macromolecules.[2,6] For example, the availability of condensed phase data, including the heat of vaporization (ΔH_{vap}) and free energy of solvation, of N-methylacetamide (NMA) is key for the optimization of the force field parameters associated with the nonbond interactions of the peptide bond with the environment. Careful optimization of the nonbond force field parameters for NMA (e.g., the Lennard-Jones 6−12 and electrostatic parameters) to reproduce such condensed phase data lays the groundwork for a protein force field that accurately models energetic differences associated with changes in the environment, such as moving of the peptide

* Corresponding author. E-mail: alex@outerbanks.umaryland.edu.

10.1021/ct8000969 CCC: $40.75
Published on Web 06/28/2008

Table 1. Enthalpies of Vaporization, kcal/mol, of NMA from Various Sources and Methods of Analysis[d,e]

temperature	Aucejo et al., 1993[15]	Manczinger and Krotüm, 1975[17]	Kortüm and Biedersee, 1970[13]	averages
		1) From Reported Antoine Constants[a]		
373 K	14.6	13.3	12.8[c]	np
410 K	13.0	12.9	12.8[c]	np
		2) From Antoine Constants Refit to Full Experimental Range[a,b]		
373 K	14.6, 14.5, 13.9	13.3, 13.3, 13.3	12.3, 12.3, 12.3	np
410 K	13.0, 13.0, 13.7	13.0, 13.0, 13.0	12.5, 12.5, 12.5	np
		3) From Original Experimental Data via Eq 2 (Temperature Range)		
	13.8 (353−428K)	13.2 (333−443K)	12.6 (353−479)	np
		4) From Original Experimental Data via Eq 2 from 390 to 430 K		
	13.1	12.9	13.0	13.0 ± 0.1
		5) From Antoine Constants Fit to 390 to 430 K[a,b]		
373 K	14.7, 13.6, 13.2	14.3, 13.4, 13.0	14.3, 13.4, 13.0	13.7 ± 0.6
410 K	13.1, 13.1, 13.1	12.9, 12.9, 12.9	12.9, 13.0, 13.0	13.0 ± 0.1

[a] Calculated from eq 4 at the reported temperatures. [b] The three individual values for each data set are from the Antoine constants from the three refits of each of the data sets as shown in Table 3. [c] Constant C set to zero in Antoine equation leading to no temperature dependence in the calculated ΔH_{vap}. [d] Statistical analysis performed for the data in the first 3 columns of the final two rows. np indicates that the calculation was not performed and errors are standard deviations. [e] NMA: bp: 479, mp: 303.7 K.

backbone from an aqueous environment to the protein interior during protein folding.

Successful development of accurate empirical force fields, therefore, requires the availability of experimental thermodynamic data for small molecules that are both accurate and precise. Experimentally, the most accurate way of determining the heat of vaporization is through calorimetry performed at the boiling point of the neat liquid. An alternative is the use of vapor pressure-temperature (P-T) data, as has been performed for NMA. When available, P-T data may be used directly to fit empirical force field parameters as has been done in a number of cases where accurate data over a wide temperature range are available.[7−10] However, in the case of neat NMA, the corresponding data reported in the literature[11−16] are not sufficiently accurate at low temperatures, making them unsuitable for use as target data for force field optimization. Presented in row 1 of Table 1 are experimental ΔH_{vap} values for NMA based on the reported Antoine constants from a number of sources.[12,15,17−19] As may be seen the values range from 12.8 up to 14.6 kcal/mol when temperatures of both 373 and 410 K are considered. In addition, values of 14.2[11] and 16.5[12] kcal/mol have been reported. No Antoine constants were presented in the former case, while the latter value can readily be excluded if one examines the raw experimental data, as performed below. Consistent with the range of reported values are the different values that have been used as target data for the optimization of empirical force fields for NMA and, by extension, peptides and proteins. For example, the OPLS and AMBER force fields were optimized targeting a value of 13.3 kcal/mol at 373 K;[20] a value also used by Caldwell and Kollman,[21] Gao and co-workers,[22,23] and Kaminski et al.[24] for the development of polarizable force fields. CHARMM22[25] and GROMOS[26] targeted a value of 14.2 at 373 K as did work by Patel and Brooks in the development of a polarizable force field based on a fluctuating charge model.[27] In addition, a recent study of a polarizable force field based on a classical Drude oscillator cited four values: −12.7, −13.1, −14.2, and −14.8 kcal/mol.[28] Thus, it is evident that in order to develop accurate force field parameters for the peptide backbone it is essential to determine the proper experimental ΔH_{vap} of NMA.

In the present report we reanalyze the original experimental data used to determine ΔH_{vap} of NMA, including available experimental vapor pressure-temperature (P-T) curves for the neat liquid. From this analysis the source of discrepancies in the original data is identified, allowing for an understanding of the source of the range of previously reported ΔH_{vap} values. This is followed by determination of a consensus value for ΔH_{vap} of NMA, a value which we suggest should be the target for future force field development efforts.

Computation of the Heat of Vaporization from Vapor Pressure−Temperature Data. Typically, calculation of the heat of vaporization from experimental data is based on the Clausius-Clapeyron equation

$$\frac{dP}{dT} = \frac{\Delta H_{vap}}{(V - V_{liq})T} \approx \frac{\Delta H_{vap}}{VT} = \frac{P\Delta H_{vap}}{RT^2} \quad (1)$$

where P is the vapor pressure of the liquid, T is the temperature, R is the universal gas constant, V is the molecular volume of the gas phase, and V_{liq} is the molecular volume of the liquid phase. Since we deal with a phase transition from liquid to gas phase, V_{liq} is many times smaller than the volume of the gas phase, V, which justifies the use of the approximate form of the Clausius-Clapeyron equation shown in the right-hand side of eq 1.

There are two approaches for determining the heat of vaporization from experimental liquid-vapor pressure data. Assuming that the heat of vaporization is constant over the selected temperature range, this equation can be integrated by separating the independent variables:

$$\frac{dP}{P} = \frac{\Delta H_{vap}}{R} \cdot \frac{dT}{T^2} \quad \text{yielding } \ln\frac{P_2}{P_1} = \frac{\Delta H_{vap}}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \quad (2)$$

This approach allows ΔH_{vap} to be readily determined from the slope of the vapor pressure-temperature (P-T) curve, though limited to a situation where the heat capacity, the

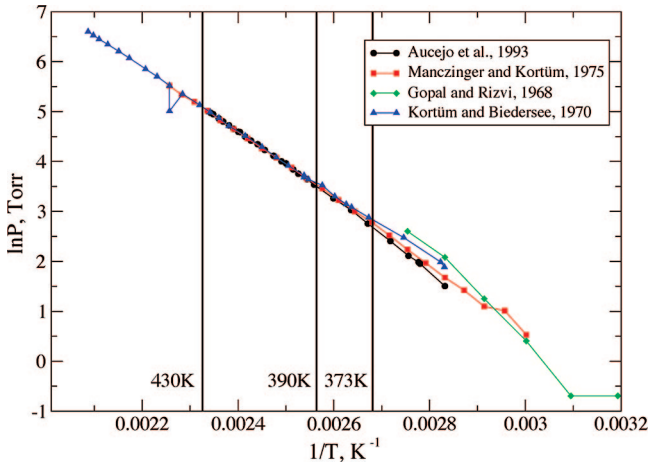


Figure 1. Experimental inverse temperature versus the natural log of the vapor pressure (P-T) curves for N-methylacetamide from Aucejo et al.,[15] Manczinger and Kortüm,[17] Gopal and Rizvi,[12] and Kortüm and Biedersee.[13] Experimental data were converted to common units of ln$P$ and Torr. Lines connecting the data points were included to facilitate visual inspection and do not represent fits of the data.

temperature dependence of the heat of vaporization, of the neat liquid is zero.

An alternative method for solving eq 1 assumes obtaining the derivative of the vapor pressure with respect to temperature. Different formulas have been suggested for this purpose. Among these, the Antoine equation is used extensively and has been found to be reliable except where the data are limited to very small temperature ranges or for low-boiling substances. The Antoine equation[29] is a simple 3-parameter fit to experimental vapor pressure measured over a given temperature range

$$\ln P = A - B\left(\frac{1}{(T+C)}\right) \tag{3}$$

where A, B, and C are the fitted parameters. This function allows rearrangement of eq 1 in the following form:

$$\frac{dP}{PdT} = \frac{d(\ln P)}{dT} = \frac{\Delta H_{vap}}{RT^2} = \frac{B}{(T+C)^2} \quad \text{or } \Delta H_{vap} = RB\frac{T^2}{(T+C)^2} \tag{4}$$

This equation explicitly takes into consideration the temperature dependence of the heat of vaporization and should be valid for a wider range of temperatures than eq 2.

Analysis of Antoine Constants and Experimental Vapor Pressure-Temperature (P-T) for N-Methylacetamide. Step one of the analysis of the discrepancies in the ΔH_{vap} values of neat NMA was inspection of available experimental P-T data. Presented in Figure 1 are the P-T data in the form of $1/T$ versus ln$P$ from four studies. Immediately evident is the oldest data set from Gopal and Rizvi.[12] Given the significant difference in this data set as compared to the remaining three sets along with the significant difference in ΔH_{vap} allows the value of 16.5 kcal/mol reported in that study to readily be discarded as can the experimental data. The remaining three data sets from Aucejo et al., Manczinger and Kortüm, and Kortüm and Biedersee appear to be in reasonable agreement, with all the curves sampling a wide range of temperatures and including a large number of data points. However, inspection of Table 1 shows the ΔH_{vap} values at 373 K from the reported Antoine constants from those studies to differ by over 1 kcal/mol.

This difference suggested that the method of analysis of the original experimental data may be leading to the discrepancy. The original data were treated via the Antoine equation, eq 3 above. Presented in Table 2 are the Antoine equation constants as originally reported in the cited studies as well as following conversion to common units. The Dykyj constants are a refit of the experimental data from Manczinger and Kortüm; those values and the values from Gopal are included for completeness, though they will not be discussed further. As is evident, significant differences are present including the impact of constraining C to 0 (i.e., assuming the heat capacity = 0). Notable are the differences in the constants from Aucejo et al.,[15] Manczinger and Kortüm,[17] and Kortüm and Biederseede[13] despite the similarity of the curves shown in Figure 1. The impact of this difference is observed in the ΔH_{vap} at different temperatures (Table 1, row 1), including the lack of temperature dependence due to C being constrained to zero. The different values of ΔH_{vap} as a function of both the particular study and temperature along with analysis of the discussion of the reported ΔH_{vap} in the original publications indicate that the discrepancy of the values reported in the more recent force field development literature is due, in part, to a lack of clarity in the original publications on the temperature associated with the reported ΔH_{vap}. In addition, it appears that when available, the temperatures were often not correctly noted when citing the original ΔH_{vap}, further compounding the problem. However, the differences in the Antoine constants indicate that the data analysis and/or experimental data contribute to the discrepancies.

To check the previous data analysis the available experimental data were refit using a modified version of the FITCHARGE[30] module in CHARMM.[31] In all cases, each data set was fit three times using the original Antoine constants from the Aucejo et al., Manczinger and Kortüm, and Kortüm and Biedersee studies as initial guesses (Table 3). Fitting was initially performed over the full range of temperatures used in the respective experiments. Analysis of the Antoine constants for the three data sets (Table 3) reveals the impact of the initial guesses on the resulting constants. The fitting results are not surprising as the objective function is nonlinear and may have multiple local minima. In this case the parameter set showing the least RMSE represents the best fit. From the present fitting the lowest RMSE values were 0.0117, 0.0379, and 0.0926, for the Aucejo et al., Manczinger and Kortüm, and Kortüm and Biedersee data, respectively. Comparison of those values with the RMSE values from the original reported Antoine constants, 0.0118, 0.0383, and 0.0988 for the Aucejo et al., Manczinger and Kortüm, and Kortüm and Biedersee data, respectively, shows the refitting to yield only marginal improvement. These results suggest that the original fitting of the Antoine constants was satisfactory, and the noted differences in ΔH_{vap} originate from the inherent differences in the experimental data sets.

Table 2. Reported Antoine Constants for NMA[1]

reported[a]	*A*	*B*	*C*	eq
Aucejo, 1993[15]	12.49715	2658.377	−148.3473	$\ln P(\text{kPa}) = A\text{-}B(1/(T+C))$
Manczinger, 1975[17]	7.7377	2043.37	−60.75	$\log P(\text{Torr}) = A\text{-}B(1/(T+C))$
Kortüm, 1970[13]	7.8259	2793.3	0	$\log P\ (\text{kPa}) = A\text{-}B(1/(T+C))$
Gopal, 1968[12]	11.063	3606	0	$\log P(\text{Torr}) = A\text{-}B(1/(T+C))$
Dykyj, 1984[18b]	6.60575	1868.206	−75.963	$\log P(\text{kPa}) = A\text{-}B(1/(T+C))$
	reported data converted to ln*P* and Torr			$\ln P(\text{Torr}) = A\text{-}B(1/(T+C))$
Aucejo, 1993	14.51214	2658.377	−148.3473	
Manczinger, 1975	17.81671	4705.0333	−60.75	
Kortüm, 1970	20.034786	6431.81094	0	
Gopal, 1968	25.473499	8303.121845	0	
Dykyj, 1984	17.225287	4301.70329	−75.963	

[a] Reported values are directly from the literature before and after conversion to common units. [b] Dykyj et al. Antoine constants represent a previously published refitting of the Manczinger and Kortüm data.

Table 3. Antoine Constants Following Refitting to the Original Experimental Data

fit over the full range of experimental data[a]	initial guess[b]	*A*	*B*	*C*	RMSE
Aucejo, 1993	i	14.5106	2658.4542	−148.3022	0.0118
	ii	14.5888	2696.4742	−146.5925	0.0117
	iii	20.2896	6183.4048	−22.2746	0.0304
Manczinger, 1975	i	18.4156	5107.5211	−47.0957	0.0379
	ii	18.4156	5107.5211	−47.0957	0.0379
	iii	18.4156	5107.5213	−47.0957	0.0379
Kortüm, 1970	i	20.7029	7157.6343	27.4688	0.0927
	ii	20.7080	7162.0838	27.6048	0.0927
	iii	20.8027	7245.8239	30.1571	0.0926
fit over 390 to 430 K	initial guess[b]	*A*	*B*	*C*	RMSE
Aucejo, 1993	i	14.6435	2703.0937	−147.3946	0.0071
	ii	17.9300	4698.0665	−64.1666	0.0075
	iii	20.2324	6422.3774	−5.8062	0.0084
Manczinger, 1975	i	14.9198	2887.2697	−136.9291	0.0077
	ii	17.8465	4706.7286	−61.4594	0.0066
	iii	20.1255	6430.2474	−2.6391	0.0067
Kortüm, 1970	i	15.0520	2956.6394	−133.4645	0.0117
	ii	17.8602	4704.1907	−61.7037	0.0110
	iii	20.1509	6428.3894	−3.1271	0.0112

[a] Refitted constants are over the full temperature ranges reported in the respective studies (353 to 428 K for Aucejo et al., 333 to 443 for Manczinger and Kortüm, and 353 to 479 for Kortüm and Biedersee) using eq 3. Units: pressure in Torr and temperature in K. [b] Initial guesses from the reported Antoine constants from i) Aucejo et al., ii) Manczinger and Kortüm, and iii) Kortüm and Biedersee as reported in Table 2.

To further verify that the original discrepancies in ΔH_{vap} values were associated with the experimental P-T data, ΔH_{vap} values were calculated from the P-T data based on the Clausius-Clapeyron equation,[29] eq 2, from the slopes of the $1/T$ versus ln*P* plots. It should be reiterated that this approach assumes that the heat capacity is zero (i.e., $C = 0$ in the Antoine equation). When this approach was applied to the data included in Figure 1, it yielded high quality fits ($R^2 > 0.99$ in all cases). Based on the resulting slopes ΔH_{vap} values of 13.8, 13.2, and 12.6 kcal/mol for the three data sets are obtained (Table 1, row 3). The level of agreement is similar to the ΔH_{vap} values obtained from the Antoine equation. Thus, the present analysis indicates that the discrepancy in the reported ΔH_{vap} values is dominated by contributions from limitations in the experimental data.

Inspection of the experimental data in Figure 1 shows the agreement to be good for the Aucejo et al., Manczinger and Kortüm, and Kortüm and Biedersee data sets at the higher temperatures. However, the data sets diverge at lower temperatures. The presence of such divergence is not unexpected. Given that the pressures at these lower temperatures become quite small it may be assumed that the ability to measure them accurately becomes limiting. Indeed the extremely low temperatures of the Gopal experiments, to a point where the vapor pressures are a fraction of a Torr (Figure 1), is suggested to contribute to the significant problems with that data set. It is the divergence of the experimental data sets at the lower temperatures that leads to differences in the refit Antoine constants discussed above (Tables 2 and 3) and to the significant differences in ΔH_{vap} values.

Based on the limitations with the experimental data at lower temperatures, the experimental P-T data were reanalyzed over a higher, though limited, range of temperatures (390 to 430 K, Figure 1). This analysis included 1) calculation of the ΔH_{vap} values using eq 2 over the selected temperature range and 2) refitting the Antoine constants over the selected temperature range (390 to 430 K) following which ΔH_{vap} values were obtained from eq 4.

Results from these analyses are included in Table 1 for the ΔH_{vap} values and in Table 3 for the Antoine constants. Based on the calculation of ΔH_{vap} using eq 2 values close to 13 kcal/mol were obtained for all three studies (Table 1, row 4), with an average and standard deviation of 13.0 ± 0.1 kcal/mol. Next, refitting of the three experimental data sets over the range 390 to 430 K leads to Antione constants that more accurately reproduce the experimental data as compared to fits of the full temperature ranges used in the experimental studies (Table 3, compare the RMSE values for the top and bottom sections), though the RMSE are similar for each of the subset Antoine constants. The corresponding ΔH_{vap} values using eq 4 for the three fits of the three data sets (Table 1, row 5) show the values to range over 1.7 kcal/mol at 373 K, while all the values are in excellent agreement at 410 K. Averaging over these values yields a mean ΔH_{vap} value of 13.0 ± 0.1 kcal/mol at 410 K, which is in ideal agreement with that obtained via eq 2 over the same data range. Thus, it is evident that limitations in the experimental data at low temperatures contribute to the discrepancies in the ΔH_{vap} values of NMA reported in the literature. Moreover, the present data analysis indicates that a ΔH_{vap} value at 13.0 ± 0.1 kcal/mol at 410 K is reliable and should be used as the target value (and temperature) for the development of theoretical models of NMA.

As discussed in the Introduction a number of force field development efforts have been based on calculation of the heat of vaporization at 373 K. Accordingly, the Antoine constants fit to 390−430 K data were used to predict ΔH_{vap} at 373 K. The results in row 5 of Table 1 show the derived values to range over 1.7 kcal/mol with an average and standard deviation of 13.7 ± 0.6 kcal/mol. Thus, it is not possible to determine a sufficiently accurate value of the heat of vaporization at 373 K for use in force field development due to the inherent limitations in the available experimental P-T data sets.

With many liquids it may be difficult to obtain the original experimental data to perform the analysis presented above; however, two or more sets of Antoine constants may be available in many cases. To test the possible utility of the Antoine constants alone, the reported constants for NMA (Table 2) were used to generate P-T data, with the results presented in Figure 2. Inspection of the curves shows them to agree well in the range of 390 to 430 K, with significant divergence at lower temperatures, consistent with the original experimental data. Such behavior is not unexpected as the Antoine constants are simply fit to the original data, but the behavior does indicate that if discrepancies exist in ΔH_{vap} values for a liquid, inspection of the P-T curves calculated from the Antoine constants may be of utility to select a temperature range where significant agreement between the different experiments occur, from which more reliable ΔH_{vap} values may be obtained. Applying this type of analysis in the present case using eq 2 applied to the calculated P-T data in Figure 2 yields ΔH_{vap} values of 13.0, 12.9, 12.9, and 12.8, respectively, for the four data sets in Figure 2, yielding an average of 12.9 ± 0.1 kcal/mol. This is within experimental error of that calculated from the original experimental data (Table 1, rows 4 and 5).

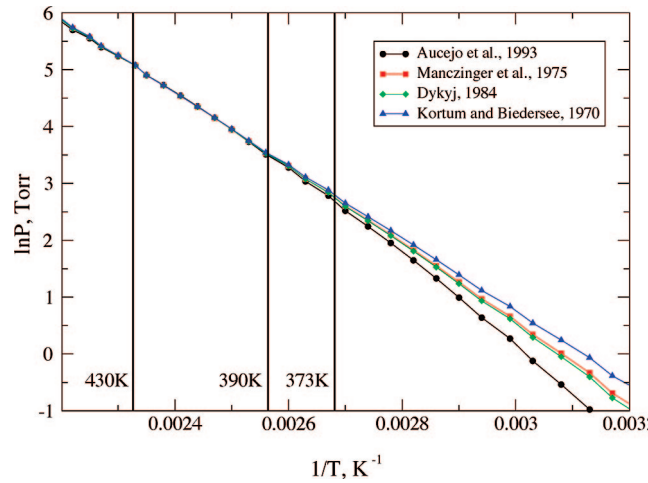


Figure 2. Calculated inverse temperature versus the natural log of the vapor pressure (P-T) curves for N-methylacetamide from the reported Antoine constants of Aucejo et al.,[15] Manczinger and Kortüm,[17] Dykyj,[19] and Kortüm and Biedersee.[13] Lines connecting the data points were included to facilitate visual inspection and do not represent fits of the data.

In summary, the wide range of ΔH_{vap} values reported in the literature for liquid NMA is shown to be due to 1) inaccuracies in reporting the temperatures at which the experiments were performed and 2) limitations in the experimental P-T data associated with decreased accuracy in the data obtained at lower temperatures due to the low vapor pressures of NMA. Taking these problems into account allows for the extraction of consistent ΔH_{vap} values from the data for the three studies for which experimental data are available. This analysis suggests that the most reliable value for ΔH_{vap} is 13.0 ± 0.1 at 410 K, the value and temperature recommended for use in force field optimization studies. The present results also indicate that similar analysis may be appropriate for other neat liquids for which reported ΔH_{vap} are used for empirical force field development.

Acknowledgment. Financial support from the NIH (GM051501 and GM072558), Dr. Jirí Sponer for access to ref 18, and Drs. Edward Harder and Benoit Roux for helpful discussions are acknowledged.

References

(1) Baldwin, R. L. Energetics of protein folding. *J. Mol. Biol.* **2007**, *371*, 283–301.

(2) MacKerell, A. D., Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25*, 1584–1604.

(3) Wolfson, H. J.; Shatsky, M.; Schneidman-Duhovny, D.; Dror, O.; Shulman-Peleg, A.; Ma, B. Y.; Nussinov, R. From structure to function: Methods and applications. *Curr. Prot. Pept. Sci.* **2005**, *6*, 171–183.

(4) Chen, J.; Brooks, C. L. I. Can Molecular Dynamics Simulations Provide High-Resolution Refinement of Protein Structure. *Proteins* **2007**, *67*, 922–930.

(5) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Prot. Chem.* **2003**, *66*, 27–85.

(6) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. Free Energies of Hydration and Pure Liquid Properties of Hydrocarbons from the OPLS All-Atom Model. *J. Phys. Chem.* **1994**, *98*, 13077–13082.

(7) Siepmann, J. I.; Karaborni, S.; Smit, B. Simulating the critical behaviour of complex fluids. *Nature* **1993**, *365* (6444), 330–332.

(8) Chen, B.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 3. Explicit-Hydrogen Description of Normal Alkanes. *J. Phys. Chem. B* **1999**, *103* (25), 5370–5379.

(9) Errington, J. R.; Panagiotopoulos, A. Z. A New Intermolecular Potential Model for the n-Alkane Homologous Series. *J. Phys. Chem. B* **1999**, *103* (30), 6314–6322.

(10) Ungerer, P.; Beauvais, C.; Delhommelle, J.; Boutin, A.; Rousseau, B.; Fuchs, A. H. Optimization of the anisotropic united atoms intermolecular potential for n-alkanes. *J. Chem. Phys.* **2000**, *112*, 5499–5510.

(11) Reynolds, W. L.; Weiss, R. H. Iron(III) Complexes in Nonaqueous Solvents. I. The Solvolysis and Chloride Complex Constants in N-Methylacetamide. *J. Am. Chem. Soc.* **1959**, *81*, 1790–1793.

(12) Gopal, R.; Rizvi, S. A. Vapour Pressures of some Mono- and Di-Alkyl Substituted Aliphatic Amides at Different Temperatures. *J. Indian Chem. Soc.* **1968**, *45*, 13–16.

(13) Kortüm, G.; Biedersee, H. v. Vapor–liquid equilibrium (boiling-point diagrams) of binary systems: High relative volatility. *Chem.-Ing.-Tech.* **1970**, *42*, 552–560.

(14) Lemire, R. H.; Sears, P. G. N-Methylacetamide as a Solvent. *Top. Curr. Chem.* **1978**, *78*, 45–91.

(15) Aucejo, A.; Montón, J. B.; Muŏz, R.; Sanchotello, M. Isobaric Vapor-Liquid Equilibrium Data for the Cyclohexanone + N-Methylacetamide System. *J. Chem. Eng. Data* **1993**, *38*, 160–162.

(16) Chickos, J. S.; Acree, W. E., Jr. Enthalpies of Vaporization of Organic and Organometallic Compounds. *J. Phys. Chem. Ref. Data* **2003**, *32*, 519–878.

(17) Manczinger, J.; Kortüm, G. Thermodynamische Mischungseffekte im System Wasser(1)/N-Methylacetamid(2). *Z. Phys. Chem. Neue Folge* **1975**, *95*, 177–186.

(18) Dykyj, J.; Repas, M.; Svoboda, J. *Tlak Nasytenej Pary Organickych Zlucenin*; Vydavatelstvo Slovenskej Akademie Vied: Bratislava, Czechoslovakia, 1984; Vol. 1, p 2.

(19) Stephenson, R. M.; Malanowski, S. *Handbook of the Thermodynamics of Organic Compounds*; Elsevier: New York, 1987.

(20) Jorgensen, W. L.; Swenson, C. J. Optimized Intermolecular Potential Functions for Amides and Peptides. Structure and Properties of Liquid Amides. *J. Am. Chem. Soc.* **1985**, *107*, 569–578.

(21) Caldwell, J. W.; Kollman, P. A. Structure and Properties of Neat Liquids Using Nonadditive Molecular Dynamics: Water, Methanol, and N-Methylacetamide. *J. Phys. Chem.* **1995**, *99*, 6208–6219.

(22) Gao, J.; Pavelites, J. J.; Habibollazadeh, D. Simulation of Liquid Amides Using a Polarizable Intermolecular Potential Function. *J. Phys. Chem.* **1996**, *100*, 2689–2697.

(23) Xie, W.; Pu, J.; MacKerell, A. D., Jr.; Gao, J. Development of a Polarizable Intermolecular Potential Function (PIPF) for Liquid Amides and Alkanes. *J. Chem. Theory Comput.* **2007**, *3*, 1878–1889.

(24) Kaminski, G.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from ab Initio Quantum Chemistry. *J. Phys. Chem. A* **2004**, *108*, 621–627.

(25) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I. W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(26) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(27) Patel, S.; Brooks, I. C. L. A nonadditive methanol force field: Bulk liquid and liquid-vapor interfacial properties via molecular dynamics simulations using a fluctuating charge model. *J. Chem. Phys.* **2005**, *122*, 024508.

(28) Harder, E.; Anisimov, V. M.; Whitman, T.; MacKerell, A. D., Jr.; Roux, B. Understanding the Dielectric Properties of Liquid Amides from a Polarizable Force Field. *J. Chem. Theory Comput.* **2008**, *112*, 3509–3521.

(29) Thomson, G. W. M. The Antoine Equation for Vapor-Pressure Data. *Chem. Rev.* **1946**, *38*, 1–39.

(30) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Jr. Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.

(31) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

CT8000969

JCTC Journal of Chemical Theory and Computation

Development of a Q2MM Force Field for the Asymmetric Rhodium Catalyzed Hydrogenation of Enamides

Patrick J. Donoghue,[†] Paul Helquist,[†] Per-Ola Norrby,[‡] and Olaf Wiest*,[†]

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556-5670, and Department of Chemistry, University of Gothenburg, SE-412 96 Göteborg, Sweden

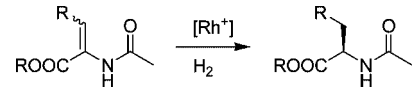
Received April 17, 2008

Abstract: The rhodium catalyzed asymmetric hydrogenation of enamides to generate amino acid products and derivatives is a widely used method to generate unnatural amino acids. The choice of a chiral ligand is of utmost importance in this reaction and is often based on high throughput screening or simply trial and error. A virtual screening method can greatly increase the speed of the ligand screening process by calculating expected enantiomeric excesses from relative energies of diastereomeric transition states. Utilizing the Q2MM method, new molecular mechanics parameters are derived to model the hydride transfer transition state in the reaction. The new parameters were based off of structures calculated at the B3LYP/LACVP** level of theory and added to the MM3* force field. The new parameters were validated against a test set of experimental data utilizing a wide range of bis-phosphine ligands. The computational model agreed with experimental data well overall, with an unsigned mean error of 0.6 kcal/mol against a set of 18 data points from experiment. The major errors in the computational model were due either to large energetic errors at high e.e., still resulting in qualitative agreement, or cases where large steric interactions prevent the reaction from proceeding as expected.

Introduction

The asymmetric hydrogenation of enamides has often been used in the synthesis of α-amino acids.[1] A representative example of this reaction is shown in Scheme 1. In order to perform this reaction asymmetrically, the choice of chiral ligand is essential but often left to trial and error. High throughput screening is often expensive both in cost of materials and in the time required to screen a large library. A computational method of ligand screening would be able to screen a large library of ligands quickly and without the need to purchase or synthesize the entire library ahead of time. A virtual library can also be more flexible than an experimental library as the virtual library can also include novel ligands that may not be immediately available. The ability to add new derivatives to the library or to design completely new ligands with a few mouse clicks makes the virtual screening more robust than if the library were limited to previously synthesized ligands. Furthermore, since the bulk of the screening is done *in silico*, a large library could be screened by a single person without requiring an inordinate amount of time. This computational screening method would ultimately supplement the experimental screening by focusing the experimental screening on a smaller set of promising ligands. It is unlikely that any *in silico* screening method would be accurate in all cases. However, even with a small number of false negatives or false positives, the experimentalist would be led to a small set of useful ligands more efficiently than by current empirical methods.

Scheme 1. Rh-Catalyzed Hydrogenation of Enamides



A computational model of a metal catalyzed asymmetric reaction would have to involve a comparison of the energies of diastereomeric transition states. Specifically, this requires

* Corresponding author e-mail:owiest@nd.edu.
[†] University of Notre Dame.
[‡] University of Gothenburg.

10.1021/ct800132a CCC: $40.75
Published on Web 07/26/2008

calculating the $\Delta\Delta G^{\ddagger}$ of the diastereomeric transition states that lead to the enantiomeric products. In order to calculate the enantioselectivity of a reaction, a full conformational search of the stereoselecting transition state must be performed with the energies being derived from a Boltzmann distribution. An exhaustive Monte Carlo (MC) search of even a small molecule would be computationally impractical if done with quantum mechanical (QM) methods, especially in comparison to the time required to screen the library experimentally. Molecular mechanics (MM) methods offer a vast speed advantage over QM methods and implicitly include dispersive interactions, which are poorly described by single-determinant methods such as DFT, in their empirical parametrization of the nonbonded terms. However, MM methods do not have adequate parameters for most metals, nor are the underlying algorithms designed to determine local maxima on the potential energy surface (PES). Finally, the functional form used to describe bonds and angles is not designed to describe transition states involved in the forming and breaking of bonds, as there is no maximum point along these potentials. Transition states and transition metals are therefore typically calculated using the much slower QM methods. One solution to this problem is the use of a QM/MM model,[2] wherein the metal catalyst and any forming or breaking bonds that relate to the reaction coordinate are treated with QM, and the rest of the molecule is treated with MM. However, this still requires a QM calculation to be performed at each step in the MC search of the conformational space and can become computationally unfeasible. In order to maximize the speed advantage of MM over QM methods and to circumvent the well-known problems of the QM/MM cross terms,[3] the model used to screen a large library should be purely MM in nature. To this end, this study employs the use of QM guided molecular mechanics or Q2MM. The Q2MM method derives MM parameters for any point on a PES, usually a transition state, from QM data and treats the state being modeled as a minimum by mathematically inverting the curvature along one dimension of the PES.[4] While this method does require some selected QM calculations to be run, they are restricted to only the force field development and are not required for the actual library screening process.

The use of MM methods to model transition states has been fairly well established. As mentioned previously, MM methods cannot directly handle maxima on the PES, so early uses of MM methods used various ground state analogues for the transition state. To model the transition state of ester hydrolysis, DeTar and co-workers used an ortho-acid as a geometric model for the transition structure.[5] More recently, a similar approach was used to determine structure-based activity of an enzyme through docking a transition state analogue to the enzyme.[6] This type of approach assumes that energetic differences at the transition state are due to the steric nature of the surrounding environment and not due to the energy involved in the reaction coordinate itself.[7] However, not all transition states do have appropriate ground-state analogues. To circumvent this, Houk and co-workers calculated transition states using *ab initio* methods and then utilized the fixed geometry of the breaking and forming bonds in MM calculations to determine selectivity in reactions.[8] This method ignores the possibility of the surrounding environment playing a role in the transition state geometry. While ground-state analogues would not respond to external forces the same way that real transition structures would, the ground-state analogues are still fully optimized structures. Houk followed this up by developing new parameters to model the transition state geometry as an energy minimum, thus making a transition state force field (TSFF). Bond lengths and angles of the new parameters were set to QM derived values, and the force constants were manually determined to reproduce MP2 energies.[9] Menger criticized the validity of these force fields, suggesting that the observed trends by Houk and co-workers was due to ground-state interactions which could be better modeled by existing parameters,[10] and then suggested an alternative method that could increase the efficiency of developing new parameters for future work.[11]

The Goddard group has developed a unique method of MM, ReaxFF, wherein energy is not determined through explicit bonds but through bond orders based on atomic distances.[12] In other MM type calculations, the energy is based on the bonding specified by the input. Energies at transition structure geometries would be abnormally high, since the geometry is highly distorted from the expected ground-state geometry. ReaxFF determines bonding based on distance between atoms; therefore, a specific bonding scheme is not fixed throughout a single calculation. While this has been used for molecular dynamic simulations involving reactions,[13] the optimization algorithms are still not designed to specifically locate transition states. The SEAM method, developed by Jensen, determines transition structures at the intersect of the reactant PES with the product PES.[14] This method requires only ground-state parameters from which the transition state is extrapolated using Morse potentials in specialized software.

All of the previous methods are still reliant on existing parameters, but transition metal parameters are generally not available, although the ReaxFF force fields have included some metal parameters.[15] Some attempts have been made to develop parameters for all elements, most notably in the UFF force field.[16] Transition metal parameters, however, must take into account changes in the coordination sphere, and therefore general parameters may not be appropriate in all situations.

The Q2MM method is designed to avoid some of these shortcomings. The general philosophy behind this method is to rapidly create a reaction specific force field, thus trading the generality of a force field to achieve higher accuracy while maintaining the speed advantage of a MM method. The new parameters are derived from QM data and are designed to be specific to the reaction being studied.[4] The transition structure is treated as a minimum and fully optimized similar to Houk's later work,[8] but the force constants are derived from a modified Hessian matrix from the QM calculations.[4] This approach circumvents the problems in the MM optimization algorithms that prevent optimization to transition structures by effectively reducing the dimensionality of the problem, and the highly specific

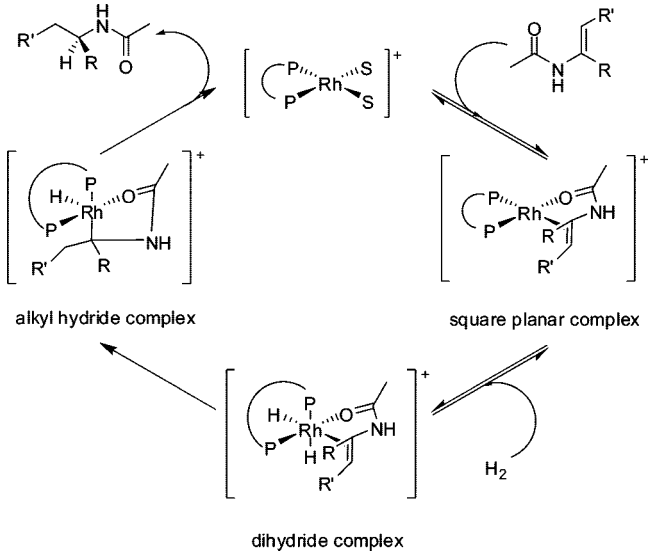


Figure 1. Overall mechanism of rhodium-catalyzed hydrogenation of enamides.

new parameters are able to describe the transition metal and transition structure appropriately.[17]

In order to develop a computational model for the asymmetric, rhodium catalyzed hydrogenation of enamides, the mechanism of the reaction must be known so that the correct stereodifferentiating transition states for the reaction are modeled. There has been a considerable amount of experimental[18–21] and theoretical[22,23] work done previously to determine the important transition states in the reaction. The overall reaction scheme is shown in Figure 1.

The substrate binds to the catalyst to form a square planar substrate-catalyst complex. This is followed by oxidative addition of H_2 to the complex to presumably form an octahedral dihydride species, although this species has not been observed experimentally. Transfer of one of the hydrides to the substrate yields an alkyl hydride complex that can be observed at low temperatures.[18] Transfer of the second hydride to the substrate yields the product, which dissociates and regenerates the catalyst. The most notable feature of the asymmetric reaction is that the relative energies of the initial square planar catalyst-substrate complex do not correlate with the observed enantiomeric excess in the product. Specifically, the major diastereomeric square planar complex gives rise to the minor enantiomer of the product. The observed enantioselectivity is based on the relative reactivity of the two substrate-catalyst complexes with hydrogen to generate the product. This anti-lock-and-key mechanism is a distinguishing feature of the reaction.[18] It has also been observed that formation of the alkyl hydride species after the first hydride transfer is irreversible.[21]

Previous computational studies have further elaborated on the mechanism of the reaction.[22,23] There are four possible orientations of hydrogen addition to the square planar substrate-catalyst complex, as shown in Figure 2. Two of these paths, labeled B and D in the nomenclature used by Landis and co-workers,[22a,b] involve the hydrogen adding parallel to the P−Rh−O bond, as opposed to the orthogonal P−Rh−alkene bond. Addition along these pathways has been determined to involve a very high barrier, and only the A and C pathways, adding the hydrogen parallel to the P−Rh−alkene bond, would be mechanistically accessible. The A and C pathways differ in the addition of hydrogen to the substrate based on proximity. Along the A pathway, the first hydride addition is made to the β-carbon of the substrate, distal to the enamide, while on the C pathway, the first hydride addition is made to the α-carbon, proximal to the enamide. Other studies from our laboratory[23] and others[22] have identified that the bias for A and C pathways is based on the electronics of the substrate. For dehydro-α-amino acids, the A pathway is the only energetically feasible pathway for the reaction. If the substrate is electronically reversed to a dehydro-β-amino acid, the C pathway is the only energetically feasible pathway.[23] This is due to the anion stabilizing ability of the substrate at the transition state due to the location of various electron withdrawing groups. Due to the change in hydrogen addition as mentioned before, the alkyl hydride intermediate that can be observed at low temperatures should reflect this difference. Experimental studies have confirmed this regioselectivity in that only the A pathway alkyl hydride is observed for dehydro-α-amino acid substrates[18a] and the C pathway alkyl hydride is observed for dehydro-β-amino acid substrates.[18b]

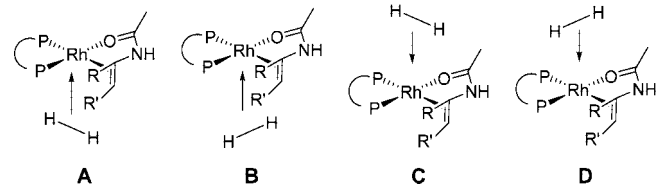


Figure 2. Mechanistic pathways derived from the orientation of H_2 addition.

The previous computational studies also identified the relative energies of the transition states between the square planar substrate-catalyst complex and the formation of the alkyl hydride species. The two key transition states involved are the oxidative addition of H_2 to form the dihydride species and the first hydride transfer to form the alkyl hydride intermediate. For the dehydro-α-amino acid substrates, these two transition states are close in energy, although the second transition state is generally slightly higher in energy. Coupled with the experimental studies that identify the formation of the alkyl hydride species to be irreversible, this study focuses solely on the second transition state as the key transition state to model for this reaction. As previously mentioned for the dehydro-α-amino acids studied here, only the A pathway is energetically feasible, and consequently there is only one transition state that needs to be modeled in this reaction, despite the complexity of the mechanism.

In the present manuscript, we will present the development of a Q2MM force field for the Rh(I)-catalyzed hydrogenation of enamides.[4] First, a QM training set has been calculated, determining selected transition structures along with the corresponding Hessian matrices and partial charges. Second, we will discuss the adaptation of the basic force field, in this case MM3*, to include any missing parameters so that the force field can be used to describe the structures calculated via QM. This will be followed by a description of the optimization of the new force field parameters in stages

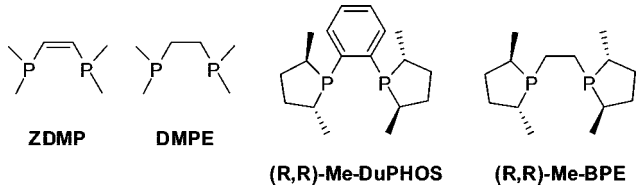


Figure 3. Ligands used for the QM training set.

to reproduce the QM data. Fourth, and finally, the new parameters are validated by calculation of experimentally observed enantiomeric excesses. Analysis of errors between theory and experiment can help further refine the force field parameters in an iterative way.

Computational Details

All QM calculations were performed using Jaguar 5.5[24] with all structures fully optimized at the B3LYP level of theory using the LACVP** basis set. This basis set corresponds to a combination of the Los Alamos double-ζ LANL2DZ ECP[25] for rhodium and the 6−31G** basis set for all other atoms. This basis set is identical to that used by our laboratory previously in mechanistic investigations[23] of the reaction and similar to that used by Landis and co-workers in both DFT[22a] and ONIOM,[22b,22c] level computations previously reported. All reported energies are enthalpies calculated at 298 K and 1 atm. Transition structures were located through the eigenvector following routine implemented in Jaguar and were confirmed as transition structures through the calculation of frequencies and identification of exactly one negative eigenvalues in the Hessian matrix. Partial charges were calculated through electrostatic potential fitting.[26] Molecular mechanics calculations were performed with MacroModel 8.6[27] using the MM3* force field as a functional form,[28] supplemented with additional terms as derived using the Q2MM method as described in the text and elsewhere.[4] Enantiomeric excesses were calculated from Boltzmann weighted energies at 298 K calculated through Monte Carlo conformational search of the transition structure. At least 15000 MC steps were used for each diastereomer of each ligand-substrate combination, with additional 15000-step searches performed, if necessary, until the MC search had converged. MC searches were considered converged when no new low energy structures (within 1 kcal/mol of the global minimum) were located within the last 5000 steps of the search. Details of the input files used are given in the Supporting Information.

Results and Discussion

Calculation of the QM Training Set. The QM training set for this system is composed of four ligands in the respective diastereomeric transition structures in several conformations. All systems were calculated using the previously studied α-formamidoacrylonitrile as a model substrate for a dehydro-amino acid.[22,23] The ligands used, shown in Figure 3, are comprised of two achiral ligands, *Z*-dimethylphosphinoethene (ZDMP) and dimethylphosphinoethane (DMPE), and two chiral ligands, (*R,R*)-Me-DuPHOS and (*R,R*)-Me-BPE. The two chiral ligands were calculated in both the *pro-R* and *pro-S* orientations of the dehydroamino acid. The two ligands with saturated backbones, DMPE and BPE, can undergo a pseudo ring flip when bound to the rhodium catalyst. This allows for a conformational change that adds an additional relative energy data point. In total, nine QM transition structures were calculated: one ZDMP structure, two DMPE conformations, two DuPHOS diastereomers, and two conformations of each of two BPE diastereomers.

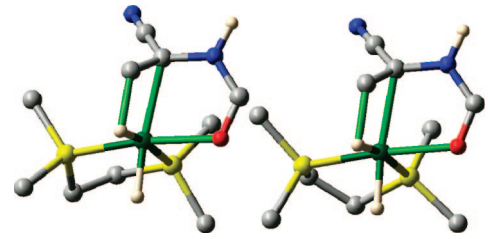


Figure 4. Optimized QM structures for the hydride transfer transition state using the DMPE ligand.

Table 1. Average Bond Lengths for Select Parameters in the Transition Structures

bond	av length (Å)	std. dev.
Rh−O	2.22	0.005
Rh−C	2.34	0.017
Rh−P_t	2.29	0.011
Rh−P_c	2.39	0.012
Rh−H_a	1.56	0.002
Rh−H_e	1.63	0.010
C−C	1.42	0.001
C−H	1.71	0.020

Table 2. Average Angles for Select Parameters in the Transition Structures

angle	av value (°)	std. dev.
O−Rh−C	74.6	0.411
O−Rh−P_t	171.4	2.608
O−Rh−P_c	89.7	0.894
O−Rh−H_a	93.0	1.074
O−Rh−H_e	93.0	0.790
C−Rh−P_t	113.6	3.076
C−Rh−P_c	108.1	3.146
C−Rh−H_a	162.1	1.180
C−Rh−H_e	83.0	1.006
P_t−Rh-P_c	86.4	0.354
P_t−Rh-H_a	79.2	2.098
P_t−Rh-H_e	89.5	0.495
P_c−Rh-H_a	84.2	2.532
P_c−Rh-H_e	168.8	4.162
H_a−Rh-H_e	84.9	1.902
C−C−H	117.4	0.709

The calculated QM structures showing the conformational switch in the DMPE ligand are shown in Figure 4. The BPE ligand undergoes a similar conformational change due to the saturated carbon linker between the two phosphorus atoms. The nine structures have very similar geometries around the rhodium catalyst and substrate. The average bond lengths and angle measurements of the structures involving the rhodium catalyst or forming and breaking bonds are shown in Table 1 for the bonds and Table 2 for the angles. A more complete table of geometric data, as well as Cartesian coordinates, is listed in the Supporting Information. The nine

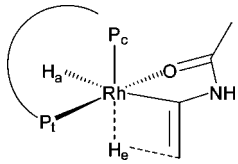


Figure 5. Rhodium catalyst and core atoms with labels used in the text.

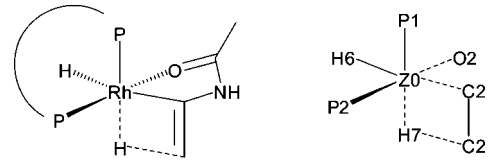


Figure 6. Substructure specification for new MM parameters. The overall structure is shown on the left, and the corresponding new atom types are shown on the right. All bonds in the substrate are single bonds, except for dashed bonds which are defined as zero order.

calculated transition structures share structural similarities, with all bond lengths and angles involving the rhodium catalyst, and the changing bonds are fairly consistent between all nine structures. The substrate binding to the rhodium shows an almost constant Rh−O bond length with an average of 2.22 Å, but the Rh−C bond length shows some variation based on the size of the ligand. The Rh−C bond length for the two achiral ligands averages 2.32 Å. However, the larger DuPHOS and BPE ligands have the Rh−C bond length slightly longer at an average of 2.35 Å, for an overall average of 2.34 Å. The ligand binds asymmetrically to the rhodium catalyst with two different Rh−P bond lengths. For clarity, the two phosphorus atoms have been named P_c and P_t, with the phosphorus proximal to the oxygen denoted as P_c and the distal phosphorus as P_t. These labels are shown in Figure 5. The Rh−P_c bond has been calculated to have a length approximately 0.1 Å longer than the Rh−P_t bond, with average bond lengths of 2.39 Å and 2.29 Å, respectively, in good agreement with the large *trans* influence of hydrides compared to neutral oxygen ligands.[19a,29]

The two hydrides bound to the Rh are also differentiated since the one in the phosphine plane is involved in the reaction coordinate. The reacting hydride in the equatorial plane of the phosphine ligand is labeled H_e, and the unreactive hydride that is out of the phosphine plane is labeled H_a and also described in Figure 5. The equatorial hydride has an expected longer average Rh−H bond distance of 1.63 Å compared to the average 1.59 Å for the axial hydride. The equatorial hydrides are also differentiated based on the size of the ligand as well. The achiral ligands have an average R−H_e bond length of 1.64 Å, whereas the larger chiral ligands have a slightly shorter Rh−H_e bond length of 1.62 Å. Furthermore, the forming C−H transition state bond length is slightly shorter for the achiral ligands, averaging 1.69 Å, whereas the chiral ligands have an average C−H bond length of 1.72 Å. This combination of data suggests that the larger steric demands of the chiral DuPHOS and BPE ligands enforce a slightly earlier transition state than the smaller achiral model ligands. Interestingly, the reacting C═C double bond does not show a noticeable bond length difference between the structures, with a 1.42 Å bond length for all calculated structures.

Modification of the MM3* Force Field. The existing MM3* force field supplied with MacroModel is insufficient to describe this transition state because it lacks suitable atom types and molecular mechanics parameters. As mentioned previously, the rhodium atom is not defined in the force field file. In addition to this, the existing parameters for hydrogen atoms are not sufficient for describing the two hydrides in the structure, and the two phosphorus atoms need to be differentiated. MacroModel includes a Z0 atom type that is designed for user definition; this atom type was used for the rhodium. The two hydrides were defined as atom types H6 (axial) and H7 (equatorial), which were amended to the atom type files as atom type numbers 46 and 47, respectively. The existing parameters for phosphorus atoms, previously labeled P0, were used in the creation of two more phosphorus atom types numbered 104 and 105, labeled as P1 and P2, and were not further optimized. The phosphorus proximal to the oxygen of the substrate was defined as being atom type P1, and the phosphorus distal to the oxygen was defined as atom type P2. The new parameters were then added to the force field file with the use of the substructure section that is utilized in MacroModel. The labeling of the substructure and given bonds are shown in Figure 6.

Scheme 2. Penalty Function for New MM Parameter Optimization

$$X^2 = \sum_i w_i^2 (y_{iQM} - y_{iMM})^2$$

New MM Parameter Optimization. The newly defined parameters were optimized through the minimization of the penalty function, as described above and in Scheme 2. In the penalty function, the weighting values are the inverse of the tolerance for each type of data. The geometric tolerances are initially set to 0.01 Å for bonds, 0.5° for angles, and 1° for torsions. The tolerance for partial charges is set at 0.02 electrons. Various tolerances are used for the Hessian matrix to emphasize important interactions, like 1,4-interactions, while reducing the impact of longer range interactions. The tolerance for relative energies was initially set at 1 kJ/mol, but this tolerance was lowered to 0.1 kJ/mol to increase accuracy, as described later in the text. The minimization of the penalty function was performed through a combination of a Newton−Raphson and Simplex optimizations. The Simplex procedure is more efficient when the penalty function is far from a minimum, and therefore this is only used at early stages of the optimization. The optimization of the penalty function is considered converged when the function is improved by less than 0.01% between steps, with only Newton−Raphson steps used when the function improvement is less than 1% between optimization steps.

Initial bond lengths and angle measurements for parameters to be optimized were set to average QM values of the corresponding parameter, and the force constants were set to initial values of 5.0 mdyn/Å for bonds and 0.5 mdyn Å/rad^2 for angles. The definition of this substructure and added parameters for new atom types are given in the Supporting Information. The optimization of the new pa-

rameters was then undertaken in several steps. The first step was the optimization of the dipole moments in the new bond parameters to fit the partial charges of the MM structures to the calculated electrostatic potential (ESP) charges of the DFT structures. During this optimization, the only parameters optimized were the dipole moments, and the reference data were only the partial charges. Once the penalty function had converged for the dipole optimization, the next step of the optimization was the initial optimization of the force constants, which was done independently for each structure. In this step, only the force constants were optimized, with the penalty function convergence criteria set to 1% overall instead of 0.01%, and the comparison data in the penalty function were based on the geometric data and the full Hessian matrix for each structure independently.

After this optimization, it was noted that several force constants tended to zero consistently among all nine structures. Some other force constants were noted to be very large. These large values corresponded to forming or breaking bonds. Since the Hessian matrix was altered to change the single negative eigenvalue to a large positive value, this generates an artificially large force constant corresponding to the reaction coordinate.[4] Therefore, these large force constants were allowable as long as they were involved in forming or breaking bonds. A single force field was then generated based on the optimized parameters for a better "initial guess" of force constants. These parameters were then optimized to reproduce the geometrical data of all nine structures along with the Hessian matrices of the achiral ligands and the relative enthalpies of diastereomeric structures. This optimization was done in two steps. First, the force constants alone were optimized with a 1% tolerance for convergence in the penalty function. Second, all geometric parameters and force constants were optimized with a 0.01% tolerance in the penalty function for convergence. In order to emphasize the importance of the relative energies in the parametrization, two force fields were derived from the QM data. The first force field, denoted as RhH, has the energy tolerance set at 1 kJ/mol for the relative energies. The second force field set the penalty threshold at 0.1 kJ/mol or a 10 times higher weighting for the error in the parametrization. This force field is denoted as the RhH-E. The resulting new parameters are listed in the Supporting Information. The bond lengths, angles, and force constants are given to four decimal places because of the number of significant figures required by the MacroModel program and is not meant to imply the level of accuracy of the parameters themselves.

The new bond parameters show excellent agreement with the average QM values for the corresponding bonds, as shown in Table 3 for the RhH and the RhH-E force fields, respectively. Furthermore, the bond parameters between the RhH and RhH-E force fields are extremely similar. The Rh−O bond length in the QM calculations had an average length of 2.22 Å. The new parameters for the Rh−O bond are 2.2213 Å for the RhH force field and 2.2091 Å for the RhH-E force field. The force constants for both fields are also similar, being 1.7982 and 1.7138 mdyn/Å, respectively. The average QM Rh−C bond length is 2.34 Å, and both

Table 3. New Bond Parameters for the RhH and RhF-E Force Fields

	RhH		RhH-E	
bond	length (Å)	force constant (mdyn/Å)	length (Å)	force constant (mdyn/Å)
Rh−P_c	2.3921	5.1238	2.3914	5.0700
Rh−P_t	2.2844	3.6079	2.2886	1.5293
Rh−O	2.2213	1.7982	2.2091	1.7138
Rh−C	2.3388	0.0000	2.3388	0.0000
Rh−H_e	1.6009	1.8394	1.6169	1.6059
Rh−H_a	1.5642	2.5618	1.684	2.5782
C−C	1.4317	5.5008	1.4281	4.1874
C−H	1.7686	18.4542	1.7492	21.9463

force fields have optimized parameters of 2.3388 Å, which is the actual average bond length when taken out to four decimal places. This parameter was not actually optimized, as the force constant for this bond was set to zero from the initial force constant optimization of each individual structure, as noted above. The parameter could not simply be omitted or the program would crash due to a missing parameter, even if that parameter does not contribute to the total energy due to the zero force constant. The Rh−P bonds were different based on their orientation to the substrate. The Rh−P_c bond length is 2.39 Å, and the Rh−P_t bond length is 2.29 Å in the QM structures. The force constants for the two bonds also varied slightly based on position and optimization procedure. The corresponding force field parameters were 2.3931 Å and 2.3914 Å, for the RhH and RhH-E force fields, respectively, for the Rh−P_c bond and 2.2844 Å and 2.2886 Å for the Rh−P_t bond. For the Rh−P_c bond, the two force fields had similar force constants, 5.1238 and 5.0700 mdyn/Å for the RhH and RhH-E force fields, respectively. The Rh−P_t had a force constant of 3.6079 mdyn/Å for the RhH force field but a weaker 1.5293 mdyn/Å for the RhH-E force field.

The two Rh−H bonds were also differentiated, with the axial, unreacting Rh−H bond being a slightly shorter 1.56 Å compared to the 1.63 Å of the reacting Rh−H bond. The parameters for the axial Rh−H bond were 1.5642 Å and 1.5684 Å, and for the reactive Rh−H bond the parameters were 1.6009 Å and 1.6169 Å for the RhH and RhH-E force fields, respectively. The force constants for the axial Rh−H bond were also larger than those for the equatorial Rh−H bond. In the RhH force field, the force constants were 2.5618 mdyn/Å for the axial bond and 1.8394 mdyn/Å for the equatorial bond, and in the RhH-E force field they were a similar 2.5782 and 1.6059 mdyn/Å.

The new parameters for the reacting C═C double bond are 1.4317 Å for the RhH force field and 1.4281 Å for the RhH-E force field, compared to an average value of 1.42 Å in the QM structures. This bond length suggests that the reacting C═C double bond is much more like a C−C single bond. The existing parameters for C−C single and double bonds utilizing the sp^2 hybridized C2 atom type in the MM3* force field have bond lengths of 1.3430 Å for a double bond and 1.4700 Å for a single bond. The force constant for the reacting C═C double bond was optimized to be 5.5008 mdyn/Å in the RhH force field and 4.1874 mdyn/Å in the

Table 4. Select New Angle Parameters for the RhH Force Field

	RhH		RhH-E	
angle	angle (°)	angle (°)	force constant (mdyn Å/rad^2)	force constant (mdyn Å/rad^2)
O−Rh−C	82.4448	71.8505	0.0009	0.0009
O−Rh−P_t	175.3066	169.9754	0.1113	0.1161
O−Rh−P_c	90.3323	90.4351	0.6735	0.6365
O−Rh−H_a	92.4769	93.0980	0.2785	1.2412
O−Rh−H_e	92.4818	92.2411	0.9752	0.9907
C−Rh−P_t	114.0324	113.9045	0.8134	1.2435
C−Rh−P_c	107.8407	108.8086	0.1108	0.1934
C−Rh−H_a	167.8640	168.5483	0.3168	0.2779
C−Rh−H_e	85.8487	85.5483	0.5178	0.8171
P_t−Rh−P_c	85.8789	86.0062	5.5794	4.7580
P_t−Rh−H_a	80.6824	84.3121	0.5802	0.4871
P_t−Rh−H_e	89.4721	89.4721	0.0000	0.0000
P_c−Rh−H_a	86.5384	85.4339	0.6378	0.7558
P_c−Rh−H_e	168.7540	168.7540	0.0000	0.0000
H_a−Rh−H_e	84.8617	84.8617	0.0000	0.0000
C−C−H	113.3287	113.3287	0.0000	0.0000

RhH-E force field. Both of these force constants are lower than for the existing ground-state bonds, being 7.5000 and 6.0000 mdyn/Å for double and single bonds, respectively.

The forming C−H bond has an average QM length of 1.71 Å, but the new values are slightly longer at 1.7686 Å for the RhH force field and 1.7492 Å for the RhH-E force field. The larger chiral ligands showed slightly longer C−H bonds at the transition states, and these larger parameters do not penalize this sterically larger environment for forcing a slightly earlier transition state. The force constant for this parameter is also rather large, 18.4542 mdyn/Å in the RhH force field and 21.9463 mdyn/Å in the RhH-E force field. However, this is an expected result of the Hessian inversion technique used in the Q2MM method. The negative eigenvalue of the Hessian matrix was replaced with a large positive value.[4] Therefore, the force constants corresponding to the reaction coordinate will be unnaturally large as a result. Interestingly, only the forming C−H bond contains an abnormally large force constant, and the reacting C=C double bond and the breaking Rh−H bonds show force constants that are within the normal range for other parameters in the MM3* force field. Since the large force constant corresponds to the reaction coordinate, this suggests that the reaction coordinate is localized in this forming C−H bond and is not coupled to a corresponding C−C bond change or Rh−H bond breaking. Analysis of the QM calculated frequencies confirms this, with the negative frequency affecting almost exclusively the forming C−H bond.

There were also 22 new angle parameters optimized in the process. As with the bonds, the angle parameters were similar between the two force fields. Selected parameters are shown in Table 4 for the RhH and RhH-E force fields. The entire set of new parameters is listed in the Supporting Information. Out of these, four parameters had force constants set to zero after the initial individual parameter optimization. As with the Rh−C bond parameter, the angle parameters with zero force constants could not be omitted from the force field to allow for the program to run correctly. These parameters were the angles from the reacting, equatorial hydrogen to both phosphorus atoms, the angle between the two hydrides on the rhodium, and the C−C−H angle of the substrate and reacting hydride. A fifth angle, the O−Rh−C angle, from the carbonyl oxygen to the α carbon of the substrate, had force constants that optimized to an extremely small 0.0009 mdyn Å/rad^2 for both force fields. The angle parameter for the O−Rh−C angle optimized to 82.4448° in the RhH force field and 71.8505° in the RhH-E force field. Compared to the average QM value of 74.6°, these optimized parameters are not in complete agreement. Since the O−Rh−C angle is part of a five-membered ring, and since the corresponding force constant is extremely low, the MM calculated angle can be defined through other parameters in the ring without a large energy penalty resulting from this discrepancy. This parameter could also be reset to the average QM angle with a zero force constant, as with the other four angles with zero force constants, but this is unnecessary since parameters with zero force constants do not affect the overall energy of the system.

The positioning of the carbonyl oxygen with respect to the phosphine ligand is determined through two O−Rh−P angle parameters corresponding to the two differentiated phosphorus atoms. The O−Rh−P_c angle was calculated to be 89.7°, and the optimized parameters are 90.3323° for the RhH force field and 90.4351° for the RhH-E force field. The O−Rh−P_t angle was calculated to be 171.4°, and the optimized parameters are 175.3066° for the RhH force field and 169.9754° for the RhH-E force field. Since the oxygen has a larger steric interaction with P_c, the agreement is more important with P_c than P_t. Furthermore, the corresponding force constants are higher for the O−Rh−P_c angle than for the P_t angle. The RhH force field has force constants of 0.6735 mdyn Å/rad^2 and 0.1113 mdyn Å/rad^2, for O−Rh−P_c and O−Rh−P_t, and the RhH-E force field has very similar values of 0.6365 mdyn Å/rad^2 and 0.1161 mdyn Å/rad^2.

The P−Rh−P ligand bite angle is also important in this system. Since the ligands were all bis-phosphine ligands with two carbon linkers, the bite angles of all the ligands were similar with an average 86.4° with very little variation between the QM calculated structures. The optimized MM parameters of 85.8789° in the RhH force field and 86.0062° in the RhH-E force field are within one degree of the average QM value, and the force constants are rather large, 5.5794 mdyn Å/rad^2 in the RhH force field and 4.7580 mdyn Å/rad^2 in the RhH-E force field. The lack of variation in bite angle in the QM data set and the resulting highly rigid angle parameter may be a source of error if a ligand with a much different bite angle is used. This error is likely to be canceled since enantiomeric excess is based on a relative energy between diastereomers, specifically the $\Delta\Delta G^{\ddagger}$. It is unlikely that the diastereomeric *pro-R* and *pro-S* transition states would involve drastically different ligand bite angles, and therefore any energy penalty due to ligand bite angle difference would be systematic. The rest of the angle parameters were optimized to within 5 degrees of the corresponding average QM value, and the force constants were within the range of angle force constants seen elsewhere in the MM3* force field. These additional parameters are described in detail in the Supporting Information.

Table 5. Comparison of Average Bond Lengths (Å) Between QM and MM Optimized Structures for Selected Parameters in the Transition State

		RhH		RhH-E	
bond	QM	av	rmsd	av	rmsd
Rh−O	2.22	2.23	0.00	2.22	0.01
Rh−C	2.34	2.34	0.01	2.34	0.01
Rh−P_t	2.29	2.30	0.01	2.32	0.03
Rh−P_c	2.39	2.40	0.01	2.39	0.01
Rh−H_{ax}	1.56	1.56	0.00	1.57	0.00
Rh−H_{eq}	1.63	1.64	0.02	1.66	0.03
C−C	1.42	1.45	0.03	1.45	0.03
C−H	1.71	1.63	0.08	1.61	0.10

Table 6. Comparison of Average Angle Measurements (deg) between QM and MM Optimized Structures for Selected Parameters in the Transition State

		RhH		RhH-E	
angle	QM	av	rmsd	av	rmsd
O−Rh−C	74.6	74.2	0.5	74.7	0.3
O−Rh−P_t	171.4	170.3	1.9	170.6	2.0
O−Rh−P_c	89.7	89.3	0.7	89.1	0.9
O−Rh−H_{ax}	93.0	92.9	0.5	92.3	1.1
O−Rh−H_{eq}	93.0	93.5	0.9	93.4	0.9
C−Rh−P_t	113.6	115.1	2.3	114.3	2.2
C−Rh−P_c	108.1	109.2	1.2	109.9	1.8
C−Rh−H_{ax}	162.1	161.1	1.5	160.8	1.7
C−Rh−H_{eq}	83.0	81.2	2.0	80.7	2.5
P_t−Rh−P_c	86.4	85.3	1.2	85.2	1.3
P_t−Rh−H_{ax}	79.2	78.6	1.5	79.7	1.3
P_t−Rh−H_{eq}	89.5	90.5	1.2	90.9	1.7
P_c−Rh−H_{ax}	84.2	83.8	1.4	83.3	1.8
P_c−Rh−H_{eq}	168.8	169.5	1.3	169.4	1.5
H_{ax}−Rh−H_{eq}	84.9	86.0	1.4	86.4	1.7

To include all parameters required by the program, ten torsion parameters were included in the force fields. Due to the pseudo-octahedral nature of the catalyst substructure, torsions are redundant when combined with the necessary bond and angle parameters. Furthermore, since the substrate and ligand are both bidentate in their binding, there is a lack of conformational flexibility that would require torsional parameters. The ten added torsions are all general parameters and are all set with $V_1 = V_2 = V_3 = 0.0000$, therefore not adding any torsional component to the energy of the substructure.

Comparison of QM and MM Optimized Structures. All nine structures showed excellent agreement between the QM and MM methods with regards to the bonds and angles defined by the new parameters. The statistics for the comparisons of the average bond lengths are shown in Table 5 and average angle measurements in Table 6. For the bond lengths, the rmsd between the QM and MM structures were 0.03 Å or less for all new bonds with both force fields, except for the forming C−H bond, where the rmsd was 0.08 Å in the RhH force field and 0.10 Å in the RhH-E force field. This slightly higher error is due to the extremely large force constant described previously that results in an almost constant value of 1.63 Å in the RhH force field and 1.61 Å in the RhH-E force field compared to the QM average value of 1.71 Å. As has been discussed earlier, the Q2MM technique results in a negative response to steric crowding of the breaking and forming bonds; the utilization of a high force constant serves to minimize the error by minimizing the bond length variation.[4,30]

Table 7. Comparison of Calculated Enthalpies (kcal/mol) between QM and MM Optimized Diastereomers and Conformers

ligand	QM	RhH	RhH-E
DMPE	−0.1	−0.3	−0.2
DuPHOS	2.3	2.1	2.2
BPE 1	2.2	2.0	2.0
BPE 2	1.9	2.5	2.4
BPE 3	3.0	2.2	2.7

The newly defined angles generally have RMSDs under 2 degrees, with a small number of exceptions. The C−Rh−H angle averages 83.0° in the QM structures, but only 81.2° in the RhH force field with an rmsd of 2.0°, and 80.7° in the RhH-E force field with an rmsd of 2.5°. Since the forming C−H bond is systematically shorter in the MM structures than in the QM structures, it follows that that this angle would also be smaller as this angle is opposite the forming C−H bond in a four-membered ring. Two other angles with large deviations are the O−Rh−P_t and C−Rh−P_t angles that have RMSDs between 1.9° and 2.3°. These deviations result from the effect of the ligand bite angle. The P−Rh−P angle averages 86.4° in the QM structures, but a slightly smaller 85.3° from the RhH force field, and 85.2° from the RhH-E force field. Since the force constant on this angle is much larger than those on the O−Rh−P_t or C−Rh−P_t angles, the bite angle is enforced at the expense of other angles. The result is a slightly smaller O−Rh−P_t angle and a slightly larger C−Rh−P_t angle. Overall, it does not grossly distort the structures and the errors are systematic and relatively small. It is also worth noting that the parameters with zero force constants did not show any excessively large deviations. The Rh−C bond has RMSDs of 0.01 Å for both force fields, and the five angles with zero, or vanishingly small, force constants had RMSDs ranging from 0.3° to 1.7°. From these data, it can be concluded that the force constants can go to zero in the optimization due to the parameters being defined through the use of other parameters within a ring.

The accurate reproduction of the relative energies between diastereomers and conformers is very important to the predictive ability of the new force field parameters. The two force fields performed moderately well in reproducing the relative energies afforded by the training set. The relative QM enthalpies and corresponding MM energies are shown in Table 7. The RhH force field generally had poorer energy reproduction than the RhH-E force field, with an rmsd of 0.5 kcal/mol compared to 0.3 kcal/mol for the RhH-E force field. The largest unsigned error for the RhH force field was between two structures that were diastereomers utilizing the BPE ligand where the RhH force field underestimated the energy by 0.8 kcal/mol. The largest unsigned error for the RhH-E force field was half of that for the RhH force field, 0.4 kcal/mol, also for diastereomers involving the BPE ligand. The energy reproductions for the conformers of the DMPE ligand and the diastereomers of the DuPHOS ligand were under 0.2 kcal/mol for both force fields. The emphasis on energy values in the RhH-E force field is thus evident in

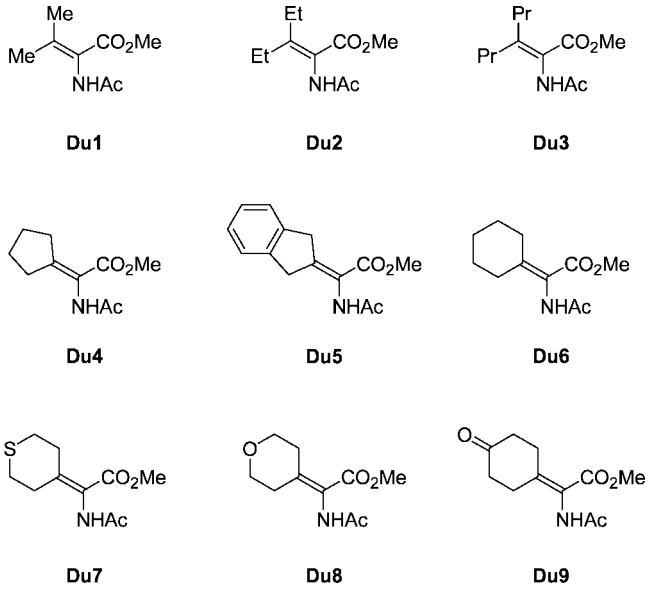


Figure 7. Substrates in the test set.

the reduction of errors in the training set. Nevertheless, the differences are fairly subtle, as can be seen in the similar optimized parameters in the two force fields. This also indicates that the optimization procedure is fairly robust at least in the present case.

Validation of the MM Parameters. Although the focus of the present manuscript is on the development rather than the application of the force field, it is important that the optimized parameters are validated using a test set that was not part of the training set. At the same time, the number of possible data sets that were studied under identical conditions and cover a range of substrates and e.e. in the literature is surprisingly small. Finally, an ideal data set would also allow a test of how well the underlying DFT calculations reproduce the interactions in the stereodetermining transition structures, even though in the absence of conformational sampling it is not expected that they accurately reproduce the experimentally observed e.e.s. As a compromise between all these boundary conditions, we chose a data set utilizing the DuPHOS and BPE ligands and a set of dehydro-α-amino acid substrates with symmetric alkyl groups at the β position, as shown in Figure 7.[31] The DuPHOS and BPE ligands were studied by DFT and used in the parameter optimization, but the experimentally observed e.e.s were not used in the parametrization. The substrates are not the same as the model substrate that was used in the QM training set and correspond to the substrates that would be used in a real-world study.

The computational results and reported experimental values are given as percent enantiomeric excesses, as shown in Figure 8 for the DuPHOS ligand and in Figure 9 for the BPE ligand. It can be seen in these two figures that the agreement between experimental and computa-

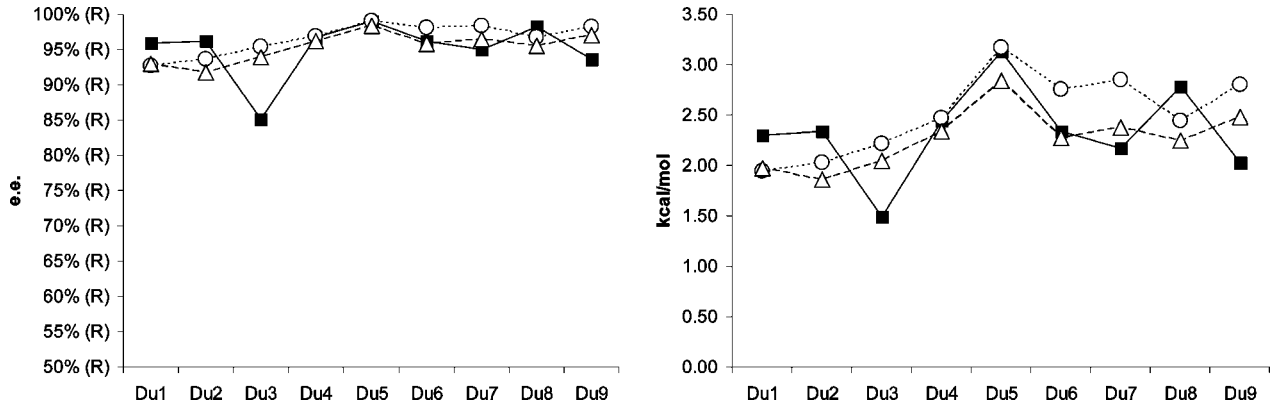


Figure 8. Comparison of computed e.e. (left) and energies (right) from the RhH force field (circles) and RhH-E force field (triangles) to experimental values (solid squares) for reactions using the DuPHOS ligand.

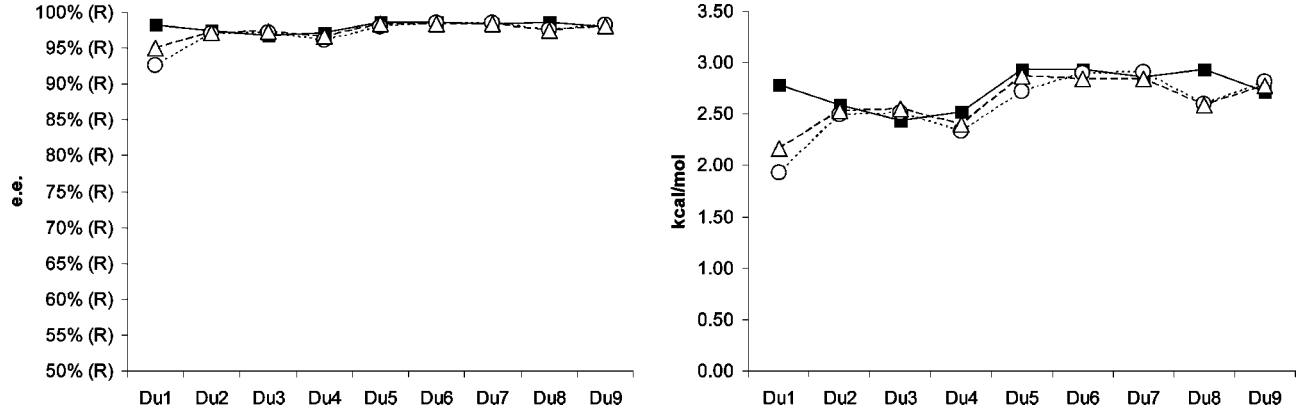


Figure 9. Comparison of computed e.e. (left) and energies (right) from the RhH force field (circles) and RhH-E force field (triangles) to experimental values (solid squares) for reactions using the BPE ligand.

tional values is very good. Because the experimentally observed e.e.s in this data set, like in most published cases, are very high, a statistical analysis in terms of e.e. is not very useful, and a discussion in terms of energies gives more insight into the performance of the method (for full listing of the energies, see tables in the Supporting Information). For the DuPHOS ligand using the RhH force field, the mean unsigned error is 0.4 kcal/mol with a maximum error of 0.7 kcal/mol, and the RhH-E force field performed slightly better with a mean unsigned error of 0.3 kcal/mol, with a maximum absolute error of 0.6 kcal/mol. The BPE ligand faired even better with a mean unsigned error of 0.2 kcal/mol for both force fields and a maximum error of 0.9 kcal/mol for the RhH force field and 0.6 kcal/mol for the RhH-E force field. The poorer performance of the DuPHOS ligand is mostly attributed to a disagreement between theory and experiment using substrate **Du3**, which is the dipropyl derivative of the substrate. This substrate gave an 85.1% e.e. experimentally, but the rest of the substrates showed e.e.s between 93.7% and 99.0%. For the chirally similar BPE ligand, **Du3** was experimentally determined to generate the reduced product in 96.8% e.e. From a statistical viewpoint, the error contributed by **Du3** is minor and does not diminish the performance of the force fields to reproduce these experimental data. Excluding this data point from the error analysis does not affect the unsigned mean error. Energetically, this point is the largest error for the DuPHOS ligand using the RhH-E force field, but the maximum unsigned error using the RhH force field is with substrate **Du9**, where the experimental e.e. of 93.7% was calculated to be a slightly higher 98.2% e.e.

Conclusions

The newly derived transition state force field parameters for the rhodium-catalyzed hydrogenation of enamides perform very well, both in comparison to the QM data and a variety of substrates. The reaction-specific force field parameters for the hydrogenation of dehydro-α-amino acid substrates with chiral bis-phosphine ligands describe the enantioselectivity of the synthesis of amino acids well and can find wide applicability to this important reaction[32] but can most likely not be used for other types of reactions or substrates. The two force fields, derived from different tolerance levels of relative energies in the Q2MM method, had new parameters that were very similar to each other and, as a result, demonstrate the robustness of the fitting procedure and performed similarly as well. At the present state of development, the new parameters meet the objective stated at the outset of this work to provide a computational tool for rapid *in silico* screening of a library of ligands to identify a smaller subset for experimental screening. This goal is met, even with the small number of false positives and false negatives that arise. Nonetheless, further modifications of the parameters may lead to improved performance of this computational screening tool.

Acknowledgment. We gratefully acknowledge the allocation of computer resources by the Center for Research Computing at the University of Notre Dame. P.J.D. is the recipient of a Schmitt Fellowship by the University of Notre Dame.

Supporting Information Available: Cartesian coordinates, imaginary frequencies, SCF energies and enthalpies of all QM transition structures calculated, expanded tables of geometric data comparing QM and MM structures, force field parameters in MacroModel format, and numerical results for Figures 8 and 9. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Burk, M. J. *Acc. Chem. Res.* **2000**, *33*, 363. (b) Genet, J.-P. *Acc. Chem. Res.* **2003**, *36*, 908. (c) Tang, W.; Zhang, X. *Chem. Rev.* **2003**, *103*, 3029. (d) Gridnev, I. D.; Imamoto, T. *Acc. Chem. Res.* **2004**, *37*, 633. (e) Cui, X.; Burgess, K. *Chem. Rev.* **2005**, *105*, 3272. (f) de Vries, J. G.; Lefort, L. *Chem. Eur. J.* **2006**, *12*, 4722. (g) Jäkel, C.; Paciello, R. *Chem. Rev.* **2006**, *106*, 2912. (h) Wu, J.; Chan, A. S. C. *Acc. Chem. Res.* **2006**, *39*, 711. (i) de Vries, J. G.; Lefort, L. *Chem. Eur. J.* **2006**, *12*, 4722. (j) Minnaard, A. J.; Feringa, B. L.; Lefort, L.; de Vries, J. G. *Acc. Chem. Res.* **2007**, *40*, 1267. (k) Zhang, W.; Chi, Y.; Zhang, X. *Acc. Chem. Res.* **2007**, *40*, 1278. (l) Johnson, N. B.; Lennon, I. C.; Moran, P. H.; Ramsden, J. A. *Acc. Chem. Res.* **2007**, *40*, 1291. (m) Roseblade, S. J.; Pfaltz, A. *Acc. Chem. Res.* **2007**, *40*, 1402.

(2) For recent reviews, compare:(a) Shurki, A.; Warshel, A. *Adv. Protein Chem.* **2003**, *66*, 249. (b) Estiu, G.; Suárez, D.; Merz, K. M. *J. Comput. Chem.* **2006**, *27*, 1240.

(3) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.

(4) (a) Norrby, P.-O. *J. Mol. Chem. (THEOCHEM)* **2000**, *506*, 9. (b) Norrby, P.-O.; Brandt, P. *Coord. Chem. Rev.* **2001**, *212*, 79. (c) Jensen, F.; Norrby, P.-O. *Theor. Chem. Acc.* **2003**, *109*, 1.

(5) DeTar, D. F.; Tenlos, C. J. *J. Am. Chem. Soc.* **1976**, *98*, 7903.

(6) Hermann, J. C.; Marti-Arbona, R.; Fedorov, A. A.; Fedorov, E.; Almo, S. C.; Shoichet, B. K.; Raushel, F. M. *Nature* **2007**, *448*, 775.

(7) For a recent implementation of this approach, compare:(a) Corbeil, C. R.; Thielges, S.; Schwartzentruber, J. A.; Moitessier, N. *Angew. Chem. Int. Ed.* **2008**, *47*, 2635.

(8) (a) Houk, K. N.; Rondan, N. G.; Wu, Y. D.; Metz, J. T.; Paddon-Row, M. N. *Tetrahedron* **1984**, *40*, 2257. (b) Houk, K. N.; Paddon-Row, M. N.; Rondan, N. G.; Wu, Y. D.; Brown, F. K.; Spellmeyer, D. C.; Metz, J. T.; Li, Y.; Loncharich, J. R. *Science* **1986**, *231*, 1108.

(9) (a) Masamune, S.; Kennedy, R. M.; Petersen, J. S.; Houk, K. N.; Wu, Y. D. *J. Am. Chem. Soc.* **1986**, *108*, 7404. (b) Dorigo, A. E.; Houk, K. N. *J. Org. Chem.* **1988**, *53*, 1650.

(10) Sherrod, M. J.; Menger, F. M. *J. Am. Chem. Soc.* **1989**, *111*, 2611.

(11) Menger, F. M.; Sherrod, M. J. *J. Am. Chem. Soc.* **1990**, *112*, 8071.

(12) (a) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., III *J. Phys Chem. A.* **2001**, *105*, 9396. (b) Chenoweth, K.; Duin, A. C. T.; Goddard, W. A., III *J. Phys. Chem. A.* **2008**, *112*, 1040.

(13) (a) Strachan, A.; van Duin, A. C. T.; Chakraborty, D.; Dasgupta, S.; Goddard, W. A., III *Phys. Rev. Lett.* **2003**, *91*, 098301. (b) Chenoweth, K.; Cheung, S.; van Duin, A. C. T.;

Goddard, W. A., III; Kober, E. M. *J. Am. Chem. Soc.* **2005**, *127*, 7192. (c) van Duin, A. C. T.; Zeiri, Y.; Dubnikova, F.; Kosloff, R.; Goddard, W. A., III *J. Am. Chem. Soc.* **2005**, *127*, 11053. (d) Bueher, M. J.; van Duin, A. C. T.; Goddard, W. A., III *Phys. Rev. Lett.* **2006**, *96*, 095505.

(14) (a) Jensen, F. *J. Am. Chem. Soc.* **1992**, *114*, 1596. (b) Jensen, F. *J. Comput. Chem.* **1994**, *15*, 1199. (c) Olsen, P. T.; Jensen, F. *J. Chem. Phys.* **2003**, *118*, 3523.

(15) (a) Nielson, K. D.; van Duin, A. C. T.; Oxgaard, J.; Deng, W.-Q.; Goddard, W. A., III *J. Phys. Chem. A* **2005**, *109*, 493. (b) Han, S. S.; van Duin, A. C. T.; Goddard, W. A., III; Lee, H. M. *J. Phys. Chem. A* **2005**, *109*, 4575. (c) Zhang, Q.; Qi, Y.; Hector, L. G., Jr.; Çağin, T.; Goddard, W. A., III *Phys. Rev. B* **2005**, *72*, 045406. (d) Ludwig, J.; Vlachos, D. G.; van Duin, A. C. T.; Goddard, W. A., III *J. Phys. Chem. B* **2006**, *110*, 4274. (e) Goddard, W. A., III; van Duin, A.; Chenoweth, K.; Cheng, M.-J.; Pudar, S.; Oxgaard, J.; Merinov, B.; Jang, Y. H.; Persson, P. *Top. Catal.* **2006**, *38*, 93.

(16) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

(17) For recent applications, see for example:(a) Rydberg, P.; Hansen, S. M.; Kongsted, J.; Norrby, P.-O.; Olsen, L.; Ryde, U. *J. Chem. Theory Comput.* **2008**, *4*, 673. (b) Rydberg, P.; Olsen, L.; Norrby, P.-O.; Ryde, U. *J. Chem. Theory Comput.* **2007**, *3*, 1765.

(18) (a) Halpern, J. *Science* **1982**, *217*, 401. (b) Chua, P. S.; Roberts, N. K.; Bosnich, B.; Okrasinski, S. J.; Halpern, J. *J. Chem. Soc., Chem. Commun.* **1981**, 1278. (c) Schmidt, T.; Baumann, W.; Drexler, H.-J.; Arrieta, A.; Heller, D. *Organometallics* **2005**, *24*, 3842. (d) Reetz, M. T.; Meiswinkel, A.; Mehler, G.; Angermund, K.; Graf, M.; Thiel, W.; Mynott, R.; Blackmond, D. G. *J. Am. Chem. Soc.* **2005**, *127*, 10305.

(19) (a) Chan, A. S. C.; Pluth, J. J.; Halpern, J. *J. Am. Chem. Soc.* **1980**, *102*, 5952. (b) Landis, C. R.; Brauch, T. W. *Inorg. Chim. Acta* **1998**, *270*, 285. (c) Landis, C. R.; Halpern, J. *J. Am. Chem. Soc.* **1987**, *109*, 1746. (d) Brown, J. M.; Parker, D. *J. Org. Chem.* **1982**, *47*, 2722. (e) Brown, J. M.; Parker, D. *Organometallics* **1982**, *1*, 950.

(20) (a) Chen, A. S. C.; Halpern, J. *J. Am. Chem. Soc.* **1980**, *102*, 838. (b) Yasutake, M.; Gridnev, I. D.; Higashi, N.; Imamoto, T. *Org. Lett.* **2001**, *3*, 1701.

(21) (a) Scott, J. W.; Keith, D. D.; Nix, G., Jr.; Parrish, D. R.; Remington, S.; Roth, G. P.; Townsend, J. M.; Valentine, D., Jr.; Young, R. *J. Org. Chem.* **1981**, *46*, 5086. (b) Koenig, K. E.; Knowles, W. S. *J. Am. Chem. Soc.* **1978**, *100*, 7561. (c) Detellier, C.; Gelbard, G.; Kagan, H. B. *J. Am. Chem. Soc.* **1978**, *100*, 7556.

(22) (a) Landis, C. R.; Hilfenhaus, P.; Feldgus, S. *J. Am. Chem. Soc.* **1999**, *121*, 8741. (b) Feldgus, S.; Landis, C. R. *J. Am. Chem. Soc.* **2000**, *122*, 12714. (c) Feldgus, S.; Landis, C. R. *Organometallics* **2001**, *20*, 2374.

(23) Donoghue, P. J.; Helquist, P.; Wiest, O. *J. Org. Chem.* **2007**, *72*, 839.

(24) *Jaguar 5.5*; Schrödinger L.L.C.: Portland, OR, 1991−2003.

(25) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.

(26) (a) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894. (b) Woods, R. J.; Khalil, M.; Pell, W.; Moffat, S. H.; Smith, V. H., Jr *J. Comput. Chem.* **1990**, *11*, 297. (c) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(27) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440–467.

(28) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(29) Chan, A. S. C.; Pluth, J. J.; Halpern, J. *Inorg. Chim. Acta* **1979**, *37*, L477.

(30) Norrby, P.-O.; Rasmussen, T.; Haller, J.; Strassner, T.; Houk, K. N. *J. Am. Chem. Soc.* **1999**, *121*, 10186.

(31) Burk, M. J.; Gross, M. F.; Martinez, J. P. *J. Am. Chem. Soc.* **1995**, *117*, 9375.

(32) For applications of this force field, see: Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. submitted for publication.

CT800132A

JCTC Journal of Chemical Theory and Computation

G3//BMK and Its Application to Calculation of Bond Dissociation Enthalpies

Wen-Rui Zheng, Yao Fu,* and Qing-Xiang Guo

Department of Chemistry, University of Science and Technology of China, Hefei 230026, China

Received March 2, 2008

Abstract: On the basis of systematic examinations it was found that the BMK functional significantly outperformed the other popular density functional theory methods including B3LYP, B3P86, KMLYP, MPW1P86, O3LYP, and X3LYP for the calculation of bond dissociation enthalpies (BDEs). However, it was also found that even the BMK functional might dramatically fail in predicting the BDEs of some chemical bonds. To solve this problem, a new composite *ab initio* method named G3//BMK was developed by combining the strengths of both the G3 theory and BMK. G3//BMK was found to outperform the G3 and G3//B3LYP methods. It could accurately predict the BDEs of diverse types of chemical bonds in various organic molecules within a precision of *ca.* 1.2 kcal/mol.

1. Introduction

Bond dissociation enthalpy (BDE) is defined as the enthalpy change of the gas-phase reaction A-B → A• + B• at 298 K, 1 atm.[1] A sound knowledge of BDE is fundamental to understanding diverse chemical and biochemical processes.[2] For that reason, there have been considerable efforts devoted to the determination of BDEs of various molecules.[3] Nonetheless, because of the experimental difficulties in dealing with highly reactive radical species, it remains difficult to reliably measure the BDEs of many important compounds. Much controversy has also taken place in the literature over the past 20 years concerning the BDEs of some simple molecules.[3]

An alternative approach to obtain BDE is to use quantum chemistry theories. This approach has proven to be useful and important in many fields because it is fast and relatively inexpensive. However, because open-shell radical systems are involved, one should be careful to choose theoretical methods for BDE calculation. Generally speaking, unrestricted Hartree−Fock and perturbation methods are not recommended for BDE calculation due to their spin-contamination problems.[4] More sophisticated *ab initio* methods such as QCISD and CCSD tend to outperform HF and MP2, but these advanced methods may also seriously underestimate some BDEs.[5]

The density functional theory (DFT) is currently a more practicable method for BDE calculation.[6,7] This method usually does not show serious spin-contamination and, therefore, is believed to be desirable for open-shell systems. The relatively low CPU-cost of the DFT method is also advantageous. Nevertheless, Jursic's[8] and our previous studies[9] showed that some popular DFT methods tend to underestimate BDEs by *ca.* 2−6 kcal/mol! This conclusion was supported by several more recent studies.[10] Some of the latest studies also showed from a more general perspective that popular DFT models (e.g., B3LYP) could systematically underestimate various reaction energies in treatments of organic molecules.[11,12]

The reliability problem of the DFT methods calls for the development and performance test of new functionals for each particular computational target.[13] Here we report on the basis of systematic tests that the newly developed BMK functional considerably outperforms many other popular functionals including B3LYP, B3P86, KMLYP, MPW1P86, O3LYP, and X3LYP for the BDE calculation. From this finding we further develop a new composite *ab initio* method named G3//BMK. This new method can accurately and reliably predict the BDEs of diverse types of chemical bonds in various organic molecules within a precision of *ca.* 1.2 kcal/mol. All calculations were performed with the Gaussian 03 suite of programs.[14]

* Corresponding author e-mail: fuyao@ustc.edu.cn.

10.1021/ct800070y CCC: $40.75
Published on Web 07/03/2008

Table 1. Correlation between the Experimental BDEs and Theoretical Predictions by Different DFT Methods[c]

compound	exp[a]	B3LYP[b]	B3P86[b]	KMLYP[b]	X3LYP[b]	O3LYP[b]	MPW1P86[b]	BMK[b]
CH_3-CH_3	90.2 ± 0.2	84.2	88.5	93.2	84.8	85.1	87.9	90.1
$CH_3-C_2H_5$	88.5 ± 0.5	81.2	85.5	91.3	81.9	81.6	85.3	88.6
CH_3-nC_3H_7	88.9 ± 0.7	81.4	85.7	91.6	82.1	81.7	85.5	88.5
CH_3-iC_3H_7	88.2 ± 0.9	78.4	82.8	89.7	79.3	78.0	82.9	86.6
CH_3-nC_4H_9	88.8 ± 0.7	81.3	85.6	91.5	82.0	81.6	85.4	88.8
$CH_3-C\equiv CH$	126.5 ± 1.0	124.1	128.3	134.7	124.8	123.6	128.1	129.7
$C_2H_5-C\equiv CH$	124.5 ± 1.0	120.7	124.7	131.9	121.5	119.7	124.8	127.2
$CH_3-CH=CH_2$	101.9 ± 1.5	94.8	99.2	105.2	95.5	94.9	98.9	101.0
$CH_3-C_6H_5$	102.0 ± 1.0	95.8	100.3	107.2	96.6	95.3	100.1	102.3
$HC\equiv C-C_6H_5$	141.2 ± 1.4	136.8	140.7	148.4	137.7	134.9	140.8	142.3
$CH_2=CH$-C_6H_5	115.2 ± 1.3	108.1	112.6	120.4	108.9	106.3	112.5	114.5
$C_2H_5-C_6H_5$	100.2 ± 1.0	92.3	96.8	104.8	93.2	91.1	97.0	100.3
MD	-	−6.4	−2.1	+4.5	−5.7	−6.9	−2.2	+0.3
RMSE	-	6.7	2.8	4.9	6.0	7.1	2.9	1.4

[a] Experimental values are obtained from ref 15. [b] For each method, the geometry optimization and single-point energy calculation are performed by using the same density functional as designated. The basis set for the geometry optimization is 6−31G(d), and the basis set for the single-point energy calculation is 6−311++G(2df,2p). [c] Unit: kcal/mol.

2. Performance of BMK in BDE Calculation

2.1. BMK Is Superior to the Other Functionals. Our finding initially arises from the study on the C−C BDEs, where we seek to systematically re-examine all the available experiment BDEs for carbon−carbon bonds by using theoretical methods. According to our previous work,[9] although almost all of the DFT methods could not reliably calculate the absolute BDEs, some DFT methods could accurately predict the relative BDEs between analogous compounds. Thus we have tested a number of different functionals for calculating C−C BDEs of 12 representative compounds (Table 1). Not surprisingly, it is found that the popular B3LYP[16] and B3P86[17] methods systematically underestimate the C−C BDEs by 6.4 and 2.1 kcal/mol, respectively (as indicated by the mean deviations (MD) in Table 1). The root of mean square errors (RMSE) of the two methods are also as high as 6.7 and 2.8 kcal/mol.

Evidently the B3LYP and B3P86 functionals are not ideal for the BDE calculation. Consequently we turn to several newly developed functionals. These include the following: (1) KMLYP that uses a mix of Slater exchange and exact exchange for the exchange functional and a mix of the correlation functional of VWN and LYP;[18] (2) X3LYP that is constructed for general-purpose applications in thermochemistry;[19] (3) O3LYP that uses the OPTX exchange functional developed by Handy and Cohen;[20] (4) MPW1P86 that uses MPW1 for the exchange functional and the P86 nonlocal correction functional;[21] and (5) BMK recently developed by Boese and Martin.[22] Unfortunately it is found that except for BMK, all these new functionals fail to predict the C−C BDEs within chemical accuracy (∼1 kcal/mol). The KMLYP functional overestimates the C−C BDEs by *ca.* 5 kcal/mol, whereas X3LYP, O3LYP, and MPW1P86 underestimate the C−C BDEs by *ca.* 2−7 kcal/mol.

Note that the BMK functional was developed to simulate a variable exact exchange by combining exact exchange (42%) and terms depending on the kinetic energy density.[22] The purpose of this strategy was to generate a back-correction for excessive HF exchange in systems where that would be undesirable. Previously BMK has been shown to predict accurate heats of formation, reaction barriers, and enthalpies.[22] In a very recent study by Van Speybroeck and coworkers it was also reported that BMK could reliably predict the C−H BDEs with a mean deviation of *ca.* 0.5 kcal/mol.[10] Here, our calculations of C−C BDEs show that BMK has a mean deviation of +0.3 kcal/mol and a root of mean square error of 1.4 kcal/mol. Thus we conclude that BMK is superior to the other functionals for BDE calculation.

2.2. Scope and Limitation of BMK in BDE Calculation. The above work has shown that BMK can reliably predict C−C BDEs. Here we further examine whether or not BMK can accurately calculate the BDEs of other types of chemical bonds in organic molecules. To this end we systematically select a number of representative compounds that possess the C−H, C−C, C−F, C−Cl, O−H, O−O−, C−O, N−O, O−S, F−O, Cl−O, P−H, N−H, N−N, N−C, F−N, Cl−N, S−H, S−S, S−C, Si−H, Si−Si, Si−C, Si−O, Si−N, Si−F, and Si−Cl bonds (Table 2). The BDEs of these compounds are calculated by the BMK/6−311++G(2df,2p)//BMK/6−31G(d) method. For comparison we also show the predictions by the B3LYP/6−311++G(2df,2p)//B3LYP/6−31G(d) method.

Analysis of the calculation results indicates that both the B3LYP and BMK predictions exhibit good correlations with the experimental BDE values (Figure 1). However, it is evident from Figure 1 that the B3LYP method systematically underestimates the BDEs by *ca.* 5 kcal/mol (MD = −4.8 kcal/mol). The RMSE error of the B3LYP method is also as high as 5.7 kcal/mol. More seriously, from Table 2 we identify 22 compounds for which the B3LYP predictions differ from the corresponding experimental results by over 8 kcal/mol! The compounds possess chemical bonds including the C−Cl, O−O, C−O, S−O, C−N, N−N, S−S, Si−C, Si−O, Si−N, and Si−Cl bonds. The serious failure of the B3LYP method to calculate these particular bond energies casts more doubts on the reliability of using B3LYP to model organic chemistry.[23]

In comparison to B3LYP, the BMK method dramatically improves the MD to −1.4 kcal/mol. The RMSE error of the BMK method is also significantly reduced to 2.7 kcal/mol. These parameters again indicate that BMK constitutes a good-quality, yet low-cost theoretical method for the predic-

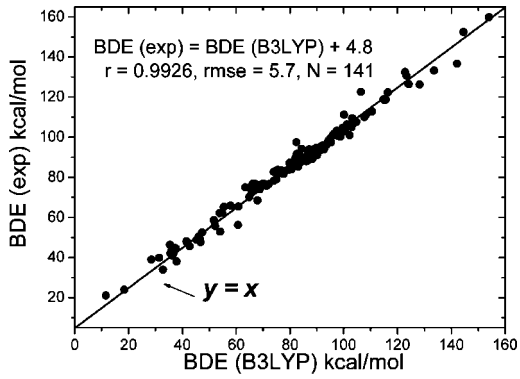
Table 2. Comparison between the Experimental BDEs and Theoretical Predictions by the B3LYP, BMK, and G3//BMK Methods[a]

compounds	exp	B3LYP[b]	BMK[c]	G3//BMK	G3	G3//B3LYP
CH_3-H	105.0 ± 0.1	102.9	104.4	104.6	104.2	104.3
$H-C_2H_5$	100.5 ± 0.3	98.3	100.1	101.1	101.2	101.0
H-nC_3H_7	100.9 ± 0.5	98.5	100.7	101.7	101.5	101.3
H-nC_4H_9	100.7	98.5	100.3	101.3	101.5	101.3
H-iC_4H_9	100.2 ± 1	98.8	100.8	101.8	101.8	101.6
H-sC_4H_9	98.3 ± 0.5	94.9	97.2	98.9	99.2	98.9
H-tC_4H_9	95.7 ± 0.7	91.9	95.0	97.2	97.4	97.2
HC≡C−H	133.3 ± 0.1	133.6	135.0	133.5	132.9	132.2
H-$CHCH_2$	111.2 ± 0.8	108.3	109.7	110.3	110.3	110.1
H-CH_2CHCH_2	88.2 ± 0.7	83.8	86.1	87.4	86.9	87.4
H-$CH(CH_3)CHCH_2$	83.8	79.7	82.6	84.7	84.2	84.6
H-$CH_2CHCHCH_3$(*trans*)	85.3	83.1	85.5	87.2	86.8	87.2
H-$CH_2CHCHCHCH_2$	83 ± 3	77.7	80.2	82.4	81.5	82.4
H-$CH(C_2H_3)_2$	76.6 ± 1.0	72.1	74.9	77.8	77.1	78.2
H-$CH(CH_3)CCCH_3$	87.3 ± 2.3	82.8	86.3	88.6	88.6	88.4
H-cyclopropylmethyl	97.4 ± 1.6	95.3	97.3	98.7	98.6	98.6
H-cyclopenta-1,3-dien-5-yl	83.9 ± 0.5	79.4	81.3	82.8	83.4	82.6
H-cyclopentyl	95.6	92.5	94.9	96.5	96.3	96.2
H-cyclohexa-1,4-dien-3-yl	76.0 ± 1.2	70.7	73.7	76.4	75.7	76.3
H-cyclohexyl	99.5	95.1	97.7	99.2	99.5	99.2
C_6H_5-H	112.9 ± 0.5	110.5	111.7	114.7	114.6	114.3
H-CN	126.3 ± 0.2	128.2	131.2	126.5	127.5	126.4
H-CH_2CN	96	92.3	95.3	96.2	96.1	96.2
H-$CH(CH_3)CN$	94.0 ± 3	87.2	90.7	92.6	92.8	92.6
$H-C(CH_3)_2CN$	91.9	**(83.5)**	87.5	89.9	90.3	89.8
H-CHCO	107.1	103.4	106.3	105.8	105.5	105.7
H-$COCHCH_2$	89.1	88.3	88.7	90.2	91.2	90.1
H-COC_2H_5	88.7	87.0	87.2	89.7	89.6	89.7
H-CH_2CHO	94.3 ± 2.2	91.8	94.2	95.2	95.4	95.2
H-CH_2COCH_3	95.9 ± 0.7	93.0	95.4	96	96.3	95.9
H-$CH(CH_3)COCH_3$	92.3 ± 1.7	86.7	89.6	91.2	91.6	91.4
H-CH_2OCH_3	96.1	93.3	94.7	96.5	96.4	96.4
H-tetrahydrofuran-2-yl	92.1 ± 1.6	89.9	91.6	93.7	93.8	93.7
H-CH_2OH	96.1 ± 0.15	93.6	94.8	96.4	96.3	96.3
H-$CH(OH)CHCH_2$	81.6 ± 1.8	77.6	79.7	82.8	80.1	80.8
H-CH_2F	101.3 ± 1	98.7	100.1	101.2	101.3	101.1
H-CHF_2	103.2 ± 1	100.3	101.8	103.7	101.8	101.8
H-CHFCl	100.8 ± 2.4	96.0	98.2	99.4	99.2	99.2
H-CF_2Cl	100.7 ± 2	98.4	100.4	102.4	102.1	102.2
H-$CFCl_2$	98.2 ± 2	94.2	96.6	98.4	97.9	98.0
H-$CH(CH_3)Cl$	97.2 ± 0.4	94.3	96.9	98.1	97.8	97.6
$H-C(CH_3)Cl_2$	95.1 ± 1.2	91.5	94.5	95.9	95.4	95.4
$H-C_2F_5$	102.7 ± 0.5	99.7	102	104.3	104.2	104.2
H-CCl_2CHCl_2	94 ± 2	88.9	93.2	94.8	94.5	94.6
H-CH_2SCH_3	93.7 ± 1.4	91.1	94.0	94.5	94.0	94.0
H-CF_3	106.4 ± 0.7	103.3	105.2	108.3	107.0	107.1
H-$COCF_3$	91 ± 2	90.1	90.7	93.2	93.0	93.1
H-CHO	88.0 ± 0.2	86.1	86.3	88.5	88.4	88.4
H-CH_2OH	96.1 ± 0.2	93.4	94.7	96.4	96.3	96.3
H-CH_2SH	93.9 ± 2.0	92.6	95.3	95.6	95.2	95.2
CH_3-CH_3	90.2 ± 0.2	84.2	90.2	88.8	88.4	88.7
$CH_3-C≡CH$	126.5 ± 1.0	124.1	129.7	126.0	125.3	124.7
CH_3-CH_2CCH	76.6 ± 1.2	69.5	77.4	77.9	77.7	77.9
$CH_3-CH_2C(CH_3)CH_2$	74.1 ± 1.0	67.3	75.5	76	75	75.7
$CH_3-CH(CH_3)CCCH_3$	76.7 ± 1.5	**(66.1)**	75.3	77.5	77.2	77.2
$nC_3H_7-CH_2CHCH_2$	75.0 ± 1.0	**(63.4)**	72.2	74.5	73.9	74.4
CH_3-CH_2CN	83.2 ± 3	**(74.6)**	82.8	82.9	82.5	82.9
CH_3-CH_2OH	87.2 ± 1.0	79.8	86.1	86.8	86.3	86.5
CN-CN	136.7 ± 1.6	142.1	**147.5**	138.1	139.8	137.8
CH_3-CH_2SH	82.5 ± 2.2	75.9	83.9	83.6	82.7	82.9
F-CF_3	130.7 ± 0.5	123.3	128.7	131.8	130.6	130.3
F-CF_2Cl	122.3	116.4	121.5	123.9	123.6	123.5
F-CH_3	110.0 ± 2.0	107.7	111.3	109.9	109.7	109.8
Cl-CN	101.0 ± 2.0	102.2	**107.7**	102.3	103.8	102.4
Cl-CF_3	87.3 ± 0.9	79.9	86.5	88.1	87.4	87.0
Cl-CHFCl	82.7 ± 3.2	**(74.0)**	81.3	81.3	81.3	80.9
Cl-$CFCl_2$	76.7 ± 2	**(67.2)**	75.0	77.1	76.9	76.6
Cl-CH_3	83.7 ± 0.4	79.2	86.0	82.5	82.5	82.4
HO-H	118.8 ± 0.1	115.6	116.4	118.1	118.0	118.1
CH_3O-H	104.6 ± 0.7	99.6	101.1	105.0	105.1	104.8

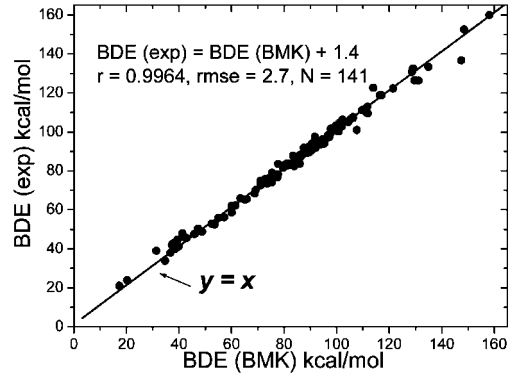
Table 2. Continued

compounds	exp	B3LYP[b]	BMK[c]	G3//BMK	G3	G3//B3LYP
CF_3O-H	118.8	114.8	117.2	119.3	119.2	119.2
$H-OC(CH_3)_3$	106.3 ± 0.7	101.1	102.3	106.8	106.8	106.7
H-OOH	87.8 ± 0.5	82.6	83.4	87.1	86.7	87.1
H-ONO	78.3 ± 0.5	74.1	77.7	77.8	77.6	78.0
HO-OH	50.4 ± 0.1	46.1	47.3	48.6	48.2	48.7
$HO-OCH_2C(CH_3)_3$	46.3 ± 1.9	**(35.4)**	41.4	46.4	42.8	42.5
HO-OF	45.6 ± 2	42.7	42.9	45.2	44.8	45.4
$HO-OCH_3$	44.7 ± 1	37.5	**39.5**	43.6	43.4	43.4
CH_3O-OCH_3	39.0 ± 1.5	**(28.5)**	**31.4**	38.7	39.0	38.4
$HO-CH_3$	92.0 ± 0.2	87.1	92.1	91.0	90.5	90.9
CH_3-OCH_3	83.6 ± 1.0	**(75.4)**	81.7	83.6	83.3	83.2
$CH_3-OC(O)H$	91.7 ± 3	**(82.6)**	90.4	91.7	92.8	90.5
CH_3-ONO	58.6	51.7	60.1	58.2	57.8	58.5
HO-NO	49.3	46.5	47.2	48.3	47.6	48.5
CH_3O-NO	41.8 ± 1.0	36.7	38.6	42.7	42.4	42.8
nC_3H_7O-NO	42.8 ± 1.5	35.8	37.9	42.6	42.2	42.6
iC_3H_7O-NO	41.9 ± 1	35.8	37.8	43	42.9	43.1
nC_4H_9O-NO	42.5 ± 1.5	35.9	37.6	42.4	42.3	42.8
iC_4H_9O-NO	42.0 ± 1.5	35.6	37.4	42.4	42.2	42.6
sC_4H_9O-NO	41.5 ± 0.8	36.1	37.6	42.9	46.3	43.1
$HO-NO_2$	48.8	45.3	48.8	48.5	48.2	48.7
HO_2-NO_2	23.9 ± 0.7	18.4	20.3	23.9	23.6	24.2
HO-SH	70.1 ± 4	64.9	69.2	68.9	68.2	68.4
$HO-SCH_3$	72.6 ± 3	65.7	71.2	72.1	70.7	70.5
HO-SOH	74.9 ± 3	**(66.6)**	71.0	72.2	71.7	71.6
$F-OCF_3$	48 ± 1	41.5	**41.3**	45.6	45.7	45.6
$Cl-OCH_3$	47.5	41.8	45.8	48.5	48.7	47.9
HO-Cl	55.8	52.1	55.0	54.8	54.8	54.6
$H-PH_2$	83.9 ± 0.5	80.6	81.5	81.5	81.6	81.7
$H-NH_2$	107.6 ± 0.1	104.7	106.2	106.8	106.8	106.7
$H-NHCH_3$	101.6 ± 2	96.2	97.8	99.7	99.6	99.4
$H-N(CH_3)_2$	94.6 ± 2	89.8	91.7	94.2	94.2	93.9
$H-NF_2$	75.7 ± 2.5	71.3	72.7	74.2	73.8	74.3
H-NO	47.7	46.8	46.1	48.2	47.5	48.4
H_2N-NH_2	65.5 ± 0.4	60.8	65.8	64.5	64.1	64.3
NH_2-NHCH_3	65.9 ± 2	**(57.9)**	63.4	64	63.6	63.7
$NH_2-N(CH_3)_2$	62.1 ± 2	**(53.9)**	60.1	62.5	62.3	62.1
F_2N-NF_2	21 ± 1	**(11.6)**	17.3	20.7	19.6	20.8
CH_3-NO	41.1	36.1	39.8	39.1	38.0	39.0
tC_4H_9-NO	39.9	**(31.4)**	38.6	41.6	41.0	41.6
CH_3-NO_2	62.3 ± 0.5	54.8	61.5	60.5	59.9	60.3
$CH_3-N_2CH_3$	52.5	47.4	53.6	53.7	53.1	53.5
F-NO	56.2	60.7	57.3	61.2	61.7	61.7
$F-NO_2$	52.9	54.1	52.4	54.8	55.3	55.2
$F-NH_2$	68.5	67.9	68.9	69.3	69.2	69.3
Cl-NO	38.0 ± 0.2	37.9	36.8	37.7	37.4	37.5
$Cl-NO_2$	33.9 ± 0.3	32.8	34.7	34.7	34.6	34.2
HS-H	91.2 ± 0.1	89.0	89.4	90.4	90.3	90.4
$H-SCH_3$	87.4 ± 0.5	83.8	85.7	87.3	86.2	86.0
H-SOH	79 ± 3.5	75.0	75.4	77.5	77.2	77.3
PhS-H	83.5 ± 1.1	76.7	**77.8**	81.2	80.8	81.2
CH_3S-SCH_3	65.2 ± 0.9	**(55.4)**	65.1	64.7	63.1	62.2
$HS-CH_3$	74.7 ± 1	68.7	75.0	73.2	72.8	72.9
HS-C(O)H	74 ± 2	68.8	72.5	72.7	72.9	72.7
CH_3S-CH_3	73.6 ± 0.8	65.7	73.6	73.2	71.9	71.8
SiH_3-H	91.7 ± 0.5	90.0	91.6	92.1	91.7	91.6
$H-SiH_2CH_3$	92.7 ± 1.2	90.5	91.5	92.2	92.7	92.6
$H-SiH(CH_3)_2$	93.5 ± 1.2	91.3	93.0	93.9	93.7	93.6
$H-Si_2H_5$	89.1 ± 2	87.1	87.7	88.4	88.9	88.8
Me_3Si-H	94.6 ± 1.7	91.9	93.3	94.3	94.6	94.5
SiF_3-H	103.2 ± 1.2	97.4	99.8	100.7	101.0	101.1
SiH_3-SiH_3	76.7 ± 1	70.4	75.5	76.0	76.0	75.9
SiH_3-CH_3	89.6 ± 1.2	82.1	89.6	87.8	87.6	87.6
Me_3Si-CH_3	94.2 ± 2	**(84.3)**	92.3	93.0	92.5	92.4
Me_3Si-OH	132.6 ± 2.0	**(122.8)**	129.1	131.9	131.8	131.7
$Me_3Si-OMe$	122.6 ± 2.6	**(106.4)**	**113.9**	120.5	120.6	119.9
$Me_3Si-NMe_2$	97.5 ± 2	**(82.4)**	**91.7**	98.0	98.2	97.6
SiH_3-F	152.5 ± 1.2	**(144.5)**	148.5	150.8	150.7	150.3
Me_3Si-F	159.9 ± 4.8	154.0	158.2	161.5	161.6	161.2
SiH_3-Cl	109.5 ± 1.7	103.2	111.9	108.2	108.4	107.9
$SiCl_3-Cl$	111.3 ± 1	**(100.1)**	110.6	110.7	110.9	110.5
MD		−4.8	−1.4	−0.004	−0.2	−0.3
RMSE		5.7	2.7	1.2	1.5	1.4

[a] Unit: kcal/mol. For the B3LYP and BMK methods, the basis set for the geometry optimization is 6−31G(d), and the basis set for the single-point energy calculation is 6−311++G(2df,2p). [b] Values in bold and in parentheses are theoretical predictions that differ from the corresponding experimental BDEs by more than 8 kcal/mol. [c] Values in bold with underlines are theoretical predictions that differ from the corresponding experimental BDEs by more than 5 kcal/mol.



(a)



(b)

Figure 1. Correlation of the experimental BDE values with the B3LYP (a) and BMK (b) predictions.

tion of BDEs for most types of chemical bonds commonly found in organic molecules. Therefore, if a DFT method *has to* be used to estimate the strength of a chemical bond of an organic compound (either because of the limitation of CPU resource or due to the bulkiness of the molecule), we suggest that BMK be considered with priority.

Nonetheless, it should also be pointed out that BMK, like all the other DFTs tested previously, is *NOT* free from failure in predicting BDEs. As shown in Table 2, the BMK predictions for eight compounds are different from the corresponding experimental values by over 5 kcal/mol. These compounds include NC-CN (error = 10.8 kcal/mol), Cl-CN (error = 6.7 kcal/mol), HO-OCH_3 (error = 5.2 kcal/mol), CH_3O-OCH_3 (error = 7.6 kcal/mol), F-OCF_3 (error = 6.7 kcal/mol), PhS-H (error = 5.7 kcal/mol), Me_3Si-OMe (error = 8.7 kcal/mol), and Me_3Si-NMe_2 (error = 5.8 kcal/mol). Evidently these magnitudes of errors are not acceptable for any meaningful theoretical prediction.

Note that the failure of the BMK method in predicting the above eight BDEs is not due to the limited size of the basis set. As demonstrated in Table 3, BMK still erratically predicts the BDE values of the same eight compounds even when the basis set is extrapolated to the infinite size using the following empirical function as developed by Kahn and Bruice.[24]

$$E_x = E_\infty + A_3X^{-3} + A_5X^{-5} \quad (1)$$

The similar extrapolations to infinite basis set based on DFT calculations to O−H and O−O BDEs were reported.[25,26] This finding prompts us to develop the G3//BMK method, which hopefully may combine the strengths of both the composite *ab initio* approach (strength in handling thermodynamics as compared to the other *ab initio* approaches) and the BMK method (strength in handling organic radicals as compared to the other DFT methods).

3. Development of G3//BMK Theory

3.1. Scale Factor for BMK. As demonstrated in several previous studies, the use of harmonic frequency scaling factors may improve the accuracy in the calculation of enthalpies.[27,28] For the purpose of developing a high-quality G3//BMK theory, we need to obtain the scale factor for the BMK method (specifically, BMK/6−31G(d)) at first.

To save computing time we select 28 molecules from the G2/97 test set,[29] for which 169 experimental harmonic frequencies are available with high precision (Supporting Information). The BMK/6−31G(d) method is used to calculate these frequencies. Subsequently we determine the optimum scaling factor λ through a least-squares procedure by minimizing the residuals

$$\Delta = \sum_i^{all} (\lambda\varpi_i^{theor} - \tilde{\nu}_i^{exp})^2 \quad (2)$$

leading to

$$\lambda = \sum_i^{all} \varpi_i^{theor}\tilde{\nu}_i^{exp} / \sum_i^{all} (\varpi_i^{theor})^2 \quad (3)$$

where ϖi^{theor} and $\tilde{\nu}_i^{exp}$ are the ith theoretical harmonic and ith experimental fundamental frequencies (in cm^{-1}), respectively. Using this approach the optimum scaling factor λ is determined to be 0.9414 for the BMK/6−31G(d) method.

3.2. G3//BMK Protocol. The steps in G3 theory have been discussed in detail in ref 30. In the BMK-based method proposed here, we have made two modifications. First, the geometries are obtained at the BMK/6−31G(d) level instead of MP2(FU)/6−31G(d) (Note: FU = all electrons included in the correlation calculation). Second, the zero-point energies are obtained at the BMK/6−31G(d) level (At the same time, we also calculated the B3LYP/6−31G(d) scale factor using these 28 molecules, and the result is 0.9615, which is nearly the same with Radom's result 0.9614.[27] Considering that in G3//B3LYP theory the B3LYP/6−31G(d) scale factor 0.96 is used, so we used 0.94 in G3//BMK.) instead of HF/6−31G(d). All of the other steps remain the same with the exception of the values of the higher-level correction parameters (see below). The G3 theory modified in this manner are referred to as G3//BMK theory (Table 4).

The higher-level correction (HLC) takes into account remaining deficiencies in the energy calculations and is similar to that used in G3 theory except that the parameters have been reoptimized. The HLC is $-An_\beta - B(n_\alpha - n_\beta)$ for molecules and $-Cn_\beta - D(n_\alpha - n_\beta)$ for atoms (including atomic ions). The n_β and n_α are the number of β and α valence electrons, respectively, with $n_\alpha \geq n_\beta$. The A, B, C, and D values are chosen to give the smallest average absolute deviation for the G2/97 test set of experimental energies. Note that the G2/97 test set contains 148 enthalpies of

Table 3. Effect of Basis Sets on BMK Calculation[a,c]

compounds	exp	6−311++G (2df,2p)	cc-pVDZ	cc-pVTZ	cc-pVQZ	∞[b]
CN-CN	136.7 ± 1.6	147.5	146.7	146.9	146.7	146.5
Cl-CN	101.0 ± 2.0	107.7	102.6	105.9	106.6	107.1
HO-OCH_3	44.7 ± 1	39.5	41.9	40.8	39.9	39.0
CH_3O-OCH_3	39.0 ± 1.5	31.4	33.9	32.8	31.8	30.8
F-OCF_3	48 ± 1	41.3	40.5	42.1	41.8	41.3
PhS-H	83.5 ± 1.1	77.8	75.2	77.3	78.1	78.8
Me_3Si-OMe	122.6 ± 2.6	113.9	108.6	114.0	114.5	114.6
Me_3Si-NMe_2	97.5 ± 2	91.7	88.3	91.7	91.9	91.8

[a] The basis set for the geometry optimization is 6−31G(d). [b] Extrapolated values to infinite basis set. [c] Unit: kcal/mol.

Table 4. Steps in G3//BMK Theory[a]

step	G3//BMK
geometry	BMK/6−31G(d)
single-point energies	MP4(FC)/6−31G(d)
	MP4(FC)/6−31+G(d)
	MP4(FC)/6−31G(2df,p)
	QCISD(T,FC)/6−31G(d)
	MP2(FU)/G3large
spin−orbit correction (ΔSO)	atomic species
high-level correction (ΔHLC)	molecules (A = 6.652, B = 3.039)
	atoms (C = 6.589, D = 1.281)
zero-point energy (ZPE)	BMK/6−31G(d)

[a] Note: FC = frozen core approximation for the correlation calculation.

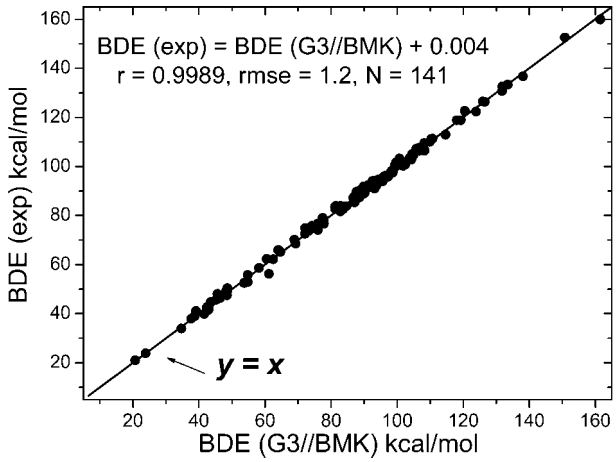


Figure 2. Correlation of the experimental BDE values with G3//BMK predictions.

formation of neutrals at 298 K ($\Delta_f H^0$), 88 ionization potentials (IP_0), 58 electron affinities (EA_0), and 8 proton affinities (PA_0) for a total of 302 reaction energies.[29] In our assessment, we have used the G2/97 test set less three ionization potentials ($C_6H_5CH_3 \rightarrow C_6H_5CH_3^+$, $C_6H_5NH_2 \rightarrow C_6H_5NH_2^+$, $C_6H_5OH \rightarrow C_6H_5OH^+$) resulting in a total of 299 energies. These energies are calculated by the following equations:

$$\Delta_f H^0(A_xB_yH_z, 298\ \mathrm{K}) = \Delta_f H^0(A_xB_yH_z, 0\ \mathrm{K}) + [H^0(A_xB_yH_z, 298\ \mathrm{K}) - H^0(A_xB_yH_z, 0\ \mathrm{K})] - x[H^0(A, 298\ \mathrm{K}) - H^0(A, 0\ \mathrm{K})]_{\mathrm{st}} - y[H^0(B, 298\ \mathrm{K}) - H^0(B, 0\ \mathrm{K})]_{\mathrm{st}} - z[H^0(H, 298\ \mathrm{K}) - H^0(H, 0\ \mathrm{K})]_{\mathrm{st}} \quad (4)$$

where $\Delta_f H^0(A_xB_yH_z, 0\mathrm{K}) = x\Delta_f H^0(\mathrm{A}, 0\mathrm{K}) + y\Delta_f H^0(\mathrm{B}, 0\mathrm{K}) + z\Delta_f H^0(\mathrm{H}, 0\mathrm{K}) - \Sigma D_0$

$$\mathrm{EA}_0 = E_0(\mathrm{neutral}) - E_0(\mathrm{anion}) \quad (5)$$

$$\mathrm{IP}_0 = E^0(\mathrm{cation}) - E^0(\mathrm{neutral}) \quad (6)$$

and

$$\mathrm{PA}_0 = E^0(\mathrm{neutral}) - E^0(\mathrm{cation}) \quad (7)$$

For G3//BMK theory, our calculations suggest that A = 6.652 mhartrees, B= 3.039 mhartrees, C = 6.589 mhartrees, and D = 1.281 mhartrees. Thus the final G3//BMK energy can now be calculated by the following equation

$$E_o(\mathrm{G3//BMK}) = E[\mathrm{MP4(FC)/6-31G(d)}] + \Delta(+) + \Delta(\mathrm{2df,p}) + \Delta(\mathrm{QCI}) + \Delta + \Delta\mathrm{SO} + \Delta\mathrm{HLC} + \mathrm{ZPE} \quad (8)$$

where

$$\Delta(+) = E[\mathrm{MP4(FC)/6-31+G(d)MP4(FC)/6-31G(d)}]\Delta(\mathrm{2df,p}) = E[\mathrm{MP4(FC)/6-31G(2df,p)MP4(FC)/6-31G(d)}]\Delta(\mathrm{QCI}) = E[\mathrm{QCISD(T,FC)/6-31G(d)MP4(FC)/6-31G(d)}]\Delta = E[\mathrm{MP2(FU)/G3large\ MP2(FC)/6-31G(2df,p)MP2(FC)/6-31+G(d)+MP2/6-31G(d)}] \quad (9)$$

4. BDEs Predicted by G3//BMK

To test the performance of the G3//BMK method, we use it to recalculate the 141 BDE values of the diverse types of chemical bonds as listed in Table 2. It is found that the G3//BMK predictions are in excellent agreement with all the experimental data (Figure 2). The correlation coefficient is as high as 0.9989, and the mean deviation is as low as −0.004 kcal/mol. Remarkably the RMSE error of the G3//BMK predictions is only 1.2 kcal/mol, which is comparable to the error bar of most of the experimental values (*ca.* 1−3 kcal/mol). Furthermore, there is only one compound (namely, F-NO) for which the G3//BMK prediction (61.2 kcal/mol) disagrees with the corresponding experimental value (56.2 kcal/mol) by over 3 kcal/mol.

In comparison with the G3//BMK method, the standard G3 method exhibits a worse performance in predicting the BDEs. As shown in Table 2, the mean deviation and the RMSE error of the G3 method are −0.2 and 1.5 kcal/mol, respectively. More importantly, there are six compounds (namely, CN-CN, HO-SOH, CH_3−NO, HO-$OCH_2C(CH_3)_3$, sC_4H_9O-NO, and F-NO) for which the G3 predictions (139.8, 71.7, 38.0, 42.8, 46.3, and 61.7 kcal/mol) disagree with the

corresponding experimental values (136.7, 74.9, 41.1, 46.3, 41.5, and 56.2 kcal/mol) by over 3 kcal/mol. The relatively poor performance of the G3 method was previously attributed to the deficiency of the MP2 theory in the geometry optimization for organic radicals.[31]

To improve the performance of the G3 method Baboul et al. once proposed the G3//B3LYP theory where the geometries and zero-point energies are obtained from the B3LYP method.[31] Here our calculations show that the G3//B3LYP predictions of the BDEs have a mean deviation of −0.3 kcal/mol and a RMSE error of 1.4 kcal/mol. These quantities indicate that G3//BMK is also superior to G3//B3LYP in handling BDEs. In agreement with this argument, four compounds (HO-$OCH_2C(CH_3)_3$, HO-SOH, F-NO, and CH_3S-SCH_3) are found in Table 2 for which the G3//B3LYP predictions (42.5, 71.6, 61.7, and 62.2 kcal/mol) differ from the corresponding experimental values (46.3, 74.9, 56.2, and 65.2 kcal/mol) by over 3 kcal/mol.

It should be noted, the difference between G3//BMK and G3//B3LYP is the use of BMK versus B3LYP for the geometries (and harmonic frequencies). We selected HO-$OCH_2C(CH_3)_3$ and CH_3S-SCH_3 from Table 2 for example, which of whose BDEs by G3//BMK disagrees with G3//B3LYP by over 2 kcal/mol to make a comment on the geometries from B3LYP versus BMK.We compared geometries of these two compounds and the corresponding radicals using BMK/6−31G(d) and B3LYP/6−31G(d), respectively, and the bond lengths, the bond angles, and the dihedral angles were shown in the Supporting Information.

From these data, we can make a conclusion that there are large differences between BMK and B3LYP geometries for molecules and radicals as well. Note that, in $OCH_2C(CH_3)_3$ radical, the O−C−C angle is about 8° difference between BMK and B3LYP.

It also should be noted that the modification for different purposes based on Gaussian-*n* theory and its development[32] with the use of new test set[33] is still promising.

5. Conclusions

The ability to reliably and efficiently calculate the bond dissociation enthalpies of various organic molecules is valuable to a number of subdisciplines of chemistry. In the present study we report our findings on the basis of systematic examinations that the BMK functional significantly outperforms the other popular density functional theory methods including B3LYP, B3P86, KMLYP, MPW1P86, O3LYP, and X3LYP for the BDE calculation.

However, it is also found that even the BMK functional may dramatically fail in predicting the BDEs of some chemical bonds. To solve this problem, we develop a new composite *ab initio* method named G3//BMK by combining the strengths of both the G3 theory and BMK. G3//BMK is found to outperform the G3 and G3//B3LYP methods. It can accurately predict the BDEs of diverse types of chemical bonds in various organic molecules within a precision of *ca.* 1.2 kcal/mol.

Acknowledgment. This study was supported by the National Natural Science Foundation of China (Grant No.20602034), Anhui Provincial Natural Science Foundation (Grant No.070416237), and the Excellent Young Scholars Foundation of Anhui Province (Grant No. 08040106829). We also thank the USTC-HP, Wuxi and Shanghai Supercomputer Center for the computational resources.

Supporting Information Available: 169 vibrational frequencies used for the optimization of the scale factor, 299 energies used for the optimization of *A*, *B*, *C*, and *D*, and the geometry comparison of two compounds and corresponding radicals of BMK/6−31G(d) and B3LYP/6−31G(d). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) McMillen, D. F.; Golden, D. M. *Annu. Rev. Phys. Chem.* **1982**, *33*, 493.

(2) Laarhoven, L. J. J.; Mulder, P.; Wayner, D. D. M. *Acc. Chem. Res.* **1999**, *32*, 342.

(3) Blanksby, S. J.; Ellison, G. B. *Acc. Chem. Res.* **2003**, *36*, 255.

(4) Henry, D. J.; Parkinson, C. J.; Mayer, P. M.; Radom, L. *J. Phys. Chem. A* **2001**, *105*, 6750.

(5) Song, K.-S.; Cheng, Y.-H.; Fu, Y.; Liu, L.; Li, X.-S.; Guo, Q.-X. *J. Phys. Chem. A* **2002**, *106*, 6651.

(6) Johnson, E. R.; Clarkin, O. J.; DiLabio, G. A. *J. Phys. Chem. A* **2003**, *107*, 9953.

(7) Yao, X.-Q.; Hou, X.-J.; Jiao, H.; Xiang, H.-W.; Li, Y.-W. *J. Phys. Chem. A* **2003**, *107*, 9991.

(8) Jursic, B. S. *J. Chem. Soc., Perkin Trans.* **1999**, *2*, 369.

(9) Feng, Y.; Liu, L.; Wang, J.-T.; Huang, H.; Guo, Q.-X. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2005.

(10) Speybroeck, V. Van.; Marin, G. B.; Waroquier, M. *Chem. Phys. Chem.* **2006**, *7*, 2205.

(11) Check, C. E.; Gilbert, T. M. *J. Org. Chem.* **2005**, *70*, 9828.

(12) Wodrich, M. D.; Corminboeuf, C.; Schleyer, P. v. R. *Org. Lett.* **2006**, *8*, 3631.

(13) Wodrich, M. D.; Corminboeuf, C.; Schreiner, P. R.; Fokin, A. A.; Schleyer, P. v. R. *Org. Lett.* **2007**, *9*, 1851.

(14) *Gaussian 03, revision D.01*; 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, Inc.: Wallingford, CT, 2004.

(15) Luo, Y.-R. *Handbook of bond dissociation energies in organic compounds*; CRC Press: Boca Raton, FL, 2003.

(16) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(17) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.

(18) Kang, J. K.; Musgrave, C. B. *J. Chem. Phys.* **2001**, *115*, 11040.

(19) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673.

(20) Hoe, W. M.; Cohen, A. J.; Handy, N. C. *Chem. Phys. Lett.* **2001**, *341*, 319.

(21) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533.

(22) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.

(23) Schreiner, P. R. *Angew. Chem., Int. Ed.* **2007**, *46*, 4217.

(24) Kahn, K.; Bruice, T. C. *Theor. Chem. Acc.* **2004**, *111*, 18.

(25) Costa Cabral, B. J.; Canuto, S. *Chem. Phys. Lett.* **2005**, *406*, 300.

(26) Agapito, F.; Costa Cabral, B. J.; Martinho Simões, J. A. *J. Mol. Struct. (Theochem)* **2005**, *729*, 223.

(27) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502.

(28) Halls, M. D.; Schlegel, H. B.; Velkovski, J. *Theor. Chem. Acc.* **2001**, *105*, 413.

(29) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42.

(30) Curtiss, L. A.; Raghhavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(31) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650.

(32) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, 084108.

(33) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123*, 124107.

CT800070Y

JCTC Journal of Chemical Theory and Computation

An Accurate and Transferable Intermolecular Diatomic Hydrogen Potential for Condensed Phase Simulation

Jonathan L. Belof, Abraham C. Stern, and Brian Space*

Department of Chemistry, University of South Florida, 4202 East Fowler Avenue, CHE205, Tampa, Florida 33620-5250

Received May 7, 2008

Abstract: An anisotropic many-body H_2 potential energy function has been developed for use in heterogeneous systems. The intermolecular potential has been derived from first principles and expressed in a form that is readily transferred to exogenous systems, e.g. in modeling H_2 sorption in solid-state materials. Explicit many-body polarization effects, known to be important in simulating hydrogen at high density, are incorporated. The analytic form of the potential energy function is suitable for methods of statistical physics, such as Monte Carlo or Molecular Dynamics simulation. The model has been validated on dense supercritical hydrogen and demonstrated to reproduce the experimental data with high accuracy.

1. Introduction

The development of intermolecular H_2 potential energy functions has a long history.[1] The topic has generated renewed interest in recent years due to the increased theoretical study of hydrogen storage materials.[2] In this work, we present our own revisitation of the issue to include transferability, necessary for the practical simulation of hydrogen in heterogeneous materials as well as anisotropic many-body effects which have been shown to be important in modeling hydrogen at high density. For example, in modeling hydrogen sorption in Metal-Organic Framework materials (MOFs), a transferable potential is required that includes the ability to capture anisotropic many-body effects in a consistent, effective manner.[3,4]

To date, most hydrogen sorption studies of materials have focused on isotropic H_2 potentials[5–8] in particular that of Buch since it is easily transferable and accurately reproduces the bulk thermodynamic properties of hydrogen; in addition, it has been shown to calculate the correct uptake of hydrogen in weakly interacting materials. However, in heterogeneous environments where electrostatic quadrupole and induced dipole effects are non-negligible such an approach cannot reproduce the correct behavior: the isotropic potential includes these effects in a mean-field fashion, appropriate only for systems in which dispersion interactions dominate.

Previous anisotropic potentials[9–15] have focused on neat H_2, with the most notable being the highly accurate potential of Diep and Johnson; however, it is not clear how such specialized forms can be systematically mixed with chemically different environments without extensive reparameterization.[16] While the anisotropic potential of Darkrim and Levesque is both transferable and includes the quadrupole term, it has been shown to overestimate the attractive part of the potential and neglects induced many-body effects.[3]

Herein we apply high-level electronic structure calculations to the H_2-H_2 dimer and then fit an anisotropic functional form to the *ab initio* surface; electron correlation methods have been utilized as they have been shown to be crucial in accurately describing the relatively weak H_2-H_2 interaction.[9,10] This functional form contains quadrupole as well as electronic repulsion/dispersion terms that are transferable using the Lorentz−Berthelot mixing rules.

Several existing potentials[8,13,17] include nonadditive three-body dipole terms *via* Axilrod-Teller-Muto[18] dispersion, which are known to influence the structure of hydrogen at high density. In this work we apply a many-body polarization term to hydrogen and demonstrate improvement in the equation of state at high density. Further, explicit many-body polarization interactions have been demonstrated to be essential in modeling H_2 sorption in polar MOFs,[19] and the present polarization model is also transferable to such systems.

* Corresponding author e-mail: space@cas.usf.edu.

10.1021/ct800155q CCC: $40.75
Published on Web 07/17/2008

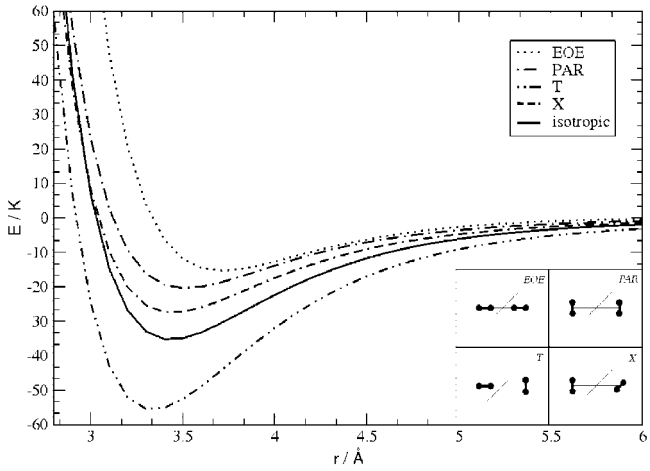


Figure 1. *Ab initio* energy curves of four distinct relative H_2 orientations (as a function of the rotor center-of-mass separation) along the Born−Oppenheimer surface. The orientations chosen were end-on-end (EOE), parallel (PAR), T-configuration (T), and X-configuration (X).

This paper is organized as follows. Section 2.1 presents the *ab initio* methods used in calculating the energy surface for the H_2-H_2 dimer. Sections 2.2 and 2.3 review the many-body polarization theory and form of the potential, respectively. Section 3 elaborates on the application of this potential toward reproducing bulk hydrogen properties. The paper is concluded in Section 4 that includes proposed applications of the transferable and highly accurate potential in predicting and retrodicting hydrogen sorption properties.

2. Methods

2.1. Born−Oppenheimer Surface. Construction of the Born−Oppenheimer surface for the H_2 dimer proceeds from optimizing the RHF[20–24] wave function parametrically as a function of the nuclear coordinates. The hydrogen molecule was approximated as a rigid rotor with a bond length of 0.742 Å, corresponding to the value determined by rotational spectra.[25,26] The surface was taken over the domain of center-of-mass separation for the rotors from 2.0 to 10.0 Å in 0.1 Å increments. A subspace of the complete surface was taken along four unique relative orientations; the orientations were chosen as being energetically and geometrically distinct (shown pictorially in Figure 1) and the four computed surfaces produce an average in agreement with experiment.[8] All electronic structure calculations were performed using the quantum chemistry code PC-GAMESS.[27,28]

Since the H_2-H_2 dimer interaction is dominated by electron correlation,[9,10] the RHF energy calculations included Møller−Plesset perturbation terms[29,30] to fourth-order,[31] including triplet states;[32] prior work in this area has sufficiently demonstrated that, for the hydrogen dimer, MP4 energies are in agreement with the CCSD(T) level of theory.[9] The set of basis functions employed in the solution of the wave function were those of Dunning[33] (aug-cc-pVTZ/QZ) with the energy eigenvalues then extrapolated to the complete basis set limit.[34] The effect of basis set superposition error was corrected by the counterpoise method.[35] The computed energy surfaces are depicted in Figure 1 along with the isotropically averaged surface.

Table 1. H_2 Molecular Polarizability Tensor (Molecular Bond Axis Aligned along the Z Axis) Calculated *via* Time-Dependent Hartree-Fock along with its Rescaled Components[a]

component	TDHF/Å^3	rescaled/Å^3
XX	0.6831	0.6945
YY	0.6831	0.6945
ZZ	0.9561	0.9720
isotropic	0.7741	0.7870

[a] The components were rescaled such that the isotropic polarizability, $^1/_3Tr\{\alpha\}$, matched experiment subject to the constraint that the ratio of TDHF values, *XX*/*ZZ*, be preserved.

The electric quadrupole value for H_2 was calculated from the electronic wave function (MP4-SDTQ/aug-cc-pVQZ) and found to be 0.664 D·Å, a value comparable to other investigations.[25,36] The molecular (static) polarizability tensor of the monomer was determined *via* TDHF[37] since it has been shown to accurately reproduce nonlinear optical properties at frequencies away from resonance; the resultant components are listed in Table 1. The components of the TDHF polarizability tensor were then rescaled such that the isotropic polarizability matched the experimental value[38] of 0.787 Å^3 subject to the constraint that the ratio of the *XX* to *ZZ* TDHF components remains invariant. Prior work has shown that, for H_2, the nonadiabatic contributions to both the quadrupole and polarizability are negligible.[36,39]

2.2. Many-Body Polarization. Here we briefly review the Thole-Applequist polarization model−details can be found elsewhere.[40–42] Briefly, the atomic point dipole, $\vec{\mu}$, of the i^{th} atom is

$$\vec{\mu}_i = \alpha_i \vec{E}_i^{stat} \tag{1}$$

where α_i is an atomic site polarizability tensor, and $\vec{E}_i^{stat}$ is the electrostatic field vector. We can decompose the dipole into separate *static* and *induced* contributions as follows

$$\vec{\mu}_i = \alpha_i^{\circ}\left(\vec{E}_i^{stat} + \vec{E}_i^{ind}\right) = \alpha_i^{\circ}\left(\vec{E}_i^{stat} - T_{ij}\vec{\mu}_i\right) \tag{2}$$

where α_i° is a *scalar* point polarizability, and T_{ij} is the dipole field tensor which, when contracted with the dipole $\vec{\mu}_j$, represents the electrostatic contribution of the j^{th} dipole toward inducing the i^{th} dipole.

The dipole field tensor is constructed based upon the positions and scalar point polarizabilities of the system and can be derived from classical electrostatic principles as

$$T_{ij}^{\alpha\beta} = \frac{\delta_{\alpha\beta}}{r_{ij}^3} - \frac{3x^{\alpha}x^{\beta}}{r_{ij}^5}$$

with an additional aspect of the model being that the second dipole may be taken to be a model distribution, with the intent that the discontinuity at short-range will be avoided; here we employ the common exponential charge distribution and associated damping parameter of 2.1304.

Equation 2 is a self-consistent field equation with respect to the dipoles and hence must be solved iteratively. However, it is possible to recast the field equations in matrix form as

$$A\vec{\mu} = \vec{E}^{\text{stat}} \tag{3}$$

where the matrix A is constructed from block matrices according to

$$A = \left[(\alpha^{\circ})^{-1} + T_{ij}\right] \tag{4}$$

The advantage to eq 4 is that the solution is then exact (to numerical precision), without the additional complication of iteration and convergence. Upon inversion of the A matrix, the molecular polarizability tensor can then be easily determined by

$$\alpha_{\alpha\beta}^{mol} = \sum_{i,j} \left(A_{ij}^{-1}\right)_{\alpha\beta} \tag{5}$$

The polarizable aspect of the hydrogen model was developed as follows. With the molecule aligned along the Z axis, the polarizability tensor was calculated with the Thole model (i.e., eq 5) and compared with the *ab initio* TDHF tensor; the scalar point polarizabilities being varied until the two tensors agree. The scalar polarizabilities, α°, were assigned at the same nuclear coordinates as the partial charges, Q, used in producing the quadrupole. The matrix in eq 4 was constructed, inverted, and then summed according to eq 5 yielding a trial polarizability tensor. The resulting trial tensor was then compared to the rescaled tensor in Table 1, and the α°'s were adjusted until the difference was minimized. The site polarizabilities were completely converged (i.e., the molecular polarizability tensor determined *via* eq 5 matched the rescaled tensor in Table 1 to within all significant digits) and are listed in Table 2.

2.3. Potential Energy Function. The interest of this work is the development of a potential energy function for use in Monte Carlo or Molecular Dynamics simulation of hydrogen interacting with heterogeneous systems. In this context, the following functional form was chosen for the potential energy function

$$U = U_{rd} + U_{es} + U_{pol} \tag{6}$$

where U_{rd} is the energy of electronic repulsion/dispersion, U_{es} is the electrostatic energy, and U_{pol} is the many-body polarization energy given by

$$U_{\text{pol}} = -\frac{1}{2}\sum_{i}^{N} \vec{\mu}_i \cdot \vec{E}_i^{\text{stat}} \tag{7}$$

where the site dipoles are found by either matrix inversion of eq 3 or iterative solution of the dipole field equations given by expression 2. For the purpose of fitting the potential energy function, we employed the matrix inversion method for maximal accuracy in the many-body potential since the computations are fast for a two-rotor system (however, in application toward condensed-phase systems the iterative method is used almost exclusively).

All of the free parameters of these functions have well-established ways of being "mixed", thereby imparting transferability. It is the main contribution of this work that these functions have been fit to the first principles data presented in Section 2.1.

Table 2. Parameter Fits for the Intermolecular Hydrogen Potential, including Many-Body Polarization Terms[a]

site	R/Å	Q/e	α°/Å^3	ε/K	σ/Å
H2GP	0.000	−0.7464	0.69380	12.76532	3.15528
H2NP	0.363	0.0000	0.00000	2.16726	2.37031
H2EP	0.371	0.3732	0.00044	0.00000	0.00000

[a] H2GP corresponds to the center-of-mass site, H2EP coincides with the true atomic locations (which, when combined with H2GP, provides the quadrupole), while H2NP contains the additional Lennard-Jones sites.

The electrostatic energy follows from the quadrupole-quadrupole interactions between the hydrogen monomers. Given the quadrupole value of 0.664 D·Å calculated from first principles and a bond length of 0.742 Å, this corresponds to partial charges of $Q = +0.3732e$ at the atomic sites and $-2Q$ at the center-of-mass.

The short-range electronic repulsion and long-range dispersion energies are included in U_{rd} by use of the Lennard-Jones 12−6 potential function. Since the Lennard-Jones r^{-6} part includes a mean-field long-range polarization, it was necessary that U_{rd} be determined. With U_{es} and U_{pol} functionally held fixed, U_{rd} was varied *via* simulated annealing such that $(U\text{-}U_{BO})^2$ was minimized (where U_{BO} is the Born−Oppenheimer energy surface found in Section 2.1). The results of the potential energy function fitting are shown in Figure 2. Given the anisotropic nature of the Born−Oppenheimer potential demonstrated in Figure 1, Figure 2 demonstrates that our relatively simple and transferable potential can capture the essential anisotropy. Interestingly, while the T orientation shows the largest deviation from the *ab initio* surface, even dramatically increasing the number of Lennard-Jones parameters did not significantly improve the fit. Note, given the large variation in the potential it is critical in modeling hydrogen in highly anisotropic environments that the potential capture these structural characteristics.

In the course of the parameter-space search, the Lennard-Jones sites on the hydrogen molecule were allowed to move off their nuclear centers. In addition, searches were made that constrained the sites to be both on and off the bond axis, with as few as two and as many as nine simultaneous sites in an attempt to improve the fit; it was found, however, that three sites constrained to the bond axis were optimal in terms of having the minimized error (with only marginal improvement in the fit upon increasing the number of sites further). At the end of the fitting process, the parameters that minimized the error were found and are presented in Table 2. In order to also develop a nonpolarizable potential, the process was repeated with $U_{pol} = 0$. In physical systems where the magnitude of electrostatic polarization is negligible, the nonpolarizable parameters presented in Table 3 are desirable for the reduced computational cost.

3. Model Validation

3.1. Second Virial Coefficient. Quantum mechanical corrections to the second virial coefficient of a gas can be systematically derived using the Wigner-Kirkwood distribution function.[43] To order h^4, we have the semiclassical series

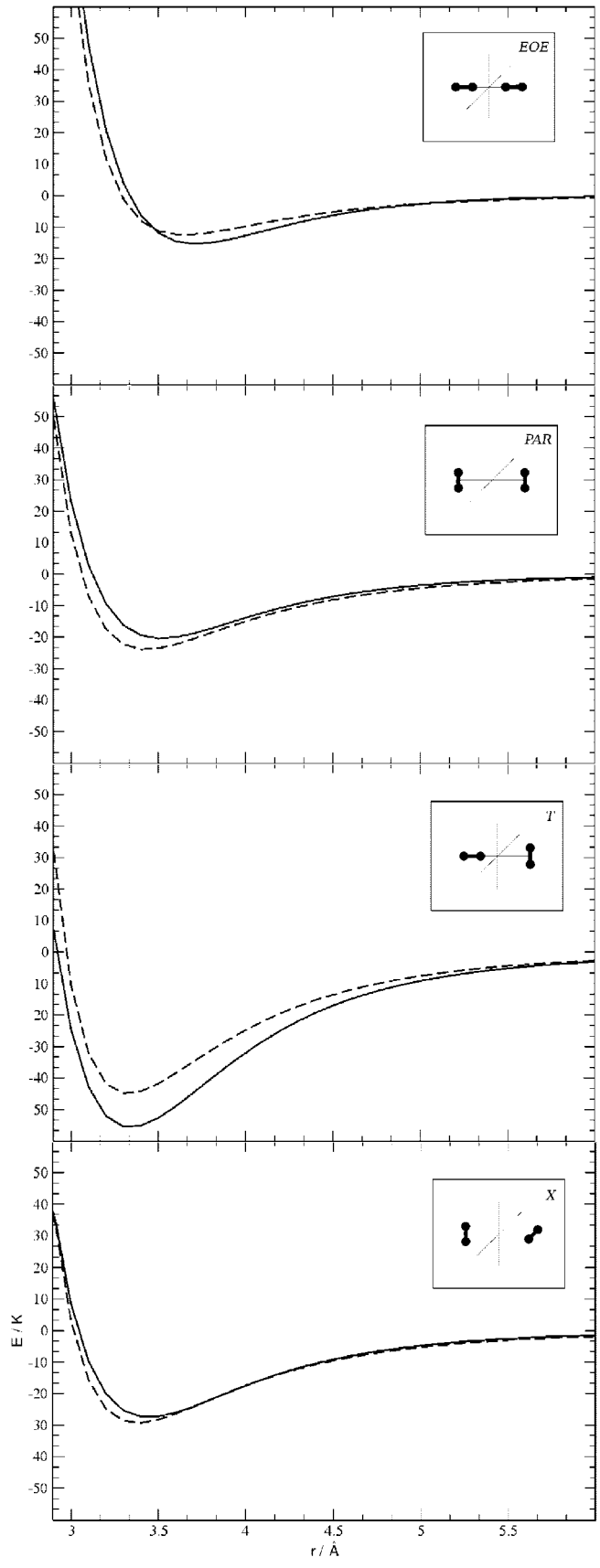


Figure 2. Potential energy curves found by fitting to eq 6 (dashed) versus the *ab initio* curves from Figure 1 (solid). The alternative nonpolar form ($U_{pol} = 0$) that was also fit to the *ab initio* data is not shown for clarity since, with a slight exception to the T configuration, the curves are visually indistinguishable.

Table 3. Potential Parameters for the Pairwise Model (Neglecting Polarization Terms)

site	R/Å	Q/e	ε/K	σ/Å
H2G	0.000	−0.7464	8.8516	3.2293
H2N	0.329	0.0000	4.0659	2.3406
H2E	0.371	0.3732	0.0000	0.0000

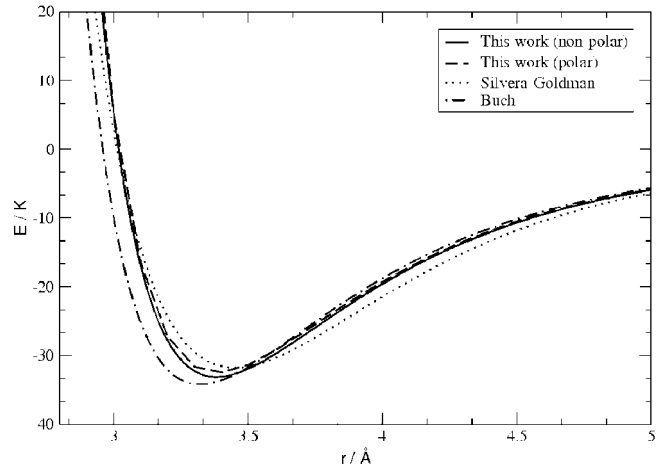


Figure 3. The isotropic projection $\phi(r)$ of the many-body potential U is shown along with the SG and Buch potentials.

$$B_2(T) = 2\pi\int_0^\infty dr\, r^2(1-e^{-\beta\varphi}) + \frac{h^2}{24\pi\beta^3}\int_0^\infty dr\, r^2 e^{-\beta\varphi}\left[\frac{\partial\varphi}{\partial r}\right]^2 + \frac{h^4}{960\pi^3\beta^4}\int_0^\infty dr\, r^2 e^{-\beta\varphi}\times\left\{\left[\frac{\partial^2\varphi}{\partial r^2}\right]^2 + \frac{2}{r^2}\left[\frac{\partial\varphi}{\partial r}\right]^2 + \frac{10\beta}{9r}\left[\frac{\partial\varphi}{\partial r}\right]^3 - \frac{5\beta^2}{36}\left[\frac{\partial\varphi}{\partial r}\right]^4\right\} \quad (8)$$

where $\phi(r)$ is an isotropic potential as a function of the separation. The intermolecular many-body potential developed here has been spherically averaged to produce an isotropic pair potential, $\phi(r)$, which is plotted in Figure 3 along with the well-known isotropic potentials of Silvera-Goldman[8] and Buch,[7] both of which are known to closely match experiment.

The second virial coefficient, $B_2(T)$, was then calculated *via* eq 8. Numerical integration was performed for r = 0.001−25 Å across a temperature range spanning from 50 to 500 K, in 10 K increments. The results are compared with experiment[38] (and the SG potential) in Figure 4.

3.2. Equation of State. The equation of state was ascertained by calculating the average number of hydrogen molecules, $\bar{N}$, via sampling of the grand canonical ensemble for a corresponding range of chemical potential. The following statistical mechanical expression was numerically estimated by Grand Canonical Monte Carlo[44,45]

$$\bar{N} = \frac{1}{\Xi}\sum_{N=0}^{\infty} e^{\beta\mu N}\left\{\prod_{i=1}^{3N}\int_{-\infty}^{\infty} dx_i\right\} N e^{-\beta U(x_1,\ldots,x_{3N})}$$

where the chemical potential of the gas reservoir, μ, was determined through empirical fugacity functions[46,47] for

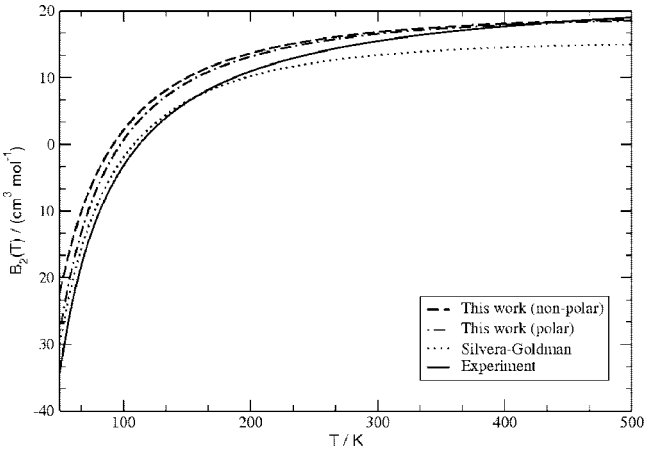


Figure 4. $B_2(T)$ from 50−500 K for the polarizable and nonpolarizable potentials including Wigner-Kirkwood quantum corrections to order $\hbar^4$.

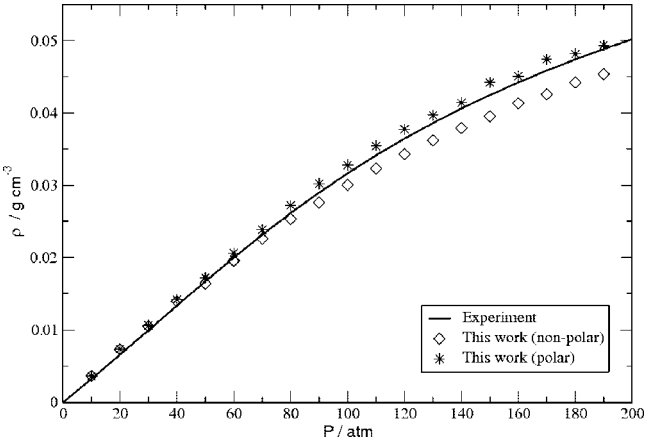


Figure 5. P-ρ plot of hydrogen at 77 K using the first principles-derived potential energy function. The densities were calculated using GCMC over the corresponding pressure range, including empirical fugacity corrections. The electronic repulsion/dispersion energy term included Feynman-Hibbs quantum effects to order $\hbar^4$; the data generated without this correction were systematically higher by about 10%. All data points have a maximum variance of ±0.001 g cm $^{-3}$.

hydrogen at both high and low temperature (the high-pressure densities have also been verified *via* NPT molecular dynamics). Isothermal P-ρ curves were generated at temperatures of 77 K (pressure range of 0−200 atm) and 298.15 K (pressure range of 0−2000 atm), and the results are compared with experimental data[48,49] in Figures 5 and 6. For the simulations at 77 K, Feynman-Hibbs quantum corrections[50] were applied to the energetically dominant electronic repulsion/dispersion part of the potential to order $\hbar^4$ *via*

$$U_{rd}^{FH} = U_{rd} + \frac{\beta\hbar^2}{24\mu}\left(U_{rd}'' + \frac{2}{r}U_{rd}'\right) + \frac{\beta^2\hbar^4}{1152\mu^2}\left(\frac{15}{r^3}U_{rd}' + \frac{4}{r}U_{rd}''' + U_{rd}''''\right) \quad (9)$$

where U_{rd}', U_{rd}'',... are the derivatives of the electronic repulsion/dispersion term of eq 6 with respect to the displacement r. These corrections were not applied in the generation of the 298.15 K data since these effects are negligible at room temperature. To reduce the computational cost, these quantum corrections were only applied to the repulsion/dispersion energy because at 77 K, under the pressure conditions demonstrated here, the repulsion/dispersion energy is greater than 85% of the total energy.

It should be pointed out that the molecular hydrogen is in a supercritical phase under the conditions reported. Note the liquid hydrogen density at boiling (20 K) is 0.07 g cm $^{-3}$; thus the state points considered include those representative of relatively strong intermolecular interactions for H_2. In both instances, we see improved agreement with experiment at high-density when many-body effects are explicitly considered by inclusion of U_{pol}. It is known that many-body polarization effects are of even greater importance when considering high-density hydrogen interacting strongly with, for example, a polar adsorbing material.[4]

4. Conclusions

An anisotropic, many-body hydrogen potential has been derived from first principles and expressed in a functional form suitable for mixing with heterogeneous systems consisting of partial charges, Lennard-Jones sites, and atomic point polarizabilities. It has been shown to reproduce the properties of bulk hydrogen under conditions of current interest in materials research. Electrostatic quadrupolar and many-body polarization interactions are included anisotropically, thus making this model useful for high-density studies where orientation-dependence is of interest.

Next, it is planned to use the potential in modeling H_2 sorption in highly polar MOFs that have been shown to sorb hydrogen at near liquid densities at 77 K.[4,19] The potential should also be useful in modeling hydrogen in any condensed phase system where the essential physics of the composite system is captured by the flexible potential energy function presented.

Acknowledgment. Computations were performed on the Ranger supercomputer located at the Texas Advanced Computing Center under a Teragrid Grant (Grant No. TG-CHE070098N) and also at the USF Research Computing Center. The authors acknowledge funding from the Depart-

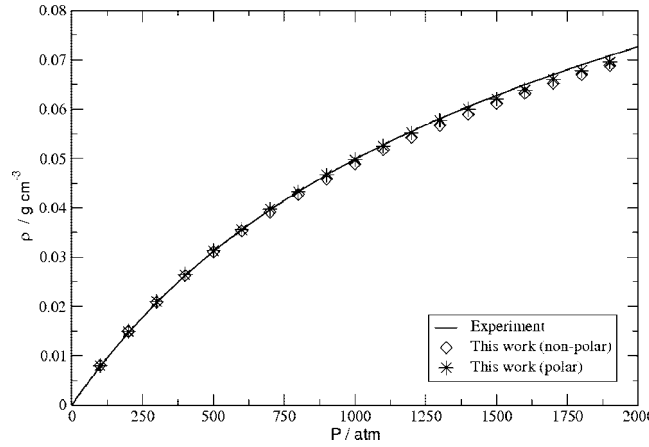


Figure 6. P-ρ plot of hydrogen at 298.15 K using the first principles-derived potential energy function. All data points have a maximum variance of ±0.0008 g cm $^{-3}$.

ment of Energy (Grant No. DE0GG02-07ER46470) and the National Science Foundation (Grant No. CHE-0312834). The authors also thank the Space Foundation (Basic and Applied Research) for partial support.

References

(1) Van Kranendonk, J. *Solid Hydrogen: Theory of the Properties of Solid H_2, HD, and D_2*, 2nd ed.; Plenum Press: New York, 1983; p 45.

(2) Long, J. R.; Alivisatos, A. P.; Cohen, M. L.; Frechet, J. M. J.; Head-Gordon, M.; Louie, S. G.; Zettl, A.; Richardson, T. J.; Mao, S. S.; Read, C.; Alkire, J. A Synergistic Approach to the Development of New Classes of Hydrogen Storage Materials, 2005;Department of Energy Hydrogen Program. http://www.hydrogen.energy.gov (accessed Mar 3, 2008).

(3) Garberoglio, G.; Skoulidas, A. I.; Johnson, J. K. *J. Phys. Chem. B* **2005**, *109*, 13094.

(4) Belof, J.; Stern, A.; Eddaoudi, M.; Space, B. *J. Am. Chem. Soc.* **2007**, *129*, 15202.

(5) Schaefer, J.; Watts, R. *Mol. Phys.* **1982**, *47*, 933.

(6) Knaap, H.; Beenakker, J. *Physica* **1961**, *27*, 523.

(7) Buch, V. *J. Chem. Phys.* **1994**, *100*, 7610.

(8) Silvera, I. F.; Goldman, V. V. *J. Chem. Phys.* **1978**, *69*, 4209.

(9) Diep, P.; Johnson, J. K. *J. Chem. Phys.* **2000**, *112*, 4465.

(10) Diep, P.; Johnson, J. K. *J. Chem. Phys.* **2000**, *113*, 3480.

(11) Gallup, G. *Mol. Phys.* **1977**, *33*, 943.

(12) Etters, R.; Danilowicz, R.; England, W. *Phys. Rev. A* **1975**, *12*, 2199.

(13) Wind, P.; Røeggen, I. *Chem. Phys.* **1996**, *211*, 179.

(14) Mulder, F.; van der Avoird, A.; Wormer, P. *Mol. Phys.* **1979**, *37*, 157.

(15) Darkrim, F.; Levesque, D. *J. Chem. Phys.* **1998**, *109*, 4981.

(16) Zhang, Z.; Duan, Z. *J. Chem. Phys.* **2005**, *122*, 214507.

(17) Kostov, M. K.; Cole, M. W.; Lewis, J. C.; Diep, P.; Johnson, J. K. *Chem. Phys. Lett.* **2000**, *332*, 26.

(18) Axilrod, B.; Teller, E. *J. Chem. Phys.* **1943**, *11*, 299.

(19) Liu, Y.; Eubank, J.; Cairns, A.; Eckert, J.; Kravtsov, V.; Luebke, R.; Eddaoudi, M. *Angew. Chem., Int. Ed.* **2007**, *46*, 1433.

(20) Hartree, D. *Proc. Camb. Phil. Soc.* **1928**, *24*, 89.

(21) Fock, V. *Z. Phys. A* **1930**, *62*, 795.

(22) Slater, J. *Phys. Rev.* **1932**, *42*, 33.

(23) Slater, J. *Phys. Rev.* **1950**, *81*, 385.

(24) Roothaan, C. *Rev. Mod. Phys.* **1951**, *23*, 69.

(25) Johnson, R. D. NIST Computational Chemistry Comparison and Benchmark Database. http://srdata.nist.gov/cccbdb (accessed Mar 3, 2008).

(26) Herzberg, G. *Molecular Spectra and Molecular Structure*, 2nd ed.; Krieger Publishing Co.: Malabar, FL, 1989; Vol. I, p 532.

(27) Nemukhin, A. V.; Grigorenko, B. L.; Granovsky, A. A. *Moscow University Chem. Bull.* **2004**, *45*, 75.

(28) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

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

(30) Binkley, J.; Pople, J. *Int. J. Quantum Chem.* **1975**, *9*, 229.

(31) Krishnan, R.; Pople, J. *Int. J. Quantum Chem.* **1978**, *14*, 91.

(32) Krishnan, R.; Pople, J. *J. Chem. Phys.* **1980**, *72*, 4244.

(33) T.H. Dunning, J. *J. Chem. Phys.* **1989**, *90*, 1007.

(34) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639.

(35) Boys, S.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(36) Poll, J.; Wolniewicz, L. *J. Chem. Phys.* **1978**, *68*, 3053.

(37) Karna, S.; Dupuis, M. *J. Comput. Chem.* **1991**, *12*, 487.

(38) Lide, D. R. *Handbook of Chemistry and Physics*, 84th ed.; CRC Press: Boca Raton, FL, 2003.

(39) Kolos, W.; Wolniewicz, L. *J. Chem. Phys.* **1967**, *46*, 1426.

(40) Bode, K. A.; Applequist, J. *J. Phys. Chem.* **1996**, *100*, 17820.

(41) Thole, B. *Chem. Phys.* **1981**, *59*, 341.

(42) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, 102, 2399.

(43) Brack, M.; Bhaduri, R. K. *Semiclassical Physics*; Westview Press: Boulder, CO, 2003; p 148.

(44) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *Phys. Lett. B* **1953**, *21*, 1087.

(45) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: New York, 2002; p 129.

(46) Shaw, H.; Wones, D. *Am. J. Sci.* **1964**, *262*, 918.

(47) Zhou, L.; Zhou, Y. *Int. J. Hydrogen Energy* **2001**, *26*, 597.

(48) Lemmon, E.; McLinden, M.; Friend, D. NIST Chemistry WebBook. Linstrom, P., Mallard, W., Eds. http://webbook.nist.gov (accessed Mar 5, 2008).

(49) Younglove, B. *J. Phys. Chem. Ref. Data* **1982**, *11*, 1.

(50) Feynman, R.; Hibbs, A. *Quantum Mechanics and Path Integrals*; McGraw-Hill: New York, 1965; p 281.

CT800155Q

JCTC Journal of Chemical Theory and Computation

What Electronic Structures and Geometries of Carborane Mono- and *ortho*-, *meta*-, and *para*-Diradicals are Preferred?

Luis Serrano-Andrés,*,† Douglas J. Klein,‡ Paul von Ragué Schleyer,§,|| and Josep M. Oliva*,⊥

Instituto de Ciencia Molecular, Universitat de València, Apartado 22085, ES-46071 Valencia, Spain, Texas A&M University at Galveston, Galveston, Texas 77553-1675, Center for Computational Chemistry, University of Georgia, Athens, Georgia 30602, Institut für Organische Chemie der Universität Erlangen-Nürnberg, Henkestrasse 42, D-91054 Erlangen, Germany, and Instituto de Química-Física Rocasolano, Consejo Superior de Investigaciones Científicas, Serrano, 119, ES-28006 Madrid, Spain

Received May 6, 2008

Abstract: Structures, relative stabilities, singlet-triplet gaps, and the ground-state character of mono- and diradicals derived from the three icosahedral carborane cage isomers have been computed by unrestricted broken-symmetry DFT and by CASPT2 methods. Whereas the bond dissociation energies (BDE) leading to the carborane monoradicals are close to the benzene BDE, the most stable carborane radicals are derived from dissociations of hydrogens farthest away from the carbon atoms. All the monomeric carborane diradicals are determined to have singlet ground states. However, the energetic accessibility of triplet states in some of the species offers the potential of building diradical multidimensional carborane network architectures with interesting magnetic properties.

1. Introduction

Carborane clusters, proposed[1] as building-blocks in a "molecular tinkertoy" set, have emerged as candidates for the design of nanosize materials with tunable electronic, optical, and magnetic properties.[2] Icosahedral *o*-, *m*-, and *p*-carboranes are commercially available. Their 12 cage atoms, serving as anchoring points, can be connected to build linear, sheetlike, and 3D molecular architectures.[3] The ion−molecule reactions of icosahedral *o*-, *m*-, and *p*-carborane in an ion trap showed that under ultrahigh vacuum conditions carborane ions reacted sequentially with their neutral counterparts to form series of cluster-of-clusters ions, i.e. 'supercluster' ions with linear, sheetlike, and 3D networks. Such clustering reactions proceeded with elimination of carborane hydrogens *but without detachment of any skeletal boron or carbon atom*. Hence, the thermal stability of the individual carborane cages was unaltered.[4] We have shown previously[5–8] how the properties of carborane cages can be tuned by appropriate modifications. This opens new possibilities for devising promising multidimensional carborane networks, e.g., for the design of cutting-edge molecular-based magnetic nanostructures for spin electronics.[9,10]

We aim to clarify the properties of (finite) 1D, 2D, and 3D constructs, based on icosahedral *o*-, *m*-, and *p*-carboranes in order to devise new materials with potential nanotechnological applications. Even expansion of the carborane cage to more than 12 vertices is nowadays possible leading to the possibility of atom encapsulation.[11] How do charge, spin (S),[5,6] and endohedral-atom inclusion[7,8,12] (M@cage) influence their properties, both in the ground and excited states?

* Corresponding author e-mail: Luis.Serrano@uv.es (L.S.-A.) and iqrjmoe@iqfr.csic.es (J.M.O.).

† Universitat de València.

‡ Texas A&M University at Galveston.

§ University of Georgia.

|| Institut für Organische Chemie der Universität Erlangen-Nürnberg.

⊥ Consejo Superior de Investigaciones Científicas.

10.1021/ct800150h CCC: $40.75
Published on Web 06/26/2008

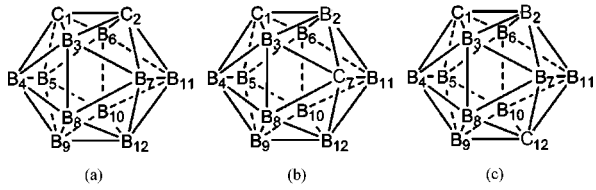


Figure 1. Labels for cage atoms in (a) *o*-carborane, (b) *m*-carborane, and (c) *p*-carborane, having point-group symmetries C_{2v}, C_{2v}, and D_{5d}, respectively. Hydrogen atoms, bound to every cage atom, are not shown for clarity.

Toward this goal, we now report a comprehensive study of the thermodynamic stabilities of carborane mono- and diradicals as well as the ground states and singlet-triplet gaps of the latter. We relate differences in behavior to their geometric and electronic peculiarities.

Previous computational investigations involved the $CB_9H_{10}^{\bullet}$ and $CB_{11}H_{12}^{\bullet}$ free radicals at the B3LYP/6−31G(d) level of theory[13] and "wheel" systems, where *closo*-carborane diradicals on opposite "poles" can either be joined to give cross-cluster bonds or can be open in singlet or triplet states.[14] The finding in the latter study that the 1,12-$C_2H_{10}B_{10}$ diradical prefers an open singlet state is confirmed and extended here. Relevant recently published experimental results include a negatively charged π-$(C_{60}^{-})_2$ dimer with a diradical state at room temperature[15] and the synthesis of dimeric diradicals derived from carborane cages connected through acetylenic and ethylenic moieties.[16]

2. Computational Methods

MP2 and B3LYP were among the methods[17] employed to compute ground states of the various carboranes radicals as well as adiabatic singlet-triplet gaps of diradical carboranes. Due to wave function instability and the need for multireference treatments, singlet state diradical geometries and energies were computed by the unrestricted (U)B3LYP broken symmetry approach (UBS),[18] an approximate but practicable method judged to be suitable for this purpose.[19] The performance of computations on selected diradicals were carefully calibrated against benchmark results employing multiconfigurational perturbation theory, CASSCF/CASPT2, a high-level method which has proven its applicability in diverse electronic structure computations.[6,7,20–28] Unless otherwise indicated, the CASPT2 multireference wave functions are based on an active space of six electrons distributed in six molecular orbitals (6,6). However smaller (2,2) and larger (12,12) active spaces were checked to ensure the stability of the results in single-point energy computations. The multiconfigurational computations used the Molcas 6.0 program[29–31] and involve CASSCF geometry optimizations and CASPT2 single-point energy computations. The 6−31G(d) basis set was employed generally. Test computations using the aug-cc-pVDZ basis sets also are discussed in order to show that increasing the size of the basis set only has minor effects. All states are described at their optimized geometries, which have been characterized as energy minima at the same level of computation through analytical second derivatives. The obtained harmonic frequencies have been used to include zero-point vibrational corrections when indicated.

3. Results and Discussion

Figure 1 displays the IUPAC numbering scheme (hydrogen atoms bound to each cage atom are not shown) for *o*-carborane, *m*-carborane, and *p*-carborane, where *o*- (*ortho*), *m*- (*meta*), and *p*- (*para*) describe the relative positions of the two carbon atoms. Single H atom removal can give five isomeric radicals from *o*-carborane, five from *m*-carborane, and two from *p*-carborane. In our designation, *r*-(*j*) is an *r*-carborane with the radical at position *j*, *j* being the lowest number among equivalent alternatives. The possible isomers are summarized below.

$$\text{Radicals derived from o} - \text{carborane} \rightarrow \{\text{o} - (1), \text{o} - (3), \text{o} - (4), \text{o} - (8), \text{o} - (9)\} \quad (1)$$

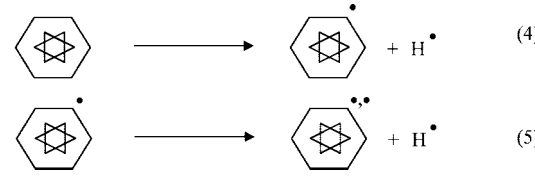
$$\text{Radicals derived from m} - \text{carborane} \rightarrow \{\text{m} - (1), \text{m} - (2), \text{m} - (4), \text{m} - (5), \text{m} - (9)\} \quad (2)$$

$$\text{Radicals derived from p} - \text{carborane} \rightarrow \{\text{p} - (1), \text{p} - (2)\} \quad (3)$$

The nomenclature scheme for diradicals is similar.

It is worth mentioning that the neutral B−H/C-H bond dissociation (homolytic bond-dissociation) is a lower energy process by 10−13 eV as compared to the ionic dissociation or deprotonation of the species derived from *o*-carborane, *m*-carborane, and *p*-carborane (see details in the Supporting Information).

3.1. Energetics of Carborane Radicals and Diradicals. The energetics of the various radicals and diradicals derived from *o*-, *m*-, and *p*-carborane (*r*-$C_2B_{10}H_{12}$, $r = o, m, p$) can be estimated from the following carbon/boron−hydrogen bond dissociation reactions (eqs 4 and 5):



The relative energy of monoradicals (ΔE_1) and diradicals (ΔE_2) can be obtained through the corresponding relationships (eqs 6 and 7, where X represents the carborane fragment):[32]

$$\Delta E_1 = E(HX^{\bullet}) + E(H^{\bullet}) - E(HX - H) \quad (6)$$

$$\Delta E_2 = E(X^{\bullet,\bullet}) + E(H^{\bullet}) - E(HX^{\bullet}) \quad (7)$$

No experimental C−H or B−H bond dissociation energies of carboranes are available, to our knowledge. Hence, and assuming similar bonding effects in both systems, we calibrated the performance of our B3LYP/6−31G(d) level against experimental C−H bond dissociation energies (CH BDE, D_0) of benzene and the phenyl radical,[33–36] obtained from the experimental C−H bond dissociation enthalpy at 298 K (DH_{298}) and the heat capacity contribution.[36] The error in our computational level, revealed by this procedure, was corrected by evaluating eqs 6 and 7 and using an adjusted parameter, K (−308.05 kcal/mol or −0.490916 au), instead of the known exact energy of the hydrogen atom (−0.5 au or −313.75 kcal/mol). The value of K was deduced from eq 8 and experimental data[33,36]

$$\Delta E_{exp} = (D_0)_{exp} = E(C_6H_5^{\bullet}) + K - E(C_6H_6) = 112.4 \text{ kcal/mol} \quad (8)$$

where the computed B3LYP/6−31G(d) energy for benzene $E(C_6H_6)$ and the phenyl radical $E(C_6H_5^{\bullet})$ includes the zero-point energy (ZPE) correction. The experimental value for D_0, 112.4 kcal/mol, (eq 8) can be compared with the computed B3LYP/6−31G(d) and MP2/6−31G(d) values of 118.1 and 132.7 kcal/mol, respectively. Because of the large discrepancy, and also because the erratic behavior of the method for the diradicals, the MP2 data are disregarded in our discussion.

This procedure was used to compute the energies of all the isomeric carborane monoradicals shown comparatively in Figure 2. These derive from H atom dissociation from all distinguishable positions of the 1,2- (ortho), 1,7- (meta), and 1,12- (para) $C_2H_{12}B_{10}$ 12-vertex carborane isomers. Computations of ΔE with the basis set 6−31G(d,p) showed energy differences of only 1 kcal/mol (see the Supporting Information). The same strategy can be employed by using eq 7 to compute the three isomeric diradicals ("benzynes") derived from benzene: *ortho* (*o*-), *meta* (*m*-), and *para* (*p*-)-$C_6H_4^{\bullet,\bullet}$. Taking the corresponding experimental $(D_0)_{exp}$ values for the hydrogen dissociation leading from the phenyl radical ($C_6H_5^{\bullet}$) to the corresponding benzyne (r-$C_6H_4^{\bullet,\bullet}$), we have

$$\Delta E_{exp}^{\ r-\text{benzyne}} = (D_0)_{exp}^{\ r-\text{benzyne}} = E(r - C_6H_4^{\bullet,\bullet}) + K_r - E(r - C_6H_5^{\bullet}) \quad (9)$$

The three experimental D_0 energies 76.5 (*o*-), 92.5 (*m*-), and 107.5 kcal/mol (*p*-)[34,36] were used to deduce three values of K: K_o: −324.5 kcal/mol, K_m: −308.8 kcal/mol, and K_p: −330.2 kcal/mol. These three values, employed to take into account the geometrical arrangement of the analogous diradical centers, were used in the calculations of the respective isomers of the carborane diradicals. The benzyne energies were computed at the UB3LYP/6−31G(d) level of theory for the *ortho* and *para* isomers and at the BLYP/6−31G(d) level for *meta*-benzyne. The exception for the latter was made because of the known trend of hybrid functionals to yield a highly strained bicyclic olefin structure for *meta*-benzyne instead of the correct monocyclic diradical conformation.[37,38] ZPE corrections were excluded in the calculations of the diradicals due to the poor convergence of some of the DFT broken-symmetry solutions (see below).

Figures 2 and 3 display and relate graphically the dissociation energies of the different radicals and diradicals derived from *o*-carborane, *m*-carborane, and *p*-carborane obtained by applying eqs 10 and 11

$$\Delta E_1 = E(HX^{\bullet}) + K - E(HX - H) \quad (10)$$

$$\Delta E_2 = E(X^{\bullet,\bullet}) + K_r - E(HX^{\bullet}) \quad r = o, m, p \quad (11)$$

whereas Table 1 summarizes the relative energies of the singlet carborane diradical ground states in terms of the CH or BH dissociation energies from the carborane monoradicals. Notice that some carborane diradicals can be obtained from two different consecutive H-atom abstractions; therefore, two dissociation energy values are given. For instance, *o*-(1,3)

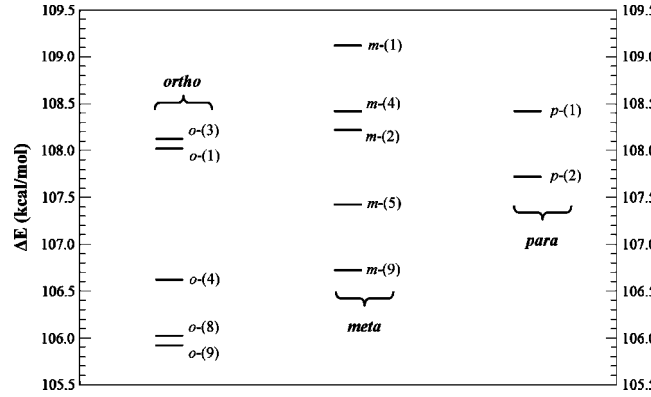


Figure 2. Comparison of the dissociation energies of monoradical derived from *o*-carborane, *m*-carborane, and *p*-carborane. ΔE computed using eq 10 at the (U)B3LYP/6−31G(d) level of theory.

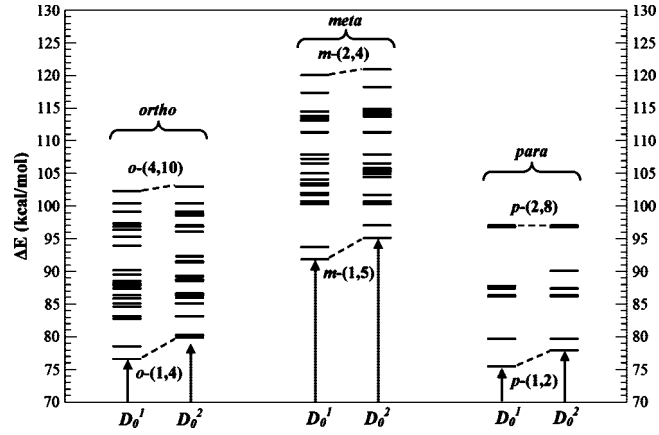


Figure 3. Plot of the dissociation energies of diradicals derived from *o*-carborane, *m*-carborane, and *p*-carborane monoradicals. ΔE computed using eq 11. Level of theory: UBS(B3LYP)/6−31G(d).

can be obtained from the radical *o*-(1) by removing the hydrogen from boron B_3 or from *o*-(3) by removing the H from carbon C_1. The energies of the carborane diradicals were based on the B3LYP broken-symmetry solution for the singlet state of the systems (see the next section), assuming, as verified by the CASPT2 calibrations, that all of them have a singlet ground state. The value for the singlet state energy was obtained by adding to the UB3LYP/6−31G(d) solution for the triplet state of the diradical the computed ΔE_{ST}(BS) singlet-triplet gap (see section 3.2).

The relative stabilities of the monoradicals (see Figure 2) obviously are influenced by the energies of the parent carboranes where the stability follows the order E(*ortho*) < E(*meta*) < E(*para*).[39] The most stable carborane radicals, with the smallest ΔEs, follow the series E(*o*-(9)) < E(*m*-(9)) < E(*p*-(2)) and correspond to the isomers where the radical center is farthest away from the cage carbons. The bond dissociation energy of the most stable radical, *o*-(9), is similar to that of the phenyl radical. In general (see Figure 3), the *para* and *meta* isomers of the carborane diradicals are the most and least stable, respectively. Structures with apical carbon atoms in the carborane cage form diradicals preferentially. Whereas the range of CH/BH monoradical bond dissociation energies is only 3 kcal/mol (Figure 2), the

Table 1. Comparison of the CH or BH Bond Dissociation Energies D_0 (in kcal/mol at the UBS(B3LYP)/6-31G(d) Level) for the Singlet Diradical Ground States Derived from *r*-Carborane Monoradicals Computed with Eqs 10 and 11[a]

diradical	D_0^1	D_0^2	diradical	D_0^1	D_0^2	diradical	D_0^1	D_0^2
o-(1,2)	84.3	84.3	*m*-(1,2)	104.4	106.8	*p*-(1,2)	76.6	79.1
o-(1,3)	79.7	81.4	*m*-(1,7)	115.0	115.0	*p*-(1,7)	88.9	91.3
o-(1,4)	77.8	81.0	*m*-(1,4)	94.9	98.2	*p*-(1,12)	87.5	87.5
o-(1,7)	95.1	98.2	*m*-(1,5)	93.0	96.3	*p*-(2,3)	80.9	80.9
o-(1,8)	88.6	92.5	*m*-(1,9)	108.4	112.5	*p*-(2,4)	97.9	97.9
o-(1,9)	83.9	87.8	*m*-(1,12)	103.2	106.5	*p*-(2,7)	88.6	88.6
o-(1,12)	85.8	89.7	*m*-(2,3)	101.5	101.5	*p*-(2,8)	98.2	98.2
o-(3,4)	89.0	90.4	*m*-(2,6)	104.7	105.6	*p*-(2,9)	87.3	87.3
o-(3,5)	98.5	99.9	*m*-(2,4)	121.2	122.1			
o-(3,6)	101.6	101.6	*m*-(2,5)	114.5	115.3			
o-(3,8)	91.4	93.5	*m*-(2,10)	114.3	116.0			
o-(3,10)	90.7	92.7	*m*-(4,8)	109.1	109.1			
o-(3,9)	97.5	99.7	*m*-(4,6)	115.6	115.6			
o-(4,5)	86.2	86.2	*m*-(4,11)	107.7	107.7			
o-(4,7)	100.3	100.3	*m*-(4,5)	102.9	102.9			
o-(4,8)	89.3	89.9	*m*-(4,12)	114.9	114.9			
o-(4,10)	103.5	104.2	*m*-(4,9)	105.2	106.1			
o-(4,9)	89.7	90.5	*m*-(4,10)	118.5	119.4			
o-(4,12)	96.5	97.2	*m*-(5,12)	112.5	112.5			
o-(8,10)	98.0	98.0	*m*-(5,10)	106.2	107.0			
o-(8,9)	87.0	87.1	*m*-(9,10)	101.9	101.9			
o-(9,12)	87.5	87.5						

[a] Two D_0's are computed for those diradicals whose parent monoradicals differ; thus, the diradical *r*-(*i*, *j*) can be derived through {*r*-(*i*) → *r*-(*i*, *j*) + H} or through {*r*-(*j*) → *r*-(*i*, *j*) + H} leading to D_0^1 and D_0^2, respectively, with $i < j$.

range for carborane diradicals is nearly ten times larger (Figure 3). Finally, it is worth mentioning that the fact that some radicals have almost degenerate energies does not imply that they coexist or interconvert at room temperature, because the corresponding conversion barrier is generally high. As an example, computations on the *o*-(8) and *o*-(9) radical show an energy barrier of about 60 kcal/mol for the 1,2-hydrogen shift. The absolute and relative energies of the ground states of the compounds used to estimate relative stabilities and an example of an interconversion barrier are included in the Supporting Information.

The electronic structures of the monoradicals have a common feature: the singly occupied molecular orbital (SOMO) corresponding to the unpaired electron in all radicals is axially delocalized—the axis corresponds to the former X-H bond at the X = B or C site where the H was removed. These SOMOs always have four nodal surfaces. Figure 4 depicts the MO ϕ_e corresponding to the unpaired electron in the radical *o*-(1) and the total spin density of the same radical. In all other species the SOMO and the spin density have very similar shapes, with ϕ_e, even for nonsymmetric radicals, having approximate axial symmetry. As expected, the spin density in the radical is largely localized around the cage atom site from which the hydrogen atom has been removed. More information about the electronic structures of the diradicals is provided below.

3.2. Electronic Structure and singlet-triplet Gaps for Carborane Diradicals. *Methodological Considerations.* Diradical systems are characterized by the existence of two electrons nominally localized at specific sites of a molecule. The energy gap between the lowest energy singlet and triplet states is governed by the effective interaction between the

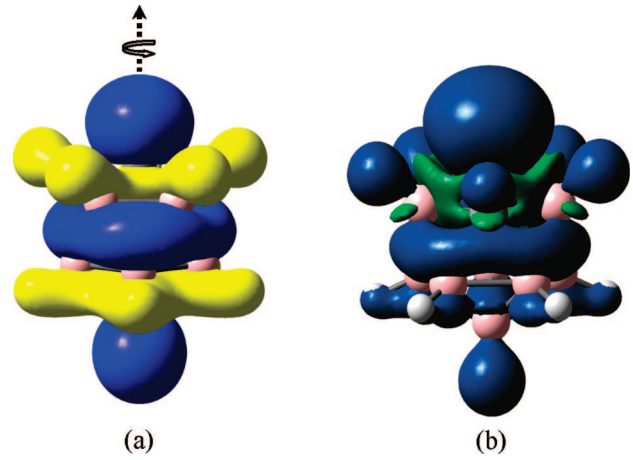


Figure 4. Electronic structure of *o*-carborane radical *o*-(1): (a) singly occupied molecular orbital (SOMO, $\phi_e = \pm 0.02$). (b) Total spin density ($\rho_s = \pm 0.0004$); the orientation of the radical corresponds to the carbon atom at the apical position (top large spin-density lobe).

electron spins of sites *i*, *j*, mapped to a Heisenberg−Dirac-Van Vleck spin Hamiltonian[40–42]

$$\hat{H}_{ex} = -J_{ij}\hat{S}_i \cdot \hat{S}_j \qquad (12)$$

Here J_{ij} is the spin-exchange coupling parameter, and a sum over *i* and *j* occurs for more than two sites. For the diradical case, J_{ij} equals the singlet-triplet splitting, with a negative value for a singlet ground state (antiferromagnetic behavior) and positive in the opposite case (ferromagnetic behavior).[19] The simplest derivations of such a spin Hamiltonian presupposes that spins are localized in each center, a phenomenon called valence trapping, but this model is often inaccurate. Still, eq 12 is generally adequate for diradicals comprised of just two unpaired electrons. For systems with several unpaired and typically delocalized valence electrons, corrections can entail further double-exchange[42,43] or cyclic (spin) permutations.[44,45] The calculation of singlet-triplet splittings, and therefore effective exchange coupling constants, J_{ij}, in almost degenerate cases by quantum chemical methods is quite delicate, requiring high accuracy to evaluate subtle correlation energy effects and to take the intrinsic multideterminantial nature of the spin manifold into account, conditions only fulfilled by computationally expensive multiconfigurational ab initio procedures with limited applicability to large complexes. A natural tendency is to move toward the widely used and less expensive Density Functional Theory (DFT) procedures.

DFT has clear limitations when applied to degenerate states, in particular because of its single-determinant character.[46,47] This circumstance occurs with diradicals, which are molecules with two weakly interacting electrons, each formally associated with different atomic centers. The standard DFT representation is approximately valid in computing triplet states but is totally inadequate for spin contaminated singlet diradicals.[19] Along with other techniques,[42] unrestricted broken symmetry (UBS) DFT approximations have been developed[18] to deal with open-shell or multiconfigurational situations. The broken spin symmetry yields a wave function which is a mixture of different spin

Table 2. Adiabatic singlet-triplet Gaps and Geometrical Parameters for Selected Carborane Diradicals at Different Levels of Theory[a]

		adiabatic singlet-triplet energy gap[b]/eV					distance between diradical centers/Å					
		B3LYP/UBS geometry		MP2 geometry		CASSCF geometry	B3LYP/UBS geometry		MP2 geometry		CASSCF geometry	
system	class	B3LYP/UBS($<S^2>$)	CASPT2	MP2	CASPT2	CASPT2	S_0	T_1	S_0	T_1	S_0	T_1
o-(1,2)	I*a*	0.93/0.93(0.00)	1.15	1.60	1.17	1.17	1.361/1.361	1.636	1.400	1.639	1.364	1.602
o-(3,6)	II*c*	−1.05/0.03(0.99)	−0.05[c]	0.19	0.03	0.06	2.800/2.880	2.887	2.892	2.885	2.899	2.901
o-(1,7)	II*b*	0.18/0.28(0.54)	−0.02[c]	0.24	−0.04[c]	0.17	2.586/2.607	2.694	2.579	2.693	2.624	2.691
o-(1,12)	III*b*	0.98/0.98(0.00)	0.85	0.99	0.87	0.76	3.097/3.097	3.090	3.100	3.101	3.121	3.118
m-(1,2)	I*a*	0.63/0.63(0.00)	0.75	0.93	0.82	0.82	1.463/1.463	1.688	1.479	1.689	1.509	1.686
m-(4,8)	I*c*	−0.01/0.38(0.75)	0.38	0.31	0.36	0.37	1.570/1.664	1.763	1.614	1.764	1.666	1.771
m-(1,7)	II*a*	−0.22/0.27(0.85)	0.36	0.94	0.33	0.37	2.292/2.462	2.517	2.432	2.517	2.438	2.517
m-(2,4)	II*c*	−0.91/-0.19(-)[e]	−0.14[c]	0.12	−0.03[c]	0.04	2.718/2.607[e]	2.838	2.837	2.843	2.842	2.855
m-(2,10)	II*c*	−0.77[d]/0.07(0.97)	−0.11[c]	0.61[d]	−0.10[c]	0.04[d]	2.707/2.806	2.819	2.793	2.828	2.853	2.840
m-(4,10)	II*c*	−0.77/-0.10(-)[e]	−0.02[c]	0.15	0.01	0.07	2.680/2.680	2.840	2.800	2.840	2.860	2.873
m-(1,12)	III*b*	0.49/0.70(0.44)	0.77	0.74	0.60	0.74	3.086/3.103	3.094	3.160	3.104	3.163	3.137
p-(1,2)	I*b*	0.91/0.91(0.00)	0.94	0.19	0.96	0.98	1.466/1.466	1.703	1.476	1.702	1.492	1.709
p-(1,7)	II*b*	−0.03/0.25(0.78)	0.27	0.24	0.22	0.34	2.454/2.596	2.663	2.378	2.669	2.615	2.659
p-(2,4)	II*c*	−0.93/-0.09(-)[e]	−0.07[c]	0.12	0.01	0.04	2.700/2.700	2.845	2.805	2.850	2.859	2.872
p-(1,12)	III*a*	0.17/0.62(0.68)	0.72	1.77	0.72	0.67	2.920/2.919	2.865	2.973	2.875	2.934	2.885
p-(2,9)	III*c*	−0.20/0.43(0.83)	0.42	1.21	0.47	0.48	3.300/3.328	3.315	3.360	3.324	3.359	3.352

[a] CASSCF computations: active space 6 electrons in 6 orbitals. [b] A positive singlet-triplet gap indicates a singlet (ground) state lower in energy. [c] At T_1 geometry the S_0 state is lower in energy than at S_0 geometry or vice versa. [d] Results increasing the basis set: B3LYP/aug-cc-pVDZ −0.71 eV; MP2/aug-cc-pVDZ 0.10 eV; CASPT2/aug-cc-pVDZ//CASSCF/6−31G(d) 0.06 eV, CASPT2(12,12)/aug-cc-pVDZ//CASSCF/6−31G(d) 0.05 eV. [e] Poor convergence. $\langle S^2 \rangle$ close to one. Less reliable result.

states.[48] singlet-triplet gaps (ΔE_{ST}) are therefore based on the standard DFT triplet state (E_T) result and the UBS (unrestricted-DFT) solution for the singlet state (E_{BS}). The gap is computed as[49]

$$\Delta E_{ST} = \frac{2(E_{BS} - E_T)}{\langle \hat{S}^2 \rangle_T - \langle \hat{S}^2 \rangle_{BS}} = \frac{2(E_{BS} - E_T)}{2 - \langle \hat{S}^2 \rangle_{BS}} \qquad (13)$$

The expression derives from a UBS calculation with a mixed wave function and considers the state of the highest multiplicity (E_T, triplet) to be a pure spin state ($\langle \hat{S}^2 \rangle_T = 2$), whereas the lower spin (E_{BS}, singlet) state, computed with the UBS method, is contaminated by higher spin state components, and its solution will not be an eigenfunction of the spin operator. The magnitude of $\langle \hat{S}^2 \rangle_{BS}$ is a measure of the degree of mixing between both states. Equation 13 is equivalent to that obtained within the spin projector technique.[42,48]

Table 2 compiles the singlet-triplet energy gaps, the expectation values for the spin operators, and the distances between diradical centers for selected carborane diradicals at different levels of theory. A positive gap indicates that the singlet (ground) state is lower in energy. Detailed discussions of the performance of the different approaches to deal with degenerate situations are available.[19] We give relevant benchmarking information regarding the level of computation employed here to obtain the energies and structures of the lowest singlet and triplet states and analyze the theoretical problems related with the diradical computations.

Results on Selected Carboranes. From the full set of carborane diradicals, Table 2 includes the different types of species identified by a nomenclature scheme. The parent *o*-, *m*-, and *p*-carboranes differ in the position of the carbon atoms. Each can give rise to different locations of the diradical centers. For instance, the carbons are adjacent in *o*-carboranes, but the radical centers can be contiguous, as in *o*-(1,2), or can be separated by one atom, *o*-(1,7), or by two atoms, *o*-(1,12). The latter has the diradical centers at opposite "poles" of the near-spherical carborane cage. We identify these three structure types as Class I, II, and III carboranes, respectively, and we combine this with labels *a*, *b*, and *c*, for diradical centers located on two carbon atoms (*a*), one carbon and one boron atom (*b*), or two boron atoms (*c*) (*cf.* Table 2).

The singlet-triplet gaps depend more on the location of the diradical centers than on any other structural parameter. For instance, at all levels of theory the computed gap in (1,2) carboranes (Classes Ia and Ib) is large (close to one eV). In such molecules, the diradical centers are contiguous, and the singlet state CASSCF wave function can be described by a predominant closed-shell electronic configuration. Consequently, standard DFT/B3LYP theory provides singlet-triplet gaps within 0.2 eV of the more accurate CASPT2 results. In contrast, the behavior of the single-reference (U)MP2 theory is erratic, deviations sometimes reach ±1 eV, essentially reflecting the large spin contamination of the unrestricted HF wave function.

The UBS results are equivalent to the standard B3LYP results for Class Ia and Ib carboranes, as expected from the single-reference character displayed by the CASSCF wave function. The expectation value of the spin operator, 0.00, indicates that the UBS solution converges to the standard B3LYP solution, with states of pure singlet spin character. Class Ic carboranes like *m*-(4,8), in which the diradical centers are two boron atoms, are different. Whereas B3LYP leads to unphysical degenerate situations, UBS and MP2 gaps are close to the CASPT2 results. The symmetry-broken (UBS) situation with $\langle \hat{S}^2 \rangle_{BS} \approx 1$ indicates diradical character. The CASPT2 gaps are quite similar to the UBS values here. The CASSCF wave function reflects a situation in which two electronic configurations are predominant, and the

Table 3. Adiabatic singlet-triplet Energy Gaps (ΔE_{ST}, eV) Computed at the Different Levels for the Singlet (S) and Triplet (T) Monomeric Diradicals Derived from *o*-, *m*-, and *p*-Carborane (See Figure 1 and Text)[a]

o-carborane	B3LYP/BS	*m*-carborane	B3LYP/BS	*p*-carborane	B3LYP/BS
label/class	ΔE_{ST}	label/class	ΔE_{ST}	label/class	ΔE_{ST}
(1,2)/Ia	0.93/0.93	(1,2)/Ib	0.63/0.63	(1,2)/Ib	0.91/0.91
(1,3)/Ib	1.07/1.07	(1,7)/IIa	−0.22/0.27	(1,7) /IIb	−0.03/0.33
(1,4)/Ib	1.07/1.07	(1,4)/Ib	1.05/1.05	(1,12) /IIIa	0.17/0.62
(1,7)/IIb	0.18/0.28	(1,5)/Ib	1.12/1.12	(2,3) /Ic	0.58/0.76
(1,8)/IIb	0.53/0.53	(1,9)/IIb	0.21/0.36	(2,4) /IIc	−0.93/−0.09[b]
(1,9)/IIIb	0.75/0.75	(1,12)/IIIb	0.49/0.70	(2,7) /Ic	−0.08/0.35
(1,12)/IIIb	0.98/0.75	(2,3)/Ic	0.73/0.83	(2,8) /IIc	−0.87/−0.11[b]
(3,4)/Ic	0.32/0.60	(2,6)/Ic	0.34/0.61	(2,9) /IIIc	−0.20/0.43
(3,5)/IIc	0.89/0.11[b]	(2,4)/IIc	−0.91/−0.19[b]		
(3,6)/IIc	−1.05/0.03	(2,5)/IIc	−0.72/0.11		
(3,8)/Ic	0.08/0.43	(2,10)/IIc	−0.77/0.07		
(3,10)/IIIc	−0.03/0.49	(4,8)/Ic	−0.01/0.38		
(3,9)/IIc	−0.58/0.13	(4,6)/IIc	−0.87/0.06		
(4,5)/Ic	0.60/0.76	(4,11)/IIIc	−0.11/0.47		
(4,7)/IIc	−0.92/0.03[b]	(4,5)/Ic	0.52/0.72		
(4,8)/Ic	0.28/0.55	(4,12)/IIc	−0.69/0.10		
(4,10)/IIc	−0.76/−0.14[b]	(4,9)/Ic	0.25/0.54		
(4,9)/Ic	0.25/0.52	(4,10)/IIc	−0.77/−0.10[b]		
(4,12)/IIc	−0.50/0.18	(5,12)/IIc	−0.51/0.21		
(8,10)/IIc	−0.67/0.11	(5,10)/Ic	0.18/0.49		
(8,9)/Ic	0.48/0.67	(9,10)/Ic	0.54/0.72		
(9,12)/Ic	0.43/0.64				

[a] A positive singlet-triplet gap indicates a singlet (ground) state lower in energy. [b] Geometry converged with lower force and displacement thresholds. Less reliable result.

HOMO (H) and LUMO (L) natural orbitals, with respective bonding and antibonding character between the contiguous boron atoms (see below), display occupations of 1.52 and 0.48 electrons, respectively. We rationalize the different behavior below based on orbital structures and interactions. However, despite the poor account of energy gaps in some cases, the computations in Table 2 indicate that the geometries obtained at the different levels of theory, B3LYP, UBS, MP2, and CASSCF, give reasonably consistent structural parameters. The single point CASPT2 gaps, based on these geometries, hardly differ more than 0.1 eV.

Class II carborane diradicals are much more interesting in terms of their potential in materials applications because of their smaller singlet-triplet gaps, especially when the diradical centers are boron atoms. Thus, Class IIc compounds like *o*-(3,6), *m*-(2,4), *m*-(2,10), *m*-(4,10), and *p*-(2,4) in Table 2 have gaps smaller than 0.1 eV at the CASSCF/CASPT2 level. Standard DFT/B3LYP theory is particularly poor in such cases. The energy of the singlet state is heavily underestimated; this leads to a negative singlet-triplet gap indicating incorrectly a triplet ground state. Errors as large as 1 eV are found. MP2 also shows erratic behavior, although increasing the one-electron basis set to aug-cc-pVDZ for *m*-(2,10) improved the results considerably (*cf.* Table 2). In contrast, UBS performs quite well compared with CASPT2 in most cases, yielding degenerate situations with small energy gaps. Except discarded computations when the wave function did not converge properly due to technical problems, singlet states always were lower in energy adiabatically. The expectation value of the spin operator for the UBS "singlet-like" solution is ≈1 for Class IIc compounds, an indication of an average mixing between the pure singlet (0.0) and triplet (2.0) cases. Analysis of the CASSCF wave function illustrates the behavior of the quantum chemical methods again. Class IIc singlet states are pure diradicals; two almost equally predominant configurations describe the wave function, and two orbitals have occupations near unity. In such situations, the UBS approach behaves much better than standard DFT, although the accuracy is, as expected, limited.[42,47] Class IIa and IIb carboranes exhibit intermediate character. Whereas the B3LYP and MP2 results are somewhat erratic, UBS gives gaps within 0.15 eV of the CASPT2 results.

The behavior of Class III carborane diradicals seems surprising at first glance. All their singlet-triplet gaps are larger than those of Class II diradicals. Class III diradicals have more stable singlet ground states, despite the larger distance found between the diradical centers (which generally favors degeneracy). This difference is rationalized below in terms of the character of the singlet state involved. The behavior of the various levels of theory with Class III carboranes is intermediate between the Class I and II behavior. Although the closed-shell HF configuration predominates in the description of the CASSCF wave function, the solution can be considered multiconfigurational. The value of the spin operator in the UBS solution reveals differences between the UBS and B3LYP descriptions. Larger deviations of the expectation value from 0.00 (see Tables 3 and 4) correspond to larger discrepancy between both UBS-DFT and DFT results. For instance, the expectation value of 0.00 for *o*-(1,12) corresponds to a singlet state well described by a closed-shell reference in the CASSCF wave function; UBS and B3LYP are equivalent. For *p*-(2,9), the UBS expectation value is $\langle \hat{S}^2 \rangle_{BS} = 0.83$, and the CASSCF wave function of the singlet state is composed of two predominant configurations: H (2.0) - L (0.0), 69%, and H (1.0) - L (1.0), 29%, with natural occupations H (1.60) and L (0.40). This intermediate diradical character, especially for

Table 4. Expectation Values for $\langle \hat{S}^2 \rangle_{BS}$ in the Broken-Symmetry BS (U)B3LYP/6-31G(d) Computations for the Singlet States of Diradical Carboranes

o-carborane	B3LYP	*m*-carborane	B3LYP	*p*-carborane	B3LYP
diradicals	$\langle S^2 \rangle$	diradicals	$\langle S^2 \rangle$	diradicals	$\langle S^2 \rangle$
(1,2)	0.00	(1,2)	0.00	(1,2)	0.00
(1,3)	0.00	(1,7)	0.85	(1,7)	0.78
(1,4)	0.00	(1,4)	0.00	(1,12)	0.68
(1,7)	0.54	(1,5)	0.00	(2,3)	0.36
(1,8)	0.00	(1,9)	0.52	(2,4)	-[a]
(1,9)	0.00	(1,12)	0.44	(2,7)	0.79
(1,12)	0.00	(2,3)	0.21	(2,8)	-[a]
(3,4)	0.55	(2,6)	0.55	(2,9)	0.83
(3,5)	-[a]	(2,4)	-[a]		
(3,6)	0.99	(2,5)	0.95		
(3,8)	0.69	(2,10)	0.97		
(3,10)	0.78	(4,8)	0.75		
(3,9)	0.94	(4,6)	0.97		
(4,5)	0.34	(4,11)	0.80		
(4,7)	-[a]	(4,5)	0.41		
(4,8)	0.58	(4,12)	0.75		
(4,10)	-[a]	(4,9)	0.60		
(4,9)	0.59	(4,10)	-[a]		
(4,12)	0.92	(5,12)	0.91		
(8,10)	0.95	(5,10)	0.64		
(8,9)	0.42	(9,10)	0.39		
(9,12)	0.46				

[a] Geometry converged with looser force and displacement thresholds. $\langle S^2 \rangle$ close to one.

Class IIIc carboranes, explains the discrepancy between the B3LYP gap (−0.20 eV) and the UBS gap (0.43 eV), the latter being quite close to the CASPT2 result (0.48 eV).

Table 3 lists the singlet-triplet (ΔE_{ST}) energy gaps for all monomeric diradicals derived from *o*-carborane, *m*-carborane, and *p*-carborane computed at the B3LYP and UBS levels of computation. The performance of different levels of theory, discussed above, leads to the conclusion that UBS results should be more trustworthy than standard B3LYP results, except when convergence is poor. The larger the gap, the closer the correspondence between both DFT-type computations. Class Ia and Ib carborane diradicals exemplify this behavior which can be rationalized by analysis of the $\langle S^2 \rangle_{BS}$ operator values (compiled in Table 4).

Only Class IIc carborane diradicals, with the diradical centers located in noncontiguous boron atoms (two connections/bonds), display degeneracy between the lowest singlet and triplet states. Such species may have interesting (temperature-dependent and readily controllable) magnetic properties. Such possibilities extend to 1D, 2D, and 3D multidimensional polyradical carborane structures with boron radical centers; exchange coupling between the different 0D units should be important. We have found no triplet diradical ground states at our highest level of theory, CASPT2//CASSCF. Therefore, it appears that all diradical *closo*-$C_2H_{10}B_{10}$ carboranes have singlet ground states.

As mentioned above, the singlet-triplet energy gaps of diradicals are generally expected to decrease as the distance (d_{XX}) between the two radical centers increases. However, the carborane diradicals Class I ($d_{XX} \approx 1.4-1.8$ Å), Class II ($d_{XX} \approx 2.5-2.8$ Å), and Class III ($d_{XX} \approx 2.9-3.4$ Å) do not behave this way. Only Class II carborane diradicals are nearly degenerate. Class III singlet-triplet gaps are greatest, despite having the largest d_{XX}'s. Clearly, the magnitude of the splitting also depends on the orbital overlap contact region of the radicaloid electron spins. As judged from the radical spin densities of Figure 4, evidently this overlap can be considerable. If the orbital overlap between the two (radical) centers were weak, then the main dependence should be that of the "exchange charge density" appearing in a traditional exchange integral, and one would anticipate a large *R (i.e.,* d_{XX} distance) dependence

$$J = J_0 e^{-\alpha R} \tag{14}$$

where α is a constant, and J_0 has a weaker polynomial dependence on R (and $1/R$) along with a possible dependence on angle of orbital orientation. Evidently, the orbital overlap is significant and may be delicately balanced with important directional dependences. Consequently, our carborane diradicals do not fulfill the rule that the exchange parameters decrease with distance. The "effective" exchange interaction is the root cause of the singlet-triplet splitting, which in turn typically dominates the magnetic properties.

Figure 5 displays the CASSCF natural HOMO- and LUMO-like orbitals of three different carborane diradicals in their lowest singlet and triplet states together with their computed ΔE_{ST} gaps. Class I carboranes have their two diradical centers on neighboring atoms. Such structures have short d_{XX} distances and large orbital overlap, which maximizes the singlet-triplet splitting. The diradical centers in Class II carboranes are on noncontiguous (two connections/bonds) atoms. Not only is the distance (d_{XX}) greater than in Class I but also the orbital overlap is smaller. Although the diradical centers in Class III carboranes, on opposite sides of the cage, are furthest apart, the orbital overlap is through the cage (the "antipodal effect")[50] and is quite large. The nature of the lowest singlet state is important. It is clearly multiconfigurational in Class I and III carboranes; the orbital occupation numbers indicate the mixed localized-delocalized character of their wave function. This can be analyzed best in Class Ic carboranes, when the diradical centers are two contiguous boron atoms. Class Ia carboranes with contiguous carbon "diradical" centers form a C=C double bond; in *o*-(1,2) the C=C distance is 1.361 Å at the DFT level. The distance between contiguous B−B diradical centers in Class Ic compounds ranges between 1.593 Å for *m*-(2,3) to 1.666 Å in *p*-(2,7) at the UBS level. These values are similar to the CASSCF distances (e.g., 1.664 Å (UBS) and 1.666 Å (CASSCF) for *m*-(4.8)) but are nearly 0.1 Å longer than the B3LYP values. The computed bond length falls within the range 1.56−1.68 Å determined as the distance of a boron−boron double bond in recently synthesized stable diborenes.[51] In Class Ic carboranes with contiguous boron radical centers, the nature of the bonding is complex. This is shown by the shape of the "HOMO" (bonding B−B) and "LUMO" (antibonding) orbitals and by their natural occupation numbers, 1.53 and 0.47, respectively. On the other hand, the occupation numbers in Class II carboranes point toward a two-configurational wave function and a more clearly diradical character for both singlet and triplet states. Illustrations with Mulliken population differences and spin densities

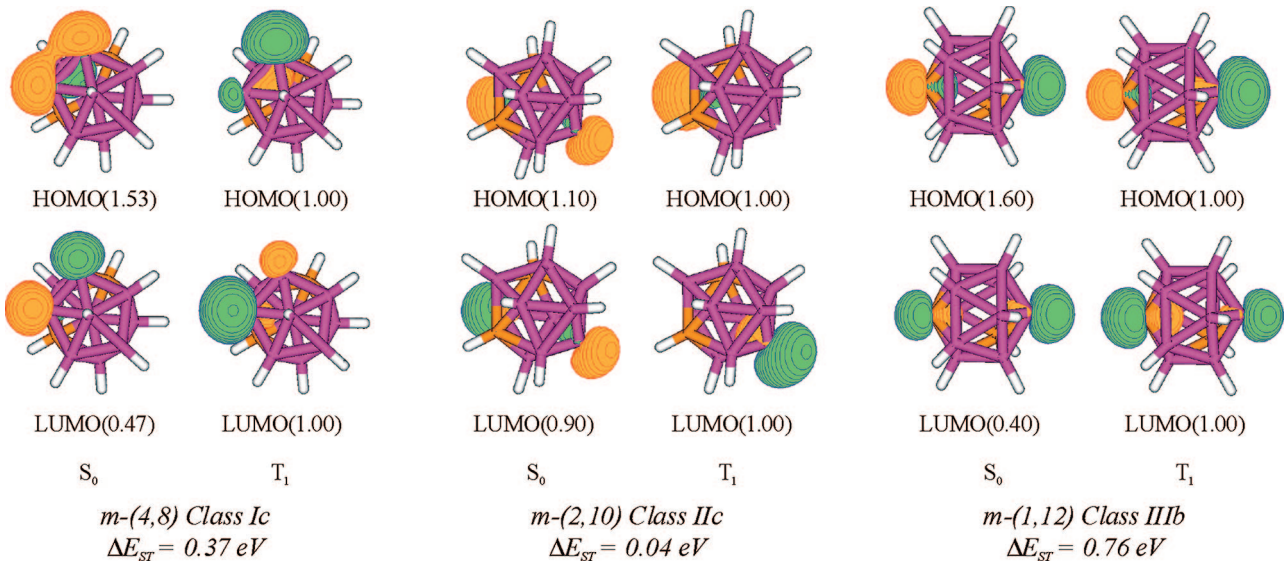


Figure 5. CASSCF HOMO- and LUMO-like natural orbital occupation for the lowest singlet and triplet states and CASPT2 singlet-triplet gaps for *m*-carborane diradical *m*-(i,j) derived from hydrogen abstraction in positions (4,8), (2,10), and (1,12).

for some of the carboranes can be found in the Supporting Information.

We conclude that it is possible to use the expectation value of $\langle \hat{S}^2 \rangle_{BS}$ in the UBS solution to evaluate the character of the singlet state, as it reflects the magnitude of the mixing between pure spin states. A value of 0.0 indicates a stable singlet wave function (identical with standard DFT), whereas a value approaching 1.0 indicates diradical character. Such systems have small HOMO−LUMO gaps and S-T separations. Thus, examples in Figure 5, the Class IIc *m*-(2,10), Class Ic *m*-(4,8), and Class IIIb *m*-(1,12) carboranes, have UBS singlet-triplet splittings of 0.07, 0.38, and 0.70 eV and expectation values for the spin operator of 0.97, 0.75, and 0.44, respectively.

4. Summary and Conclusions

A comprehensive analysis of all the isomeric monoradicals and diradicals, derived from icosahedral carboranes by single and by double hydrogen abstraction, has been presented by using high-level quantum chemical calculations. Whereas the bond dissociation energies (BDE) leading to the carborane monoradicals are close to the benzene BDE, the most stable carborane radicals (designated *o*-(9), *m*-(9), and *p*-(2)) are derived from dissociations of hydrogens farthest away from the carbon atoms. Most of the carborane radicals and diradicals will be expected to be highly reactive and only stabilized in the gas phase, having a strong tendency for dimerization or polymerization. The production of radicals as oxidized methylated carborane anions, the existence of a permethylated carborane radical less prone to dimerization,[52] and the recent synthesis of diradicals derived from the former[16] are experimental starting points of polyradical architectures based on carboranes.

The theoretical adiabatic singlet-triplet splittings for carborane diradicals underscores the possibility of using diradicals as operational magnet units in multidimensional carborane architectures. Double hydrogen abstraction from noncontiguous boron centers, particularly those separated by two connections/bonds in the carborane cages (described here as Class IIc carborane diradicals), gives rise to near degenerate lowest singlet-triplet states, although the ground state is always a singlet state at reliable levels of theory. Therefore single carborane diradicals have an antiferromagnetic behavior. Examples of Class IIc carborane diradicals are the systems *o*-(3,6) (e.g., double hydrogen abstraction from *o*-carborane in positions 3 and 6), *o*-(4,7), *m*-(2,4), *m*-(2,10), *m*-(4,10), and *p*-(2,4), in which very small singlet-triplet gaps (<0.1 eV) have been computed ($\Delta E_{ST} \approx 0.10$ eV = 2.30 kcal/mol). The performance of standard DFT, unrestricted broken-symmetry (UBS) DFT, MP2, and CASSCF/CASPT2 methods has been evaluated. UBS is the only practicable strategy giving reliable singlet-triplet gaps comparable with the more accurate but computationally demanding CASPT2 procedure. UBS may be successful because only two configurations are required to basically describe the reference wave function. Therefore, we plan to employ UBS to analyze general trends in larger diradical systems such as those with multicarborane polyradical architectures.

Additional supplemental information (geometries and frequencies) is available from the authors (PDF).

Acknowledgment. The authors would like to thank Prof. Manuela Merchán (Valencia), Prof. Wesley D. Allen (Athens), and Jose M[a] Castelló (Barcelona), for fruitful discussions and comments on the manuscript. The research was supported in Spain by projects CTQ2007-61260, MAT2006-13646-C03-02, and CSD2007-0010 Consolider-Ingenio in Molecular Nanoscience of the Spanish MEC/FEDER and in the U.S.A. by the Welch Foundation of Houston, Texas, BD-0894, and by the National Science Foundation Grant CHE-0716718.

Supporting Information Available: Mulliken population and spin densities in selected positions of several carboranes monoradicals and diradicals derived from *o*-, *m*-, and *p*- at the B3LYP/6−31G(d) level of theory, tables for the computation of relative stabilities of monoradicals and

diradicals, computed energy barrier for the 1,2-hydrogen shift in the interconversion *o*-(8) and *o*-(9) radicals (Figure S2), and figures and data on the stability of protonated carborane species. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Muller, J.; Base, K.; Magnera, T. F.; Michl, J. *J. Am. Chem. Soc.* **1992**, *114*, 9721.

(2) González-Campo, A.; Nuñez, R.; Viñas, C.; Boury, B. *New J. Chem.* **2006**, *30*, 546.

(3) Grimes, R. N. *J. Chem. Educ.* **2004**, *81*, 657.

(4) Hiura, H.; Kanayama, T. *J. Mol. Struct.* **2005**, *735*, 367.

(5) Oliva, J. M.; Allan, N. L.; Schleyer, P. v. R.; Viñas, C.; Teixidor, F. *J. Am. Chem. Soc.* **2005**, *127*, 13538.

(6) Oliva, J. M.; Serrano-Andrés, L. *J. Comput. Chem.* **2006**, *27*, 524.

(7) Serrano-Andrés, L.; Oliva, J. M. *Chem. Phys. Lett.* **2006**, *432*, 235.

(8) Manero, V.; Oliva, J. M.; Serrano-Andres, L.; Klein, D. J. *J. Chem. Theory Comput.* **2007**, *3*, 1399.

(9) (a) Wolf, S. A.; Awschalom, D. D.; Buhrman, R. A.; Daughton, J. M.; von Mólnar, S.; Roukes, M. L.; Chtchelkanova, A. Y.; Treger, D. M. *Science* **2001**, *294*, 1488. (b) Hirjibehedin, C. F.; Lutz, C. P.; Heinrich, A. J. *Science* **2006**, *312*, 1021. (c) Nori, F.; Tonomura, A. *Science* **2006**, *311*, 344. (d) Brune, H. *Science* **2006**, *312*, 1005.

(10) *Magnetism: Molecules to Materials*; Miller, J. S., Drillon, M., Eds.; Wiley-VCH: 2004.

(11) Ellis, D.; McIntosh, R. D.; Esquirolea, S.; Viñas, C.; Rosair, G. M.; Teixidor, F.; Welch, A. J. *Dalton Trans.* **2008**, 1009.

(12) Jemmis, E. D.; Balakrishnarajan, M. M. *J. Am. Chem. Soc.* **2000**, *122*, 7392–7393.

(13) McKee, M. L. *J. Am. Chem. Soc.* **1997**, *119*, 4220.

(14) Wang, Z. X.; Schleyer, P. v. R. *Angew. Chem., Int. Ed.* **2002**, *41*, 4082.

(15) Konarev, D. V.; Khasanov, S. S.; Otsuka, A.; Saito, G.; Lyubovskaya, R. N. *J. Am. Chem. Soc.* **2006**, *128*, 9292.

(16) Eriksson, J.; Vyakaranam, K.; Ludvík, J.; Michl, J. *J. Org. Chem.* **2007**, *72*, 2351.

(17) *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.; 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, Inc.: Wallingford, CT, 2004.

(18) (a) Yamaguchi, K.; Yoshioka, Y.; Takayuki, F. *Chem. Phys. Lett.* **1977**, *46*, 360. (b) Noodleman, L. *J. Chem. Phys.* **1981**, *74*, 5737. (c) Borden, W. T.; Davidson, E. R.; Feller, D. *Tetrahedron* **1982**, *38*, 737. (d) Noodleman, L.; Davidson, E. R. *Chem. Phys.* **1986**, *109*, 131. (e) Noodleman, L.; Peng, C. Y.; Case, D. A.; Mouesca, J. M. *Chem. Rev.* **1995**, *144*, 199. (f) Yamaguchi, K.; Kawasaki, T.; Takano, Y.; Kigarawa, Y.; Yamashita, Y.; Fujita, H. *Int. J. Quantum Chem.* **2002**, *90*, 370.

(19) Calzado, C. J.; Cabrero, J.; Malrieu, J. P.; Caballol, R. *J. Chem. Phys.* **2002**, *116*, 2728.

(20) Andersson, K.; Malmqvist, P.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218.

(21) Serrano-Andrés, L.; Merchán, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. *J. Chem. Phys.* **1993**, *98*, 3151.

(22) Roos, B. O.; Andersson, K.; Fülscher, M. P.; Malmqvist, P.; Serrano-Andrés, L.; Pierloot, K.; Merchán, M. *Adv. Chem. Phys.* **1996**, *93*, 219.

(23) Borin, A. C.; Serrano-Andrés, L. *Chem. Phys.* **2000**, *262*, 253.

(24) Serrano-Andrés, L.; Merchán, M.; Borin, A. C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8691.

(25) Serrano-Andrés, L.; Fülscher, M. P.; Roos, B. O.; Merchan, M. *J. Phys. Chem.* **1996**, *100*, 6484.

(26) Serrano-Andrés, L.; Forsberg, N.; Malmqvist, P.-Å. *J. Chem. Phys.* **1998**, *108*, 7202.

(27) Muñoz, D.; De Graaf, C.; Illas, F. *J. Comput. Chem.* **2004**, *25*, 1234.

(28) Serrano-Andrés, L.; Merchán, M.; Borin, A. C. *Chem. Eur. J.* **2006**, *12*, 6559.

(29) Andersson, K.; Barysz, M. ; Bernhardsson, A.; Blomberg, M. R. A.; Carissan, Y.; Cooper, D. L.; Cossi, M.; Fülscher, M. P.; Gagliardi, L.; de Graaf, C.; Hess, B.; Hagberg, G.; Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Nakajima, T.; Neogrády, P.; Olsen, J.; Raab, J.; Roos, B. O.; Ryde, U.; Schimmelpfennig, B.; Schütz, M.; Seijo, L.;Serrano-Andrés, L.; Siegbahn, P. E. M.; Stålring, J.; Thorsteinsson, T.; Veryazov, V.; Widmark, P.-O. MOLCAS, version 6.0; Department of Theoretical Chemistry, Chemical Centre, University of Lund: Lund, Sweden, 2004.

(30) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neograd y, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222.

(31) Veryazov, V.; Widmark, P.-O.; Serrano-Andŕes, L.; Lindh, R.; Roos, B. O. *Int. J. Quantum Chem.* **2004**, *100*, 626.

(32) Nonhebel, D. C.; Tedder, J. M.; Walton, J. C. *Radicals*; Cambridge University Press: Cambridge, 1979.

(33) Ervin, K. M.; DeTuri, V. F. *J. Phys. Chem. A* **2002**, *106*, 9947.

(34) Wenthold, P. G.; Squires, R. R. *J. Am. Chem. Soc.* **1994**, *116*, 6401.

(35) G. E. Davico, V. M.; Bierbaum, C. H.; DePuy, G. B.; Ellison, R. R. Squires. *J. Am. Chem. Soc.* **1995**, *117*, 2590.

(36) Blanksby, S. J.; Ellison, G. B. *Acc. Chem. Res.* **2003**, *36*, 255.

(37) Winkler, M.; Sander, W. *J. Phys. Chem. A* **2001**, *105*, 10422.

(38) Schreiner, P. R.; Navarro-Vazquez, A.; Prall, M. *Acc. Chem. Res.* **2005**, *38*, 29.

(39) *Gmelin Handbuch Der Organische Chemie*; Ergenzungswerek zur 8, Auflage, Band Berlin-Heidelberg, Seite 15, 1977.

(40) Eq 12 corresponds to a purely phenomenological Hamiltonian.

(41) (a) Heisenberg, W. Z. *Phys.* **1928**, *49*, 619. (b) Dirac; P, A. M. *Proc. R. Soc. London, Ser. A* **1929**, *123*, 714. (c) Van Vleck, *J. Phys. Rev.* **1934**, *45*, 405.

(42) Ciofini, I.; Daul, C. A. *Coord. Chem. Rev.* **2003**, *238*, 187.

(43) Klein, D. J.; Seitz, W. A. *Phys. Rev. B* **1973**, *8*, 2236.

(44) Mulder, J. J. C.; Oosterhoff, L. *J. Chem. Commun.* **1970**, 305.

(45) Wu, J.; Schmalz, T. G.; Klein, D. J. *J. Chem. Phys.* **2002**, *117*, 9977.

(46) Parr, R. G.; Young, W. *Density Functional Theory of Atoms and Molecules*;Oxford University Press: New York, 1989.

(47) Bersuker, I. B. *J. Comput. Chem.* **1997**, *18*, 260.

(48) Ovchinnikov, A.; Labanowski, J. K. *Phys. Rev. A* **1996**, *53*, 3946.

(49) Noodleman, L., Jr. *J. Chem. Phys.* **1979**, *70*, 4903.

(50) Bühl, M.; Schleyer, P. v. R.; Havlas, Z.; Hnyk, D.; Hermanek, S. *Inorg. Chem.* **1991**, *30*, 3107.

(51) (a) Wang, Y.; Quillian, B.; Wei, P.; Wannere, C. S.; Xie, Y.; King, R. B.; Schaefer, H. F., III; Schleyer, P. v. R.; Robinson, G. H. *J. Am. Chem. Soc.* **2007**, *129*, 12412. (b) Wang, Y. Z.; Quillian, B.; Wei, P. R. *J. Am. Chem. Soc.* **2008**, *130*, 3298.

(52) King, B. T.; Noll, B. C.; McKinley, A. J.; Michl, J. *J. Am. Chem. Soc.* **1996**, *118*, 10902.

CT800150H

JCTC Journal of Chemical Theory and Computation

A Hamiltonian Replica Exchange Approach and Its Application to the Study of Side-Chain Type and Neighbor Effects on Peptide Backbone Conformations

Chao Xu, Jun Wang, and Haiyan Liu*

School of Life Sciences, and Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China (USTC), Hefei, Anhui 230027, China

Received December 29, 2007

Abstract: We presented a Hamiltonian replica exchange approach and applied it to investigate the effects of various factors on the conformational equilibrium of peptide backbone. In different replicas, biasing potentials of varying strengths are applied to all backbone (φ,ψ) torsional angle pairs to overcome sampling barriers. A general form of constructing biasing potentials based on a reference free energy surface is employed to minimize sampling in physically irrelevant parts of the conformational space. An extension of the weighted histogram analysis formulation allows for conformational free energy surfaces to be computed using all replicas, including those with biased Hamiltonians. This approach can significantly reduce the statistical uncertainties in computed free energies. For the peptide systems considered, it allows for effects of the order of 0.5−1 kJ/mol to be quantified using explicit solvent simulations. We applied this approach to capped peptides of 2−5 peptide units containing Ala, Phe, or Val in explicit water solvent and focused on how the conformational equilibrium of a single pair of backbone angles are influenced by changing the residue types of the same and neighboring residues as well as conformations of neighboring residues. For the effects of changing side-chain types of the same residue, our results consistently showed increased preference of β for Phe and Val relative to Ala. As for neighbor effects, our results not only indicated that they can be as large as the effects of changing the side-chain type of the same residue but also led to several new insights. We found that for the N-terminal neighbors, their conformations seem to have large effects. Relative to the β conformer of an N-terminal neighbor, its α conformer stabilizes the β conformer of its next Ala disregarding the residue type of the neighbor. For C-terminal neighbors, their chemical identities seem to play more important roles. Val as the C-terminal neighbor significantly increases the PII propensity of its previous Ala disregarding its own conformational state. These results are in good accordance with reported statistics of protein coil structure libraries, proving the persistent presence of such effects in short peptides as well as in proteins. We also observed other side-chain identity and neighbor effects which have been consistently reproduced in our simulations of different small peptide systems but not displayed by coil library statistics.

Introduction

Sequence-dependences of intrinsic local conformation propensities of short peptide segments have attracted wide attentions. There has been ample evidence that such propensities may lead to sequence-specific stabilization of local structures in native proteins or nonrandom local conformations in unstructured proteins or peptides.[1–4] One interesting question is to what extent such propensities depend on not

* Corresponding author e-mail: hyliu@ustc.edu.cn.

10.1021/ct7003534 CCC: $40.75
Published on Web 06/21/2008

only the identity of individual residues but also their sequence and conformation contexts. Such context dependences lead to deviations from the "isolated-pair hypothesis" of Flory,[5] which has often been assumed in equilibrium and kinetic treatments of peptide and protein folding.[6]

A number of theoretical and experimental studies indicated that the context effects can be non-negligible. Theoretical investigations included statistical analyses of structural databases, which assume that the effects of nonlocal interactions can be averaged out over a large number of structures if only residues in coil regions are analyzed, so that conformation distributions in structure databases can reflect intrinsic preferences.[7,8] Erman and co-workers analyzed pair wise correlations of backbone torsional angles in native protein structures and found that such correlations favor the choices of native torsional angles.[9,10] Betancourt and Skolnick also showed that dihedral angle distribution for a residue can significantly depend on its sequence and conformation contexts.[11] Sosnick and co-workers performed detailed analyses of backbone conformations in a protein coil library and observed not only preferences for the polyproline II (PII) conformation but also strong influences of both the chemical identity and the conformation of neighboring residues.[12] Experimental data, especially the NMR $J^3_{\alpha HN}$ coupling constants for denatured proteins and short unstructured peptides, have shown good correlations with predictions based on database models.[13,14]

These context dependences may result from steric interactions. Srinivasan and Rose employed Monte Carlo simulations to sample conformations of unfolded peptide chains.[15] They considered only steric interactions and attractive interactions separated by no more than five residues in sequence and observed sequence-specific biases significantly anticipating native secondary structures in the sampling. Pappu et al. found that the conformation of a central residue enveloped by two residues in the helical conformation was sterically restricted by these neighbors.[16] Penkett et al. pointed out that a preceding residue with a β-branched or aromatic side chain could promote the less sterically restricted β conformation of its successive residue.[14] The coupling between solvation and electrostatic interactions has also been proposed to promote context-dependent local conformation propensities. Avbelj and Baldwin surveyed a coil library for the effects of a large or small neighboring residue on the backbone angle φ and identified correlations between neighbor effects and electrostatic free energies.[17]

Small peptides have often been employed to investigate local conformation propensities to avoid complications associated with large protein molecules. Avbelj et al. reported that the intrinsic backbone preferences are present in blocked amino acids.[18] Stenner and co-workers have analyzed structures of XA and AX dipeptides and AXA tripeptides by a range of spectroscopy techniques.[3,4] They also obtained conformational propensities in line with results from host−guest experiments on longer peptides, although they did not observe context effects in the systems they studied. Kallenbach and co-workers have employed temperature dependent far-UV CD spectra and NMR *J*-coupling constants to investigate the neighbor effects on the PII conformation in alaline peptides.[19] They found increased PII-β transitions and *J* values corresponding to more negative φ angles in alanine flanked by isoleucine residues. On the simulation side, Zaman et al. investigated conformational equilibriums in short alanine peptide models, using seven different atomic level force-fields with implicit solvent and Langevin dynamics.[20] They observed that although different force fields yield large variations in the local structure propensities, a residue's conformation was invariably sensitive to the side-chain identity and backbone conformation of its immediate sequence neighbors.

Molecular dynamics (MD) simulations with explicit solvation can be a direct way to investigate the effects of sequence and conformational contexts on backbone conformations. This may not only provide more fundamental insights into when and how such context dependences take effects but also add highly quantitative criteria for the refining and testing of biomolecular force fields. However, existing evidence suggested that context effects are generally subtle, and to quantify them would require conformational equilibriums to be estimated with much lower statistical noises achievable by standard explicit solvent simulations.

Simulations based on generalized canonical ensembles can be employed to reduce statistical errors caused by rare transitions between different local minima on complicated conformational free energy surfaces. Those methods included multicanonical algorithm,[21] simulated tempering,[22,23] and replica exchange method[24–27] (also called replica Monte Carlo,[24] multiple Markov chain,[25] or parallel tempering[26,27]). In particular, using temperature as an extended degree of freedom, temperature replica exchange molecular dynamics (TREMD) have been widely applied to avoid sampling inefficiency in peptide and protein simulations. TREMD, however, meets with difficulty for explicitly solvated systems, because to maintain reasonable exchange acceptance ratios, the temperature difference between neighboring replicas need to be very small and the number of replicas need to be large.[28]Another problem is that although high temperature enhances barrier crossing, it may shift the equilibrium of two states to favor the larger entropy one making the transition less reversible.[29]

A number of alternatives to TREMD have been investigated.[30–33] One approach to avoid the problem of TREMD is Hamiltonian replica exchange (HREMD) in which a series of biasing potentials are used instead of temperatures as the extra degree of freedom. The replica exchange umbrella sampling (REUS) presented by Sugita et al. is one kind of multidimensional replica exchange method (MREM).[34] Fukunishi et al. have proposed scaled hydrophobicity REM and phantom chain REM when studying protein structure prediction problem.[28] Other HREMD schemes include that by Jang et al.[35] who proposed a generalized effective potential to change the effective temperature of the system by modifying the torsional and nonbonded terms of the potential energy function, by Affentranger et al.[36] who proposed to variably scale interactions within the protein and between protein and solvent atoms in the generalized ensemble and leave solvent−solvent interactions unchanged,

and by Kwak et al.[37] who proposed to scale partial local conformation energy.

A critical issue for properly extending a canonical ensemble into a generalized ensemble is to maintain the physical relevance of conformations sampled by the different replicas, and this can be a major difficulty with nondiscriminative, general scaling of certain interactions in HREMD. On the other hand, umbrella sampling is another commonly used technique to overcome sampling inefficiency. An umbrella potential can be designed to compensate the free energy barriers along specific degrees of freedom (DOF) or reaction coordinates (RC), while maintaining the physical relevance of sampling along DOFs perpendicular to the RCs. Equilibrated sampling with umbrella potentials applied on a single or few RCs can still be difficult to achieve if major free energy barriers exist along the remaining DOFs. If a set of coordinates causing major difficulties for sampling could be identified, coordinate-specific modification of the Hamiltonian by umbrella potentials can be integrated with the generalized canonical ensemble scheme so that free energy barriers along a set of coordinates can be simultaneously demolished and truly converged results can be obtained. One example is given by a recent work of Kannan and Zacharias, who proposed a HREMD scheme in which a series of five specifically parametrized potentials on the backbone (φ,ψ) angles were applied to different replicas. Their simulations of peptide systems showed efficiency of this approach.[38]

In this study, we describe a HREMD approach we developed independently which also employed biasing potentials on backbone conformations. The biasing potentials are based on the (φ,ψ) free energy surface of the alanine dipeptide and a simple functional form for boosted sampling proposed by Hamelberga et al.[39] These potentials compensate the free energy barriers to varying extents on the φ-ψ surface in different replicas, allowing the simultaneous sufficient sampling of different (φ,ψ) pairs in a peptides without overextending the sampling into irrelevant parts of the (φ,ψ) spaces. An extension of the weighted histogram analysis method (WHAM) can be employed so that all replicas can be used to compute conformational free energy surfaces. We show that the resulting low statistical uncertainties enable meaningful extraction of residue identity and neighbor effects from simulations of model peptides with explicitly solvent. We will consider such effects on conformational equilibriums in several model systems containing 2−5 peptide units composed of Ala, Val, or Phe. Val has been chosen as the simplest representative for residues with β-branched side chains and Phe for residues with aromatic side chains.[12]We note that systematic deviations of biomolecular force fields from quantum mechanical/molecular mechanical calculations and database statistics for the descriptions of conformational equilibriums in the Ramachandran space have been revealed.[40] Reoptimizing the backbone torsional angle terms in several force fields have led to significantly improved models.[40–43] Because we have not tried to correct such potential systematic deviations in the force field we used, we will not consider comparing the absolute probabilities of different backbone conformers from our simulations with experiments or database statistics. Instead, we will only consider comparisons of the shifts in these probabilities upon changes of side-chain identities or sequence/conformational contexts. These shifts are mainly results of intra- and intermolecular nonbonded interactions and should not be influenced by refining the backbone torsional angle terms in the force field.

Materials and Methods

1. Biasing Potentials for HREMD Simulations of Peptides. For each replica i, a biasing term is applied to each pair of backbone (φ,Ψ) angles, resulting in a total potential energy function of the form

$$V^i_{total} = V_{system} + V^i_{bias}, i = 1, ..., n_r \quad (1)$$

Here V_{system} is the unbiased potential energy function, n_r is the total number of replicas, and

$$V^i_{bias} = \sum_{m=1}^{n_m} \Delta V(\phi_m, \psi_m, \alpha_i) \quad (2)$$

Here the summation is over the total number of n_m backbone ($\varphi\Psi$) pairs.

Different from the work of Kannan and Zacharias who used five fixed biasing potentials for their HREMD simulations,[38] we use a functional form proposed by Hamelberga et al.[39] and the free energy surface $\Delta F_{ref}(\varphi,\psi)$ of alanine dipeptide to design the biasing potential ΔV

$$\Delta V(\phi, \psi, \alpha_i) = \begin{cases} 0, \text{ if} \Delta F_{ref}(\phi, \psi) \geq E_{cut} \\ \dfrac{(E_{cut} - \Delta F_{ref}(\phi, \psi))^2}{\alpha_i + (E_{cut} - \Delta F_{ref}(\phi, \psi))}, \text{ otherwise} \end{cases} \quad (3)$$

In the above formula, the strength of the biasing term depends monotonically on the replica-specific parameter α_i, with $\Delta V \equiv 0$ for $\alpha_i \equiv \infty$ and $\Delta V = E_{cut} - \Delta F_{ref}(\varphi,\psi)$ for $\alpha_i = 0$.

The reason for designing the biasing terms based on $\Delta F_{ref}(\varphi,\psi)$ is that to achieve optimum effects on sampling efficiency, it is preferred that ΔV compensates the free energy barriers separating the major stable conformational states. The same biasing potentials, however, should not result in overextended sampling into irrelevant regions of the conformational space. Here we assume that free energy surface of a model alanine dipeptide (with solvation taken into account) would provide a good first order approximation to the free energy surfaces of (φ,ψ) surfaces in peptides. Such a biasing potential with $\alpha_i = 0$ applied to alanine dipeptide would result in a flat (φ,ψ) free energy surface for regions associated with free energies lower than E_{cut} relative to the lowest free energy minimum. With a properly chosen E_{cut} the free energy barriers separating different conformational minima can be demolished without introducing too much overly extended sampling.

2. Obtaining the Reference Free Energy Surface. The reference free energy surface is numerically represented on a grid spanning the (φ,ψ) space with a bin size of 5° × 5°. Functional values and derivatives at off-grid points have been computed using bicubical interpolation[44] as has been first

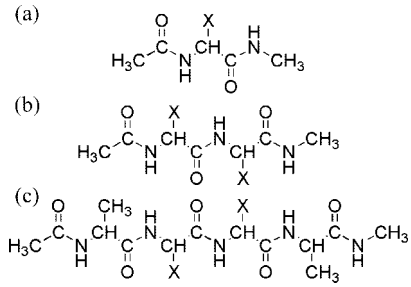


Figure 1. Peptide models containing (a) 2, (b) 3, and (c) 5 peptide units. In (a) X represents the side chain of one of alanine, phenylalanine, and valine and in (b) and (c) one of X can be alanine and the other can be one of alanine, phenylalanine, and valine.

employed in protein simulations by Mackerell et al.[41] For this purpose, a highly smooth and converged free energy surface is needed. We used a multidimensional adaptive umbrella sampling procedure[45] which makes use of WHAM[38,46] to construct this surface. Details of the dipeptide system are given in the computational details section. Repetitive calculations with altered initial configurations indicated that statistical errors can be less than 0.8 kJ/mol over the entire resulting surface and much smaller in lower free energy regions.

Based on the barrier heights on the computed $\Delta F_{ref}(\varphi,\psi)$, we choose $E_{cut} = 23$ kJ/mol. Here we have chosen the series of α, so that the barrier heights on the biased free energy surfaces decreased approximately evenly between neighboring replicas. With the number of replicas $n_r = 8$, the α_is are 10^8, 120, 50, 26.67, 15, 8, 3.33, and 0, respectively, in descending order, with the strengths of the respective biasing potentials in ascending order.

3. Applying WHAM to Trajectories of Replicas to Construct Free Energy Surfaces. Here to construct the free energy surface using a single pair of (φ,ψ) as the reaction coordinates, the effects of the biasing potentials on all (φ,ψ) pairs should also be taken into account for all replicas with a nonzero biasing potentials. In the Appendix we give the corresponding WHAM equations which allows for conformations sampled for all the replicas to be used, not only the unbiased one.

4. Peptide Systems and Simulation Details. The various peptide models considered are given in Figure 1. These are respectively X, X-Ala, Ala-X, Ala-X-Ala-Ala, and Ala-Ala-X-Ala blocked by CH_3-CO- at the N terminus and -NH-CH_3 at the C terminus, respectively. X is one of Ala, Val, or Phe. We will denote these systems by their sequence strings with "nc" attached to the left side and "cc" attached to the right side, representing the N and C terminal capping groups, respectively.

Simulations have been performed with the software package GROMOS96 with necessary modifications to implement the biasing potentials and replica exchange operations. The solute have been described by the GROMOS 53A6 force field.[47] Each peptide in its fully extended conformation ($(\varphi,\psi) = (180,180)$) was solvated in a box of SPC[48] water, size of the cubic periodic box determined by maintaining a minimum distance of 1.4 nm for any solute atom from the box boundary, with the center of mass of the solute at the box center. The weak coupling method[49] has been used to maintain a constant temperature of 300 K and a pressure of 1 atm, with relaxation times of 0.1 and 0.5 ps, respectively. All bond lengths have been constrained using SHAKE.[50] The integration time step was 2 fs. Nonbonded interactions have been treated with a twin-range cutoff method[51] with reaction field corrections for long-range interactions, using a short-range cutoff of 0.8 nm, a long-range cut off pf 1.4 nm, and a dielectric permittivity of 54 for the reaction field corrections, the pair list for short-range interactions updated every 0.02 ps.

Following initial energy minimization, each peptide system, including the dipeptide systems, has been copied into eight replicas, each with a different biasing potential applied to all backbone (φ,Ψ) pairs. Then 30.04 ns REMD simulations have been performed. The initial 40 ps has been used for equilibration and has not been used for analysis. In the HREMD simulations, every 100 steps (0.2 ps), replica exchange was attempted between neighboring replicas. At each exchange step, all the 7 pairs of neighboring replicas were considered in random order, with the constraint that a configuration can be exchanged for at most one time at a given step.

To provide comparisons with the HREMD results, one of the peptide systems, ncAAFAcc, was subjected to 8 independent normal MD simulations at 300 K. Each simulation included 10 ns sampling after 90 ps of equilibration.

Results

1. Convergences and Statistical Errors. As the effects we are interested in translated into small free energy differences, it is critical for the statistical errors produced by the HREMD process to be sufficiently small relative to the magnitudes of these effects. To estimate the statistical errors, we divided each of the 30 ns REMD trajectory sets (each set containing trajectories for all 8 replicas) into blocks of equal lengths. With block lengths of 1, 2, 5, and 10 ns, a REMD trajectory set was divided into 30, 15, 6, and 3 blocks, respectively. Each block was employed separately to construct a free energy surface. Then the averages and standard deviations of free energy surfaces computed using blocks of the same length can be computed. These standard deviations provided estimations of statistical errors and of how such errors decayed with the increasing length of the HREMD trajectories.

Among all the systems considered, ncAFAAcc and ncAAFAcc are the ones associated with the largest statistical uncertainties because of their larger numbers of backbone DOFs and the more difficult sampling of different side-chain conformations for Phe. In addition, as α conformation is the least populated among the three major local backbone conformers, the free energy surfaces for the neighboring residue in the α conformation contained the largest statistical uncertainties as they were computed with the least amount of sampled conformations. Thus to demonstrate the convergence property of the HREMD procedure, as examples for the best and worst situations, respectively, we present results for the free energy surfaces with respect to the (φ,ψ) angles of X in ncXcc and of the second Ala in the ncAAFAcc

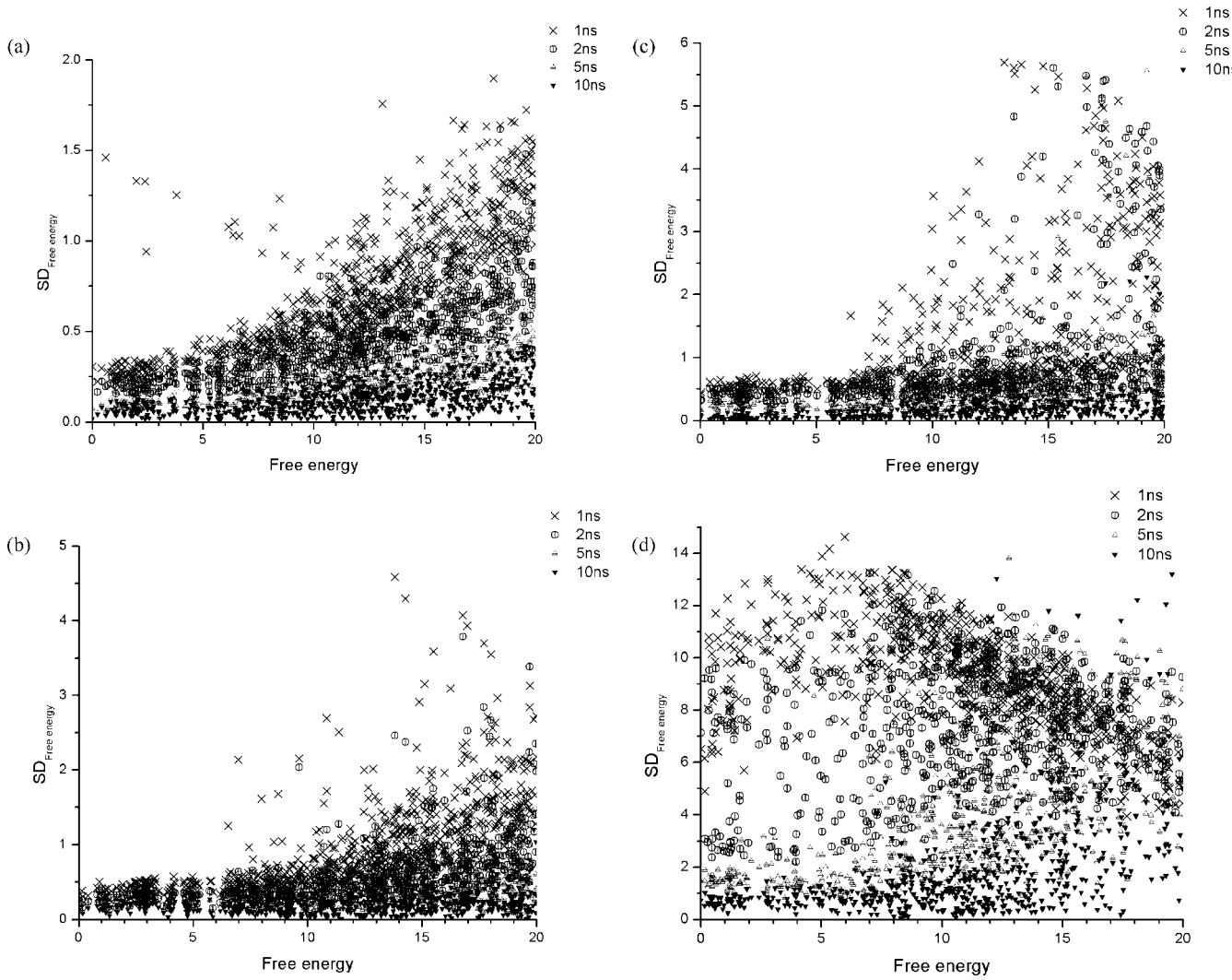


Figure 2. Standard deviations of free energy surfaces determined using different block lengths. Different points on the surfaces have been shown as functions of the free energies relative to the minimum points on the surfaces. (a)−(c) are for the ncXcc systems with X = Ala, Phe, and Val, respectively. (d) is for the second Ala in ncAAFAcc, with F in the α conformation.

system with its neighboring Phe in the α conformation. In Figure 2, we show the standard deviations of the relative free energies, computed using varying block lengths, as functions of the relative free energies themselves at different (φ,ψ) angles. The relative free energies plotted have been computed by treating the entire 30 ns trajectory set as a single block. Before used for averaging and standard deviation computing, each free energy surface computed using a smaller block has been shifted by a constant value so that the root-mean-square deviations from the free energy surface computed using the entire data set were minimized. Only the (φ,ψ) regions with free energies less than 20 kJ/mol relative to the minimum have been considered to determine the constant shifts.

Obviously, the statistical uncertainties depend not only on the block lengths but also on the relative free energies. And for our purpose only the statistical uncertainties in the lower free energy regions are of concern, as only these regions contribute to the conformational equilibriums. When the block length is increased from 1 to 10 ns, the standard deviations for regions with relative free energies below 10 kJ/mol descend from a few kJ/mol to below 0.2 kJ/mol for ncXcc and from above 10 kJ/mol to mostly below 1 kJ/mol for ncAAFAcc.

The standard deviations shown in Figure 2 have been computed for 10^o X 10^o (φ,ψ) bins. To estimate relative free energies between different conformers and the corresponding statistical uncertainties, we first transform a free energy surface into normalized probability densities, then integrated the probability densities within each basins to obtain the total probability P_c for conformer C and used the expressions $G_c = -RT\ln P_c$ to represent the free energy of conformer C. The different basins have been defined as ($-180^o<\varphi<90^o$, $60^o<\Psi<240^o$) for β, ($-90^o<\varphi<0^o$, $60^o<\Psi<240^o$) for PII, and ($-180^o<\varphi<0^o$, $-120^o<\Psi<60^o$) for α. As the free energies at the boundaries separating different basins are usually much higher than those close to the minima, the exact choices of these boundaries have negligible influences on the final results, except for the φ value separating β and PII, which are not well separated minima on the dipeptide surfaces. We used the same boundary φ values for different systems for the results to be comparable between each other. The averaged values and standard deviations of G_c have been computed using the

Table 1. Averaged Values and Standard Deviations of Conformational Free Energies (in kJ/mol) for the Second Ala in ncAAFAcc with the Neighboring Phe in Different Conformations, Determined with Different Block Lengths

	block length	1 ns	2 ns	5 ns	10 ns
Phe in	G_β	1.54 ± 0.32	1.54 ± 0.20	1.54 ± 0.15	1.53 ± 0.03
	G_{PII}	3.01 ± 0.44	2.98 ± 0.32	2.96 ± 0.17	2.96 ± 0.12
	G_α	6.12 ± 2.37	5.62 ± 1.34	5.45 ± 0.85	5.36 ± 0.48
Phe in PII	G_β	1.65 ± 0.36	1.68 ± 0.25	1.65 ± 0.18	1.67 ± 0.08
	G_{PII}	2.68 ± 0.48	2.65 ± 0.35	2.64 ± 0.13	2.63 ± 0.02
	G_α	6.75 ± 2.79	6.05 ± 1.89	5.74 ± 0.89	5.55 ± 0.48
Phe in α	G_β	1.84 ± 1.22	1.72 ± 0.62	1.70 ± 0.45	1.62 ± 0.07
	G_{PII}	3.30 ± 1.39	3.10 ± 0.85	2.90 ± 0.41	2.80 ± 0.23
	G_α	6.92 ± 4.38	5.92 ± 2.32	5.70 ± 1.77	5.55 ± 1.02

trajectory partitioning scheme described above. As expected, the statistical uncertainties in G_cs are much smaller than in the free energy surfaces.

Table 1 shows the averaged values and statistical uncertainties for G_c of the second Ala in ncAAFAcc with its neighboring Phe in different backbone conformations, computed using different block lengths. As expected, the statistical uncertainties of G_c are directly correlated with the number of sampled configurations with conformer C. At a given block length, the largest uncertainties are associated with G_αs because of its higher values or less sampling (0.48 to 1.02 kJ/mol for a block length of 10 ns), especially with the neighboring Phe also in the α conformation (1.02 kJ/mol). With a block length of 10 ns, the statistical uncertainties of other G_cs are much smaller (0.02−0.2 kJ/mol). In later sections, we will report the standard deviations computed using a 10 ns block length as the estimated statistical uncertainties of the computed G_cs.

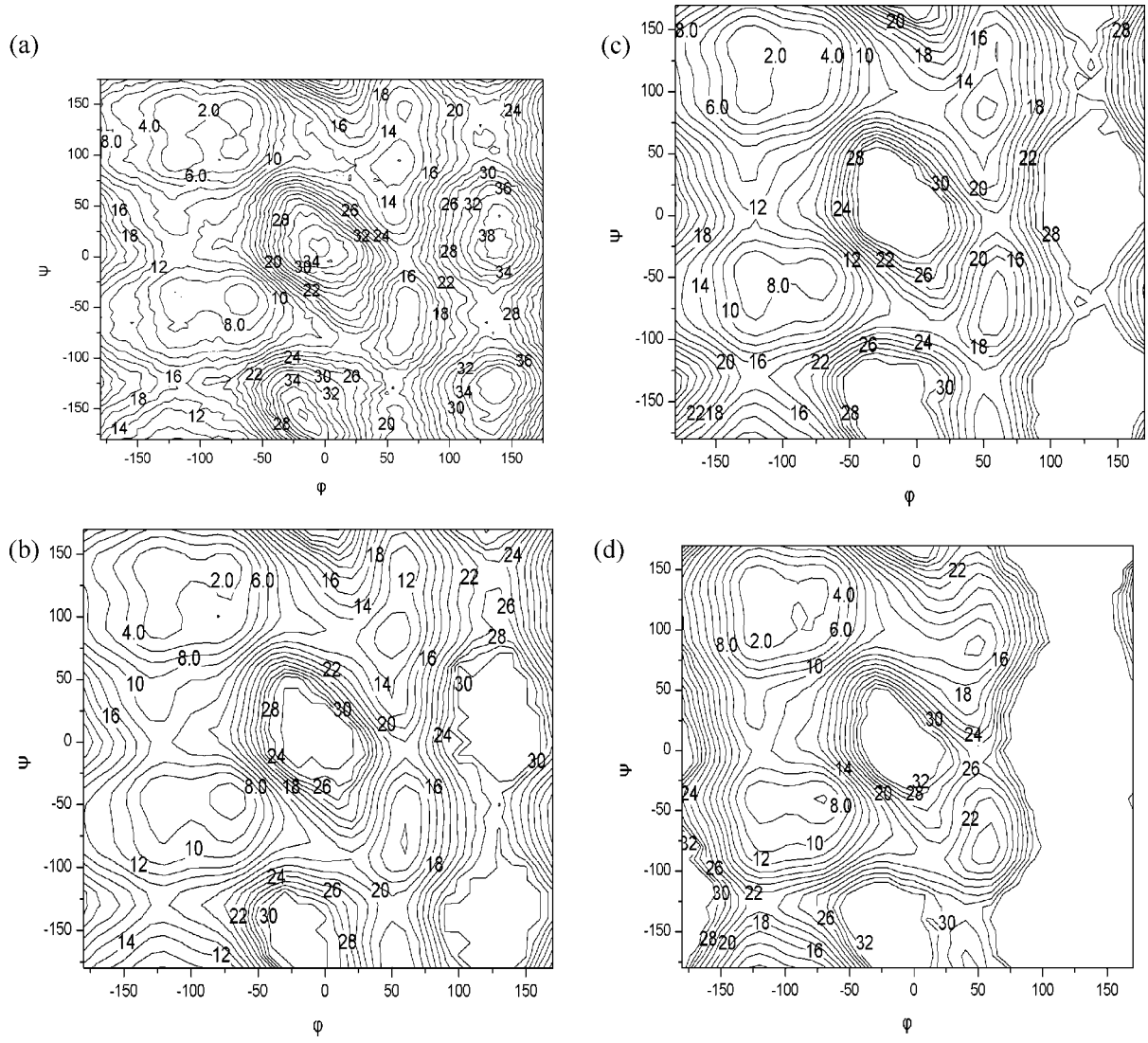


Figure 3. Contour maps of the (φ, Ψ) free energy surfaces in different model dipeptides. The numerical labels are in kJ/mol. (a) is for ncAcc computed using adaptive umbrella sampling. (b)−(d) are for ncXcc with X = Ala, Phe, and Val, respectively, computed using HREMD and the extended WHAM formulations.

Table 2. Free Energies of Different Conformers and Their Differences (in kJ/mol), To Display the Effects of Varying the Side-Chain Identity between Ala, Phe, and Val in Different Model Peptides[a]

peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncA*cc	1.40 ± 0.01	2.95 ± 0.02	5.74 ± 0.08	1.56 ± 0.02	4.34 ± 0.09
ncF*cc	1.00 ± 0.01	3.70 ± 0.08	6.11 ± 0.20	2.71 ± 0.09	5.12 ± 0.19
ncV*cc	1.27 ± 0.01	2.84 ± 0.01	6.47 ± 0.10	1.57 ± 0.01	5.20 ± 0.11
ncA*Acc	1.57 ± 0.02	2.68 ± 0.02	5.70 ± 0.10	1.11 ± 0.03	4.13 ± 0.11
ncF*Acc	1.14 ± 0.03	3.05 ± 0.04	7.10 ± 0.41	1.91 ± 0.05	5.96 ± 0.43
ncV*Acc	1.40 ± 0.06	2.56 ± 0.02	6.77 ± 0.40	1.16 ± 0.08	5.37 ± 0.46
ncAA*cc	1.46 ± 0.01	2.83 ± 0.06	5.78 ± 0.21	1.38 ± 0.05	4.32 ± 0.22
ncAF*cc	0.90 ± 0.03	4.30 ± 0.08	5.56 ± 0.03	3.41 ± 0.10	4.66 ± 0.05
ncAV*cc	1.24 ± 0.05	2.96 ± 0.04	6.29 ± 0.46	1.71 ± 0.06	5.05 ± 0.51
ncAA*AAcc	1.75 ± 0.06	2.67 ± 0.07	5.01 ± 0.07	0.92 ± 0.13	3.26 ± 0.09
ncAF*AAcc	1.16 ± 0.11	3.42 ± 0.15	5.92 ± 0.90	2.26 ± 0.10	4.76 ± 1.00
ncAV*AAcc	1.46 ± 0.06	2.77 ± 0.14	5.71 ± 0.72	1.31 ± 0.11	4.25 ± 0.77
ncAAA*Acc	1.69 ± 0.08	2.82 ± 0.02	4.94 ± 0.21	1.12 ± 0.09	3.25 ± 0.29
ncAAF*Acc	1.00 ± 0.03	3.52 ± 0.02	6.43 ± 0.20	2.51 ± 0.02	5.43 ± 0.23
ncAAV*Acc	1.27 ± 0.01	2.83 ± 0.01	6.59 ± 0.18	1.57 ± 0.02	5.32 ± 0.18

[a] The residues for which the free energies of their backbone conformations are shown have been marked with the superscript *.

The statistical uncertainties contained in the eight 10 ns normal MD simulations of ncAAFAcc were compared with those in the HREMD results. As expected, samplings by normal MD suffered from insufficient statistics because of the rare transitions between different conformers, and estimations of the relative free energies with statistical uncertainties comparable to that obtained by the HREMD approach cannot be obtained from 8 × 10 ns simulations. Considering the conformation of the second Ala without differentiating conformers of the neighboring Phe, the proportions of the α conformer varied between 2.0% and 34% in different simulations. For the β and PII conformers the values varied between 34.3% and 60.9% and between 22.1% and 33.5%, respectively. Such large variations in the computed conformational probabilities themselves would not allow meaningful extraction of the subtle relative effects of changes in the side-chain identity or in the sequence/conformation contexts.

2. Effects of Side-Chain Identities. Figure 3 shows the free energy surfaces for the model ncXcc dipeptides. Under the given set of force field parameters, all three free energy surfaces show similar qualitative features, with basins centered around (-65^o,-40^o), (-120^o,145^o), and (-70^o,-145^o) corresponding to the α, β, and PII conformers, respectively. Differences associated with different side-chain types are visible. For ncAcc, the results obtained using adaptive umbrella sampling and using HREMD are essentially the same, with the HREMD surface much smoother, testifying to the validity of the extended WHAM formulations given in the Appendix. Table 2 lists the free energies of different conformers associated with a given residue type X in different model peptides and at different positions. We note that the development of the GROMOS 53A6 force field parameters has so far mainly focused on the nonbonded interaction parameters which have been optimized to reproduce condensed phase thermodynamics data of pure and mixtures of small molecules in the liquid phase, and bonded interactions, especially the torsional angle terms associated with peptide backbone conformations have not been changed with respect to the older version. The main purpose of the current study is to look at how various factors such as side-chain identities as well as identities and conformations of neighboring residues would affect the conformational equilibriums. Such effects can be reflected by how the free energy differences between different conformers (ΔG_cs such as G_α-G_β and G_{PII}-G_β) changes across different systems. Unlike the differences themselves, these changes do not depend on the analytical torsional angles terms in the force field.

Table 2 shows that although the effects of side-chain identities on the ΔG_c values (below 1.2 kJ/mol) are quite small, many of them are still significantly larger than the corresponding statistical uncertainties (mostly 0.01−0.2 kJ/mol) and indicate definite and consistent effects of changing side-chain identities. In ncXcc, both Phe and Val have G_α-G_β values larger (by ca. 0.8 kJ/mol) than Ala. Interestingly, the G_{PII}-G_β values for Phe are significantly larger (by ca. 1.1 kJ/mol) than for Val and Ala, indicating that Phe does not prefer PII, at least in such small peptides. The effects of side-chain identities are consistently reproduced in longer peptides if we compare the G_α-G_β and G_{PII}-G_β values between Ala, Phe, and Val at the same position in a peptide of the same length (see Table 2).

It is also interesting to compare the ΔG_cs of the same residue type at different positions and in different peptide systems. For the second and third Ala in the pentapeptide model ncAAAAcc, the G_α-G_β values are ca. 3.2 kJ/mol, more than 1 kJ/mol lower than the corresponding differences in the shorter peptides ncAcc and ncAAcc. For Phe and Val, going from the shorter to the longer peptides does not result in such large changes in G_α-G_β. These combined produced much larger increases in G_α-G_β upon changing the central residues from Ala to Phe or Val in the model pentapeptides than in the shorter peptides. Effects of changing peptide length on G_{PII}-G_β are smaller.

In summary, our HREMD simulations show that Phe and Val prefer β over α relative to Ala. In addition, Phe does not prefer the PII conformation relative to β as compared with either Val or Ala. For Ala, the preference for α relative to β increases significantly when the peptide is extended into pentapeptides. For Phe and Val, such peptide length dependences are not strong.

3. N-Terminal Side Neighbor Effects. Tables 3 and 4 list the free energies G_c of different conformers of the second

Table 3. Effects of Chemical Identity and Conformation (Noted by Subscripts) of the N-Terminal Side Neighbor on Conformational Equilibriums of the Second Ala (Marked with the Superscript *) in ncXAcc Systems[a]

peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncA_βA*cc	1.49 ± 0.02	2.77 ± 0.05	5.76 ± 0.24	1.28 ± 0.04	4.27 ± 0.26
ncA_{PII}A*cc	1.42 ± 0.02	2.88 ± 0.09	5.80 ± 0.16	1.46 ± 0.10	4.38 ± 0.16
ncA_αA*cc	1.26 ± 0.04	3.05 ± 0.11	6.16 ± 0.23	1.80 ± 0.14	4.90 ± 0.23
ncF_βA*cc	1.49 ± 0.03	3.08 ± 0.09	4.88 ± 0.20	1.60 ± 0.12	3.39 ± 0.19
ncF_{PII}A*cc	1.43 ± 0.02	3.28 ± 0.07	4.66 ± 0.11	1.84 ± 0.09	3.22 ± 0.08
ncF_αA*cc	1.14 ± 0.04	3.19 ± 0.04	6.77 ± 0.34	2.05 ± 0.04	5.63 ± 0.39
ncV_βA*cc	1.69 ± 0.04	2.49 ± 0.04	5.61 ± 0.15	0.80 ± 0.08	3.92 ± 0.19
ncV_{PII}A*cc	1.63 ± 0.02	2.52 ± 0.01	5.74 ± 0.17	0.89 ± 0.01	4.12 ± 0.18
ncV_αA*cc	1.17 ± 0.05	3.08 ± 0.07	6.66 ± 0.20	1.91 ± 0.11	5.49 ± 0.25

[a] Free energies and differences in kJ/mol are given.

Table 4. As in Table *3* but for the ncAXAAcc Systems

peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncAA_βA*Acc	1.76 ± 0.11	2.72 ± 0.04	4.89 ± 0.26	0.95 ± 0.13	3.13 ± 0.36
ncAA_{PII}A*Acc	1.74 ± 0.09	2.83 ± 0.06	4.84 ± 0.33	1.08 ± 0.02	3.10 ± 0.42
ncAA_αA*Acc	1.37 ± 0.04	3.08 ± 0.12	5.27 ± 0.34	1.71 ± 0.12	3.90 ± 0.37
ncAF_βA*Acc	1.69 ± 0.13	2.75 ± 0.09	4.85 ± 0.26	1.06 ± 0.22	3.15 ± 0.38
ncAF_{PII}A*Acc	1.55 ± 0.05	2.98 ± 0.07	4.95 ± 0.24	1.44 ± 0.06	3.40 ± 0.27
ncAF_αA*Acc	1.28 ± 0.14	3.12 ± 0.04	6.20 ± 1.27	1.84 ± 0.10	4.92 ± 1.41
ncAV_βA*Acc	1.97 ± 0.05	2.53 ± 0.09	4.60 ± 0.17	0.56 ± 0.14	2.63 ± 0.18
ncAV_{PII}A*Acc	1.96 ± 0.04	2.53 ± 0.12	4.81 ± 0.34	0.57 ± 0.13	2.84 ± 0.36
ncAV_αA*Acc	1.30 ± 0.15	3.21 ± 0.23	5.50 ± 0.33	1.91 ± 0.36	4.20 ± 0.46

Table 5. Effects of Chemical Identity and Conformation (Noted by Subscripts) of the C-Terminal Side Neighbor on Conformational Equilibriums of the First Ala (Marked with the Superscript *) in ncAXcc Systems[a]

Peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncA^*A_βcc	1.61 ± 0.03	2.66 ± 0.03	5.58 ± 0.08	1.05 ± 0.06	3.97 ± 0.11
ncA^*A_{PII}cc	1.48 ± 0.02	2.70 ± 0.03	6.07 ± 0.18	1.22 ± 0.04	4.59 ± 0.20
ncA^*A_αcc	1.57 ± 0.06	2.60 ± 0.06	5.96 ± 0.22	1.03 ± 0.11	4.39 ± 0.26
ncA^*F_βcc	1.37 ± 0.02	2.96 ± 0.09	5.98 ± 0.33	1.58 ± 0.09	4.61 ± 0.34
ncA^*F_{PII}cc	1.58 ± 0.06	2.59 ± 0.05	5.99 ± 0.15	1.00 ± 0.11	4.40 ± 0.20
ncA^*F_αcc	1.36 ± 0.06	2.75 ± 0.03	6.68 ± 0.42	1.39 ± 0.08	5.33 ± 0.48
ncA^*V_βcc	1.98 ± 0.03	2.09 ± 0.05	6.11 ± 0.27	0.11 ± 0.05	4.13 ± 0.28
ncA^*V_{PII}cc	1.85 ± 0.04	2.14 ± 0.03	6.51 ± 0.24	0.29 ± 0.07	4.66 ± 0.25
ncA^*V_αcc	1.91 ± 0.11	2.00 ± 0.02	6.87 ± 0.64	0.09 ± 0.12	4.97 ± 0.75

[a] Free energies and differences in kJ/mol are given.

Ala in the ncXAcc and the third Ala in the ncAXAAcc systems. Results for X in different conformations have been listed separately.

Interestingly, the conformations of X have much larger and better-defined effects on the conformational equilibrium of its next Ala than the identity of X. Compared with X in the β conformation, X in the α conformation results in stabilization of β of the next Ala relative to α by 0.6−0.7 kJ/mol when X is Ala, and 1.2−1.8 kJ/mol when X is Phe or Val, and relative to PII by 0.5−0.7 kJ/mol when X is Ala or Phe and 1.1−1.3 kJ/mol when X is Val. Also compared with X in the β conformation, X in PII conformation does not result in much change in the conformational equilibrium of the next Ala, with the computed ΔG_cs less than 0.4 kJ/mol of varying signs.

The effects of the identity of X are less certain, and the computed changes are dependent on both the lengths of the model peptides and the conformations of X. Although the data show the trend that changing X in the α conformation from Ala to Phe or Val would further stabilize the β and PII conformers relative to the α conformer of the next Ala, the computed differences are small, and some of them suffer from large statistical uncertainties.

4. C-Terminal Side Neighbor Effects. Tables 5 and 6 list the free energies G_c of different conformers of the first Ala in the ncAXcc and the second Ala in the ncAAXAcc systems.

Unlike the N-terminal side neighbor effects, the chemical identity of X as the C-terminal neighbor produced consistent effects on the conformational equilibrium on its previous Ala. The G_α-G_β values for Ala with Phe or Val as its next residue in either α or β conformation are consistently larger than with Ala as its next residue, mostly by 0.5−1.0 kJ/mol except for ncAVcc. X in the PII conformation does not have much identity-specific effects on the α-β equilibrium of its previous Ala. Another consistent observation is that Val can significantly stabilize the PII conformer of its previous Ala as compared with Ala or Phe, reducing G_{PII}-G_β values from 0.8 to 1.5 kJ/mol to 0.0 to 0.3 kJ/mol.

As compared with the conformation effects of the N-terminal side neighbors, the conformation effects of X on its previous Ala are less obvious and less consistent across different residue types and peptide lengths. There seems to be the trend that X in either PII or α would slightly stabilize the β and PII conformers relative to α of the previous Ala.

Table 6. As Table 5*5*, but for the ncAAXAcc Systems

peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncAA*A_βAcc	1.82 ± 0.08	2.68 ± 0.09	4.71 ± 0.04	0.86 ± 0.17	2.89 ± 0.10
ncAA*A_{PII}Acc	1.65 ± 0.07	2.65 ± 0.08	5.33 ± 0.07	1.01 ± 0.15	3.68 ± 0.14
ncAA*A_αAcc	1.76 ± 0.03	2.56 ± 0.18	5.35 ± 0.57	0.80 ± 0.17	3.59 ± 0.59
ncAA*F_βAcc	1.53 ± 0.02	2.96 ± 0.12	5.36 ± 0.48	1.44 ± 0.10	3.84 ± 0.50
ncAA*F_{PII}Acc	1.67 ± 0.08	2.63 ± 0.02	5.55 ± 0.48	0.96 ± 0.06	3.88 ± 0.56
ncAA*F_αAcc	1.62 ± 0.07	2.80 ± 0.23	5.55 ± 1.02	1.19 ± 0.19	3.93 ± 1.09
ncAA*V_βAcc	2.07 ± 0.08	2.19 ± 0.01	5.84 ± 0.22	0.12 ± 0.09	3.76 ± 0.30
ncAA*V_{PII}Acc	2.14 ± 0.17	2.11 ± 0.03	5.80 ± 0.58	−0.04 ± 0.15	3.65 ± 0.75
ncAA*V_αAcc	2.15 ± 0.15	1.93 ± 0.16	6.71 ± 0.08	−0.22 ± 0.31	4.56 ± 0.11

5. Correlations with Experiments and Database Statistics. The effects of side-chain identity on conformation preferences in short peptides and in coil libraries have been well documented. The increased β propensity of Phe and Val with respect to α and PII relative to Ala are consistent with experimental studies and database statistics such as refs 3, 12, and 18. The decreased PII preferences of Phe indicated by our simulations are not displayed in reported statistical analyses of databases (for example, ref 12). It is difficult to verify this particular result with reported experimental data because of the dominance of the β conformer for Phe in short peptides[3] and also because the NMR *J*-coupling constant cannot differentiate between the PII and α conformers. Whether this would indicate limitations of the force field or of the statistical analyses may deserve further investigations.

The neighbor effects from our HREMD analysis may not only be correlated with previous experimental and statistical results but also lead to additional insights into them. For the identity changing of the N-terminal neighbor between Ala, Phe, and Val, our HREMD results indicated that conformation-specific neighbor effects dominate. The N-terminal neighbor in α significantly stabilize the β conformer of the next Ala. We found this in excellent agreement with the statistics presented as Figure 6 in ref 12, which showed separately the (φ, Ψ) distributions of Ala following residues of different types and in different conformations in a restricted coil library. We found that as the C-terminal side neighbor, Val may significantly stabilize the PII conformer of its previous Ala. We also found this in consistence with Figure 7b of ref 12, which showed clearly that Ala right before residues with β-branched side chains in a restricted coil library have higher fractions in the PII conformations than those before residues with aromatic or alanine-like side chains. It is interesting to note that for this particular result, both our HREMD simulations and the database analyses showed very low uncertainties.

However, there are also some effects implicated by our results but not displayed in the coil library analyses and vice versa. Our simulations indicated that relative to Ala as the C-terminal neighbor, Val or Phe as the C-terminal neighbor in the α or β conformation would stabilize the β conformer with respect to α of its previous Ala, but this was not indicated by Figure 7b of ref 12. On the other hand this same figure showed that alanine residues with an aromatic side C-terminal neighbor would have quite different conformation preferences with the aromatic residue in different conformations, especially that when the aromatic residue is in the PII conformation, the fraction of the α conformers of the Ala drops significantly. While our results did implicate some stabilization effects of the PII conformer of Ala with the next Phe in PII, the G_α-G_β of the Ala was essentially unchanged compared with the case with Phe in the β conformation. It is yet difficult to judge whether these differences should be attributed to the limitations of either analyses (for examples, crude classifications of residues by their chemical identities and insufficient amount of data in the statistical analyses or inaccuracies of the force field) or to different environmental contexts of the model peptides and the coil library.

Conclusions

We have presented a Hamiltonian replica exchange approach and applied it to investigate the conformational equilibriums of short peptides. A general form for constructing biasing potentials of varying strengths in different replicas has been employed. For peptide systems, the biasing potentials are based on a reference free energy surface of model alanine dipeptide and applied to all backbone (φ,Ψ) pairs to overcome major sampling barriers. Free energy surfaces using one of the biased pairs as the reaction coordinates can be constructed using conformations sampled in all replicas, including the biased ones, using an extension of the WHAM formulations. This HREMD approach can significantly reduce the statistical uncertainties in computed conformational free energies, thus allowing the subtle but important effects of factors such as side-chain identities and conformations of neighboring residues to be quantified using explicit solvent simulations.

Using the GROMOS 53A6 force field, we apply this HREMD approach to investigate the effects of these factors in model peptides containing two to five peptide units. In this work, we have considered Ala, Phe, and Val as simplest representatives of non β-branched, aromatic, and β-branched side chains, respectively. As for side-chain identity effects, our results consistently showed an increased preference of β by Phe and Val relative to Ala, in agreement with other analyses. Our results not only indicate that neighbor effects can be as large as effects of the side-chain identity but also lead to additional insights into such effects. For N-terminal neighbors, the effects of their conformations seem to be larger than those of their chemical identities. Any of Ala, Phe, and Val as the N-terminal neighbor of an Ala significantly stabilizes the β conformer of the Ala when the neighbor is in the α conformation. In the case of C-terminal neighbors, their chemical identities seem to play more important roles

in modifying the conformational preferences of its previous Ala. We found that Val may significantly increase the PII propensity of its previous Ala. Beside these results which are in accordance with reported statistics of protein coil structure libraries, our simulations also indicated that the PII conformer of Phe is less preferred compared with Ala and Val and that Phe and Val as the C-terminal neighbor in either α or β conformation would stabilize the β conformer of its previous Ala.

Previous simulations have shown that different force fields did not produce consistent results for backbone conformational equilibriums of the same system. One important reason for these is that the analytical backbone torsional angle terms, a major determinant of the backbone conformational equilibriums, have been determined using different reference data in different force fields. We have focused not on the free energy differences between different conformers themselves but on the changes in these free energy differences associated with the changes of side-chain identity of the same residue and of side-chain identity and backbone conformation of neighboring residues. These changes do not depend on the analytical backbone torsional angle terms but mainly result from complicated changes in intra- and intermolecular nonbonded interactions. Parameters for such interactions in different biomolecular force fields have been shown to reproduce a more convergent set of reference data (for example, free energies of transfer between different solvents). The HREMD approach is essential for meaningful analyses of these small changes because of sufficiently lowered statistical uncertainties in the computed conformational free energies. Given the reasonable agreements between the current simulations and other analyses, we expect that accurate computations of conformational free energies using explicit solvent simulations can help not only to delineate the relationships between local sequences and conformations but also to identify their physical origins through future well-designed computer experiments. From a computational chemistry point of view, such simulations may also be employed to test and refine a given force field model, to compare different models, and to investigate when and how the inclusion of higher order effects (for example, polarization) in force fields would be important for the description of conformational equilibriums.

Acknowledgment. Financial support from the Chinese Natural Science Foundation (grant number 30670485) and from the Chinese Ministry of Science and Technology (grant number 2006AA02Z303) are acknowledged.

Appendix

In this Appendix we present the WHAM equations[38,46] for the estimation of a probability density for a reaction coordinate using HREMD. In different replicas biasing potentials have been applied to DOFs other than the reaction coordinate. We can first consider the unbiased probability density for a certain (φ_t,ψ_t) pair from a single simulated replica j

$$P_{unbiased}(\phi_t, \psi_t) = P_{unbiased}(Q) = \frac{Z^j_{biased}}{Z_{unbiased}} P^j_{biased}(Q)\Lambda_{j,Q} \tag{A1}$$

with

$$\Lambda_{j,Q} = <\exp(\beta V^j_{bias})>_{biased,j,Q} \tag{A2}$$

Here Q refers to the reaction coordinate (φ_t,ψ_t), $P^{unbiased}(Q)$ is the unbiased probability density, $Pbiased^j(Q)$ is the probability density of replica j with biasing potential V^j_{bias}, and $<...>_{biased,j,Q}$ refers to average over members of subensemble j with (φ_t,ψ_t) at Q. $Zbiased^j$ and $Z_{unbiased}$ are the configurational partition functions of the biased and original system, respectively

$$Z_{unbiased} = \int \exp(-\beta V_{system}) d\Omega \tag{A3}$$

and

$$Z^j_{biased} = \int \exp[-\beta(V_{system} + V^j_{bias})] d\Omega \tag{A4}$$

The $P^{unbiased}(Q)$ can be estimated using each of the individual replicas based on eq A1. We denote an estimation using replica j as $\tilde{P}unbiased^j$. Here we use $\tilde{X}$ to represent estimation for X from the REMD samples. The estimation considering all n_r replicas is the weighted summation of individual estimations

$$\tilde{P}_{unbiased}(Q) = \sum_{j=1}^{n_r} w_j(Q)\tilde{P}^j_{unbiased}(Q) \tag{A5}$$

The weighting factors $w_j(Q)$ should be chosen to minimize the statistical uncertainty in $\tilde{P}_{unbiased}(Q)$ subject to a normalization constraint.[38,46] If we assume that the uncertainties in $\tilde{P}_{biased}{}^j(Q)$and in $\tilde{\Lambda}_{j,Q}$ contribute independently to the uncertainties in $\tilde{P}_{unbiased}{}^j(Q)$, we have the following squared standard error

$$\begin{aligned} <\tilde{P}_{unbiased}(Q)^2> - <\tilde{P}_{unbiased}(Q)>^2 &= \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right)^2 \times \\ &\left\{<\left(\tilde{P}^j_{biased}(Q)\tilde{\Lambda}_{j,Q}\right)^2> - <\tilde{P}^j_{biased}(Q)\tilde{\Lambda}_{j,Q}>^2\right\} \\ \approx \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right)^2 &\left[\Lambda_{j,Q}{}^2 \delta^2 \tilde{P}^j_{biased}(Q) + P^j_{biased}(Q)^2 \delta^2 \tilde{\Lambda}_{j,Q}\right] \end{aligned} \tag{A6}$$

As in standard WHAM the uncertainties in $\tilde{P}_{biased}{}^j(Q)$ is given by[38,46]

$$\delta^2 \tilde{P}^j_{biased}(Q) = \frac{\delta^2 \tilde{H}^j(Q)}{\Delta Q^2 M^{j2}} = \frac{<H^j(Q)>_{biased,j}}{\Delta Q^2 M^{j2}} = \frac{P^j_{biased}(Q)}{M^j \Delta Q} \tag{A7}$$

Here $H^j(Q)$is the number of independent conformations sampled with reaction coordinate Q in replica j, M^j is the total number of independent conformations sampled in replica j, and ΔQ is the bin size of the reaction coordinates.

We assume that the uncertainties in $\tilde{\Lambda}_{j,Q}$ is proportional to $\tilde{\Lambda}_{j,Q}$, the fluctuations in exp (βV^j_{bias}), and inversely the square root of number of independent samples for the averaging,

$$\delta^2\widetilde{\Lambda}_{j,Q} = \delta^2 < \exp(\beta V^j_{bias}) >_{j,Q}$$
$$= \Lambda^2_{j,Q}\delta^2 < \exp(\beta V^j_{bias} - \ln\Lambda_{j,Q}) >_{j,Q}$$
$$\approx \Lambda_{j,Q}{}^2 \frac{\delta_{\Lambda_{j,Q}}{}^2}{<H^j(Q)>}$$
$$= \Lambda_{j,Q}{}^2 \frac{\delta_{\Lambda_{j,Q}}{}^2}{P^j_{biased}(Q)M^j\Delta Q} \quad \text{(A8)}$$

with

$$\delta_{\Lambda_{j,Q}}{}^2 = < [\exp(\beta V^j_{bias} - \ln\Lambda_{j,Q})]^2 >_{biased,j,Q} - < [\exp(\beta V^j_{bias} - \ln\Lambda_{j,Q})] >^2_{biased,j,Q} \quad \text{(A9)}$$

then the overall uncertainty is approximately

$$< \widetilde{P}(Q)^2 > - < \widetilde{P}(Q) >^2 \approx \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right)^2 \times$$
$$\left[\Lambda_{j,Q}{}^2 \frac{P^j_{biased}(Q)}{M^j\Delta Q} + P^j_{biased}(Q)^2 \Lambda_{j,Q}{}^2 \frac{\delta_{\Lambda_{j,Q}}{}^2}{P^j_{biased}(Q)M^j\Delta Q}\right]$$
$$= \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right)^2 \Lambda_{j,Q}{}^2 \frac{P^j_{biased}(Q)}{M^j\Delta Q}\left(1 + \delta_{\Lambda_{j,Q}}{}^2\right)$$
$$= \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right) \Lambda_{j,Q} \frac{P_{unbiased}(Q)}{M^j\Delta Q}\left(1 + \delta_{\Lambda_{j,Q}}{}^2\right) \quad \text{(A10)}$$

Minimizing the above uncertainty with respect to $w_j(Q)$ under the constraint $\sum_j w_j(Q) = 1$ results in

$$w_j(Q) = \lambda \frac{1}{Z^j_{biased}} \frac{M^j}{\Lambda_{j,Q}\left(1 + \delta_{\Lambda_{j,Q}}{}^2\right)} \quad \text{(A11)}$$

with λ determined by the normalization constraint

$$\lambda = \frac{1}{\sum_k \frac{1}{Z^k_{biased}} \frac{M^k}{\Lambda_{k,Q}\left(1 + \delta_{\Lambda_{k,Q}}{}^2\right)}} \quad \text{(A12)}$$

Combining eqs A1, A5, A11, and A12 gives

$$\widetilde{P}_{unbiased}(Q) =$$
$$\sum_j \frac{\frac{1}{Z^j_{biased}} \frac{M^j}{\widetilde{\Lambda}_{j,Q}\left(1 + \tilde{\delta}_{\Lambda_{j,Q}}{}^2\right)}}{\sum_k \frac{1}{Z^k_{biased}} \frac{M^k}{\widetilde{\Lambda}_{k,Q}\left(1 + \tilde{\delta}_{\Lambda_{k,Q}}{}^2\right)}} \frac{Z^j_{biased}}{Z_{unbiased}} \frac{\widetilde{H}^j(Q)}{M^j\Delta Q} \widetilde{\Lambda}_{j,Q}$$
$$= \frac{1}{\Delta Q} \frac{\sum_j \frac{\widetilde{H}^j(Q)}{\left(1 + \tilde{\delta}_{\Lambda_{j,Q}}{}^2\right)}}{\sum_k \frac{Z_{unbiased}}{Z^k_{biased}} \frac{M^k}{\widetilde{\Lambda}_{k,Q}\left(1 + \tilde{\delta}_{\Lambda_{k,Q}}{}^2\right)}} \quad \text{(A13)}$$

Applying the normalization condition

$$\sum_Q \Delta Q \widetilde{P}^j_{biased}(Q) = \sum_Q \Delta Q \frac{Z_{unbiased}}{Z^j_{biased}} \widetilde{P}_{unbiased}(Q) \frac{1}{\widetilde{\Lambda}_{j,Q}} = 1 \quad \text{(A14)}$$

to all the biased probability densities produces the following set of self-consistent equations which can be solved iteratively to obtain $Z_{unbiased}/Z^k_{biased}$.

$$Z^j_{biased} = Z_{unbiased} \sum_Q \Delta Q \frac{1}{\widetilde{\Lambda}_{j,Q}} \frac{1}{\Delta Q} \frac{\sum_l \frac{\widetilde{H}^l(Q)}{\left(1 + \tilde{\delta}_{\Lambda_{l,Q}}{}^2\right)}}{\sum_k \frac{Z_{unbiased}}{Z^k_{biased}} \frac{M^k}{\widetilde{\Lambda}_{k,Q}\left(1 + \tilde{\delta}_{\Lambda_{k,Q}}{}^2\right)}}$$
$$= \sum_Q \frac{\frac{1}{\widetilde{\Lambda}_{j,Q}} \sum_l \frac{\widetilde{H}^l(Q)}{\left(1 + \tilde{\delta}_{\Lambda_{l,Q}}{}^2\right)}}{\sum_k \frac{1}{Z^k_{biased}} \frac{M^k}{\widetilde{\Lambda}_{k,Q}\left(1 + \tilde{\delta}_{\Lambda_{k,Q}}{}^2\right)}} \quad \text{(A15)}$$

References

(1) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526–535.

(2) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322–333.

(3) Eker, F.; Griebenow, K.; Cao, X. L.; Nafie, L. A.; Schweitzer-Stenner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10054–10059.

(4) Hagarman, A.; Measey, T.; Doddasomayajula, R. S.; Dragomir, I.; Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2006**, *110*, 6979–6986.

(5) Flory, P. J. In *Statistical Mechanics of Chain Molecules*; Wiley Interscience: New York, 1969; pp 252−256.

(6) Ohkubo, Y. Z.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13916–13921.

(7) Munoz, V.; Serrano, L. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 301–311.

(8) Swindells, M. B.; Macarthur, M. W.; Thornton, J. M. *Nat. Struct. Biol.* **1995**, *2*, 596–603.

(9) Keskin, O.; Yuret, D.; Gursoy, A.; Turkay, M.; Erman, B. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 992–998.

(10) Ormeci, L.; Gursoy, A.; Tunca, G.; Erman, B. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 29–40.

(11) Betancourt, M. R.; Skolnick, J. *J. Mol. Biol.* **2004**, *342*, 635–649.

(12) Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. *Biochemistry* **2005**, *44*, 9691–9702.

(13) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494–506.

(14) Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R. A. G.; Dobson, C. M.; Smith, L. J. *J. Mol. Biol.* **1997**, *274*, 152–159.

(15) Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14258–14263.

(16) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565–12570.

(17) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10967–10972.

(18) Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1272–1277.

(19) Chen, K.; Liu, Z. G.; Zhou, C. H.; Shi, Z. S.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2005**, *127*, 10146–10147.

(20) Zaman, M. H.; Shen, M. Y.; Berry, R. S.; Freed, K. F.; Sosnick, T. R. *J. Mol. Biol.* **2003**, *331*, 693–711.

(21) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249–253.

(22) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451–458.

(23) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsovvelyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776–1783.

(24) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.

(25) Tesi, M. C.; vanRensburg, E. J. J.; Orlandini, E.; Whittington, S. G. *J. Stat. Phys.* **1996**, *82*, 155–181.

(26) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

(27) Falcioni, M.; Deem, M. W. *J. Chem. Phys.* **1999**, *110*, 1754–1766.

(28) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.

(29) Faraldo-Gomez, J. D.; Roux, B. *J. Comput. Chem.* **2007**, *28*, 1634–1647.

(30) Cheng, X. L.; Cui, G. L.; Hornak, V.; Sinnnerling, C. *J. Phys. Chem. B* **2005**, *109*, 8220–8230.

(31) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

(32) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420–433.

(33) Christen, M.; van Gunsteren, W. F. *J. Chem. Phys.* **2006**, *124*, 154106.

(34) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.

(35) Jang, S. M.; Shin, S.; Pak, Y. *Phys. Rev. Lett.* **2003**, *91*, 058305.

(36) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. *J. Chem. Theory Comput.* **2006**, *2*, 217–228.

(37) Kwak, W.; Hansmann, U. H. E. *Phys. Rev. Lett.* **2005**, *95*, 138102.

(38) Kannan, S.; Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 697–706.

(39) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.

(40) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.

(41) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(42) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(43) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(44) Akima, H. *ACM Trans. Math. Soft.* **1996**, *22*, 357–361.

(45) Wang, J.; Gu, Y.; Liu, H. Y. *J. Chem. Phys.* **2006**, *125*, 094907.

(46) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(47) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(48) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1981; pp 331−342.

(49) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(50) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Chem. Phys.* **1977**, *23*, 327–341.

(51) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

CT7003534

JCTC Journal of Chemical Theory and Computation

Mechanism of Coupled Folding and Binding in the siRNA-PAZ Complex

Hai-Feng Chen*

College of Life Sciences and Biotechnology, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai 200240, China

Received January 29, 2008

Abstract: The PAZ domain plays a key role in gene silencing pathway. The PAZ domain binds with siRNAs to form the multimeric RNA-induced silencing complex (RISC). RISC identifies mRNAs homologous to the siRNAs and promotes their degradation. It was found that binding with siRNA significantly enhances apo-PAZ folding. However, the mechanism by which folding is coupled to binding is poorly understood. Thus, the coupling relationship between binding and folding is very important for understanding the function of gene silencing. We have performed molecular dynamics (MD) of both bound and apo-PAZ to study the coupling mechanism between binding and folding in the siRNA-PAZ complex. Room-temperature MD simulations suggest that both PAZ and siRNA become more rigid and stable upon siRNA binding. Kinetic analysis of high-temperature MD simulations shows that both bound and apo-PAZ unfold via a two-state process. The unfolding pathways are different between bound and apo-PAZ: the order of helix III and helices I & II unfolding is switched. Furthermore, transition probability was used to determine the transition state ensemble for both bound and apo-PAZ. It was found that the transition state of bound PAZ is more compact than that of apo-PAZ. The predicted Φ-values suggest that the Φ-values of helix III and sheets of β3-β7 for bound PAZ are more native-like than those of apo-PAZ upon the binding of siRNA. The results can help us to understand the mechanism of gene silencing.

Introduction

Short RNAs mediate gene silencing, a process associated with virus resistance, developmental control, and heterochromatin formation in eukaryotes.[1–6] At first, the RNase III-like enzyme Dicer cleaves dsRNA into 21−23-nucleotide small interfering RNAs (siRNAs).[7–10] Then, siRNAs bind with the PAZ domain to form the multimeric RNA-induced silencing complex (RISC).[11–13] Finally, RISC identifies mRNAs homologous to the siRNAs and promotes their degradation.[7,14] Therefore, the PAZ domain from the Argonaute 2 protein is a critical component of the gene silencing pathway.[8,9,15,16] Because of its importance in gene silencing, the PAZ-siRNA complex is a possible target to study gene silencing.

The crystal structure siRNA-PAZ complex was reported in 2004 (pdb code: 1SI2).[17] The complex has three α-helices and eight β strands. The PAZ domain consists of helix I from 228 to 237, helix II from 249 to 261, and helix III from 306 to 313. The PAZ domain in the complex adopts a heart-shaped globular topology, with a twisted β-barrel consisting of six β-strands (β1−β3, β6−β8), capped by two amino-terminal α-helices (H I and H II) on one side and connected to an αβ module (β4−β5-HIII) on the other side. The RNA is bound in the cleft between the β-sheet surface (β2, β5) and the β3, β4, H3 module of the PAZ domain (shown in Figure 1).

The NMR experiment indicates that the PAZ domain residues undergo chemical shift changes upon RNA binding.[18] The pocket of the PAZ domain elongates and consists of helix HIII and strand β4 along one face and strand β7 and loop β2−β3 along the opposite face. This suggests that siRNA binding induces significant conformational change

* Corresponding author phone: 0086-21-34204348; fax: 0086-21-34204348; e-mail: haifengchen@sjtu.edu.cn.

10.1021/ct800030s CCC: $40.75
Published on Web 07/02/2008

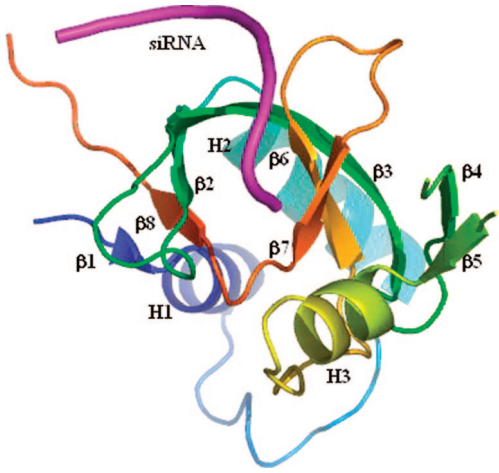


Figure 1. Ribbon representation of crystal structure of PAZ-siRNA (pdb code: 1SI2).[17] The locations of main secondary structures are indicated.

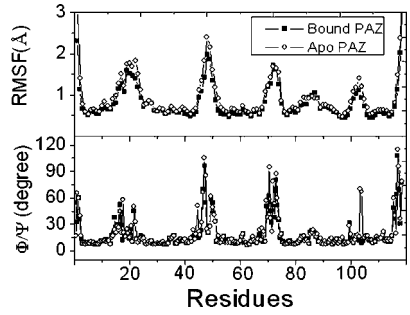


Figure 2. Cα and Ψ/Φ variations at the folded state for bound and apo-PAZ, respectively.

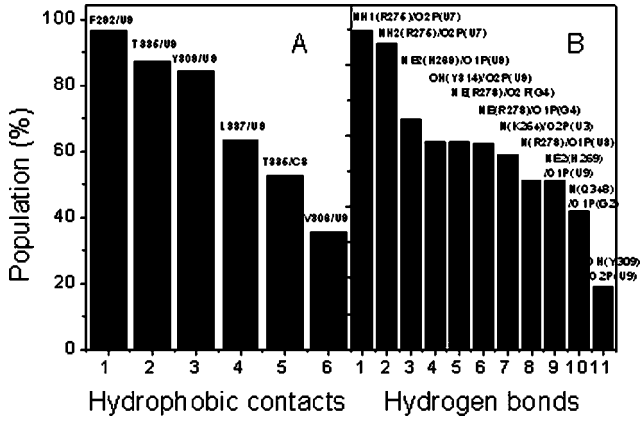


Figure 3. Hydrophobic contacts and hydrogen bonds between siRNA and PAZ. A: hydrophobic contacts and B: hydrogen bonds.

in the PAZ domain. These experimental observations raise a series of interesting questions. Some of those questions are as follows: (i) How does the folded state of PAZ change upon siRNA binding? (ii) What is the difference in the folding pathway between bound and apo-PAZ? (iii) If the binding of siRNA influences the unfolded state of PAZ? To shed light on these questions, we utilize molecular dynamics (MD) simulations in explicit solvent to analyze the coupling between binding and folding[19,20] in the siRNA-PAZ complex.

However, MD simulations are currently restricted to time scales of less than 1 μs, which is much shorter than folding and unfolding half-times of most proteins (at least 1 ms).[21,22] As we known, the rate of unfolding increases at high temperature, so most proteins unfold in the ns time scale at 498 K.[21] Therefore, simulations of protein unfolding at high

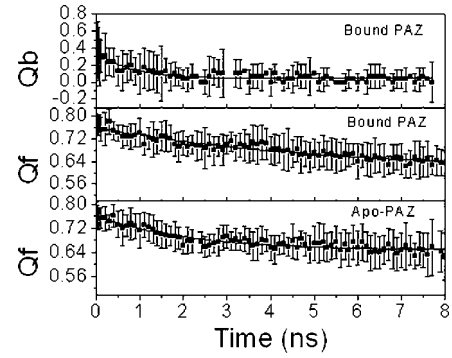


Figure 4. Kinetics fitting for bound and apo-PAZ.

Table 1. Unfolding Kinetics Constants[a]

		τ (ns)	A	B	R^2
bound PAZ	Qb	0.77	0.31	0.046	0.83
	Qf	5.26	0.15	0.61	0.91
	H (I & II)	9.70	0.28	0.51	0.72
	H (III)	12.93	0.094	0.84	0.70
Apo PAZ	Qf	2.09	0.11	0.65	0.89
	H (I & II)	8.74	0.34	0.44	0.72
	H (III)	6.97	0.35	0.63	0.67

[a] All curves are fitted by A exp(−t/τ) + B.

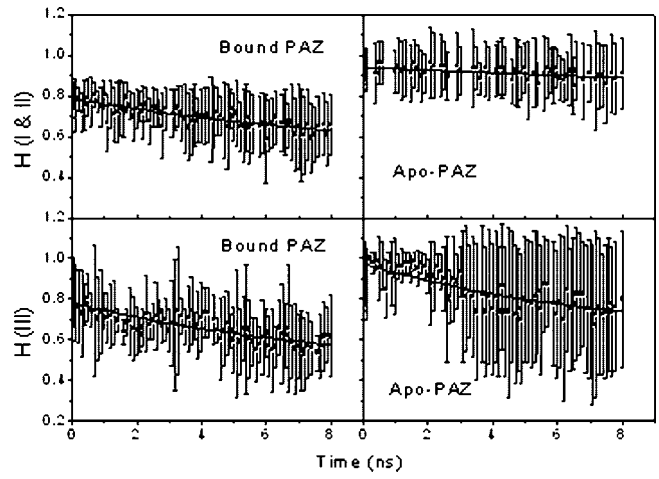


Figure 5. The unfolding kinetics of two group helices.

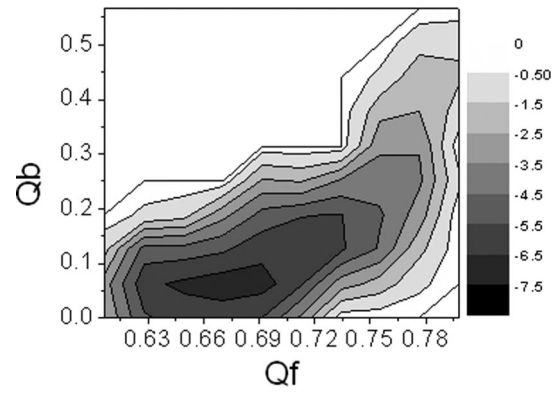


Figure 6. Unfolding landscapes with respect to Qb and Qf for bound PAZ.

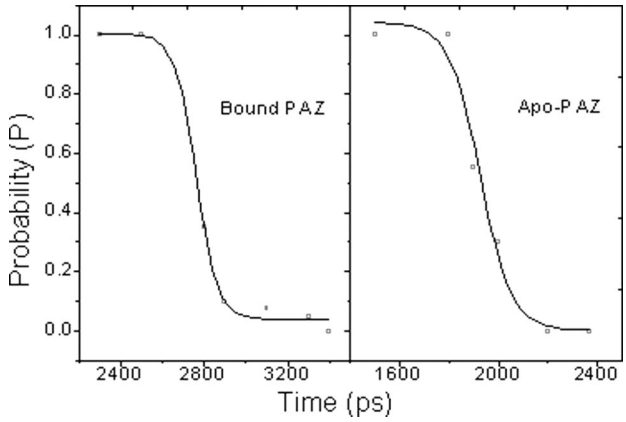


Figure 7. A representative transition probability *P* calculated at 498 K for the F ⇔ U transition for snapshot in the transition region for one of trajectories for bound and apo-PAZ, respectively. The red line is the fit to $P = 1/\{1 + \exp[(\tau - \tau_{TS})/\tau_{trans}]\}$.

temperature have been widely used[23–34] to study protein folding within a reasonable time. Furthermore, experiment and simulation confirm that the transition state for folding and unfolding is expected to be the same from the principle of microscopic reversibility.[21,35] Consequently, unfolding simulations at high temperature have been used in the current study.

In the following, we will discuss the folded states, unfolding kinetics, unfolding landscapes, transition states, prediction of Φ-values, unfolded states, and possible folding pathway for both bound and apo-PAZ to understand the influence of siRNA binding in the folding of PAZ.

Materials and Method

Room-Temperature and High-Temperature Molecular Dynamics Simulations. The atomic coordinates of the PAZ-siRNA complex were obtained from the crystal structure (pdb code: 1SI2).[17] Hydrogen atoms were added using the LEAP module of AMBER8.[36] Counterions were used to maintain system neutrality. All systems were solvated in a truncated octahedron box of TIP3P[37] waters with a buffer of 8 Å. The dimensions of the water box were 61.4 × 61.4 × 61.4 Å for bound PAZ. Particle Mesh Ewald (PME)[38] was employed to treat long-range electrostatic interactions with the default setting in AMBER8.[36] The parm99 force field was used for intramolecular interactions.[39,40] The SHAKE algorithm[41] was used to constrain bonds involving hydrogen atoms. 1000-step steepest descent minimization was performed to relieve any structural clash in the solvated systems. This was followed by heating up and brief equilibration for 20 ps in the NVT ensemble at 293 K with SANDER of AMBER8. Langevin dynamics with a time step of 2 fs was used in the heating and equilibration runs with a friction constant of 1 ps^{-1}.

To study the folded state of each solvated system, ten independent trajectories of 10.0 ns each in the NPT ensemble[42] at 293 K were then simulated with PMEMD of AMBER8. Here molecular dynamics with a time step of 2 fs was used for all production dynamics runs. To study unfolding pathways of each solvated system, ten independent unfolding trajectories of 20 ns each were performed in the NVT ensemble at 498 K but with the water density at 293 K (i.e., all high-temperature simulations were started from the end of the 10 ns 293 K trajectories). A total of 800 ns trajectories were collected for three solvated systems (bound PAZ, apo-PAZ, and siRNA) at both 293 K and 498 K, respectively, taking about 38,290 CPU hours on the in-house Xeon (1.86 GHz) cluster.

Native contacts for the bound and apo-PAZ were monitored to detect the beginning of unfolded state simulations. It was found that 8 ns at 498 K were needed to reach the equilibrium stage for both bound and apo-PAZ, so that the first 8 ns (a total of 80 ns for each system) were used to study unfolding kinetics and the remaining 12 ns (a total of 120 ns for each system) were used for equilibrium simulations at an unfolded state for each system. The siRNA is far less stable; it was found that only 3 ns at 498 K was enough to reach the unfolded state, so that the first 3 ns (a total of 30 ns) were used to study unfolding kinetics and the remaining 7 ns (a total of 70 ns) were used to study the unfolded equilibrium state.

Transition State Simulations. According to the definition of transition state (TS), 40 test MD runs for each candidate snapshot were performed to calculate the transition probability (*P*).[43–45] All test simulations have the same initial conformation for protein and solvent atoms but differing initial velocities. Because transition state structures agree with experiment under varying simulation temperatures,[46] TS simulations were done at 498 K in order to accelerate a simulated folding/unfolding rate. Each test trajectory is terminated when a conformation reaches the folded or unfolded state, as determined by Cα-rmsd. The Cα-rmsd vs simulation time at room and high temperature are shown in Figure 1s (Supporting Information). The results suggest that the folded state defines as the Cα-rmsd within 3.0 Å from the average structure at a folded state at 293 K. Up to 1 ns simulation at 498 K was found to be sufficient for each test trajectory, i.e. *P* values for tested snapshots were no longer changing when longer trajectories were run.

Free Energy Landscape Analysis. The unfolding landscapes were determined by calculating normalized probability from a histogram analysis.[43] Here we used a fraction of native tertiary contacts Qf and a fraction of native binding contacts Qb to map the unfolding landscape.

Data Analysis. Tertiary contact assignment was handled with in-house software. Two nonadjacent residues are in contact when their Cα atoms are closer than 6.5 Å. Secondary structure assignment was performed by using the DSSP program.[47] All rmsd and distance calculations were performed with PTRAJ in AMBER8.[36] The unfolding kinetics was fitted in Origin 7.0. Representative structures at unfolding half-times were used to construct unfolding pathways. Each representative structure is the closest snapshot to the average of all chosen snapshots at a given half-time (within ± its standard deviation).

Φ-values were computed with a strategy similar to those used in other studies[23,44,48]

$$\Phi_i^{calc} = \frac{N_i^{TS} - N_i^U}{N_i^F - N_i^U} \quad (1)$$

where N_i^{TS} is the number of native contacts of residue i at transition state, and N_i^F and N_i^U are the number of native contacts of residue i at folded and unfolded states, respectively. As defined by Calfisch and co-workers, native contacts were counted when the side chain heavy atoms of two nonadjacent residues are closer than 6.5 Å.[44]

Conformational clustering of bound and apo-PAZ at 293 K and 498 K was performed by using the MMTSB program.[49]

Results

Folded State. As a reference for the unfolding simulations, 10 trajectories of 10.0 ns each were simulated at 293 K to analyze the folded state of apo-PAZ, apo-siRNA, and their complex, respectively. To study the influence of siRNA binding on the stability of the folded PAZ, Cα and Φ/ψ variations for bound and apo-PAZ are illustrated in Figure 2. The Cα variation of bound PAZ is smaller than that of apo-PAZ, especially in the domain of the siRNA binding site. This suggests that bound PAZ become less flexible and more stable upon siRNA binding, consistent with experiment.[17] The Φ/ψ variation of bound PAZ is also smaller than that of apo-PAZ at the siRNA binding region, suggesting that the secondary structure stabilities have visible change upon siRNA binding.

To study the drive force for binding induced stable change in the folded state, the hydrophobic and hydrogen-bonding interactions between siRNA and PAZ were analyzed. All possible hydrophobic contacts and hydrogen bonds between siRNA and PAZ in the crystal structure were identified with Ligplot,[50] shown in Figure 2s (Supporting Information). The populations of six hydrophobic contacts in the simulation of ten trajectories are shown in Figure 3A. Five stable hydrophobic interactions can be found: F292/U9, T335/U9, Y309/U9, L337/U9, and T335/C8, with populations higher than 50%. Another one is rather unstable. Besides hydrophobic interactions, eleven possible hydrogen bonds were also identified with Ligplot.[50] Their populations in simulation are shown in Figure 3B. The results suggest that there are seven stable hydrogen bonds with a population higher than 50%. The other four hydrogen bonds are very weak. These strong hydrogen bonds between residues K264, H269, R275, R278, and Y314 and the Watson—Crick edges of the guanine and uracil bases (U9, U7, G4, and U3) and hydrophobic interactions are responsible for the stability in the PAZ domain.[17] This is consistent with the structural analysis by Ma et al.[17] These hydrophobic and hydrogen-bonding interactions play a key role in stabilizing the siRNA-protein interface. In summary, siRNA binding introduced more hydrophobic contacts and hydrogen bonds at the interface which are responsible for the higher stability in siRNA and PAZ.

Unfolding Kinetics. To investigate unfolding kinetics of bound and apo-PAZ, native tertiary contacts (Qf) and native binding contacts (Qb) are used to monitor unfolding and unbinding kinetics. The time scale of Qb and Qf for bound PAZ unfolding is illustrated in Figure 4. Apparently, the tertiary unfolding and unbinding kinetics can be represented well by single exponential functions, indicating that they are first order kinetics in the NVT ensemble at 498 K but with the room-temperature water density. The fitted kinetics data are listed in Table 1. Analysis shows that the unbinding half-time is 0.77 ns, and the unfolding half-time is 5.26 ns, suggesting that unbinding is faster than tertiary unfolding. The time evolution of Qf for apo-PAZ is also shown in Figure 4. It is found that tertiary unfolding of apo-PAZ also obeys first order kinetics, with a half-time of 2.09 ns, which is obviously faster than tertiary unfolding of bound PAZ. This suggests that the binding of siRNA significant postpones the tertiary unfolding of PAZ. This is consistent with the experimental observation.[17]

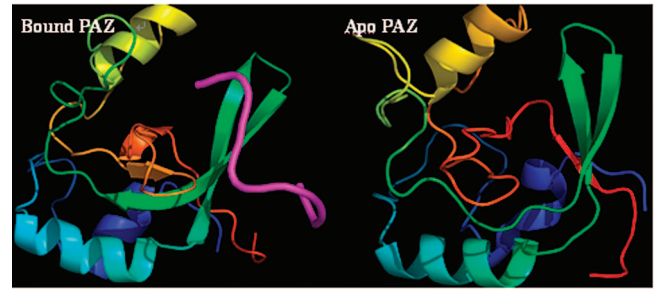


Figure 8. Average TSE structures for bound and apo-PAZ, respectively.

The unfolding kinetics of all three helices is also analyzed. The three helices can be grouped into two categories: helices (I & II) and helix III according to the position relative to siRNA in complex. Helix III is located at the binding pocket of siRNA and helices I and II are apart from siRNA. Time evolution curves are shown in Figure 5 for two groups of helixes in both bound and apo-PAZ, respectively. Our analysis shows that helical unfolding also obeys first order kinetics under the high-temperature simulation condition. THe unfolding half-time is 6.97 ns for helix III and 8.74 ns for helices (I and II), respectively, in apo-PAZ. The unfolding half-time is 12.93 ns for helix III and 9.70 ns for helices (I and II), respectively, in bound PAZ. Not surprisingly, helical unfolding is slower than tertiary unfolding and unbinding for both. This is consistent with other unfolding simulations of helical proteins, for example chymotrypsin inhibitor 2 and MDM2.[34,46] Note also that the helical unfolding half-times of the bound PAZ are longer than those of apo-PAZ, suggesting that the binding to siRNA stabilizes the helices in PAZ.[51] Furthermore, the unfolding half-time is reversed for helices (I and II) and helix III upon the binding of siRNA. This suggests that siRNA binding significantly changes the pathway of PAZ folding. Finally, since the seven β strands are very flexible, they are not monitored during unfolding.

Unfolding Landscapes. Furthermore, to understand the coupling mechanism between unfolding and unbinding, the unfolding landscape of bound PAZ was analyzed with the variables Qf and Qb (shown in Figure 6). The unfolding landscape shows that unbinding happens first while tertiary contacts are held stable and then followed by tertiary unfolding and unbinding. This is in agreement with the

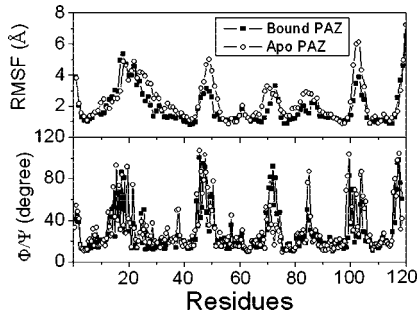


Figure 9. Ca and Ψ/Φ variations at TSE for bound and apo-PAZ, respectively.

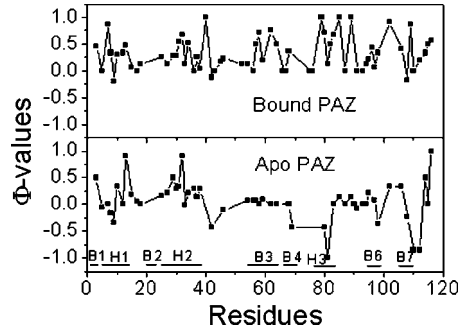


Figure 10. Predicted Φ-values of bound and apo-PAZ.

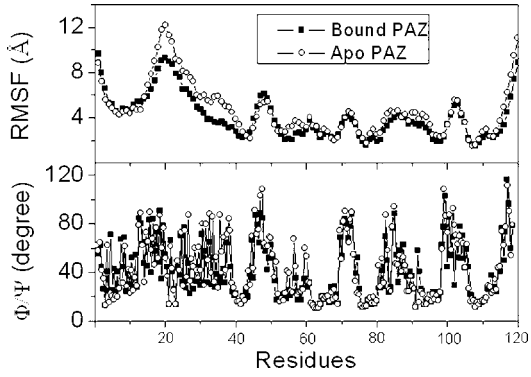


Figure 11. Structural variation of an unfolded state for bound and apo PAZ.

unfolding kinetics analysis for the bound PAZ. This suggests the formation of binding interface partly depends on formation of tertiary contacts.

The coupling between secondary and tertiary unfolding is also investigated for both bound and apo-PAZ, shown in Figure 3s (Supporting Information). Since seven β strands are too short or too flexible, only helical structures were monitored during unfolding. For bound PAZ, it is found that Qf decreases first while the content of helices (I and II) does not change. This is followed by a simultaneous decrease of tertiary contacts and the content of helices (I and II). This suggests that the tertiary unfolding is followed by secondary and tertiary unfolding. This is also consistent with the above unfolding kinetics analysis of bound PAZ. For apo-PAZ, Qf and the content of helices (I and II) simultaneously decrease, and then Qf accelerates to decrease. The coupling between the content of helix III and tertiary contacts is also monitored. We found that helix III keeps stable during tertiary unfolding for bound PAZ. For apo-PAZ, it is found that Qf decreases first, while the content of helix III does not change. Then Qf and the content of helix III simultaneously decrease. The different unfolding landscapes between Qf and helix III for bound and apo-PAZ suggest that helix III changes more stability upon siRNA binding and consistent with unfolding kinetics.

Transition State. Kinetics analysis shows that tertiary unfolding of both bound and apo-PAZ obeys first order kinetics. This suggests that bound and apo-PAZ unfold via a two-state process. Therefore, there is a transition state which corresponds to the free energy maximum between two states along each of their unfolding pathways. The transition state ensemble (TSE) structures can either fold or unfold, and the transition probability (P) will be 50%. According to the definition of TSE, we have scanned TSE structures from MD snapshots in all 10 unfolding trajectories for each of the bound and apo-PAZ, respectively.[43] The transition probability curves are further fitted by the Boltzmann equation and shown in Figure 7. The equation is $P = 1/(1 + \exp((t - \tau_{TS})/\tau_{trans}))$, where τ_{TS} is the time when $P = 50\%$, and τ_{trans} sets the period when P is between 0.4 and 0.6.[43–45] Our analysis yields 681 snapshots for bound PAZ TSE and 414 snapshots for apo-PAZ TSE, respectively.

Figure 8 illustrates the average structures for all TSE snapshots for bound and apo-PAZ, respectively. There are 60.7% native hydrophobic contacts and 72.3% native helical content for apo-PAZ TSE, while there are 73.5% native hydrophobic contacts and 88.2% native helical content for bound PAZ. Apparently, it can be concluded that the TSE of bound PAZ is more native-like than that of apo-PAZ. This indicates that the folding activation free energy for bound PAZ is probably smaller than that for apo-PAZ, leading to a relatively faster folding rate for bound PAZ. Combining with the above analysis that the unfolding rates for bound PAZ are slower than that of apo-PAZ, this leads to the conclusion that bound PAZ is more stable than apo-PAZ, based on the following relationship for two-state folding systems,[22,52] $K_{eq} = k_f/k_u$, where K_{eq} is folding equilibrium constant, and k_f and k_u are folding and unfolding rates, respectively. This conclusion on the relative stability between bound and apo-PAZ is consistent with the previous experimental observation.[17]

Figure 9 illustrates structural variations of TSE structures for bound and apo-PAZ, respectively. The Cα fluctuation of bound PAZ is smaller than that of apo-PAZ, suggesting that bound PAZ is more stable than that of apo-PAZ at the transition state. Thus TSE for bound PAZ is more native-like than that for apo-PAZ. However, the Φ/ψ fluctuation of bound PAZ is similar to that of apo-PAZ: there are some regions of helix and sheet with low fluctuation in both states.

All TSE snapshots were used to predict Φ-values of bound and apo-PAZ as shown in Figure 10. The sequence distribution of Φ-values suggests that the helices of helix I and II are more native-like than other loop regions for both bound and apo-PAZ. Furthermore, the Φ-values of helix III, β3, β4, β7, and the loop (314−334) of bound PAZ are significantly larger than those of apo-PAZ. This suggests that these regions of bound PAZ are more native-like than those

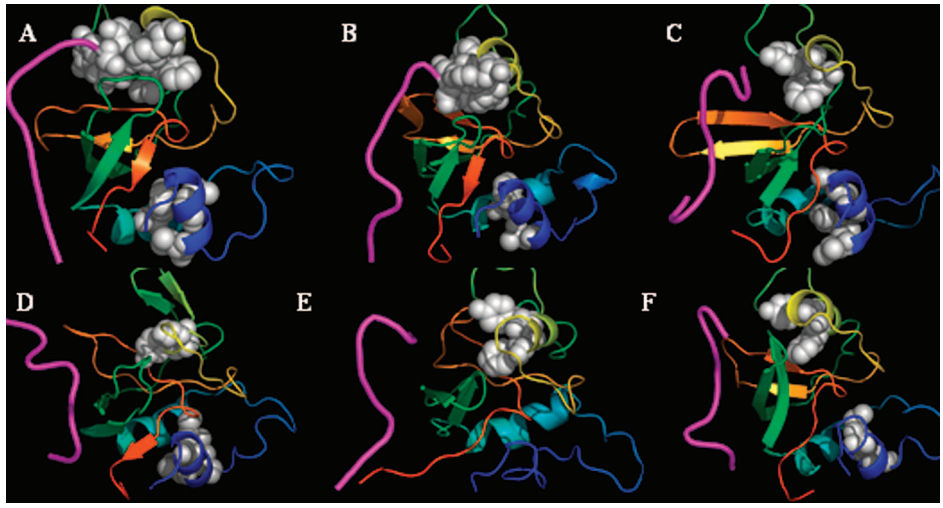


Figure 12. Unfolding pathway of bound PAZ. A: <0 ns (F), B: 0.77 ns (τ_{Qb}), C: 5.26 ns (τ_{Qf}), D: 9.70 ns (τ_{H12}), E: 12.93 ns (τ_{H3}), and F: >15 ns (U).

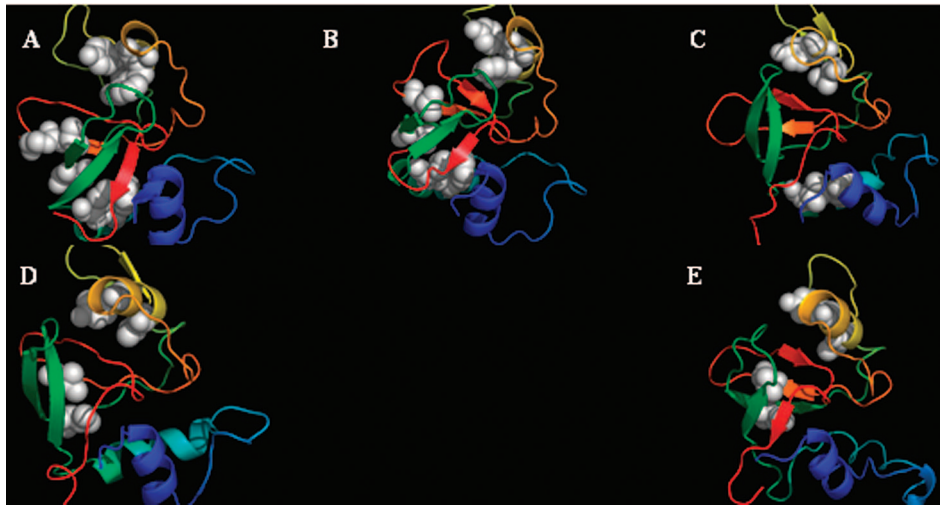


Figure 13. Unfolding pathway of Apo-PAZ. A: <0 ns (F), B: 2.09 ns (τ_{Qf}), C: 6.97 ns(τ_{H3}), D: 8.74 ns (τ_{H12}), and E: >15 ns (U).

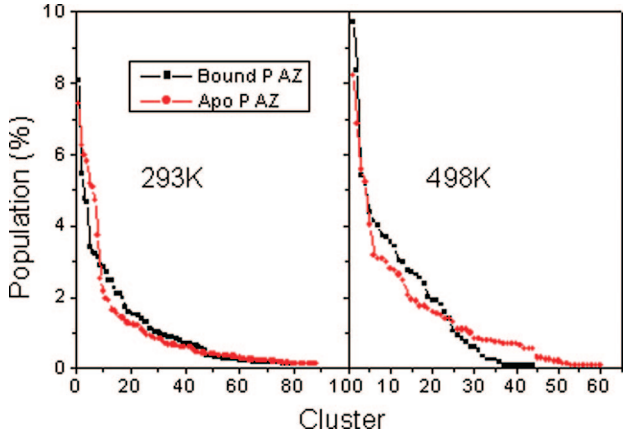


Figure 14. Cluster of bound and apo-PAZ.

of apo-PAZ upon the binding of siRNA and consistent with the unfolding kinetics analysis.

Unfolded State. All unfolding simulations eventually reach the unfolded equilibrium state under the high-temperature unfolding condition. For the unfolded state of apo-PAZ, there are 50.0% native hydrophobic contacts and 54.2% helical content remaining. Worth noting is the fact that the native hydrophobic contacts of H (I & II) all disappear. This is different from that of the bound-PAZ, for which the part hydrophobic contacts of H(I & II) remain. There are 57.8% native hydrophobic contacts and 60.4% helical content remaining.

Conformational fluctuation was also analyzed using structural variation from the average structures for the unfolded state (Figure 11). The Cα fluctuation of bound PAZ is smaller than that of apo-PAZ, indicating that bound molecules at an unfolded state are also more stable than apo molecules. However, there is little difference in the Φ/ψ fluctuation between bound and apo-PAZ.

Discussion

Comparison with Experiment. Structural analysis suggests that F292 and L337 are critical residues for efficient binding of siRNA to PAZ.[17] Our folded simulation illustrates that there are two stable hydrophobic contacts between F292 and U9 and L337 and U9. This suggests that simulation

results are consistent with structural analysis.[17] Beside these native hydrophobic contacts, there is also stable hydrophobic contact between Y309 and U9. This is in agreement with the mutation result that Y309F/Y314F causes a 46-fold reduction of binding affinity.[17]

Furthermore, it has been observed in experiment that PAZ must undergo an expansion upon siRNA binding, achieved through a rearrangement and outward expansion of the helices and β-sheets surrounding the binding pocket. In order to make quantitative comparison, the distance between the centers of β4-β5 and loop of 267−276 at the binding pocket opening in the experimental structures is measured and is found to increase by 0.48 Å upon siRNA binding. In simulation, the average distance between the centers of the same regions increases by about 0.25 Å upon siRNA binding, in qualitative agreement with experiment.[13,17]

Unfolding Pathways and Likely Folding Pathways. Based on the unfolding kinetics, landscape analysis, and transition state analysis, unfolding pathways for bound PAZ can now be constructed as shown in Figure 12. A) At the unbinding half-time, there are 36 out of 45 (folded state) native hydrophobic contacts within PAZ. The lost hydrophobic contact is within N-terminal and C-terminal. The native hydrophobic contacts between siRNA and PAZ also start to disappear: only 1 out of 4 exists. There is 81.1% native helical content remaining. B) At the half-time of tertiary unfolding, there are 32 native hydrophobic contacts within PAZ. The hydrophobic core is located mostly among H (I & II) and H III. All native hydrophobic contacts between siRNA and PAZ have disappeared. siRNA is partly structured. There is 75.9% helical content remaining. C) At the half-time of H (I & II) unfolding, there are 30 native hydrophobic contacts within PAZ. Two of the native hydrophobic contacts within H (I & II) have disappeared. siRNA moves away from the cavity of PAZ. There is 65.0% helical content remaining. D) At the half-time of H III unfolding, there are 29 native hydrophobic contacts within PAZ. The native hydrophobic contacts within H III began to unfold. There is 60.3% helical content remaining.

Similarly, the unfolding pathway of apo-PAZ is constructed and shown in Figure 13. A) At the half-time of tertiary unfolding, there are 35 out of 44 (folded state) native hydrophobic contacts. The hydrophobic core among β4-β5 almost disappears. The remaining hydrophobic interactions are within the regions of H III and H (I & II). There is 76.8% helical content remaining. B) At the half-time of H III unfolding, there are 33 native hydrophobic contacts within PAZ. Two of the native hydrophobic contacts within H III have disappeared. There is 65.3% helical content remaining. C) At the half-time of H (I & II) unfolding, there are 28 native hydrophobic contacts within PAZ. The hydrophobic core for siRNA binding disrupts. There is 64.7% helical content remaining.

If we assume that folding is the reverse of unfolding, the proposed folding/binding pathway of bound PAZ is the H III folding, H (I & II) folding, tertiary folding, and then siRNA binding. The folding order for apo-PAZ is H (I & II) folding, H III folding, and tertiary folding. The different order between the folding pathways of bound and apo-PAZ is the reverse of H III folding and H (I & II) folding upon the binding of siRNA.

Entropy Effect in the Binding/Folding Coupling. Figure 14 illustrates the cluster of bound and apo-PAZ at 293 K and 498 K, respectively. In the folded state, bound PAZ is less heterogeneous than apo-PAZ, indicating that the entropy decreases in the presence of siRNA, which induces from specific binding interactions between PAZ and siRNA. The strong and specific binding interactions make the specific binding favorable thermodynamically. In the unfolded state, however, there is no specific binding interaction between the two molecules. Nevertheless, the cluster of bound PAZ is also smaller than that of apo-PAZ, so that the entropy still decreases in the presence of siRNA. Thus the nonspecific binding of the two molecules is also favorable thermodynamically. Therefore, the entropic effect is significantly unfavorable in the coupled folding and binding in the PAZ-siRNA complex, and this unfavorable entropy effect weakens the strong contribution of hydrophobic and hydrogen bonding interactions between siRNA and PAZ. This is consistent in structural analysis that siRNA binding partly stabilizes the complex of siRNA-PAZ.[17]

Conclusion

Room-temperature MD simulations were used to study the stability of the PAZ domain uponing the binding of siRNA. The results suggest that both PAZ and siRNA become more rigid and stable upon siRNA binding. High-temperature MD simulations were used to investigate the folding kinetic of the PAZ domain. The results show that both bound and apo-PAZ unfold via a two-state process. The unfolding pathways are different between bound and apo-PAZ: the order of helix III and helices I & II unfolding is switched. Furthermore, transition probability was used to determine the transition state ensemble for both bound and apo-PAZ. It was found that the transition state of bound PAZ is more compact than that of apo-PAZ. The predicted Φ-values suggest that the Φ-values of helix III and sheets of β3-β7 for bound PAZ are more native-like than those of apo-PAZ upon the binding of siRNA. The results can help us to understand the mechanism of gene silencing.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (Grants No. 30770502 and No. 20773085).

Supporting Information Available: Cα-RMSD vs simulation time for one trajectory (Figure 1s), two-dimensional representation for the interactions mode between siRNA and PAZ (Figure 2s), and unfolding landscapes for bound and apo-PAZ (Figure 3s). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Denli, A. M.; Hannon, G. J. RNAi: an ever-growing puzzle. *Trends Biochem. Sci.* **2003**, *28*, 196.

(2) Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **2004**, *116*, 281.

(3) Voinnet, O. RNA silencing as a plant immune system against viruses. *Trends Genet.* **2001**, *17*, 449.

(4) Volpe, T. A.; Kidner, C.; Hall, I. M.; Teng, G.; Grewal, S. I.; Martienssen, R. A. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **2002**, *297*, 1833.

(5) Peters, L.; Meister, G. Argonaute proteins: mediators of RNA silencing. *Mol. Cell* **2007**, *26*, 611.

(6) Hall, I. M.; Shankaranarayana, G. D.; Noma, K.; Ayoub, N.; Cohen, A.; Grewal, S. I. Establishment and maintenance of a heterochromatin domain. *Science* **2002**, *297*, 2232.

(7) Hammond, S. M.; Bernstein, E.; Beach, D.; Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **2000**, *404*, 293.

(8) Zamore, P. D.; Tuschl, T.; Sharp, P. A.; Bartel, D. P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **2000**, *101*, 25.

(9) Bernstein, E.; Caudy, A. A.; Hammond, S. M.; Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **2001**, *409*, 363.

(10) Elbashir, S. M.; Lendeckel, W.; Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **2001**, *15*, 188.

(11) Agrawal, N.; Dasaradhi, P. V.; Mohmmed, A.; Malhotra, P.; Bhatnagar, R. K.; Mukherjee, S. K. RNA interference: biology, mechanism, and applications. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 657.

(12) Song, J. J.; Liu, J.; Tolia, N. H.; Schneiderman, J.; Smith, S. K.; Martienssen, R. A.; Hannon, G. J.; Joshua-Tor, L. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat. Struct. Biol.* **2003**, *10*, 1026.

(13) Lingel, A.; Simon, B.; Izaurralde, E.; Sattler, M. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **2003**, *426*, 465.

(14) Nykanen, A.; Haley, B.; Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **2001**, *107*, 309.

(15) Hammond, S. M.; Boettcher, S.; Caudy, A. A.; Kobayashi, R.; Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **2001**, *293*, 1146.

(16) Knight, S. W.; Bass, B. L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science* **2001**, *293*, 2269.

(17) Ma, J. B.; Ye, K.; Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **2004**, *429*, 318.

(18) Yan, K. S.; Yan, S.; Farooq, A.; Han, A.; Zeng, L.; Zhou, M. M. Structure and conserved RNA binding of the PAZ domain. *Nature* **2003**, *426*, 468.

(19) Henkels, C. H.; Kurz, J. C.; Fierke, C. A.; Oas, T. G. Linked folding and anion binding of the Bacillus subtilis ribonuclease P protein. *Biochemistry* **2001**, *40*, 2777.

(20) Henkels, C. H.; Oas, T. G. Ligation-state hydrogen exchange: Coupled binding and folding equilibria in ribonuclease P protein. *J. Am. Chem. Soc.* **2006**, *128*, 7772.

(21) Fersht, A. R.; Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **2002**, *108*, 573.

(22) Baker, D. Metastable states and folding free energy barriers. *Nat. Struct. Biol.* **1998**, *5*, 1021.

(23) Caflisch, A.; Karplus, M. Molecular-dynamics simulation of protein denaturation-solvation of the hydrophobic cores and secondary structure of barnase. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1746.

(24) Caflisch, A.; Karplus, M. Acid and thermal-denaturation of barnase investigated by molecular-dynamics simulations. *J. Mol. Biol.* **1995**, *252*, 672.

(25) Daggett, V.; Li, A. J.; Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* **1996**, *257*, 430.

(26) Ladurner, A. G.; Itzhaki, L. S.; Daggett, V.; Fersht, A. R. Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8473.

(27) Gsponer, J.; Caflisch, A. Role of native topology investigated by multiple unfolding simulations of four SH3 domains. *J. Mol. Biol.* **2001**, *309*, 285.

(28) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **2003**, *421*, 863.

(29) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518.

(30) Akanuma, S.; Miyagawa, H.; Kitamura, K.; Yamagishi, A. A detailed unfolding pathway of a (beta/alpha)8-barrel protein as studied by molecular dynamics simulations. *Proteins* **2005**, *58*, 538.

(31) Scott, K. A.; Randles, L. G.; Moran, S. J.; Daggett, V.; Clarke, J. The folding pathway of spectrin R17 from experiment and simulation: using experimentally validated MD simulations to characterize states hinted at by experiment. *J. Mol. Biol.* **2006**, *359*, 159.

(32) Oard, S.; Karki, B. Mechanism of beta-purothionin antimicrobial peptide inhibition by metal ions: molecular dynamics simulation study. *Biophys. Chem.* **2006**, *121*, 30.

(33) Tsai, J.; Levitt, M.; Baker, D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J. Mol. Biol.* **1999**, *291*, 215.

(34) Chen, H. F.; Luo, R. Binding induced folding in p53-MDM2 complex. *J. Am. Chem. Soc.* **2007**, *129*, 2930.

(35) Day, R.; Daggett, V. Direct observation of microscopic reversibility in single-molecule protein folding. *J. Mol. Biol.* **2007**, *366*, 677.

(36) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L., III; 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, 2004.

(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.

(38) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(39) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J. Comput. Chem.* **2000**, *21*, 1049.

(40) Lwin, T. Z.; Lu, Q.; Luo, R. Force field influences in protein folding simulations. *Protein Sci.* **2006**, *15*, 2642.

(41) Rychaert, 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. *Comput. Phys.* **1977**, *23*, 327.

(42) 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.

(43) Pande, V. S.; Rokhsar, D. S. Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062.

(44) Gsponer, J.; Caflisch, A. Molecular dynamics simulations of protein folding from the transition state. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6719.

(45) Chong, L. T.; Snow, C. D.; Rhee, Y. M.; Pande, V. S. Dimerization of the p53 oligomerization domain: identification of a folding nucleus by molecular dynamics simulations. *J. Mol. Biol.* **2005**, *345*, 869.

(46) Day, R.; Daggett, V. Sensitivity of the folding/unfolding transition state ensemble of chymotrypsin inhibitor 2 to changes in temperature and solvent. *Protein Sci.* **2005**, *14*, 1242.

(47) Kabsch, W.; Sander, C. Dictionary of protein secondary structure-pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577.

(48) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **2001**, *409*, 641.

(49) Feig, M.; Karanicolas, J. ; Charles, L. ; Brooks, I. I. I. *MMTSB Tool Set. MMTSB NIH Research Resource*; The Scripps Research Institute: 2001.

(50) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127.

(51) Petsko, G. A.; Ringe, D. *Protein Structure and Function. Chapter 2. From Structure to Function*; New Science Press: London, 2003.

(52) Nolting, B. *Protein Folding Kinetics, Biophysical Methods*; Springer: Berlin, 1999.

CT800030S

JCTC Journal of Chemical Theory and Computation

Sulfoxide, Sulfur, and Nitrogen Oxidation and Dealkylation by Cytochrome P450

Patrik Rydberg,[†] Ulf Ryde,[‡] and Lars Olsen*,[†]

Department of Medicinal Chemistry, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark, and the Department of Theoretical Chemistry, Lund University, P.O.B. 124, SE-22100 Lund, Sweden

Received March 25, 2008

Abstract: The oxidation and dealkylation of dimethylsulfoxide (DMSO), dimethylsulfide (DMS), and trimethylamine (TMA) by cytochrome P450 has been studied with density functional theory calculations. The results show that the oxidation reactions always occur on the doublet spin surface, whereas dealkylations can take place for both the doublet and quartet spin states. Moreover, DMS is more reactive than DMSO, and S-oxidation is more favorable than S-dealkylation, whereas N-dealkylation is more favorable than N-oxidation. This is in perfect agreement with experimental results, showing that density functional activation energies are reliable and comparable for widely different reactions with cytochrome P450.

Introduction

The cytochromes P450 (CYPs) constitute an enzyme family that is found in all types of organisms, from bacteria to mammals. In the human genome, there are 57 genes for CYPs.[1] These isoforms have functions including synthesis and the degradation of many physiologically important compounds, as well as the degradation of xenobiotic compounds, for example, drugs.[2] Numerous studies have been performed on these enzymes, because they influence the transformation of pro-drugs into their active form as well as the bioavailability and degradation of many drugs. In fact, it has been estimated that the CYPs are responsible for ~75% of the phase I metabolism of drugs.[3]

The active site in the CYPs is buried inside the protein and connected to the surface by several channels, which vary between the various isoforms.[4,5] At the bottom of the active site, there is a heme group with a central iron ion. Below the plane of the heme group, the sulfur atom of a cysteine amino acid coordinates to the iron ion, whereas the site above the heme plane may bind various extraneous small ligands during the reaction cycle. In the resting state, this position is occupied by a water molecule.

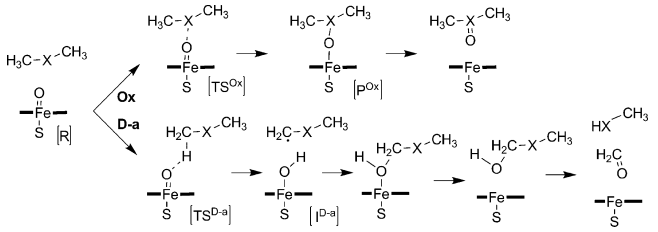
The CYPs catalyze several different types of reactions, of which the most common ones are hydroxylation of saturated C−H bonds, dealkylations, epoxidation of double bonds, oxidation of aromatic ring systems, and oxidation of heteroatoms. Most of these reactions have been thoroughly characterized using theoretical calculations,[6,7] but heteroatom oxidation has been much less studied than the other types of reactions.[8–12] The active species in the reactions is in general a high-valent (formally Fe^{V}=O) state of the active-site heme group, called compound I. It has two close-lying electronic states, a doublet and a quartet, and they may contribute differently to various reactions. For example, they give comparable activation energies for aliphatic hydroxylation, whereas the addition of compound I to the aromatic ring system is most favorable on the doublet spin surface.[6,7]

For substrates containing heteroatoms, there is the possibility of both heteroatom oxidation and dealkylation, as shown in Scheme 1. The dealkylation reaction starts with hydroxylation of the C^{α} atom, followed by a cleavage of the bond between the heteroatom and C^{α}. Interestingly, N-containing compounds seem to prefer the hydroxylation−dealkylation reaction, whereas the opposite is true for S-containing compounds.[9,13–16] It is a challenge for theoretical methods to explain these trends. Loew and Chang[9] have addressed this question using Hartree−Fock calculations, but they could rationalize the experimental observations only if they used different criteria for different reactions: The relative

* Corresponding author e-mail: lo@farma.ku.dk.
[†] University of Copenhagen.
[‡] Lund University.

10.1021/ct800101v CCC: $40.75
Published on Web 07/25/2008

Scheme 1. The Reaction Paths of Heteroatom Oxidation (Ox) and Heteroatom Dealkylation (D-a)



stability of the oxidation products could explain why N compounds prefer dealkylation. However, to explain the preference for oxidation of S compounds, they instead had to compare the energy difference between the C^{α} radical intermediate and the sulfoxide products. In this paper, we try to obtain a consistent view of the N-, S-, and SO-oxidation and dealkylation reactions by using more accurate density functional theory (DFT) calculations. We study the substrates trimethylamine (TMA), dimethylsulfide (DMS), and dimethylsulfoxide (DMSO). To the best of our knowledge, this is the first time sulfoxide oxidation and dealkylation have been studied using DFT calculations.

Computational Methodology

We have modeled the compound I species of the CYPs as iron (formally Fe^{V}) porphine (i.e., a porphyrin without side chains), with CH_3S^- and O^{2-} as axial ligands. Four different states along the reactions were studied (Scheme 1), namely, compound I and the substrate isolated from each other, the complex between compound I and the substrate (the reactant complex, **R**), the transition state (**TS**), and the product after the oxidation ($\mathbf{P^{Ox}}$) or the intermediate after the dealkylation ($\mathbf{I^{D\text{-}a}}$; cf., Scheme 1). All energies are given relative to that of the reactant complex (**R**). All four states were studied in both the doublet and quartet states.

The quantum chemical calculations were performed using the density functional method B3LYP[17–19] with the VWN(V) correlation functional[20] (unrestricted formalism for open-shell systems). In the geometry optimizations, we used for iron the double-ζ basis set of Schäfer et al.,[21] enhanced with a p function with the exponent 0.134 915. For the other atoms, the 6-31G(d) basis set[22–24] was used (this combination is denoted BSI). More accurate energies were determined by single-point calculations at the B3LYP/6-311++G(2d,2p) level[25] with the double-ζ basis set of Schäfer et al.,[21] enhanced with s, p, d, and f functions (exponents of 0.013 772 32, 0.041 843, 0.1244, 2.5, and 0.8; two f functions) on iron[26] (denoted BSII). The B3LYP functional was chosen because it has previously been shown to give very good geometries compared to crystal structures, as well as energies, compared to CCSD(T) and CASPT2 calculations for heme models.[27,28] Moreover, it has been employed in almost all recent studies of CYP reactions.[6,7,29]

To get a view of what effects a surrounding protein may have on these reactions, we have performed calculations in a vacuum, as well as with an implicit solvent model (note that the CYPs are a large enzyme family,[1] so it is not possible to model the general effect of the protein in a more detailed way). The effective dielectric constant (ε) in proteins has been much discussed, and values from 2 to 40 have been suggested.[30–32] Therefore, we tested three different values, 1, 4, and 80, to get a feeling of possible effects of the protein. Solvent calculations were carried out at the B3LYP/BSI level with the continuum conductor-like screening model (COSMO).[33] For the atomic radii, we used the optimized COSMO radii in Turbomole[34] (and 2.0 Å for Fe). For all of the other parameters, we employed the default values, implying a waterlike probe molecule. The full data are presented in Table S1 in the Supporting Information. They show that the continuum solvent stabilizes the separated reactants relative to the reactant complex by 9−11 and 16−20 kJ/mol for $\varepsilon =$ 4 and 80, respectively. The two spin states are always affected in the same way for both spin states, so the solvent gives no net effect on the spin splitting energy. The solvent effect on the activation energies depends on the substrate, but not on the reaction type: For DMSO, the activation energies increase by 3−7 kJ/mol; for DMS, they are not much affected (±2 kJ/mol), whereas for TMA, they decrease by 5−9 kJ/mol (with $\varepsilon = 4$; $\varepsilon = 80$ gives similar, but larger, effects, up to 18 kJ/mol). Thus, possible solvation effects of the enzyme will not change the general conclusions of this paper, regarding the relative activation energies of the oxidation and dealkylation reactions and the three types of substrates.

Frequency calculations were performed at the B3LYP/BSI level of theory to obtain the zero-point vibrational energy and thermal corrections to the Gibbs free energy (at 298 K and 1 atm pressure, using an ideal-gas approximation[35]). They also verified that the structures represent true minima or transition states. All calculations were carried out with the Turbomole program package, version 5.9.[36] Presented energies are those obtained at the B3LYP/BSII level, including the zero-point vibrational energy, thermal corrections to the Gibbs free energy, and solvent effects with a dielectric constant of 4. The charges and spin densities presented were calculated with Mulliken population analysis in a vacuum.

Previous studies of sulfur and nitrogen oxidation[8,10–12] used a SH^- model of the cysteine iron ligand, instead of our CH_3S^- model (which is the preferred cysteine model in the theoretical study of all other proteins). Therefore, we studied the difference between these two models for the DMS oxidation. These studies were restricted to the quartet spin state, because we failed to locate the doublet transition state with the SH^- model. The two models gave very similar geometries, with differences of 0.005, 0.008, and 0.004 Å for the Fe−O distance; 0.056, 0.003, and 0.008 Å for the $Fe-S_{Cys}$ distance; and 0.009, 0.013, and 0.001 Å for the $O-S_{DMS}$ distance for the reactant complex, the transition state, and the product complex, respectively. The only significant geometric difference was that the two models sometimes gave different $C/H_{Cys}-S_{Cys}-Fe-N$ torsion angles, owing to differing interactions between the methyl group or the hydrogen atom of the cysteine model and the porphyrin ring. The CH_3S^- model gave a 12 kJ/mol higher activation energy and a 14 kJ/mol less negative energy of the product, but the energetic effects of solvation, zero-point vibrational

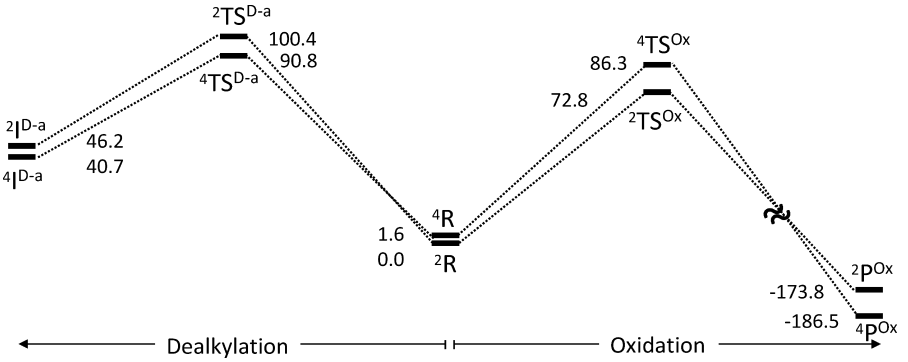


Figure 1. Relative energies (in kJ/mol) for the various states in the reactions with DMSO.

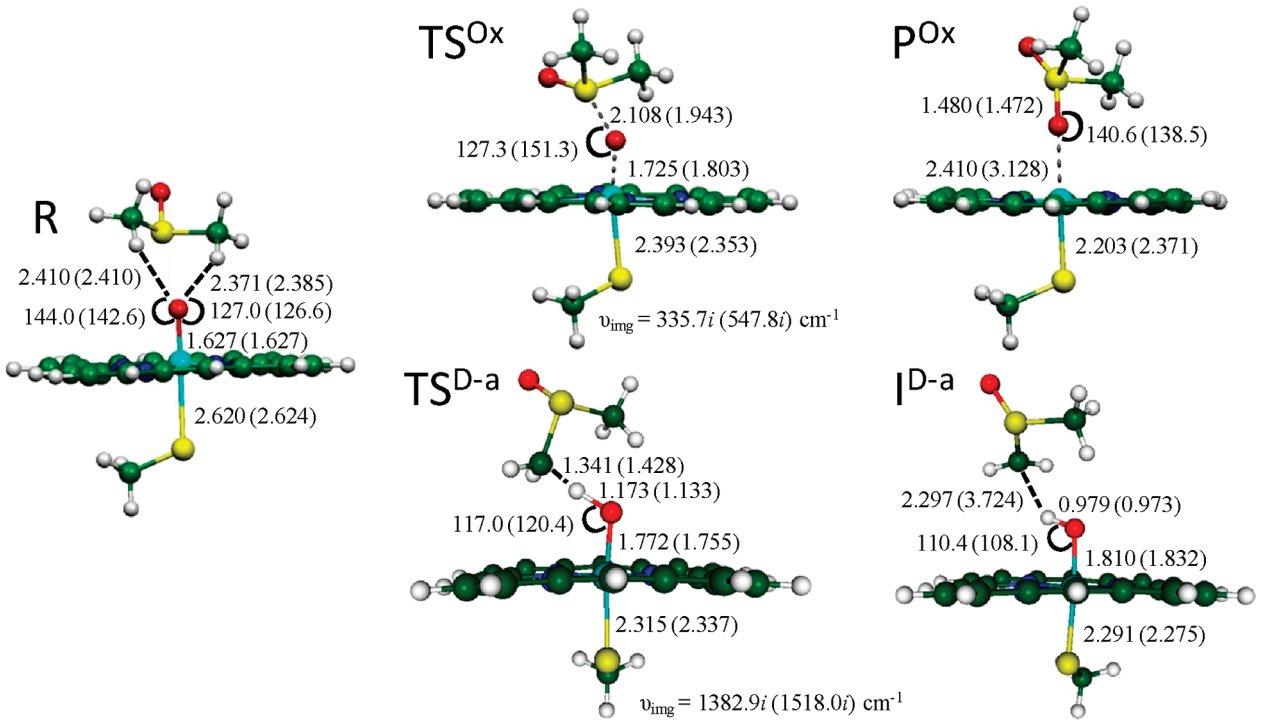


Figure 2. Structures of the various states in the DMSO reactions. Data are given for the doublet (quartet) spin state with distances in ångstroms and angles in degrees. Imaginary frequencies for the transition states are shown on the lower, right side of the respective structure.

energies, and basis set changes were very similar and always of the same sign. We also calculated the structure of compound I with the SCH_3^- and SH^- models in both the doublet and quartet spin states. Again, the geometries were very similar, and the energy splitting between the two spin states differed by less than 0.3 kJ/mol. The full data of this comparison are presented in Tables S2 and S3 in the Supporting Information.

We also tested the effect of amide NH···S_{Cys} interactions in the protein by single-point energy calculations with two added NH_3 groups at the B3LYP/BSI level, as has been done by Ogliaro et al.[37] However, we found that the energies from such calculations are unreliable, because the differences in the C/H_{Cys}−S_{Cys}−Fe−N torsions and the S_{Cys}−Fe−N angles between the reactions and between the various states in the same reaction give varying interactions between the NH_3 groups and the porphyrin ring. Therefore, these results are not presented.

Results and Discussion

Heteroatom oxidation by cytochrome P450 may occur via direct O-atom transfer or an initial electron transfer from the substrate to compound I, followed by the formation of a bond between the heteroatom and the oxygen atom.[2,6] However, we have only studied the former reaction, because experimental results have shown that this is the most likely reaction for sulfoxidation,[38] and previous theoretical investigations have given the same results for both sulfoxidation and nitrogen oxidation.[8–10]

DMSO Oxidation and Dealkylation. Sulfoxides are important groups in drugs and have therefore been thoroughly studied experimentally.[39] They are also interesting because

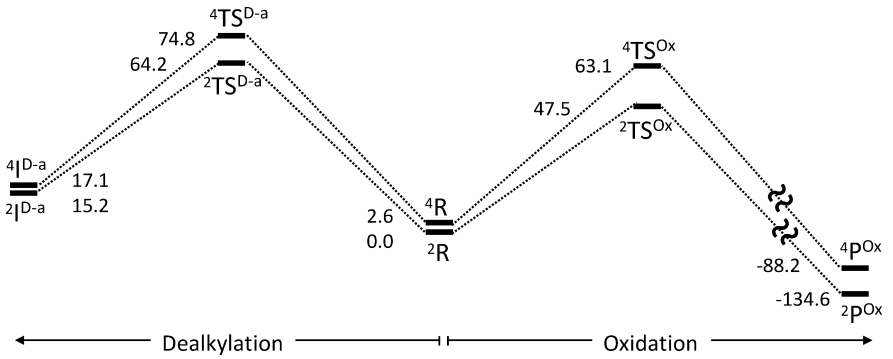


Figure 3. Relative energies (in kJ/mol) for the various states in the reactions with DMS.

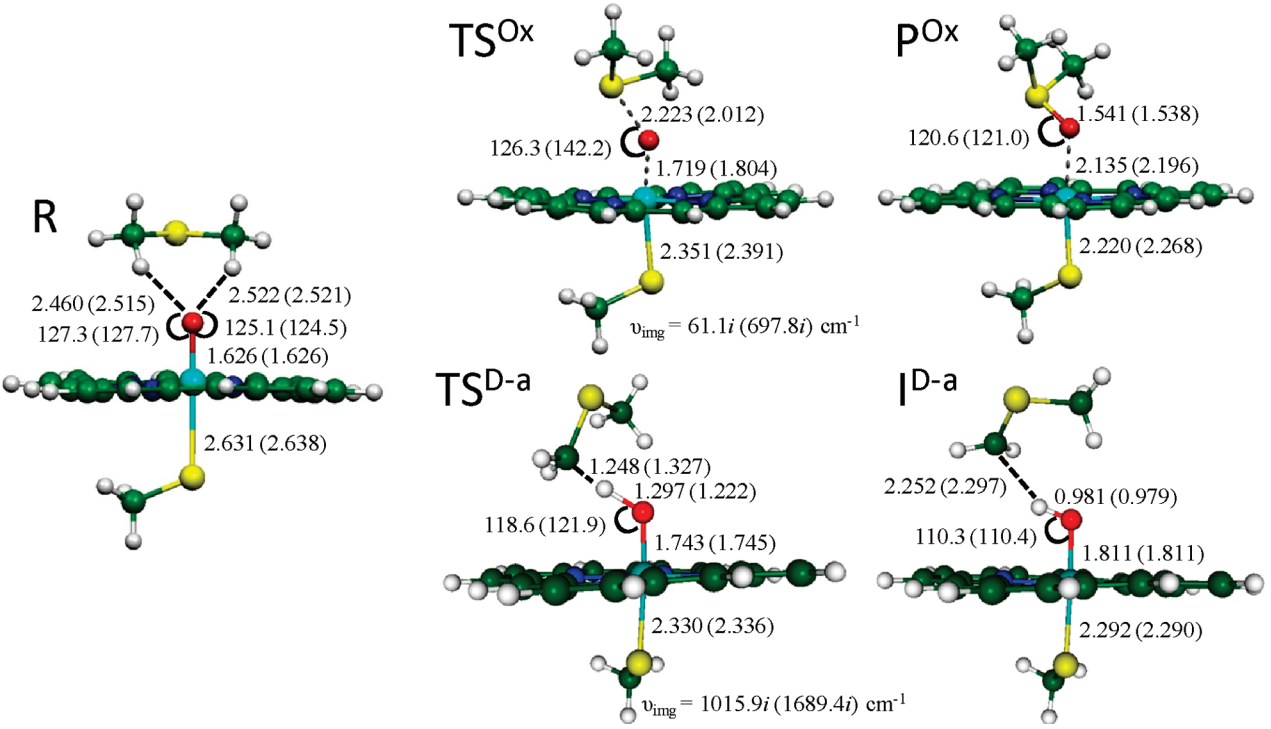


Figure 4. Structures of the various states in the DMS reactions. Data are given for the doublet (quartet) spin state with distances in ångstroms and angles in degrees. Imaginary frequencies for the transition states are shown on the lower, right side of the respective structure.

any oxidation of a sulfide can be followed by a second oxidation of the sulfoxide, resulting in a sulfone. Neither sulfoxide oxidation nor dealkylation seems to have been studied before with theoretical methods. We find that the reactant complex of DMSO and compound I is 28−29 kJ/mol less stable than the isolated substrate and compound I for both the doublet and quartet states (owing to solvation effects and the loss in translational and rotational entropy). The oxidation transition state displays a relatively large splitting (13 kJ/mol) between the doublet (73 kJ/mol) and quartet (86 kJ/mol) states, showing that this reaction is more likely to take place on the doublet spin surface (cf. Figure 1). This is also reflected in the transition-state geometries: The doublet has an early transition state with a S−O distance of 2.11 Å, whereas the quartet transition state is late with a S−O distance of 1.94 Å (cf. Figure 2).

Next, we turn to the dealkylation reaction of DMSO. The activation energies for the hydrogen abstraction step in this reaction display a similar but opposite splitting (9 kJ/mol) between the doublet and quartet states (Figure 1). This difference is caused entirely by the thermal corrections; without them, the difference is only 1 kJ/mol. Moreover, the activation energy is larger for the hydroxylation reaction than for the direct oxidation of DMSO, and the hydroxylation reaction is more likely to occur on the quartet spin surface (91 kJ/mol). The geometries of the transition states are much more similar than in the oxidation reaction. The quartet transition state is only slightly later than the doublet, with O−H distances of 1.13 and 1.17 Å, respectively.

There are two possible conformations of the DMSO molecule in the dealkylation reaction: The oxygen atom in DMSO can either point toward or away from the heme ring. However, the reactant state with the oxygen atom pointing toward heme is 14 kJ/mol less stable than the other conformation. On the other hand, the two conformations gave similar transition-state structures with similar energies for the quartet spin state and only a 6 kJ/mol difference in the

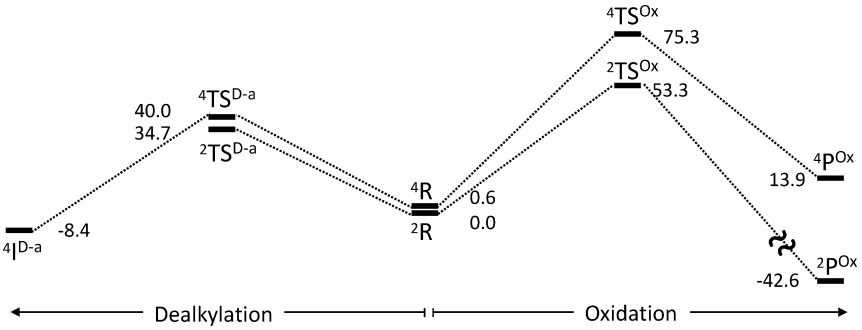


Figure 5. Relative energies (in kJ/mol) for the various states in the reactions with TMA. No energy is given for $^2\mathbf{I}^{\mathbf{D-a}}$, because this state converges directly to the hydroxylated product complex.

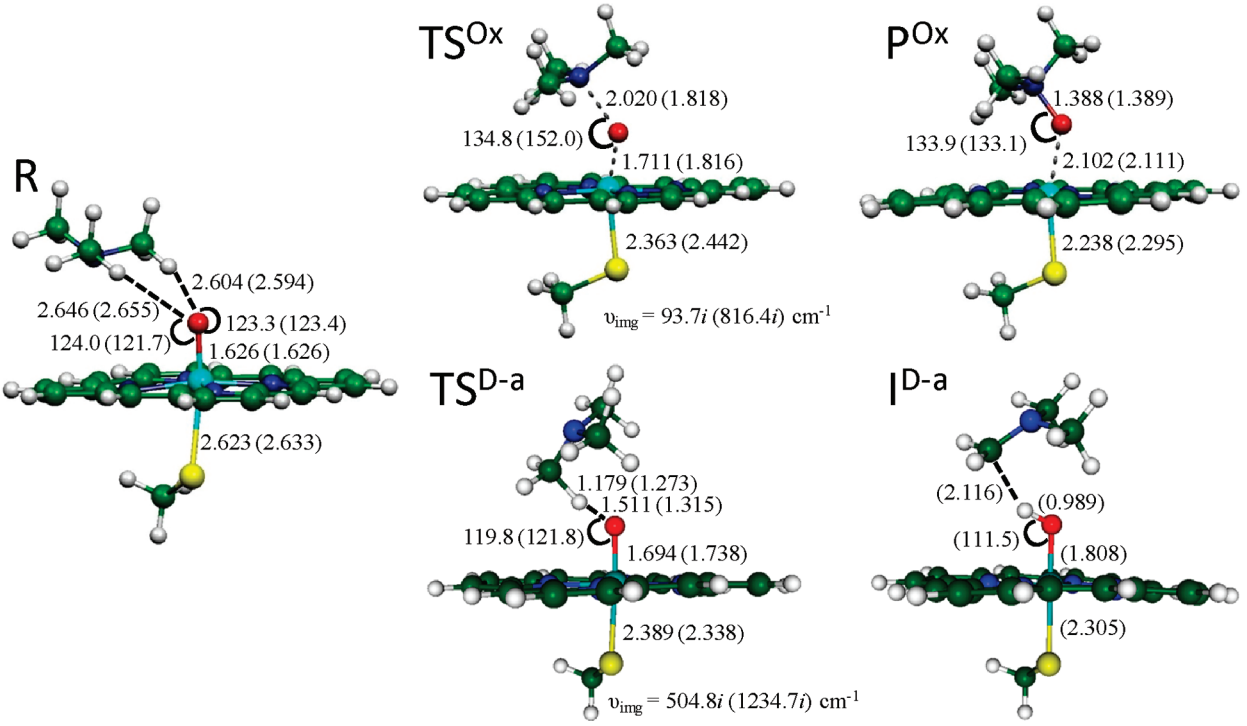


Figure 6. Structures for the various states in the TMA reactions. Data are given for the doublet (quartet) spin state with distances in ångstroms and angles in degrees. Imaginary frequencies for the transition states are shown on the lower, right side of the respective structure.

doublet spin state. Energies for the less stable conformation are shown in Table S4 in the Supporting Information.

DMS Oxidation and Dealkylation. Sulfide reactions are important reactions in drug metabolism by cytochrome P450.[13,14] They have been studied theoretically before[8–10,12,40] (but only the quartet spin state for the dealkylation reaction). The oxidation of DMS is similar to the oxidation of DMSO: The doublet state has lower activation energy (47 kJ/mol) than the quartet state (63 kJ/mol, cf. Figure 3). Thus, the activation energy of the doublet state is 25 kJ/mol lower than that for the DMSO oxidation. This is also reflected in the transition-state structure, in which the S−O distance is 2.22 Å for the doublet (Figure 4), that is, an earlier transition state than that of the sulfoxide oxidation. Moreover, the reactant complex is 41−42 kJ/mol higher in energy than the separated substrate and compound I.

In contrast to DMSO, the activation energy of the DMS dealkylation reaction is lower on the doublet surface (64 kJ/mol) than on the quartet surface (75 kJ/mol, Figure 3). Thus, they are higher than the activation energy for the direct S-oxidation reaction, but lower than those of the corresponding dealkylation step for DMSO. The intermediate in the dealkylation reaction, $\mathbf{I}^{\mathbf{D-a}}$, is also ∼30 kJ/mol more stable than for DMSO. This is because the sulfur atom in DMS has one more available orbital available to accommodate the substrate radical that is formed in this state.

TMA Oxidation and Dealkylation. Most drugs contain nitrogen atoms, which often are metabolized by cytochrome P450. Experimentally, it is found that dealkylation is preferred before N oxidation.[16] The dealkylation reaction (caused by hydroxylation of the C^{α} atom) has been the subject of several theoretical investigations,[9,11,40,41] but the N oxidation has only been studied once before with DFT,[11] namely, for the substrate *N*,*N*-dimethylaniline. Here, we study the dealkylation and N-oxidation reactions for a simpler substrate, TMA, and compare them to the corresponding reactions with DMS and DMSO.

Our results show that the reactant complex is 44−45 kJ/mol less stable than the separated substrate and compound I for both spin states. The activation energy for TMA oxidation is 53 kJ/mol on the doublet surface, but significantly higher for the quartet, 75 kJ/mol (cf. Figure 5). This is also reflected in the transition-state geometries: The doublet has an early transition state with an O−N distance of 2.02 Å, whereas the quartet transition state is late, with an O−N distance of 1.82 Å (Figure 6).

For the corresponding dealkylation reaction, the activation energies of the doublet and quartet differ by 5 kJ/mol, with that of the doublet state having the smaller barrier of 35 kJ/mol. This is lower than the N-oxidation reaction.

Comparison with Experimental Data. Our results for the heteroatom oxidation and dealkylation reactions suggest that sulfides and sulfoxides are more likely to undergo heteroatom oxidation than dealkylation, because the S- and SO-oxidation activation energies are 16−18 kJ/mol lower than that for the corresponding dealkylation reaction. Of course, the reactivity will depend on the properties of the enzyme pocket and the substituents on the substrate, but our results agree well with the experimental observation that CYPs in general degrade S-containing compounds by S oxidation rather than by dealkylation.[9,13–16] It has also been noted that CYPs perform sulfoxidation faster than sulfoxide oxidation.[39] This is in perfect agreement with the 25 kJ/mol lower activation energy for DMS oxidation than for DMSO oxidation (12 kJ/mol relative to the isolated reactants).

There is a similar difference in the activation energies between the N-oxidation and N-dealkylation reactions (19 kJ/mol). However, in contrast to the S-containing compounds, the N-dealkylation reaction has lower activation energy than the N oxidation. This suggests that the N-dealkylation reactions are more likely to occur (especially as different N compounds all give similar hydroxylation barriers[40]), again in excellent agreement with experimental results.[9,13–16]

Rationalization of the Trends. The dealkylation reactions have activation energies ranging from 35 (TMA) to 100 (DMSO) kJ/mol. This large variation can be explained by the ability of the substrate to stabilize the radical that is formed during the reaction. If we plot the percentage of the spin of the substrate on the carbon from which the hydrogen is abstracted against the activation energy, we get correlation coefficients (r^2) of 0.96 or 0.99 for the final free energies or the raw BSI energies, as shown in Figure 7 (the zero-point and thermal corrections lower the correlation somewhat). Thus, it is clear that the type of neighboring atom (i.e., the possibility to delocalize the radical over several atoms) strongly affects the activation energy of the dealkylation reactions, as has been shown before in a different manner.[40]

For the oxidation reactions, the variation of the activation energies between the three substrates is much smaller: For the doublet state, the activation energies are 48−73 kJ/mol relative to the **R** complex, but 89−101 kJ/mol relative to the isolated reactants. Thus, there are hardly any trends to explain.

On the other hand, the preference for the double state of the oxidation reactions can be easily rationalized: Previous

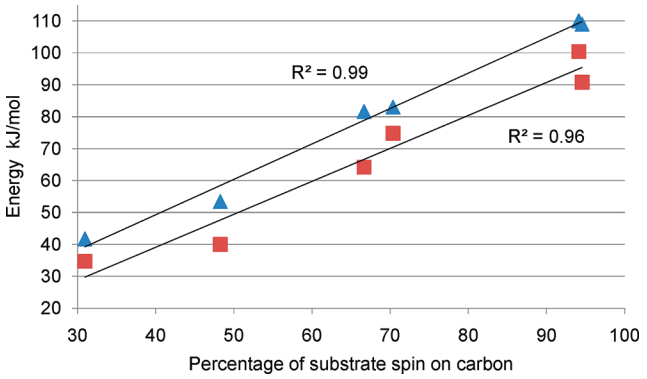
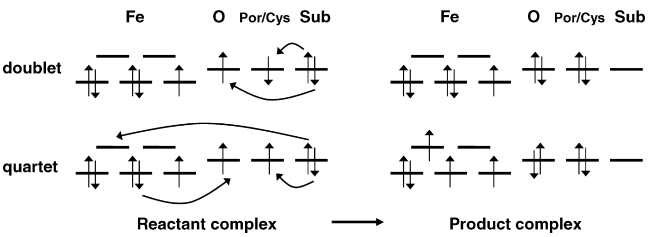


Figure 7. The correlation between the activation energy of the dealkylation reactions and the percentage of the substrate spin that resides on the carbon atom from which the hydrogen atom is abstracted. The red squares represent the final energies, and the blue triangles represent the BSI energy.

Scheme 2. Schematic Picture of the Movement of Electrons in the Heteroatom Oxidation Reaction in the Doublet and Quartet States



DFT studies have shown that compound I essentially contains three unpaired electrons, one localized on the iron ion, one on the oxy group, and one shared by the porphyrin ring and the cysteine model.[6,7,29] In the quartet state, all three spins are parallel, whereas in the doublet state, the latter spin is antiparallel to the other two. Our calculated spin densities in Tables S8−S10 in the Supporting Information confirm this also for the present calculations. In the heteroatom oxidation, two electrons with opposite spin in the same orbital are moved from the substrate to compound I. In the doublet state, this is straightforward, because the two acceptor orbitals (one on the oxy group and the other shared by the porphyrin ring and the cysteine model) are singly occupied with electrons of opposite spins (cf. Scheme 2). However, for the quartet state, the two acceptor orbitals have parallel spin, and one of the electrons must either flip the spin or end up in another orbital (an unoccupied iron 3d orbital), giving an excited product state.

This is reflected in the spin densities (Tables S8−S10, Supporting Information): The doublet transition state has only a small spin on the substrate (∼0.1*e* for DMSO and ∼0.3*e* for the other two substrates), whereas more spin is found on the substrate in the quartet state (∼0.5*e*). Moreover, the spin on the iron ion has increased in the quartet state. This shows that one of the electrons goes mainly to iron, rather than to the oxy group. Consequently, the $^2\mathbf{P}^{ox}$ state is always a pure low-spin Fe(III) state, with essentially no spin on the substrate. However, the quartet $^4\mathbf{P}^{ox}$ state contains an intermediate-spin Fe(III) state (with three unpaired electrons), with 0.2−0.4 unpaired electrons on the cysteine model, but

no spin on the product (including the former oxy group). This shows that one electron has also moved from a doubly occupied Fe orbital to the oxy group.

For the hydroxylation reaction, no such differences arise, because the intermediate involves a radical, which will have the same spin as the unpaired electron in the porphyrin ring in compound I.

Comparison with Previous Studies. As mentioned above, nitrogen oxidation by cytochrome P450 has only been studied once before,[11] with the substrate *N*,*N*-dimethylaniline. These results are quite similar to ours (disregarding the thermal corrections, which were not included in the other study). For example, the Fe−O and O−N distances for oxidation transition states differ by less than 0.03 Å between *N*,*N*-dimethylaniline and TMA (distances for the dealkylation transition state were not given[11]). For the activation energies of the two reactions and the two spin states, the differences are somewhat larger, 4−12 kJ/mol, but not larger than what could be expected from the differences in details of the calculations (especially the cysteine model) and the fact that different substrates were studied.

The oxidation of DMS has previously been studied by Shaik and co-workers.[8,10,12] The two earlier studies[8,10] suggested that the reaction takes place on the quartet spin surface, with a quite high activation energy. However, these results were recently revised, and a new and lower transition state on the doublet spin surface was presented.[12] Energetically, our results agree quite well with the latter results (again disregarding the thermal corrections and also the amide hydrogen-bond model): Their activation energies are 29 and 43 kJ/mol for the doublet and quartet, respectively, which are reasonably close to our energies of 32 and 55 kJ/mol, considering the difference in the cysteine model. The transition-state structures of the quartet states are also quite similar, with differences of 0.02−0.04 Å for the S−O, Fe−O, and Fe−S distances and 3° for the Fe−O−S angle. Likewise, the imaginary frequencies agree within 9 cm^{-1}.

However, some discrepancies are observed for the doublet transition state: Shaik and co-workers observe a S−O distance of 2.39 Å, an Fe−O distance of 1.66 Å, and an imaginary frequency of $1015i$ cm^{-1}, whereas we get a S−O distance of 2.22 Å, an Fe−O distance of 1.72 Å, and a much smaller imaginary frequency of $61i$ cm^{-1}. Therefore, we performed scans of the O−S distance for the DMS oxidation (and also TMA and DMSO oxidation for comparison, Figure S2 in the Supporting Information). These showed a very flat potential surface, explaining our low frequencies. Moreover, we also scanned the Fe−O distance for several different O−S distances around the transition state for the DMS oxidation. This two-dimensional surface is shown in Figure 8 and confirms the flat potential energy surface. We also analyzed the eigenmode of our transition state and confirmed that it corresponds to the correct reaction coordinate (shown in Figure S1 in the Supporting Information). Moreover, we have also tried to locate the transition state reported by Shaik and co-workers by restrained optimizations, using both SCH_3^- and SH^- as the cysteine model. However, we always found a low imaginary frequency (less than

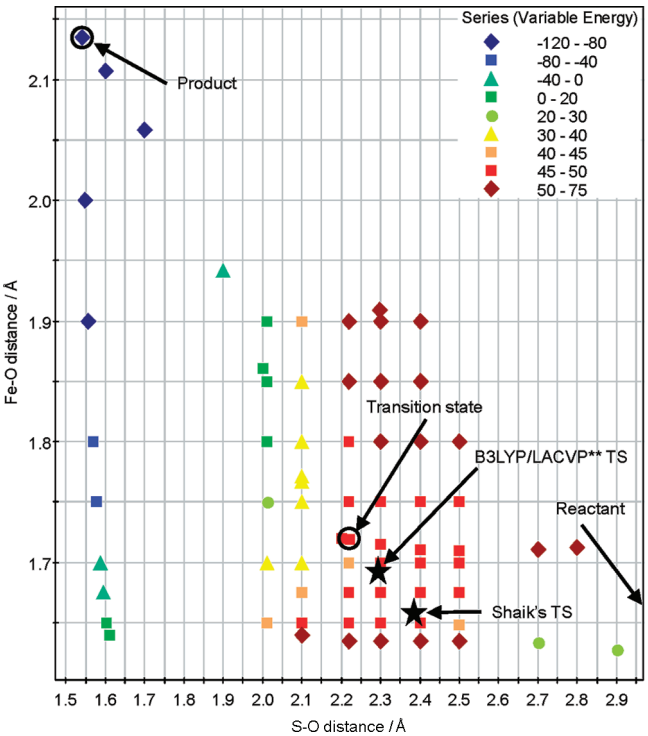


Figure 8. Scatter plot of the energies around the transition state of the DMS oxidation reaction in the doublet spin state. The points are color-coded according to the energies computed with BSI in kilojoules per mole. The geometries of the transition state calculated with B3LYP/LACVP** and by Shaik et al. are shown as black stars. The reactant state has been excluded from the figure to make it clearer. It has a S−O distance of 4.14 Å and an Fe−O distance of 1.63 Å.

$100i$ cm^{-1}), and when the restraints were removed, the geometry converged back to the structure shown in Figure 4. To understand the difference between our calculations, we finally optimized the transition state, with SH^- as the cysteine model, using the LACVP** basis set[42] and the Jaguar software[43] (version 7.0; Shaik and co-workers used this basis set and Jaguar 6.5), as well as the Turbomole software.[36] The results show geometries in between our geometry and the one they have found (as shown in Figure 8), with Fe−O distances of 1.69 and 1.68 Å and O−S distances of 2.28 and 2.31 Å with the Jaguar and Turbomole software, respectively. The imaginary frequencies of these two structures were 124 and $291i$ cm^{-1}. Therefore, we conclude that the two structures most likely represent the same transition state, but the flat potential surface makes it extremely sensitive to variations in the theoretical method. However, it should be characterized by a low imaginary frequency; the frequency reported by Shaik and co-workers is much too large for a reaction involving only heavy atoms, especially on this flat potential surface.

Conclusions

We have studied the oxidation and dealkylation of DMSO, DMS, and TMA with the same DFT approach to obtain a consistent view of these six reactions. The results clearly show that, for sulfur and sulfoxide compounds, heteroatom

oxidation is the preferred pathway (the activation energies for the oxidations of DMSO and DMS are 73 and 47 kJ/mol, respectively, whereas the activation energies for the corresponding dealkylations are 91 and 64 kJ/mol), and it will always take place on the doublet spin surface. For nitrogen-containing compounds, on the other hand, the dealkylation has lower activation energy than the oxidation, 35 and 53 kJ/mol, respectively. The results do not change if possible solvent effects in the protein are considered. These results are in excellent agreement with the experimental findings that S oxidation is more favorable than S dealkylation, that N dealkylation is more favorable than N oxidation, and that heteroatom oxidation is more common for S-containing than for N-containing molecules.[9,13–16] Moreover, the barrier for DMS oxidation is lower than that for DMSO oxidation, again in agreement with experiments.[39]

Thus, our results rationalize all experimental findings using a single criterion, namely, the activation energy for the various reactions, calculated at the DFT level. This is a most important finding, because it shows that the calculated activation energies are consistent and comparable also for different types of reactions. This is mandatory for a predictive method of the intrinsic reactivity of an arbitrary drug candidate with cytochromes P450. Previous methods, based on semiempirical calculations, have had severe problems in this respect.[41] Thus, our calculations indicate that it may be possible to estimate the intrinsic reactivity of any reaction for a drug candidate with DFT methods, and therefore probably also by other, cheaper theoretical methods,[40] an important goal in medicinal chemistry.

Acknowledgment. This work was supported by grants from the Carlsberg Foundation, the Benzon Foundation, and the Swedish Research Council. It has also been supported by computer resources from Lunarc at Lund University.

Supporting Information Available: Comparison of the SCH_3^- and SH^- cysteine models. Energies for the dealkylation of DMSO with the alternative conformation. All energy data for the reactions. Charge and spin distributions of all states studied. Figure of the eigenmodes of the DMS transition state in the doublet spin state. Figures of the S−O or N−O distance scans for all three substrates. Figures of the highest occupied natural orbitals in all the transition states studied. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Guengerich, F. P. Human Cytochrome P450 Enzymes. In *Cytochrome P450 Structure, Mechanism, and Biochemistry*; 3rd ed.; Ortiz de Montellano, P. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2004; pp 377−531.

(2) Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* **2001**, *14*, 611–650.

(3) Bertz, R. J.; Granneman, G. R. Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions. *Clin. Pharmacokinet.* **1997**, *32*, 210–258.

(4) Cojocaru, V.; Winn, P. J.; Wade, R. C. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta, Gen. Subj.* **2007**, *1770*, 390–401.

(5) Rydberg, P.; Rod, T. H.; Olsen, L.; Ryde, U. Dynamics of water molecules in the active-site cavity of human cytochromes P450. *J. Phys. Chem. B* **2007**, *111*, 5445–5457.

(6) Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* **2004**, *104*, 3947–3980.

(7) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* **2005**, *105*, 2279–2328.

(8) Kumar, D.; de Visser, S. P.; Sharma, P. K.; Hirao, H.; Shaik, S. Sulfoxidation mechanisms catalyzed by cytochrome P450 and horseradish peroxidase models: Spin selection induced by the ligand. *Biochemistry* **2005**, *44*, 8148–8158.

(9) Loew, G. H.; Chang, Y. T. Theoretical-studies of the oxidation of N-containing and S-containing compounds by cytochrome-P450. *Int. J. Quantum Chem.* **1993**, *48*, 815–826.

(10) Sharma, P. K.; de Visser, S. P.; Shaik, S. Can a single oxidant with two spin states masquerade as two different oxidants? A study of the sulfoxidation mechanism by cytochrome P450. *J. Am. Chem. Soc.* **2003**, *125*, 8698–8699.

(11) Cho, K. B.; Moreau, Y.; Kumar, D.; Rock, D. A.; Jones, J. P.; Shaik, S. Formation of the active species of cytochrome P450 by using iodosylbenzene: a case for spin-selective reactivity. *Chem.—Eur. J.* **2007**, *13*, 4103–4115.

(12) Li, C.; Zhang, L.; Zhang, C.; Hirao, H.; Wu, W.; Shaik, S. Which oxidant is really responsible for sulfur oxidation by cytochrome P450. *Angew. Chem., Int. Ed.* **2007**, *46*, 8168–8170.

(13) Afzelius, L.; Arnby, C. H.; Broo, A.; Carlsson, L.; Isaksson, C.; Jurva, U.; Kjellander, B.; Kolmodin, K.; Nilsson, K.; Raubacher, F.; Weidolf, L. State-of-the-art tools for computational site of metabolism predictions: comparative analysis, mechanistical insights, and future applications. *Drug Metab. Rev.* **2007**, *39*, 6186.

(14) Bu, H. Z. A literature review of enzyme kinetic parameters for CYP3A4-mediated metabolic reactions of 113 drugs in human liver microsomes: structure-kinetics relationship assessment. *Curr. Drug Metab.* **2006**, *7*, 231–249.

(15) de Montellano, P. R. O.; De Voss, J. J. Oxidizing species in the mechanism of cytochrome P450. *Nat. Prod. Rep.* **2002**, *19*, 477–493.

(16) Seto, Y.; Guengerich, F. P. Partitioning between N-dealkylation and N-oxygenation in the oxidation of N,N-dialkylarylamines catalyzed by cytochrome-P450-2B1. *J. Biol. Chem.* **1993**, *268*, 9986–9997.

(17) Becke, A. D. Density-functional thermochemistry. 3. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(18) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(19) Lee, C. T.; Yang, W. T.; 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.

(20) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin-density calculations - a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(21) Schafer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian-basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(22) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-consistent molecular-orbital methods. 12. Further extensions of Gaussian-type basis sets for use in molecular-orbital studies of organic-molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(23) Hariharan, P. C.; Pople, J. A. Influence of polarization functions on molecular-orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(24) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. Self-consistent molecular-orbital methods. 23. A polarization-type basis set for 2nd-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

(25) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular-orbital methods. 25. Supplementary functions for Gaussian-basis sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269.

(26) Rulisek, L.; Jensen, K. P.; Lundgren, K.; Ryde, U. The reaction mechanism of iron and manganese superoxide dismutases studied by theoretical calculations. *J. Comput. Chem.* **2006**, *27*, 1398–1414.

(27) Strickland, N.; Harvey, J. N. Spin-forbidden ligand binding to the ferrous-heme group: Ab initio and DFT studies. *J. Phys. Chem. B* **2007**, *111*, 841–852.

(28) Liao, M. S.; Watts, J. D.; Huang, M. J. Assessment of the performance of density-functional methods for calculations on iron porphyrins and related compounds. *J. Comput. Chem.* **2006**, *27*, 1577–1592.

(29) Rydberg, P.; Sigfridsson, E.; Ryde, U. On the role of the axial ligand in heme proteins: a theoretical study. *J. Biol. Inorg. Chem.* **2004**, *9*, 203–223.

(30) Sharp, K. A.; Honig, B. Electrostatic interactions in macromolecules - theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301–332.

(31) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144–1149.

(32) Schutz, C. N.; Warshel, A. What are the dielectric "constants" of proteins and how to validate electrostatic models. *Proteins* **2001**, *44*, 400–417.

(33) Klamt, A.; Schuurmann, G. Cosmo - a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799–805.

(34) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. Refinement and parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.

(35) Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons: Chichester, U. K., 1999; pp 292−304.

(36) Treutler, O.; Ahlrichs, R. Efficient molecular numerical-integration schemes. *J. Chem. Phys.* **1995**, *102*, 346–354.

(37) Ogliaro, F.; Cohen, S.; de Visser, S. P.; Shaik, S. Medium polarization and hydrogen bonding effects on compound I of Cytochrome p450: what kind of a radical is it really. *J. Am. Chem. Soc.* **2000**, *122*, 12892–12893.

(38) Baciocchi, E.; Lanzalunga, O.; Pirozzi, B. Oxidations of benzyl and phenethyl phenyl sulfides. Implications for the mechanism of the microsomal and biomimetic oxidation of sulfides. *Tetrahedron* **1997**, *53*, 12287–12298.

(39) Alvarez, J. C.; Demontellano, P. R. O. Thianthrene 5-oxide as a probe of the electrophilicity of hemoprotein oxidizing species. *Biochemistry* **1992**, *31*, 8315–8322.

(40) Olsen, L.; Rydberg, P.; Rod, T. H.; Ryde, U. Prediction of activation energies for hydrogen abstraction by cytochrome P450. *J. Med. Chem.* **2006**, *49*, 6489–6499.

(41) Jones, J. P.; Mysinger, M.; Korzekwa, K. R. Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction. *Drug Metab. Dispos.* **2002**, *30*, 712.

(42) Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations - potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* **1985**, *82*, 299–310.

(43) *Jaguar*, version 7.0; Schrodinger LLC: New York, 2007.

CT800101V

JCTC Journal of Chemical Theory and Computation

Topologically Based Multipolar Reconstruction of Electrostatic Interactions in Multiscale Simulations of Proteins

Michele Cascella,*,†,⊥ Marilisa A. Neri,† Paolo Carloni,§,|| and Matteo Dal Peraro‡

Laboratory of Computational Chemistry and Biochemistry and Laboratory for Biomolecular Modeling, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, International School for Advanced Studies (SISSA/ISAS) and CNR-INFM-DEMOCRITOS, I-34014 Trieste, Italy, and Italian Institute of Technology, Italy

Received April 07, 2008

Abstract: We present a new method to incorporate electrostatic interactions in coarse-grained representations of proteins. The model is based on a topologically reconstructed multipolar expansion of the all-atom centers of charge, specifically of the backbone dipoles and the polar or charged side chains. The reliability of the model is checked by studying different test cases, namely protein−cofactor/substrate interactions, protein large conformational changes, and protein−protein complexes. In all cases, the model quantitatively reproduces the all-atom electrostatic field in both a static and a dynamic framework. The model is of general applicability and can be used to improve both full coarse-grained simulations and hybrid all-atom/coarse-grained multiscale approaches.

1. Introduction

In past years, atomistic-detailed molecular dynamics (MD) simulations proved to be a reliable method for description of events occurring in the $10^{-12}-10^{-9}$ s time scales.[1] Although the increase in computational power allows application of such technique to larger systems, and for longer times, it remains in general too expensive to investigate dimensional and dynamical scales that are critical to most of the biological processes both *in vitro* and *in vivo*. In fact, all fundamental biological processes necessary to life (e.g., protein folding, signal transduction, DNA transcription), which are triggered by interactions at atomistic dimensionality, occur at very different time scales (from femtoseconds to seconds and even longer) and span over different sizes (from few tens to millions of atoms).[2] Therefore, vast dimensional and temporal scales should be taken into account.

To address such size/time scale issues various coarse-grained (CG) Hamiltonians have been recently developed for systems like membranes,[3–8] proteins,[9–16] and DNA[17–19] and were successfully applied to problems of great biological relevance, from membrane self-assembly and dynamics [5,7,8] to protein folding.[20–22] The drawback of CG approaches lies in the lack of an atomistic-detailed description, which is crucial when studying phenomena involving molecular recognition (e.g., receptor−ligand binding). These interactions are of fundamental importance for establishing functional protein networks entitled to control complex metabolic pathways.[2] On the experimental side, proteomics techniques are progressing very fast in enabling the characterization of such challenging systems; however, what is still missing for a complete comprehension of the biological function is the atomistic picture of these assemblies. The only accessible way to get to this level, having the opportunity to develop

* Corresponding author e-mail: michele.cascella@epfl.ch.

† Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne.

⊥ Present Address: Universität Bern, Department für Chemie und Biochemie, Freiestrasse 3, CH-3012 Bern, Switzerland.

§ International School for Advanced Studies (SISSA/ISAS) and CNR-INFM-DEMOCRITOS.

‡ Laboratory for Biomolecular Modeling, Ecole Polytechnique Fédérale de Lausanne.

|| Italian Institute of Technology.

10.1021/ct800122x CCC: $40.75
Published on Web 07/12/2008

strategies to interfere with these networks, is through current structural biology techniques. For this reason, the possibility to computationally predict the molecular interactome is very desirable for the advance of life sciences.[23–25]

Recently, considerable efforts have been put into development of novel multiscale simulation techniques that are able to couple atomistic Hamiltonians to CG models (MM/CG).[8,26–41] In particular, it has been shown that hybrid schemes, where only a portion of the system is treated at the atomistic-detailed level, can be implemented and successfully applied to proteins[30,42–44] and protein-DNA complexes.[29,46] A key issue when building hybrid multiscale models lies in retaining an accurate description of long-range interactions. Such long-range interactions can significantly influence the polarization and the structural orientation at an atomistic level and, thus, can be functional to the protein biological activity.[47] Therefore, hybrid MM/CG schemes should somehow incorporate such potentials between the CG part and the all-atom portion. In particular, long-range electrostatic interactions, crucial in many biological processes, are the most difficult to model. In fact, the electrostatic field of biopolymers is dominated by their total charge, which is usually nonzero, and varying with pH and ionic strength, and by a dipolar component, which is a function of both the residue sequence and its instantaneous spatial organization. Specifically, the secondary structure elements of a folded protein can induce strong dipolar fields by aligning the backbone dipoles (e.g., in helical structures). This feature is of great importance, for example, when assessing the topology of a helix-bundle or when determining the correct orientation of a molecule during dynamic ligand−receptor binding studies. Therefore, CG long-range interactions should be modeled to have a structure-dependent asymptotic behavior. Such behavior is in general nontrivial to reproduce by force field-like CG models.[9] For example, CG Hamiltonians based on effective potentials (typically, potentials of mean force) vanish as a power of six with the distance.[3,4]

Recently developed coarse-grained potentials, where specific polar regions are introduced (see e.g. refs 48 and 49), have been proposed. Interactions among polar groups still take a radial form, and, therefore, they are not able to reproduce the directionality of the backbone dipolar field. Therefore, to represent the different behavior of the backbone in different structural local environments, the polarity of the backbone centroid is defined according to the secondary structure of the amino acid. Such an approach is able to reproduce well the secondary structure elements of a protein but, nonetheless, biases the potential at the beginning of the run, forcing each amino acid to keep the initial secondary structure geometry and thus reducing its universality. A RESP[50] expansion of the all-atom electrostatic field on all the CG centroids represents an alternative approach to such a problem.[16] This scheme is particularly suited to improve CG-based protocols for rigid docking. Unfortunately, such an approach cannot be straightforwardly applied when treating hybrid multiscale systems, where different dimensionalities are present and interacting in the simulation. In particular, MD schemes based on RESP-based expansions of all-atom based electrostatic potentials on CG centroids suffer some inconsistencies, as shown later in this paper.

In the past few years, Voth and co-workers have shown that coarse-grained potentials can be rigorously determined from atomistic scale interactions through a force-matching approach.[51] Such a procedure implies an initial all-atom classical MD run, over which the CG potential is fitted to reproduce all-atom forces.[33,34] The CG Hamiltonian obtained by such a scheme is not based on thermodynamic quantities and, therefore is, in general, unbiased toward the specific simulated thermodynamic conditions, and it is in principle able to correctly describe long-range interactions. The method has been successfully applied to construct models of lipid bilayers,[33,52] cholesterol-lipid interactions,[34] monosaccharides,[53] and distinct peptide structural moieties.[54,55] In practice, the quality of such a coarse-graining procedure is determined by the efficacy of the underlying all-atom MD to sample the relevant phase space of the system.[34] This may be more stringent when studying large macromolecular complexes, which present long characteristic time scales.

In this work, we propose an accurate and robust scheme to evaluate electrostatic interactions in hybrid MM/CG simulations of proteins.[30] The proposed model uses topological information coming from experimental structures and from computational molecular models to reconstruct the electrostatic field for both contributions coming from the backbone and the side chains of a protein. Our method is not computationally expensive and is consistent with respect to protein motion. Moreover, it allows a reliable reconstruction of the electrostatic interactions both at an atomistic and a CG scale, allowing applications to different biological systems for the efficient investigation of crucial processes as ligand binding recognition or protein and nucleic acids network dynamics.

2. Computational Methods

The electrostatic field of a biomolecule is univocally determined by its tridimensional structure. In proteins, the electrostatic potential can be further divided into two separate contributions, coming from the backbone peptide bonds (V_{bb}) and from the polar or charged side chains (V_{sc}):

$$V_{el} = V_{bb} + V_{sc} \tag{1}$$

The presented model reconstructs the all-atom electrostatic field of a protein using the tridimensional coordinates of the C_α atoms only, which is some structural information present in most of the CG models for proteins. It will be referred to as *topologically based multipolar reconstruction* (TBMR) hereafter.

2.1. Backbone Contribution. The electrostatic potential generated by the backbone units at point $\vec{R}$ can be expressed as a sum of dipolar contributions

$$V_{bb}(\vec{R}) = \frac{1}{4\pi\varepsilon_0}\sum_i \frac{\vec{\mu}_i \cdot \vec{R}}{|\vec{R}|^3} \tag{2}$$

where the *-i* index runs over the backbone units of the protein, and $\{\vec{\mu}_i\}$ are the electrostatic dipoles associated with them. The position of the C_β of an amino acid can be

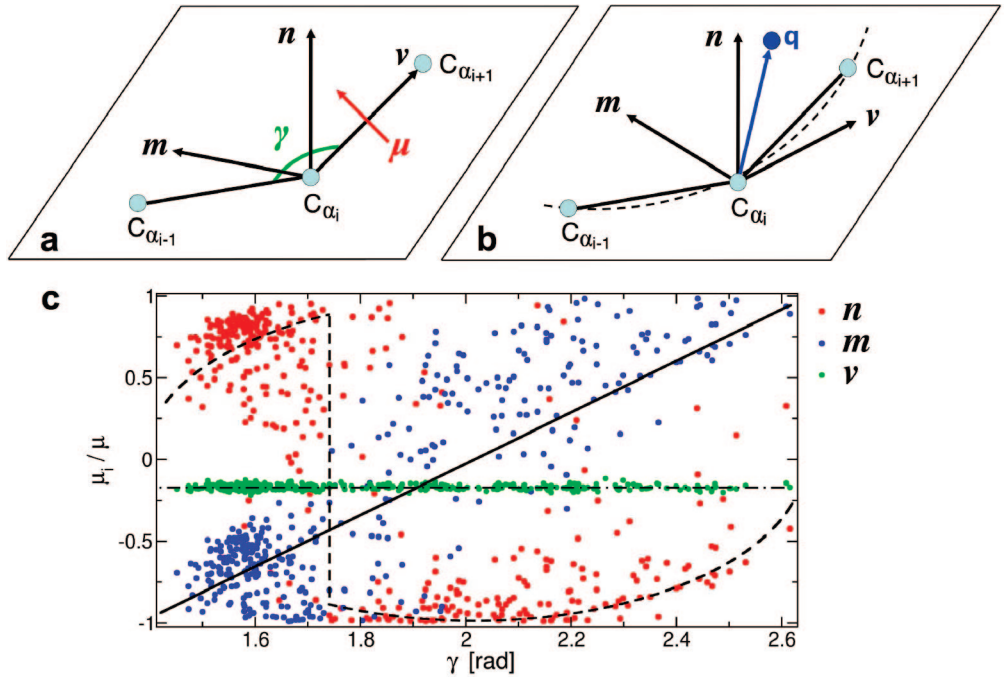


Figure 1. Definition of the internal reference systems. Panels **a** and **b**: internal reference systems used for the backbone and side-chain reconstruction, respectively. Panel **c**: statistical distribution (extracted from the PDB databank) of the projection of the backbone dipole over the vectors *v* (in green), *n* (in red), and *m* (in blue). The black lines are representing the proposed fits for the distributions.

determined, within good approximation, starting from the positions of three consecutive C_α's in the amino acids sequence.[56] Such an idea can be extended with little effort to the orientation of the peptide bond linking two amino acids. In fact, the positions of three consecutive C_α's, their chirality, and the planarity of the peptide moiety provide, in principle, a sufficient number of geometrical constraints to determine the peptide bond geometry. In practice, a good orientation of the backbone peptides can be conveniently achieved by mining statistical orientations from the PDB databank.[57] First, one has to define an internal reference system (Figure 1a) related to three consecutive amino acids. In particular, the following system is chosen: the origin is placed at C_{α_i}, the α-carbon of the second amino acid of the triplet; the three axes are defined by the unitary vectors: $\hat{\mathbf{v}}$ is the direction C_{α_i}-$C_{\alpha_{i+1}}$, $\hat{\mathbf{n}}$ is the normal to the plane defined by the three consecutive $C_{\alpha_{i-1}}$-C_{α_i}-$C_{\alpha_{i+1}}$ C_α's, and $\hat{\mathbf{m}}$ is perpendicular to the first two (Figure 1a). In such a reference system, the electrostatic dipole $\vec{\mu}_i$, associated with the peptide bond between amino acids *i* and *i*+1 is well defined: in fact, the projection of $\vec{\mu}_i$ over $\hat{\mathbf{v}}$ is, by construction, constant; the projection of $\vec{\mu}_i$ over $\hat{\mathbf{m}}$ can be fitted by a linear function of the angle γ formed by the $C_{\alpha_{i-1}}$-$C_{\alpha_{i+1}}$ atoms (Figure 1); and finally, the projection of $\vec{\mu}_i$ over $\hat{\mathbf{n}}$ is determined according to the cosine rule

$$\mu_n = \pm|\vec{\mu}_i|\left[1-\left(\frac{\mu_v}{|\vec{\mu}_i|}\right)^2-\left(\frac{\mu_m}{|\vec{\mu}_i|}\right)^2\right]^{1/2} \tag{3}$$

where $\mu_{n,m,v}$ are the projections of the dipole along the $\hat{\mathbf{n}}$, $\hat{\mathbf{m}}$, and $\hat{\mathbf{v}}$ axes, respectively. The sign of μ_n can be derived by the statistical distribution in the PDB databank (Figure 1c). In our model, it is defined as positive for $\gamma < 1.72$ *rad* and negative for $\gamma \geq 1.72$ *rad*. Following such a scheme, the electrostatic field produced by the backbone peptides in a CG region is easily reconstructed by simply determining the angles formed by all triplets of consecutive C_α's in the protein sequence.

2.2. Side-Chain Contribution. CG models usually lack structural information about the side-chain position and orientation. Therefore, it is necessary to use a serial multiscale procedure to build it. Starting from an initial information (e.g., X-ray or NMR data), we make the plausible assumption that, in a folded protein, the close packing forces the side-chain orientations to remain, on average, unchanged over time.

In this case, one can determine the side-chain positions at any time by (i) imposing that the motion of the side chain is solidal to a properly chosen internal coordinate system (defined below) and (ii) calculating the instantaneous rototranslations undergone by such an internal reference system.[58] Similarly to what is done for the backbone dipoles, the internal reference system (Figure 1b) for the side chain of an *i*th amino acid is defined according to the positions of the $C_{\alpha_{i-1}}$-C_{α_i}-$C_{\alpha_{i+1}}$ triplets. In this case, the unitary vector $\hat{\mathbf{v}}$ is defined as the vector tangent in C_{α_i} to the circle defined by the three consecutive C_α's[58] (Figure 1b).

The electrostatic potential produced by each side chain can be expressed by a multipolar expansion

$$V_{sc}(\mathbf{R}) = \frac{1}{4\pi\varepsilon_0}\left[\frac{q}{|\mathbf{R}|}+\frac{\vec{\mu}\cdot\mathbf{R}}{|\mathbf{R}|^3}+O(quadrupole)+...\right] \tag{4}$$

where $q = \sum_i q_i$ is the total charge of each side chain. The geometrical center of the single multipolar expression is defined as the center of charge $\mathbf{C}_q$ of the side chain

$$\mathbf{C}_q = \frac{\sum_i |q_i| \mathbf{r}_i}{\sum_i |q_i|} \quad (5)$$

where $\mathbf{r}_i$ are the coordinates of the side-chain atoms, and q_i are their respective MM charges. The single centers of charge and the respective multipolar contributions to the electrostatic potential are then defined with respect to their relative internal coordinates ($\hat{\mathbf{v}}$, $\hat{\mathbf{n}}$, $\hat{\mathbf{m}}$). All these data have to be evaluated on the initial all-atom structure. In practice, it is more convenient to get this set of data from an average of structures, e.g. generated from a MD run. In such a case, the multipolar expansion will be averaged on the typical mobility of each side chain. Thus, it will represent the electrostatic mean-field produced by the side chains.

In the current work, only highly polar and charged side chains are explicitly considered, and their multipolar expansion is truncated to the dipole term. This is due to the particular form of the electrostatic potential of the GROMOS96 force field,[59] which is the one implemented in the MM/CG Hamiltonian developed and used in this work. In fact, the charge-fitting protocol used in GROMOS96 makes use of charge-groups, which, in turn, make the all-atom electrostatic potential dominated by dipolar terms. Moreover, in the GROMOS96 force field, apolar/aliphatic side chains do not contribute to the electrostatic potential.

The reconstruction protocol is anyway general and can be applied with little effort to other force fields, although expansion to higher order terms (e.g., quadrupoles) may be necessary to accurately describe the electrostatic field.

The computational advantage of our scheme is directly related to the reduced number of interactions to be computed. In fact, within such a protocol, the total number of dipoles corresponds to the number of peptide bond moieties (≈number of amino acids) plus the number of polar side chains (statistically ≈0.5 times the number of amino acids).[57] Given that the total number of atoms in a protein is statistically ≈15.7 times larger than its number of amino acids,[57] our scheme is at least ≈9.5 times faster in evaluating solute−solute interactions. Such an estimate is done considering electrostatic interaction evaluated in both cases via commonly used Nlog N schemes (where N is the number of particles), like Particle Mesh Ewald algorithms.[60] Moreover, coarse-grained schemes are usually coupled to implicit solvent methods or zero-charge CG water models,[5,36,48] while all-atom simulations suffer for solvent-solvent electrostatic interaction evaluations, which can typically account for twice the CPU time of solute−solute calculations. Therefore, in a rough estimation, TBMR can be up to 30 times faster than all-atom MD. Importantly, we finally note that the routines needed to reconstruct the spatial topology of the dipoles require negligible CPU time with respect to the electrostatic calculations.

3. Results and Discussion

TBMR is applied to a set of systems, representing different test situations in which long-range electrostatics interactions can be relevant for the overall energetics. Namely, we address the performance and accuracy of TBMR in describing the following: (1) a protein-induced electrostatic field at binding/active sites, (2) an electrostatic field tuning upon conformational changes in the secondary structure, and (3) long-range electrostatic interactions at the interface of protein complexes.

3.1. Protein-Induced Electrostatic Field at Binding/Active Sites. Two representative proteins have been chosen−actin[61,62] and azurin[63]−to test TBMR using a hybrid all-atom/coarse grained model (MM/CG). As proposed in the original formulation of MM/CG,[30] the (re)active center can be described using a classical force field, whereas the rest of the system is treated using a CG scheme.

Actin is a globular protein that polymerizes to form helical filaments constituting the cytoskeleton. ATP hydrolysis drives the filament dynamics inducing the structural change in the protein.[61,64] ADP/ATP have a well-known binding site (Figure 2a) located in the core of the protein,[61] which has peculiar electrostatic interactions with actin residues.[61,62] This specific protein is meant to be representative of the vast class of proteins and enzymes where the ligand-protein interactions are the crucial ingredient for ligand recognition and catalytic activation.[47]

Azurin from *Pseudomonas aeruginosa*[63] is known to be shaped in such a way that the long-range electrostatic contribution of the protein frame is crucial for the correct tuning of the redox properties of the enzyme.[65] On a general basis, the secondary/tertiary structure motifs often have a direct influence on the electronic properties of the reactive centers in metalloenzymes.[66,67] Here, azurin is chosen as a representative target for the broad class of metalloproteins, where long-range electrostatic interactions play a crucial functional role.

The crystallographic structures of actin and azurin (PDB codes 2HF3[68] and 4AZU,[63] respectively) are used as initial models to calculate the all-atom electrostatic potential (V_{aa}) in implicit solvent (PB scheme[69,70]), which is used as a reference. Hydrogen atoms and terminal and missing side-chain atoms are reconstructed following standard protocols (see e.g. ref 71). The GROMOS96 force field is used to describe the system in the all-atom calculations. In the MM/CG calculations, the (re)active centers along with neighbor residues are described at the MM level; an interface region linking MM and CG, defined as all the residues within a 6.5 Å radius from the MM part, is described at the intermediate layer following ref 30; and the remaining part of the protein is treated at the CG level. The calculated electrostatic potential (V_{hybrid}) produced by the model on the active site agrees well with the reference as shown in Figure 2b,d. From a quantitative standpoint, the weighted deviation between the all-atom and the MM/CG field $[(|\mathbf{E}|_{hybrid} - |\mathbf{E}|_{aa})/|\mathbf{E}|_{aa})]$ is calculated for all the values of the electric field, and the covariance is used for characterizing the correlation between the vectors $\mathbf{E}_{aa}$ and $\mathbf{E}_{hybrid}$ $[\mathbf{C} = (\langle \mathbf{E}_{aa} \cdot \mathbf{E}_{hybrid} \rangle)/(\langle \mathbf{E}_{aa} \rangle \langle \mathbf{E}_{hybrid} \rangle)]$ for the possible values of the electric field in the different systems (Figure 2b,d). It is clear that the error on the module of the electric field made by TBMR is non-negligible only for very low absolute values of the field. The same trend is found for the vector covariance: except for low values of the field, the covariance is very close to one, showing also

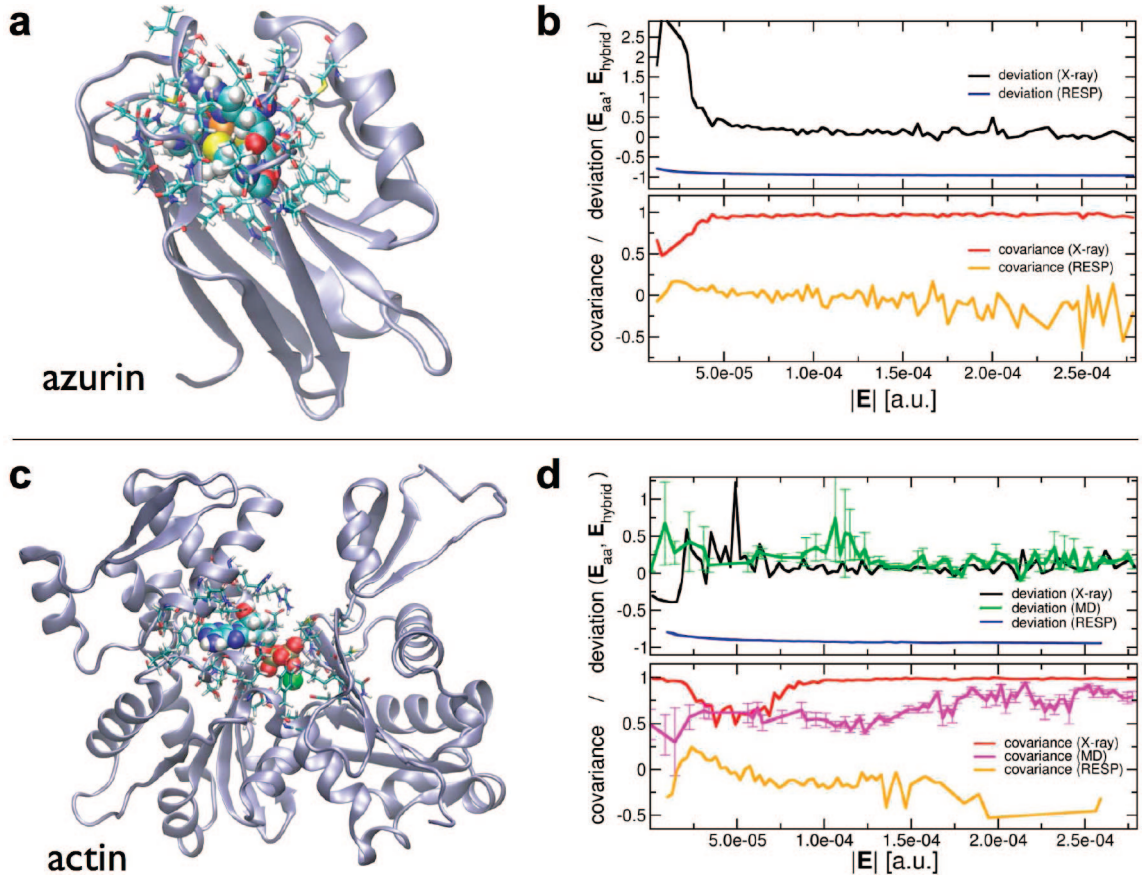


Figure 2. Reconstruction of the electrostatic field in the binding sites of azurin and actin. Panels **a** and **c**: cartoon representation of azurin and actin folding determinants, the active pockets are drawn in ball-and-stick and represent the all-atom portion in the MM/CG calculations. Panels **b** and **d**: the accuracy of the multipolar scheme is compared with all-atom simulations of azurin (b) and actin (d). Deviation and covariance between $\mathbf{E}_{aa}$ and $\mathbf{E}_{hybrid}$ are reported in the graphs for different values of the electrostatic field.

a striking match of the mutual orientation of the fields produced by the all-atom calculation and the TBMR model.

Our model is able to capture well the intensity and orientation of the all-atom electrostatic field. The usefulness of such model is straightforward for rigid docking studies, where the crystallographic structure of the protein is used as such. In fact, the TBMR model keeps its reliability also when the protein structure is allowed to relax and fluctuate, e.g. by MD. We have tested the performance of the model in such conditions by selecting snapshots along an equilibrated MD simulation of actin (one snapshot every 1 ns, averaged over 20 ns trajectory) and performing the same analysis on the electrostatic fields (Figure 2d). We find that the TBMR model produces, also in this case, good results both in terms of field intensity and orientation. The minor performance with respect to a fixed geometry relies on the fact that, in the case of moving structures, the side-chains contribution to the electrostatic field is generated by their average structure coming from an independent MD run. Therefore TBMR is expected to reproduce the average features of the electrostatic field but not to catch its instantaneous fluctuations as such.

Finally, the TBMR method is compared to the performance of a RESP-charges fitting. In this model, the all-atom electrostatic potential produced by the region of the protein that falls in the CG region is fitted by RESP-charges centered on each CG centroid. In particular, we have tested both RESP fittings assuming CG models comprise C_α centroids only or C_α and C_β centroids per amino acid. We find that the RESP fitting has an intrinsic inconsistency at different scales. In fact, when the RESP charges are fitted to reproduce the all-atom electrostatic potential on a large grid around the whole protein, we find that the electrostatic field in the smaller MM region is not even comparable to the all-atom one. If the RESP fitting is, on the contrary, performed on a grid centered at the MM region, the field there is more reliable. Nonetheless, from the graphs in Figure 2d, it is evident that even in this case RESP charges are unable to represent either the module or the orientation of the electric field at the static crystallographic structure and during the MD simulation. Such behavior (RESP systematically underestimates the field) is related to the nature of the RESP electrostatic potential. In fact, the electrostatic all-atom potential, which is generated by a series of electrostatic monopoles (the atomic charges), is fitted by a smaller number of electrostatic monopoles (the RESP centers). Moreover the spatial distribution of the RESP charges is not fully consistent with that of the all-atom charge groups. As a result, RESP-fitting can catch at best an average of the all-atom potential in a specific region, without being able to describe its local fluctuations, which lead to an underestimation of the electrostatic field and its nonaccurate alignment.

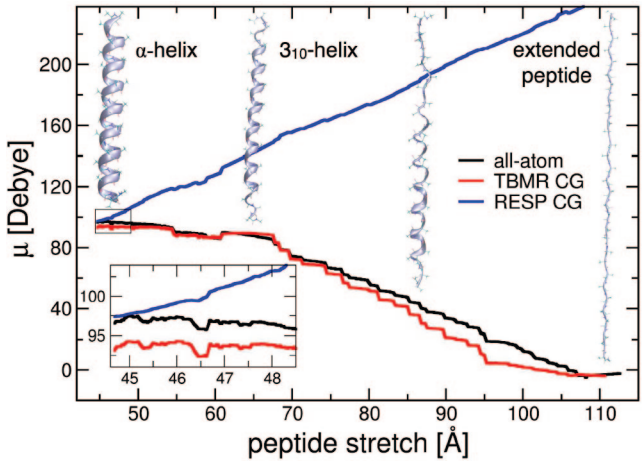


Figure 3. Change of the electrostatic dipole upon structural rearrangement. The component of the electrostatic dipole along the longitudinal direction of a poly ala α-helical peptide is reported upon progressive stretching. Snapshots of the structural determinants are superimposed indicating the progressive unfolding of the structure. The inset focuses on the initial fluctuations around the α-helix conformation.

The application of TBMR to these systems shows that the TBMR method is able to correctly describe the long-range contribution on the active center, demonstrating to be a promising solution for hybrid multiscale simulations of ligand binding recognition.

3.2. Electrostatic Field Tuning upon Conformational Changes in the Secondary Structure. An important requirement for any multiscale scheme is the need to be consistent and accurate upon extended structural rearrangements (such as secondary and/or tertiary structural changes, partial or total unfolding, etc.) that might occur during protein dynamics. The performance of our method in such conditions has been tested by studying the stretching of an α-helix (a 30 aa long poly-ala chain) using standard steering molecular dynamics. Upon mechanical stress, the polypeptide first refolds into a 3_{10} helical species and then unfolds progressively to form a β-sheet-like stretched peptide. We have calculated the projection of the total electrostatic dipole along the stretching direction produced by the TBMR model at different helical configurations, and we have compared it to the values of the all-atom system (Figure 3a). The dipole component obtained by the TBMR method matches very well the all-atom calculations. The model is able to reproduce qualitatively and quantitatively the increase in the electrostatic dipole at the conformational transition between the strained α and 3_{10} helical species. When the helix is completely unfolded, the electrostatic dipole along the stretching direction becomes negligible, as expected. Also in this case, we compare our results to those that would come from a RESP-based description of the electrostatics. In such a case, already in the regime of small oscillations, the change in the dipole is qualitatively not correct (Figure 3, inset). Further stretching of the peptide leads to a complete failure of the predicted values of the dipole. This is related to the fact that RESP charges are located at the C_α and C_β centroids, which do not correspond to the charge groups of the all-atom system and which generally follow qualitatively different spatial motions. While the helix is being stretched, the relative distances of the CG centroids increase, and this leads to an increase in the distances of the positive and negative centers of charge, that is, to an increase of the dipole. Rather, in the all-atom system, the stretching of the helix corresponds to a twist of the peptide bonds, which leads to a decrease of the total dipole along the stretching direction.

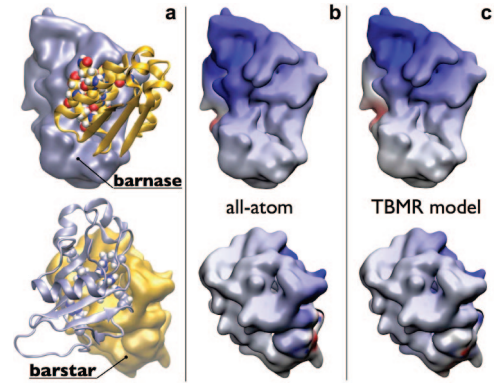


Figure 4. Protein−protein interactions in the barnase-barstar complex. Panel **a**: overall fold structure of the complex (PDB code: 1BRS), indicating the residues involved in the binding at the interface. Panels **b** all-atom and **c** hybrid electrostatic potentials are reported on a molecular surface drawn on top of barnase and barstar, respectively.

Thus, our method is able to describe well the twisting of the dipoles during helical unfolding, providing promising ground when coupled to nontopological coarse-grained force fields. In particular, the electrostatic description given by our method is general and consistent with incoming structural rearrangements of the secondary and tertiary determinants of a macromolecule. Thus, this contribution can be easily included in existing or future coarse-grained force fields developed to sample complex conformational spaces in order to address topics like protein folding and macromolecular assembly in biological networks.

3.3. Long-Range Electrostatic Interactions at Protein−Protein Interfaces. As a final test for the TBMR model, we have inferred its capability of reproducing long-range electrostatic interactions at the protein−protein interface in protein complexes. Thus, we apply the multipolar method to the barnase-barstar complex (PDB code: 1BRS[72]), which is known to be dominated by strong electrostatic interactions at the protein−protein interface[72] and which structure was determined by X-ray crystallography.

Following the MM/CG protocol, the protein−protein interface (defined as the Leu26-Val45 loop-helix motif of barstar and the loops of barnase in contact with it) is described at all-atom level, while the rest of the barnase/barstar complex is described by CG centroids. The long-range electrostatic potential on the surface of the barnase-barstar contact appears to be qualitatively well reproduced by TBMR, as depicted in Figure 4. In particular, it is able to represent the strong positive potential produced by barnase, which is functional to the long-range recognition of the negatively charged residues present on the interacting portion of the surface of barstar. The long-range contribution to the

electrostatic potential of barstar seems to be of less importance, and such an effect is also well described by TBMR (Figure 4).

On a general note, the TBMR method is able to represent quite well the electrostatic potential in a coarse-grained scheme, and for this reason it can be used in a 2-fold manner in the field of protein−protein interaction studies: (i) in a hybrid MM/CG simulation to refine the late stage of a macromolecular complex formation, when the proteins are close together, and (ii) in the early stage of the same process, when the proteins are far apart and long-range electrostatics plays a role in determining the right orientation for the formation of the complex. The latter process can be studied, for example, with full CG calculations coupled to TBMR method.

4. Concluding Remarks

We propose a new scheme for introducing long-range electrostatic contributions in multiscale simulations of proteins.[27,30] The topologically based multipolar reconstruction (TBMR) scheme can be generally adopted to describe electrostatics in the coarse-grained region of hybrid multiscale simulations and can be further extended to full coarse-grained simulations of protein dynamics.

We show that a multipolar expansion of the coarse-grained residues, truncated in the present formulation to the dipolar term, is able to capture a large part of the electrostatic contribution in a variety of biological problems and gives results appreciably more accurately than alternative methods such as RESP charge fitting. The method enhances in fact the reliability of hybrid multiscale methods at localized active sites in enzymatic complexes where the electrostatic contribution is supposed to be crucial for ligand binding recognition and for catalytic activation. It has also been shown to describe with accuracy comparable to all-atom representation the evolution of the dipoles during helical unfolding. Moreover, the method is well suited for the application in the broad field of protein−protein interaction studies. It can be used to improve the reliability of molecular interactions at the interface of a macromolecular complex in the framework of hybrid multiscale simulations, whereas in full coarse-grained simulations it can contribute in identifying the protein−protein recognition pathway when the interacting interface is not known a priori.

The TBMR scheme can be easily incorporated within existing hybrid multiscale methods and can be naturally extended to higher orders in the multipolar expansion, when the adoption of different all-atom force fields is required. In particular, we think that our method is particularly fit to be used in combination with multiscale coarse-graining techniques,[33,34] which to date represent the most general and rigorous approach to coarse-graining.

Finally, we foresee that coupling our model to force field-like coarse-grained Hamiltonians will improve the reliability of molecular simulations dealing with large structural rearrangements that might happen in macromolecular assembly and protein networks dynamics. The description of nucleic acids and lipids constituting biological membranes is also a natural and desirable expansion of the present scheme, which would improve the accessibility and reliability of multiscale molecular simulations in the relevant and complex field of DNA−protein and membrane-protein interacting networks.

Acknowledgment. Authors thank Dr. Ivano Tavernelli for useful discussions. The presented work was supported by the Swiss National Science Foundation (Grant No. PP002-_118930). M.N. acknowledges Prof. Ursula Rothlisberger for mentoring her research work at EPFL.

References

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(2) Stryer, L. *Biochemistry;* W. H. Freeman and Company: New York, U.S.A., 1995.

(3) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464–4470.

(4) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 9785–9792.

(5) Saiz, L.; Klein, M. L. *Acc. Chem. Res.* **2002**, *35*, 482–489.

(6) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.

(7) Brannigan, G.; Lin, L. C. L.; Brown, F. L. H. *Eur. Biophys. J.* **2006**, *35*, 104–124.

(8) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045–15048.

(9) Tozzini, V. *Curr. Opin. Struct. Biol.* **2008**, *15*, 144–150.

(10) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694–698.

(11) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59–107.

(12) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 945–950.

(13) Go, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535–542.

(14) Bahar, I.; Atilgan, A. R.; Erman, B. *Fold. Des.* **1997**, *2*, 173–181.

(15) Micheletti, C.; Carloni, P.; Maritan, A. *Proteins* **2004**, *55*, 635–645.

(16) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Phys. Chem. B* **2007**, *111*, 9390–9399.

(17) Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2005**, *122*, 124906.

(18) Becker, N. B.; Everaers, R. *Phys. Rev. E* **2007**, *76*, 021923.

(19) Voltz, K.; Trylska, J.; Tozzini, V.; Kurkal-Siebert, V.; Langowski, J.; Smith, J. *J. Comput. Chem.* **2008**, *29*, 1429–1439.

(20) Cieplak, M.; Hoang, T. X. *J. Biol. Phys.* **2000**, *26*, 273–294.

(21) Hoang, T. X.; Cieplak, M. *J. Chem. Phys.* **2000**, *112*, 6851–6862.

(22) Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 10712–10717.

(23) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, T.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Sali, A.; Rout, M. P. *Nature* **2007**, *450*, 684–700.

(24) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, T.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Rout, M. P.; Sali, A. *Nature* **2007**, *450*, 701–710.

(25) Trabuco, L. G.; Villa, E.; Mitra, K.; Frank, J.; Schulten, K. *Structure* **2008**, *16*, 673–683.

(26) Liu, P.; Voth, G. A. *J. Chem. Phys.* **2007**, *126*, 045106.

(27) Ayton, G. A.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.

(28) Delle Site, L.; Abrams, C. F.; Alavi, A.; Kremer, K. *Phys. Rev. Lett.* **2002**, *89*, 156103.

(29) Villa, E.; Balaeff, A.; Mahadevan, L.; Schulten, K. *Multiscale Model. Simul.* **2004**, *2*, 527–553.

(30) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. *Phys. Rev. Lett.* **2005**, *95*, 218102.

(31) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. *Phys. Rev. Lett.* **2006**, *96*, 028105.

(32) Lyman, E.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2006**, *2*, 656–666.

(33) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.

(34) Izvekov, S.; Voth, G. A. *J. Chem. Theory Comput.* **2006**, *2*, 637–648.

(35) Ayton, G. S.; Voth, G. A. *J. Struct. Biol.* **2007**, *157*, 570–578.

(36) Praprotnik, M.; Delle Site, L.; Kremer, K. *J. Chem. Phys.* **2005**, *123*, 224106.

(37) Praprotnik, M.; Delle Site, L.; Kremer, K. *Phys. Rev. E* **2006**, *73*, 066701.

(38) Praprotnik, M.; Delle Site, L.; Kremer, K. *Phys. Rev. E* **2007**, *75*, 017701.

(39) Matysiak, S.; Clementi, C.; Praprotnik, M.; Kremer, K.; Delle Site, L. *J. Chem. Phys.* **2008**, *128*, 024503.

(40) Ensing, B.; Nielsen, S. O.; Moore, P. B.; Klein, M. L.; Parrinello, M. *J. Chem. Theory Comput.* **2007**, *3*, 1100–1105.

(41) Heyden, A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 217–221.

(42) Neri, M.; Cascella, M.; Micheletti, C. *J. Phys.-Condens. Matter* **2005**, *17*, S1581−S1593.

(43) Neri, M.; Anselmi, C.; Carnevale, V.; Vargiu, A. V.; Carloni, P. *J. Phys.-Condens. Matter* **2006**, *18*, S347−S355.

(44) Neri, M.; Baaden, M.; Carnevale, V.; Anselmi, C.; Maritan, A.; Carloni, P. *Biophys. J.* **2008**, *94*, 71–78.

(45) Villa, E.; Balaeff, A.; Schulten, K. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6783–6788.

(46) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210–3235.

(47) Marrink, S.-J.; Risselada, H. J.; Yefimov, S.; Tielemann, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(48) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tielemann, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.

(49) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* 199394, 10269–10280.

(50) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896.

(51) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

(52) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.

(53) Zhou, J.; Thorpe, I.; Izvekov, S.; Voth, G. A. *Biophys. J.* **2007**, *92*, 4289–4303.

(54) Ayton, G. S.; Noid, W. G.; Voth, G. A. *MRS Bull.* **2007**, *32*, 929–934.

(55) Park, B.; Levitt, M. *Proteins* **1996**, *258*, 367–392.

(56) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.

(57) Banavar, J. R.; Hoang, T. X.; Maritan, A.; Seno, F.; Trovato, A. *Phys. Rev. E* **2004**, *70*, 041905.

(58) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberg, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 manual and user guide*; Hochschulverlag AG an der ETH Zurich: Zurich, Switzerland, 1996.

(59) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(60) Otterbein, L. R.; Graceffa, P.; Dominguez, R. *Science* **2001**, *293*, 708–711.

(61) Dominguez, R.; Graceffa, P. *Biophys. J.* **2003**, *85*, 2073–2074.

(62) Nar, H.; Messerschmidt, A.; Huber, R.; van de Kamp, M.; Canters, G. W. *J. Mol. Biol.* **1991**, *221*, 765–772.

(63) Chu, J. W.; Voth, G. A *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13111–13116.

(64) Cascella, M.; Magistrato, A.; Tavernelli, I.; Carloni, P.; Rothlisberger, U. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19641–19646.

(65) Olsson, M. H. M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.

(66) Cascella, M.; Cuendet, M. L.; Tavernelli, I.; Rothlisberger, U. *J. Phys. Chem. B* **2007**, *111*, 10239–10247.

(67) Rould, M. A.; Wan, Q.; Joel, P. B.; Lowey, S.; Trybus, K. M. *J. Biol. Chem.* **2006**, *281*, 31909–31919.

(68) Fogolari, F.; Brigo, A.; Molinari, H. *J. Mol. Recognit.* **2002**, *15*, 377–392.

(69) Baker, N. A.; Sept, D.; Joseph, S.; Holst, H. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.

(70) Vidossich, P.; Cascella, M.; Carloni, P. *Proteins* **2004**, *55*, 924–931.

(71) Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8878–8889.

(72) Lee, L.; Tidor, B. *Nat. Struct. Biol.* **2001**, *8*, 73–76.

CT800122X

JCTC Journal of Chemical Theory and Computation

Implementation of a Serial Replica Exchange Method in a Physics-Based United-Residue (UNRES) Force Field

Hujun Shen, Cezary Czaplewski, Adam Liwo, and Harold A. Scheraga*

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853

Received February 27, 2008

Abstract: The kinetic-trapping problem in simulating protein folding can be overcome by using a Replica Exchange Method (REM). However, in implementing REM in molecular dynamics simulations, synchronization between processors on parallel computers is required, and communication between processors limits its ability to sample conformational space in a complex system efficiently. To minimize communication between processors during the simulation, a Serial Replica Exchange Method (SREM) has been proposed recently by Hagan et al. (*J. Phys. Chem. B* **2007**, *111*, 1416−1423). Here, we report the implementation of this new SREM algorithm with our physics-based united-residue (UNRES) force field. The method has been tested on the protein 1E0L with a temperature-independent UNRES force field and on terminally blocked deca-alanine (Ala_{10}) and 1GAB with the recently introduced temperature-dependent UNRES force field. With the temperature-independent force field, SREM reproduces the results of REM but is more efficient in terms of wall-clock time and scales better on distributed-memory machines. However, exact application of SREM to the temperature-dependent UNRES algorithm requires the determination of a four-dimensional distribution of UNRES energy components instead of a one-dimensional energy distribution for each temperature, which is prohibitively expensive. Hence, we assumed that the temperature dependence of the force field can be ignored for neighboring temperatures. This version of SREM worked for Ala_{10} which is a simple system but failed to reproduce the thermodynamic results as well as regular REM on the more complex 1GAB protein. Hence, SREM can be applied to the temperature-independent but not to the temperature-dependent UNRES force field.

1. Introduction

For a theoretical investigation of protein structure and protein-folding dynamics, it is essential to obtain an adequate sampling of the potential energy surfaces of proteins. However, it is difficult to obtain an accurate canonical distribution from a molecular dynamics (MD) simulation at low temperature because of the rugged energy landscapes of proteins, although the overall shape of foldable proteins is funnel-like.[1] These energy landscapes contain many energy barriers between local minimum-energy states in which a simulation at low temperature is easily trapped. Overcoming these local minima is a key to extending MD simulation to search the conformational spaces of many proteins; therefore, a variety of algorithms have been proposed to overcome this trapping problem. One of them (a generalized-ensemble algorithm[2]) performs a simulation based on non-Boltzmann probability weight factors to facilitate surmounting energy barriers by a random walk in the energy space. However, the probability weight factors used in this generalized-ensemble algorithm are not known a priori and must be determined by a short trial simulation. Another algorithm, termed the replica-exchange method (REM),[3–9] is based on the Boltzmann probability weight factor, which is known a priori. Other names for this replica-exchange method are as follows: replica Monte Carlo method,[3] Markov Chain Monte Carlo,[4] exchange Monte Carlo Method,[5] multiple Markov

* Corresponding author phone: (607) 255-4034; fax: (607) 254-4700; e-mail: has5@cornell.edu.

10.1021/ct800063d CCC: $40.75
Published on Web 07/22/2008

chain method,[6] and parallel tempering.[7] The replica-exchange method has been tested with the pentapeptide Met-enkephalin system[8] and has been developed successfully for molecular dynamics simulations.[9]

In the replica-exchange method, a set of simulations of replicas is carried out simultaneously and independently at different temperatures; periodically, exchanges of temperatures between neighboring replicas are attempted and accepted with a well-defined acceptance probability which is consistent with detailed balance; with an ultimate exchange with a high temperature, the simulation at low temperature can surmount energy barriers and thereby search a larger portion of the conformational space of a protein. In addition to overcoming the trapping problem in protein folding, this method takes advantage of the use of canonical ensembles at various temperatures to calculate the thermodynamic characteristics of protein folding.

The replica-exchange method has proven to be a powerful one for studying peptides and small proteins and has been applied to investigate implicit and explicit solvent models in different force fields.[8–11] However, the REM algorithm requires synchronization between different replicas in order to exchange replicas (temperatures) on parallel computers. Therefore, the slowest processor determines the overall performance of the REM simulation. To minimize communication between different processors in the REM simulations, several methods[12–14] have been developed to facilitate efficient use of distributed computers. One of them is the serial replica exchange method (SREM),[14] which has been tested successfully on the terminally blocked alanine (also termed alanine dipeptide) solvated in 256 TIP4P water molecules by Hagen et al.[14] Because SREM can run asynchronously on distributed computers, it is obvious that SREM has the advantage over REM on an inhomogeneous system. Moreover, in the production phase, SREM can be expected to be more efficient than REM even on homogeneous systems for the following two reasons. First, rigorous implementation of REM requires synchronization of trajectories before a given exchange event; this implies overhead because individual energy and force calculations do not take exactly the same wall-clock time due to the application of cutoff of long-range interactions and other processes running on a given core. These differences in wall-clock time are not substantial, but, when synchronization of hundreds or thousands of trajectories is attempted after 10,000 or 20,000 steps and the protein studied is large, they do matter. Second, communication overhead even on a homogeneous system does matter when frequent communication is required, i.e., when the protein studied is small and many replicas are run.

In the past decade, we have been developing a physics-based united-residue force field, hereafter refered to as UNRES,[15–21] for off-lattice protein structure simulation for physics-based protein structure prediction. Reduction of the complexity of the problem (i.e., of the number of interaction sites and variables) with UNRES is necessary in order to carry out protein-folding simulations using a reasonable amount of computer time. However, in contrast to most united-residue force fields which are largely knowledge-based potentials, UNRES was carefully derived, based on the physics of interactions, as a cluster-cumulant expansion of the effective free energy of a protein plus the surrounding solvent, in which the secondary degrees of freedom had been averaged out.[17,20] Moreover, temperature dependence has been introduced recently[22] into the force field to reflect the fact that it is a restricted free energy and not a potential energy function.

In this article, we apply this new SREM algorithm to the UNRES force field. In section 2.1, we briefly describe the UNRES model and force field. In section 2.2, we describe the replica exchange method (REM). In section 2.3, we describe the SREM algorithm briefly. In section 3, we report the performance of our implementation of the SREM algorithm for the β-protein 1E0L with the temperature-independent UNRES force field and for the deca-alanine peptide (Ala_{10}) and 1GAB (a three-helix bundle) with the temperature-dependent UNRES force field. We also carried out regular REM and canonical MD simulations with the UNRES force field for each system in order to compare their performance with that of the SREM algorithm with the UNRES force field under the same conditions. We observed that, with the temperature-independent force field, the performance of SREM is comparable to that of REM. However, SREM could not reproduce results obtained with REM when we tested it on 1GAB with a temperature-dependent UNRES force field. The failure of SREM in this test on 1GAB is attributed to the invalid calculation of the acceptance probability based on the Metropolis criterion, because SREM requires the probability distribution of states as a function of energy and temperature, which is available only in an approximate form for the temperature-dependent version of the UNRES force field.

2. Methods

2.1. UNRES Force Field. In the UNRES model,[15–28] a polypeptide chain is represented by a sequence of α-carbon atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p). Each united peptide group is located in the middle between two consecutive α-carbons. Only these united peptide groups and the united side chains serve as interaction sites; the α-carbons serve only to define the chain geometry, as shown in Figure 1. The UNRES force field has been derived as a restricted free energy (RFE) function of an all-atom polypeptide chain plus the surrounding solvent, where the all-atom energy function is averaged over the degrees of freedom that are lost when passing from the all-atom to the simplified system (viz., the degrees of freedom of the solvent, the dihedral angles for rotation about the bonds in the side chains, and the torsional angles for rotation of the peptide groups about the $C^{\alpha}\cdots C^{\alpha}$ virtual bonds).[20,21] The RFE is further decomposed into factors coming from interactions within and between a given number of united interaction sites.[20] Expansion of the factors into generalized Kubo cumulants[29] enabled us to derive approximate analytical expressions for the respective terms, including the multibody or correlation terms, which are derived in other force fields from structural databases or on a heuristic basis.[30] The theoretical basis of

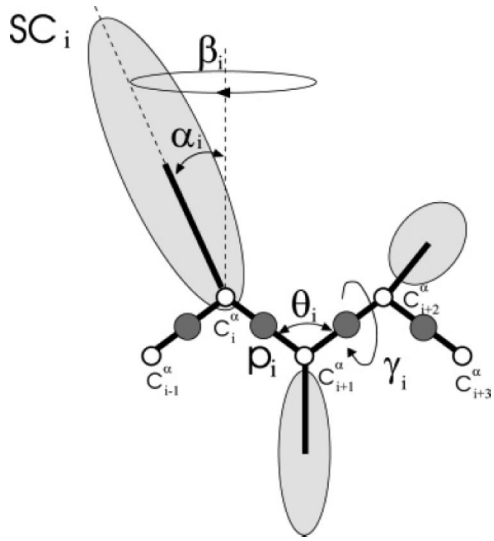


Figure 1. The UNRES model of the polypeptide chain. Dark circles represent united peptide groups (p) and open circles represent the C^{α} atoms. Ellipsoids represent side chains, with their centers of mass at the SCs. The ps are located halfway between two consecutive C^{α} atoms. The virtual-bond angles, θ, the virtual-bond dihedral angles, γ, and the angles, α_{SC} and β_{SC}, that define the location of a side chain with respect to the backbone, are also indicated.

the force field is described in detail in our earlier paper.[17] The energy function is expressed by eq 1[22]

$$U = w_{SC}\sum_{i<j} U_{SC_iSC_j} + w_{SCp}\sum_{i\neq j} U_{SC_ip_j} + w_{pp}\sum_{i<j-1} U_{p_ip_j} + w_{tor}\sum_i U_{tor}(\gamma_i) + w_{tord}\sum_i U_{tord}(\gamma_i,\gamma_{i+1}) + w_b\sum_i U_b(\theta_i) + w_{rot}\sum_i U_{rot}(\alpha_{SC_i},\beta_{SC_i}) + \sum_{m=3}^{6} w_{corr}^{(m)} U_{corr}^{(m)} + w_{bond}\sum_i U_{bond}(d_i) + w_{SS}\sum_i U_{SS;i} \quad (1)$$

where $U_{SC_iSC_j}$, $U_{SC_ip_j}$, and $U_{p_ip_j}$ denote the effective free energies of interaction between the side chains in water, between the side chains and the peptide groups, and between the peptide groups, respectively; U_{bond}, U_b, and U_{rot} are the effective potentials for stretching the virtual bonds, for bending the virtual-bond angles, and for the energetics of the rotameric states of virtual side chains, respectively; U_{tor} and U_{tord} are torsional and double-torsional potentials for rotation about the $C^{\alpha}...C^{\alpha}$ virtual bonds; $U_{corr}^{(m)}$, $m = 1,2,3,4$, are correlation or multibody terms pertaining to coupling between backbone-local and backbone-electrostatic interactions[20] (where m denotes the order of correlation), and U_{SS} is a term accounting for the energetics of disulfide bonds.[31] Each of these terms is multiplied by an appropriate weight, w.

The force field was initially applied to solve the protein-structure prediction problem as that of finding the global minimum of the effective energy function. However, recently[22] we realized that such an approach ignores conformational entropy and, strictly speaking, corresponds to "folding" proteins at the temperature of 0 K; therefore, we redefined[22] the protein-structure prediction problem as finding the most probable conformational ensemble at temperatures below the folding-transition and above the glass-transition temperature. Moreover, because UNRES is an effective free-energy and not a potential-energy function, it should depend on temperature explicitly. By taking advantage of the cumulant expansion, this temperature dependence was introduced in our recent work[22] by multiplying the energy-term weights by factors $f_n(T)$ (where T denotes absolute temperature) defined by eq 2, where applicable

$$f_n(T) = \frac{\ln(e^1 + e^{-1})}{\ln\left\{\exp\left[\left(\frac{T}{T_o}\right)^{n-1}\right] + \exp\left[-\left(\frac{T}{T_o}\right)^{n-1}\right]\right\}} \quad (2)$$

where $T_0 = 300$ K.

2.2. Replica Exchange Method (REM). The replica exchange algorithm based on a molecular dynamics (MD) method is called replica exchange molecular dynamics (REMD).[9] The basic idea of REMD is to run constant temperature MD simulations on sequential replicas of the system simultaneously and independently, each replica at a different temperature $\{T_0,T_1,...,T_i,T_{i+1},..,T_N\}$, where T_N is the maximum temperature at which an MD simulation is run. Usually, those temperatures are arranged in increasing order, e.g., $T_i < T_{i+1}$ ($i = 0, 1, 2,..., N-1$). After every fixed number of steps, an attempt is made to exchange temperatures of a pair of neighboring replicas (e.g., i and $i+1$); the velocities are rescaled following the change in temperature, and the exchange processes are then repeated during MD simulations. The neighboring replicas undergo an exchange with the acceptance given by the Metropolis criterion

$$\omega(X_i \rightarrow X_{i+1}) = \begin{cases} 1 & \Delta \leq 0 \\ \exp(-\Delta) & \Delta > 0 \end{cases} \quad (3)$$

where ω is the transition probability and

$$\Delta = (\beta_{i+1} - \beta_i)[U(X_i) - U(X_{i+1})] \quad (4)$$

where $\beta_i = 1/RT_i$, T_i is the temperature corresponding to the ith replica (trajectory), X_i denotes the variables of the UNRES conformation of the ith replica at the attempted exchange point, and $U(X_i)$ denotes the UNRES energy of conformation X_i. If $\Delta \leq 0$, T_i and T_{i+1} are exchanged; otherwise, the exchange is accepted with probability $\exp(-\Delta)$. When the temperature-dependent UNRES force field is applied, eq 4 must be replaced by eq 5;[22] this result is obtained directly by inserting the temperature-dependent energy function into eq 14 of ref 9.

$$\Delta = -[\beta_{i+1}U(X_{i+1},T_{i+1}) - \beta_i U(X_{i+1},T_i)] + [\beta_{i+1}U(X_i,T_{i+1}) - \beta_i U(X_i,T_i)] \quad (5)$$

The restricted free energy $U(X,T)$ is calculated with eq 5 of ref 22.

To enhance sampling, in this study we used the multiplexing variant of REMD introduced by Rhee and Pandé[12] in which m independent replicas are simulated at each temperature; the same extension applied to serial replica exchange is described in the next section.

2.3. Serial Replica Exchange Method (SREM). The basic idea of the serial replica exchange method is as follows. Let us assume that we know the potential energy distribu-

tions, $P_n(E;T_n)$ at the temperatures T_n ($n = 0, 1, 2,..., N$). Let us perform a constant temperature molecular dynamics simulation at a temperature T_k, $0 \le k \le N$, for a given number of time steps, the final energy being equal to $\boldsymbol{E^0}$. From the energy distribution at its neighboring temperature, $T_{k\pm1}$ (only T_{k-1} if $k = N$, and only T_{k+1} if $k = 0$), we sample an energy $\boldsymbol{E^1}$. Next, we attempt a move to this neighboring temperature, and the move is accepted or rejected according to the Metropolis criterion (eqs 3 and 4). If this move is accepted, the simulation (trajectory) is continued at the new temperature $T_{k\pm1}$ with velocities rescaled following the exchange of temperature; otherwise, it remains at the same temperature T_k.

Sampling an energy value from the distribution at a neighboring temperature replaces use of the energy of the conformation at a neighboring temperature in REM but does not require communication with the processor running the trajectory on a neighboring temperature, which is an obvious advantage. On the other hand, the energy distributions at all temperatures must be obtained prior to the production phase of the method which makes it an adaptive method as MUCA and, consequently, introduces an overhead to obtain the energy distribution. Moreover, when SREM is applied to the temperature-dependent UNRES force field, the acceptance probability cannot be computed from eq 5 given only the energy distributions at all temperatures; this problem is discussed in section 2.4.

Now the question is how to obtain the energy distributions to use in the SREM simulation. The algorithm described in ref 14 and also implemented in our work is as follows: (i) perform a short REM simulation to obtain the first approximate energy distribution at each temperature; (ii) with the approximate energy distributions, perform a set of simulations using SREM for a certain amount of time and collect the sampled potential energies at each temperature; (iii) update the potential energy distribution at each temperature by using the sampled potential energies in step (ii); and (iv) repeat steps (ii) and (iii) until the energy distributions converge at each temperature.

To investigate how quickly the potential energy distributions reach equilibrium, we employ the chi-square measure which is defined by eq 6

$$\chi^2(t) = \sum_{i=1}^{M} \frac{\left(\bar{P}_i(t) - \bar{P}_i^{ref}\right)^2}{\bar{P}_i^{ref}} \tag{6}$$

where $\bar{P}_i(t)$ and $\bar{P}_i^{ref}$ are the current and the reference value, respectively, of the energy distribution at the ith bin of the energy histogram averaged over a time window of length t, and M is the number of bins in the energy distribution histogram. In this article, we use the final converged energy distributions obtained by a regular REM simulation as the reference. The energy distributions have become converged when $\chi^2(t)$ becomes stationary. After the energy distributions converge, the update of the energy distributions can be stopped.

To sample the energy distribution, $\boldsymbol{E}$, we employed the acceptance-rejection method,[32] which can be summarized as follows. Suppose that we have an arbitrary probability distribution $P(E;T)$ at temperature T. Let E^{max}, E^{min}, P^{max}, and P^{min} denote the maximum energy, the minimum energy, the maximum probability, and the minimum probability, respectively, for this energy distribution. Then a point (e^r, y) is generated randomly, where e^r is a random value between E^{max} and E^{min}, and y is a random number between P^{max} and P^{min}. If $y \le P(e^r,T)$, then the value e^r is accepted; otherwise, it is rejected and the sampling step is repeated.

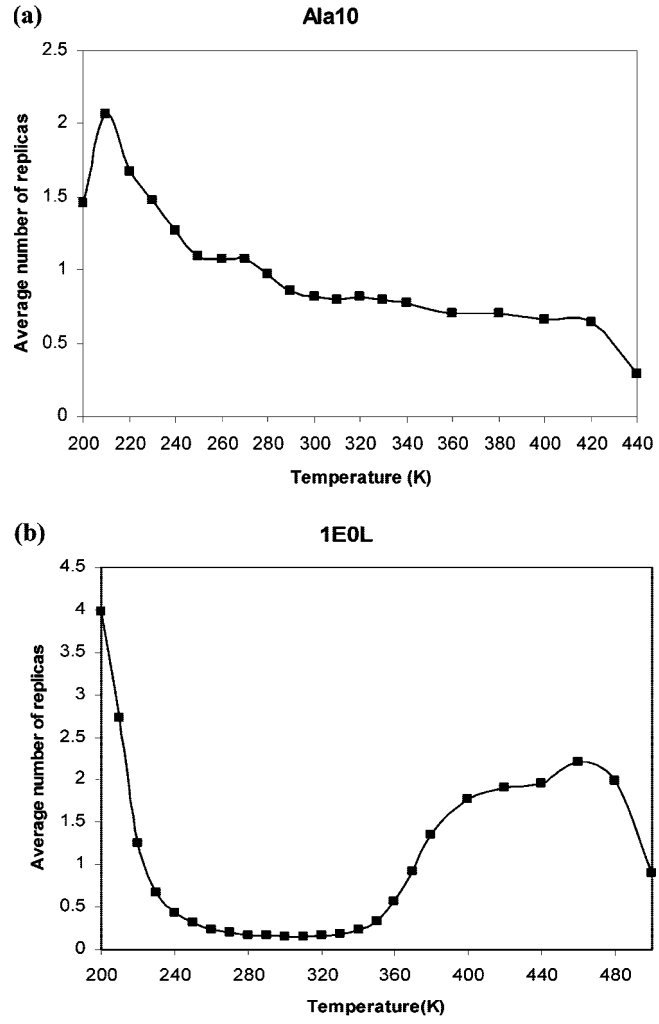


Figure 2. Temperature distributions between replicas, expressed as average number of replicas at a given temperature. (a) Ala_{10} with the temperature-dependent force field and (b) 1E0L with the temperature-independent force field, respectively, in the SREM simulation without modification for 10 million steps. The temperature-independent and temperature-dependent runs are placed in the same figure to show that the bias problem that occurred in the old SREM algorithm is caused by the algorithm itself instead of by the assumption that $U(X_{i+1},T_i)$ is approximately equal to $U(X_{i+1},T_{i+1})$].

2.4. Modification of the SREM Algorithm for the UNRES Force Field. In the original SREM algorithm,[14] attempts are made to move the temperature of a simulation randomly to its neighboring temperatures during the process. However, this could make the simulation move more frequently to some specific temperatures before the potential energy distributions reach equilibrium. For example, the simulations of Ala_{10} and 1E0L using this SREM algorithm in Figure 2 show that the trajectories preferentially move to

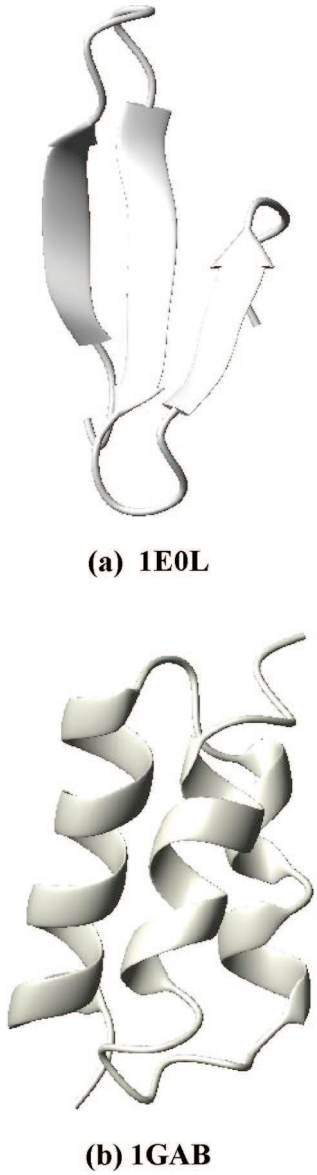


Figure 3. Ribbon diagrams of the experimental structures of (a) 1E0L[39]and (b) 1GAB.[40]

low temperatures for Ala_{10} and to some lower temperatures for 1E0L during the updating phase. Therefore, to avoid this problem, we modified the original SREM algorithm as follows: (i) the frequency of each replica (simulation) moving (up or down) from its temperature to a neighboring temperature is updated periodically during the MD simulation and (ii) based on the "up or down" frequency for each replica, simulation (trajectory) attempts are made to move to its neighboring temperatures in favor of low frequency. In a word, the distinction between the modification of SREM and the original SREM is the different ways to move a simulation to its neighboring temperatures: the SREM with modification is based on the "up or down" frequency, whereas the original SREM is based on random selection.

However, this modification gives rise to another issue: the modification may violate the detailed balance condition because a simulation is forced to exchange with its neighboring temperatures based on its exchange frequency instead of by a random choice. This feature does not downgrade the algorithm during the equilibration phase, i.e., before the energy probability distributions converge, because the detailed balance condition does not hold in SREM for nonconverged energy distributions.[14] However, the detailed balance condition should hold in the production phase when sampling is performed with the converged energy distributions, and the modification is, therefore, not applied in the production phase.

Another problem occurs when a temperature-dependent force field is used with the application of the SREM to UNRES. In the regular REM, eq 4 is simply replaced by eq 5 to compute the acceptance probability. However, the use of eq 5 is not straightforward when SREM is performed with a temperature-dependent force field because this requires knowledge of both $U(X_{i+1},T_{i+1})$and $U(X_{i+1},T_i)$ and, because $U(X_{i+1},T_{i+1})$ is sampled from the energy distribution at temperature T_{i+1}, the conformation X_{i+1} to which it corresponds is unknown. Consequently, we cannot compute $U(X_{i+1},T_i)$ given $U(X_{i+1},T_{i+1})$. Hence, in the applications described in this paper (section 3.2), we assumed that $U(X_{i+1},T_i) \approx U(X_{i+1},T_{i+1})$. To use eq 5 accurately, we would need to construct the joint (and hence multidimensional) energy distribution of all groups of energy terms that depend on temperature in the same way; then, after sampling all energy components from the distribution at T_{i+1}, and by taking advantage of eq 5 of ref 22, we could compute the energy at temperature T_i. The energy components fall into 4 groups as far as its temperature dependence is concerned:[22]

(i) $\sum_{i<j}U_{SC_i}(X)$, $\sum_{i\neq j}U_{SC_ip_j}(X)$, $\sum_{i<j-1}U_{p_ip_j}^{VDW}(X)$, $\sum_i U_b(\theta_i)$, $\sum_i U_{rot}(\alpha_{SC_i},\beta_{SC_i})$, and $\sum_{i=1}^{nbond}U_{bond}(d_i)$ are independent of temperature.

(ii) $\sum_{i<j-1}U_{p_ip_j}^{el}(X)$ and $\sum_i U_{tor}(\gamma_i)$ are multiplied by $f_2(T)$ of eq 2.

(iii) $U_{corr}^{(3)}(X)$, $U_{turn}^{(3)}(X)$, and $\sum_i U_{tord}(\gamma_i,\gamma_{i+1})$ are multiplied by $f_3(T)$ of eq 2.

(iv). $U_{corr}^{(4)}(X)$ and $U_{turn}^{(4)}(X)$ are multiplied by $f_4(T)$ of eq 2.

However, the construction of a four-dimensional distribution of these four groups of energy components requires a much greater, prohibitively expensive, effort to obtain an acceptable convergence compared to obtaining a converged one-dimensional distribution of the total energy.

2.5. Evaluation of Thermodynamic Quantities from Replica Exchange Simulation. To compute thermodynamic quantities (such as the partition function, total energy, and heat capacity) from the results of the simulations carried out at different temperatures, the weighted histogram analysis method (WHAM)[33] was used, as described in ref 22. The expressions for the thermodynamic quantities are given by eqs 17−21 of ref 22.

2.6. Simulation Details. All UNRES/SREM MD simulations were carried out with the Berendsen thermostat,[34] the implementation of which in UNRES/MD has been described in our earlier papers.[27,35] Consequently, the random and stochastic forces were not included. In our earlier work,[27,35,36] we have shown that Newton's dynamics with the Berendsen thermostat leads to faster simulated folding compared to that of Langevin dynamics, which justifies its use in the present work, where the purpose of the simulations was to obtain converged thermodynamic functions and ensemble averages and not to determine the kinetics of folding. The coupling

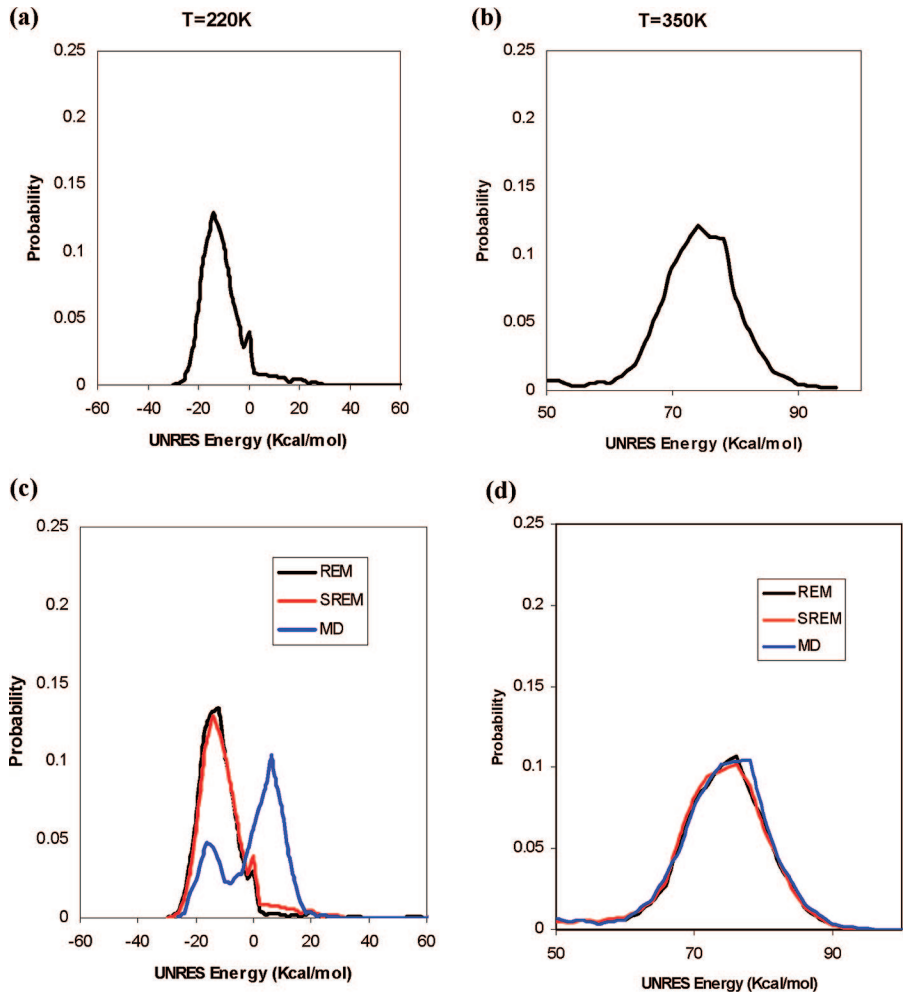


Figure 4. The UNRES energy distributions for 1E0L at two representative temperatures, 220 and 350 K (with the temperature-independent force field);[22] curves in the top panels correspond to the initial rough UNRES energy distributions from the trial REM simulations for 1,000,000 steps at (a) 220 K and (b) 350 K; curves in the bottom panels correspond to the converged UNRES energy distributions from the REM (black), SREM (red), and canonical MD (blue) simulations at the two representative temperatures (c) 220 K and (d) 350 K.

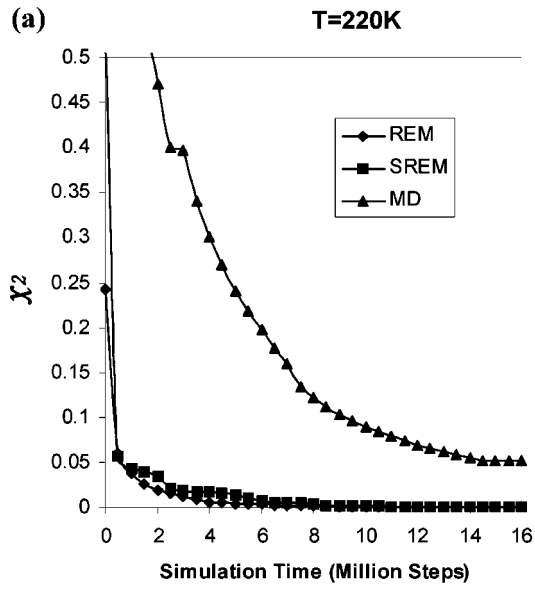
parameter for the Berendsen thermostat was assumed to be $\tau = 1$ mtu (where mtu stands for molecular time unit, and 1 mtu = 48.9 fs), as in our earlier work.[35] The velocity Verlet (VV) algorithm,[37] with the variable time step extension developed in our earlier work,[28] was used, and the basic time step in integrating the equations of motion was 5 fs. The drawings of the structures of the proteins considered in this work were prepared with the MOLMOL program.[38]

3. Results and Discussion

3.1. Test of the SREM with the Temperature-Independent UNRES Force Field. First, we tested the SREM algorithm on the 37-residue protein 1E0L[39] using the temperature-independent UNRES force field parametrized on 1E0L.[22] As in our earlier study,[22] we selected the central 28-residue fragment of this protein, which corresponds to a three-stranded β-sheet (Figure 3a).

To generate initial approximate UNRES energy distribution functions, we ran a regular REM simulation for 1,000,000 steps using 25 replicas over a range of temperatures from 200 to 500 K with 4 independent trajectories at each temperature. The folding-transition temperature of 1E0L with the force field used is 339 K.[22] Replica exchanges were attempted every 20,000 steps during the simulation. The initial UNRES energy distributions at two representative temperatures (220 and 350 K) are plotted in Figure 4a,b. Then, 100 SREM simulations (4 independent simulations at each temperature) were performed using the initial energy distributions obtained from the short REM simulation. During the SREM simulation, the energy distributions were updated every 500,000 steps for 16,000,000 steps, after which the updated energy distributions roughly converged. A subsequent SREM simulation of another 8,000,000 steps was performed with converged energy distributions to collect the thermodynamic data. To compare the results with the REM and canonical MD, we also ran regular REM simulations and canonical MD simulations at 25 temperatures with 4 independent trajectories at each temperature for 24,000,000 steps.

The converged UNRES energy distributions of the SREM simulations at each temperature nearly converged to those of the REM simulations. For example, the converged UNRES energy distributions of the SREM and REM simulations at two representative temperatures (220 and 350 K) are plotted



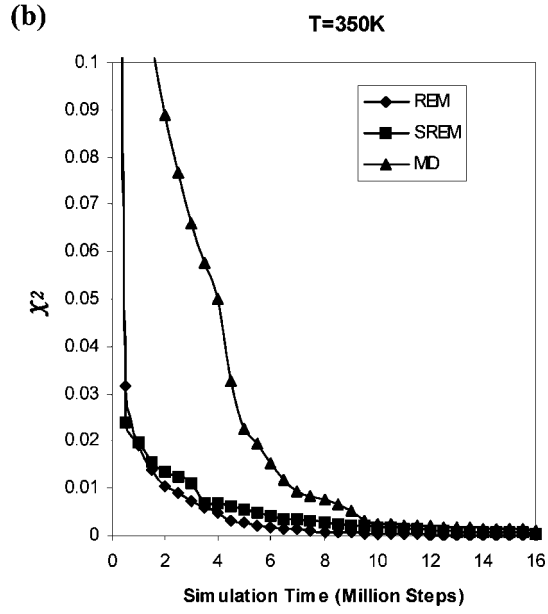


Figure 5. Convergence measure of the UNRES energy distributions for the REM (diamonds), SREM (squares), and canonical MD (triangles) simulations of 1E0L, at temperature (a) 220 K and (b) 350 K (with the temperature-independent force field),[22] as a function of the simulation time. The χ^2 curves are plotted by using the converged UNRES energy distributions from the REM run for 24,000,000 steps as the reference.

in Figure 4c,d; the black curves show the converged UNRES energy distributions obtained from the REM run, and the red curves show the converged UNRES energy distributions obtained from the SREM run. However, we observed that the converged UNRES energy distributions from the canonical MD simulations (blue curves) at low temperatures did not converge to those of the REM simulation, but they converged to those of the REM simulation only at high temperatures. For example, at $T = 220$ K, the plot of the converged UNRES energy distribution obtained from the canonical MD run has two peaks (one in the low-energy region and the other in the high-energy region, and the high-energy region made a major contribution to the energy distribution; Figure 4c), which means that canonical MD simulations are trapped in the high-energy region, whereas the canonical MD simulation at 350 K avoided this trapping in the high-energy region.

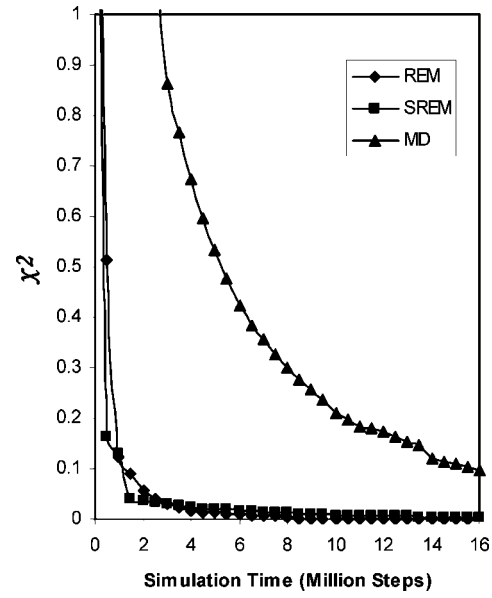


Figure 6. Convergence measure of the UNRES energy distributions for REM (diamonds), SREM (squares), and canonical MD (triangles) simulations of 1E0L as a function of the simulation time with the temperature-independent force field.[22] The χ^2 curves are plotted by using the converged UNRES energy distributions from the regular REM run for 24,000,000 steps as the reference, and the χ^2 values are averaged over 25 temperatures.

The convergence of the UNRES energy distributions of the REM, SREM, and canonical MD simulations was assessed quantitatively by using the $\chi^2(t)$ measure defined in eq 6, at these two representative temperatures, 220 and 350 K. The values of $\chi^2(t)$ are plotted against the simulation time in Figure 5. At high temperature in Figure 5b, $T = 350$ K, the values of $\chi^2(t)$ of the canonical MD and SREM simulations have nearly converged to that of the REM simulation after 16 million steps. At lower temperature, $T = 220$ K, the χ^2 values of only the SREM simulation has converged to that of the REM simulation, but the χ^2 values of the canonical MD simulation has not converged to them, which confirms that the canonical MD simulation at lower temperature is trapped (see Figure 4c). To investigate the convergence properties of the SREM UNRES energy distributions at all 25 temperatures, all the simulations and the χ^2 values are averaged over all 25 temperatures, and the results are shown in Figure 6. The χ^2 result in Figure 6 shows that the convergence properties of REM and SREM are the same for all temperatures as those at two representative temperatures, 220 and 350 K (Figure 5), and the averaged result of the canonical MD simulations has not converged yet because of the poor convergence behavior at lower temperatures.

To investigate the thermodynamic properties in these simulations, we have plotted heat-capacity curves. The ensemble averages of the heat capacity calculated from

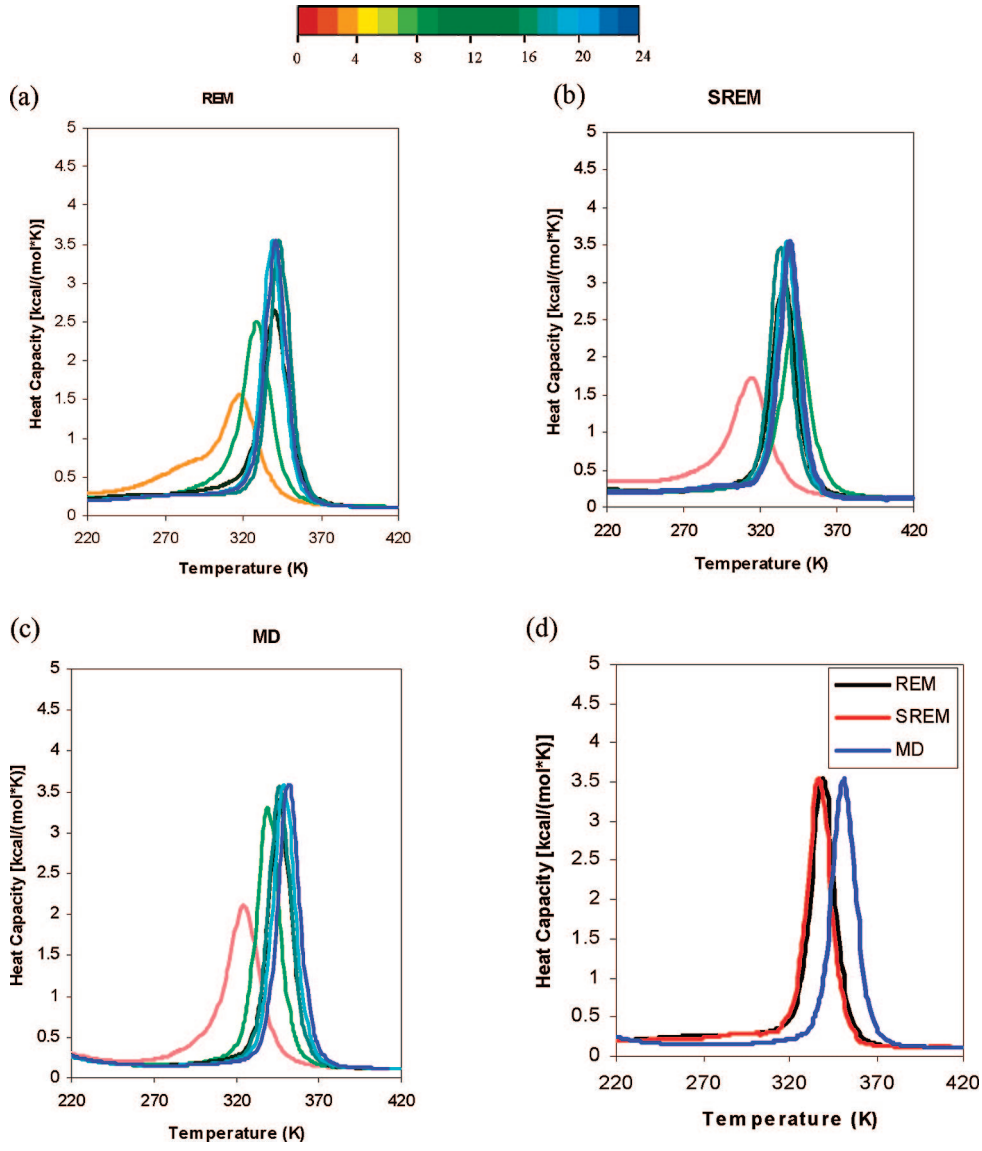


Figure 7. Plots of heat capacity using 4,000,000 MD consecutive steps/trajectory windows taken from (a) the REM run, (b) the SREM run, and (c) the canonical MD run of 1E0L with the optimized force field (the temperature-independent force field).[22] The curves in parts (a)-(c) are colored from red to blue according to the number of MD steps; the color codes are shown in the color bar with numbers indicating the number of millions of MD steps. (d) Two converged heat capacities are obtained from the REM (black) and SREM (red) simulations, and the obvious shift of the heat capacity is obtained from the canonical MD (blue) simulation.

SREM, REM, and canonical MD simulations converge in 20,000,000 steps. The heat capacities, calculated for these simulations with 4 million consecutive steps/trajectories, are shown in Figure 7a-c. When the results of the converged heat capacities obtained from the REM, SREM, and canonical MD simulations were compared, it was observed that the two heat capacity curves of the REM and SREM nearly fit to each other, whereas the heat capacity curve of the canonical MD is obviously shifted from these curves (in Figure 7d).

3.2. Test of the SREM with the Temperature-Dependent UNRES Force Field. To test the SREM with the temperature-dependent UNRES force field,[22] we first chose a simple system, a ten-residue polyalanine (Ala_{10}) chain. The folding-transition temperature of Ala_{10} with the force field used is 311 K. To obtain initial energy distribution functions, we ran a regular REM simulation (it should be noted that eq 4 is replaced by eq 5 in the regular REM run with the temperature-dependent force field) for 400,000 steps using 20 replicas over a range of temperatures from 200 to 440 K with 1 trajectory per temperature. Replica exchanges were attempted every 10,000 steps during the simulation. Then, 20 SREM simulations (1 simulation per temperature) were performed using the energy distributions obtained from the initial REM simulation. Next, the energy distributions were updated every 400,000 steps for 10,000,000 steps, at which point the UNRES energy distributions became roughly steady, which indicates that they converged. A subsequent SREM simulation of another 10,000,000 steps was performed with converged energy distributions to collect the thermodynamic data. To compare the SREM to the REM and the canonical MD simulations, we also ran a regular REM simulation and a canonical MD simulation at 20 temperatures

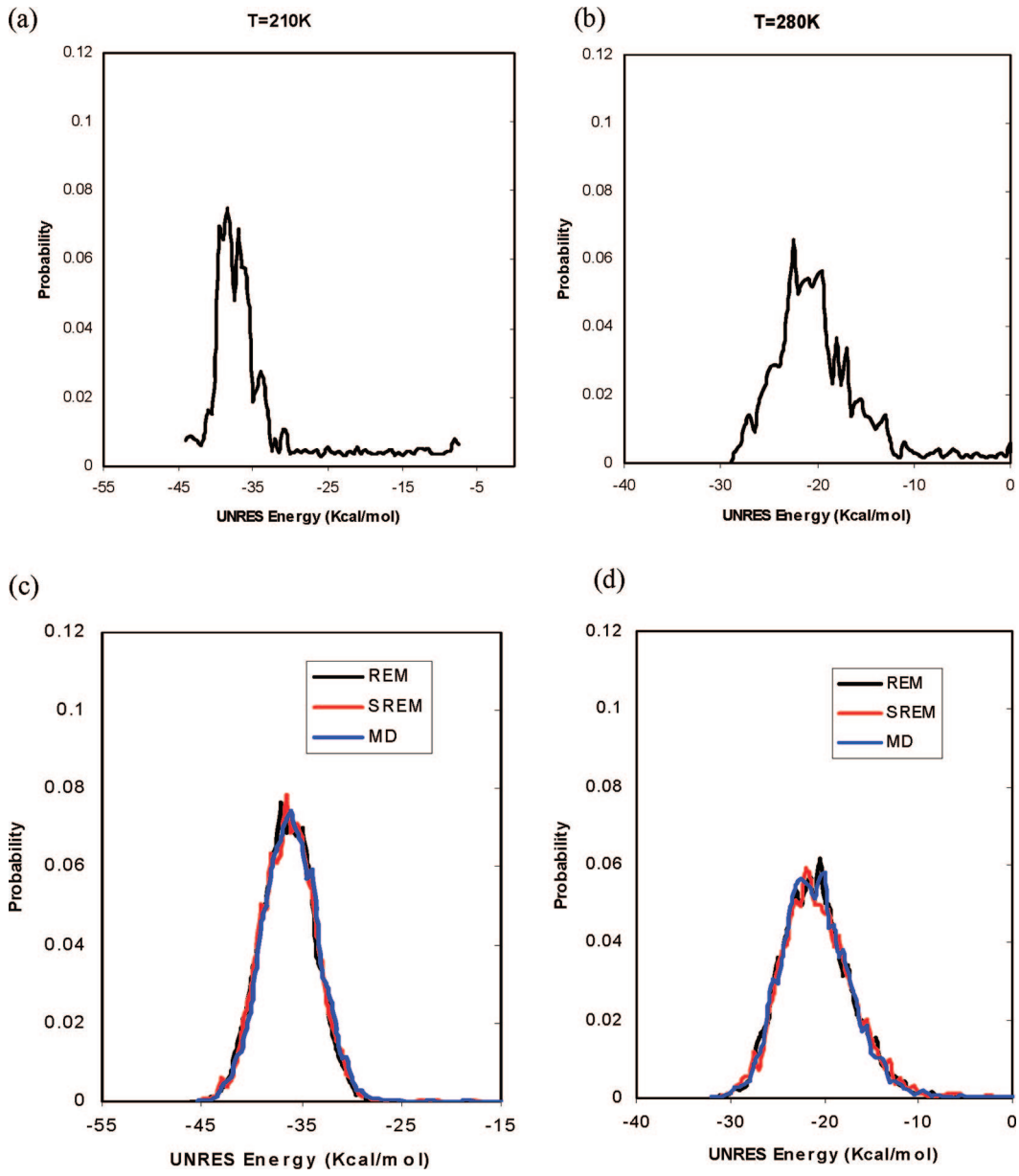


Figure 8. The UNRES energy distributions for Ala_{10} at two representative temperatures, 210 and 280 K (with the temperature-dependent force field):[22] curves in the top panels correspond to the initial rough UNRES energy distributions for the trial REM simulation for 400,000 steps at temperatures (a) 210 K and (b) 280 K. Curves in the bottom panels correspond to the converged UNRES energy distributions at temperatures (c) 210 K and (d) 280 K obtained from the REM (black), SREM (red), and canonical MD (blue) simulations.

also for 20,000,000 steps each with 1 trajectory per temperature.

The initial UNRES energy distributions obtained from the trial REM simulation at two representative temperatures (210 and 280 K) are plotted in Figure 8a,b. During the updating phase in the SREM simulation, the energy distributions obtained from the SREM simulation of 10,000,000 steps have nearly converged to those of the REM simulation (see Figure 8c,d). The results show that the converged UNRES energy distributions of the REM, MD, and SREM simulations, at low and high temperatures, fit each other very well, and no trapping problem was observed in the MD runs at other low temperatures.

The convergence properties of the energy distributions obtained from the REM, SREM, and canonical MD simulations, by using the $\chi^2(t)$ measure, defined in eq 6, were assessed quantitatively at two representative temperatures, 210 and 280 K. Plots of $\chi^2(t)$ against simulation time are shown in Figure 9. From Figure 9, it is observed that the χ^2 values for SREM and canonical MD at temperatures 210 and 280 K converge very well to those of the REM. The convergence properties of the energy distributions obtained from SREM, canonical MD, and REM, averaged over 20 temperatures, are shown in Figure 10, which illustrates the same behavior (averaged over all temperatures) as those at the two representative temperatures, 210 and 280 K (Figure 9).

To investigate the thermodynamic properties from these simulations, heat-capacity curves for Ala_{10} were plotted. Heat capacities were calculated for those simulations using 2

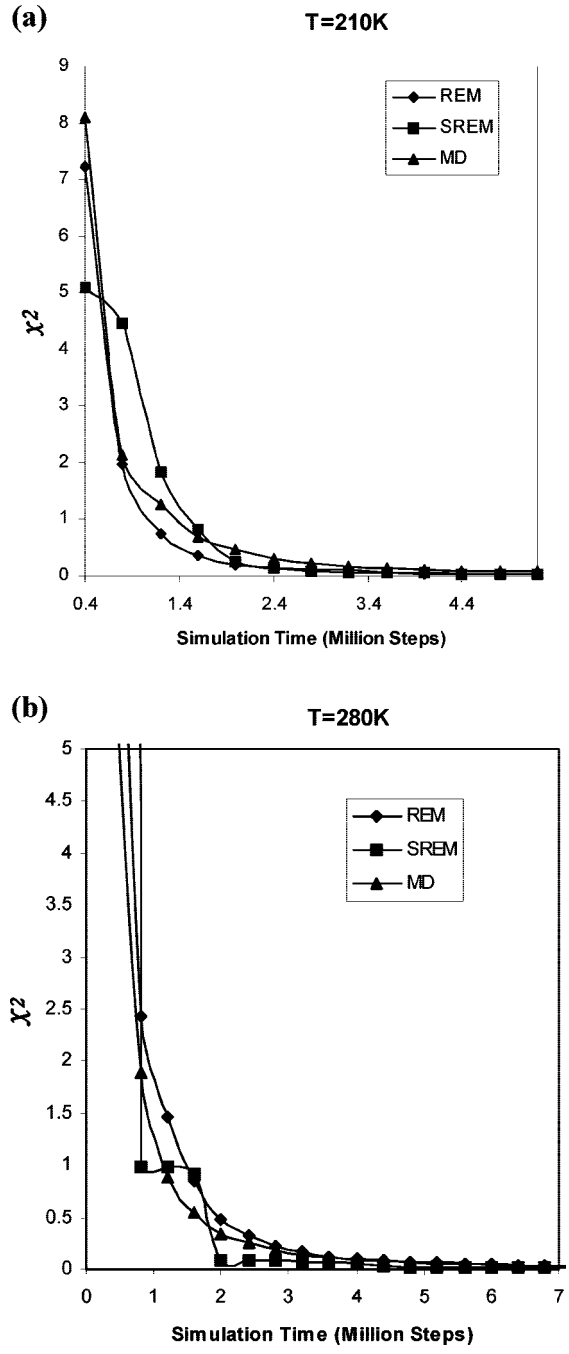


Figure 9. Convergence measure of the UNRES energy distributions for the REM (diamonds), SREM (squares), and canonical MD (triangles) simulations of Ala_{10} at temperatures (a) 210 K and (b) 280 K, as a function of the simulation time with the temperature-dependent force field.[22] The χ^2 curves are plotted by using the converged UNRES energy distributions from the regular REM run for 20,000,000 steps as the reference.

million consecutive steps per trajectory, which are shown in Figure 11a-c. The ensemble averages of the heat capacity calculated from the SREM, REM, and canonical MD converged in 12,000,000 steps. When the results of the converged heat capacities obtained from REM, SREM, and canonical MD simulations were compared, it was observed that the three heat capacities for REM, MD, and SREM fit each other very well (as shown in Figure 11d).

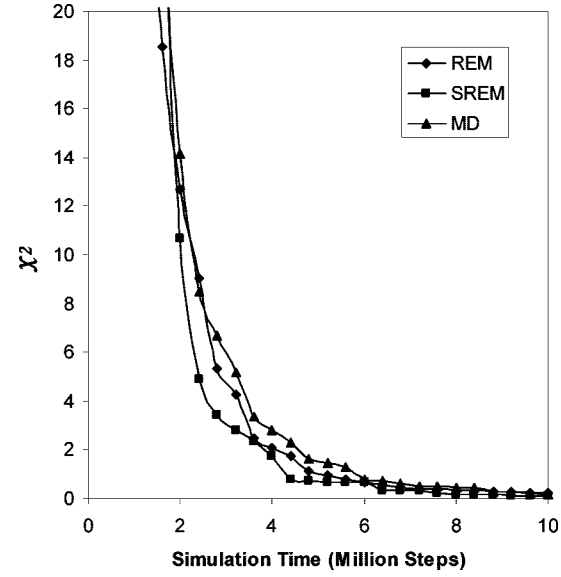


Figure 10. Convergence measure of the UNRES energy distributions for the REM (diamonds), SREM (squares), and canonical MD (triangles) simulations of Ala_{10} as a function of the simulation time with the temperature-dependent force field.[22] The χ^2 curves are plotted by using the converged UNRES energy distributions from the regular REM run for 20,000,000 steps as the reference, and the χ^2 values are averaged over all 20 temperatures.

However, the Ala_{10} system is a simple one, so that even canonical MD simulations give as good results as REM does with the temperature dependent force field. Hence, we tried to test SREM on a more complex protein system, 1GAB (see Figure 3b), but we were not able to obtain good results, as shown in Figures 12 and 13. From Figure 12, it is observed, at low temperature (i.e., $T = 220$ K), that the maximum of the UNRES energy distribution obtained from SREM is shifted to higher energies compared to that obtained from REM, while the plot of the UNRES energy distribution obtained from the canonical MD run has two peaks (one in the low-energy region and the other in the high-energy region, and the high-energy region made a major contribution to the energy distribution; Figure 12a). The χ^2 result in Figure 12b shows that the averaged results of the SREM and the canonical MD simulations have not converged to that of the REM simulation because of the poor convergence behavior at lower temperatures. Heat capacities were calculated for those simulations using 4 million consecutive steps per trajectory, which are shown in Figure 13a-c. The ensemble average of the heat capacity calculated from the REM converges in 12,000,000 steps, but the results from the SREM and canonical MD did not converge even in 20,000,000 steps. From Figure 13d, the heat capacities obtained from the SREM and canonical MD simulations for 20,000,000 steps have broad and multiple peaks around the transition temperature.

Therefore, SREM has failed to reproduce the thermodynamics of folding of 1GAB with the temperature-dependent UNRES force field because of the invalid calculation of the acceptance probability in eq 4 for SREM. Apparently, the assumption that $U(X_{i+1},T_i) \approx U(X_{i+1},T_{i+1})$ (see section 2.4) works for the simple Ala_{10} system but does not work for a

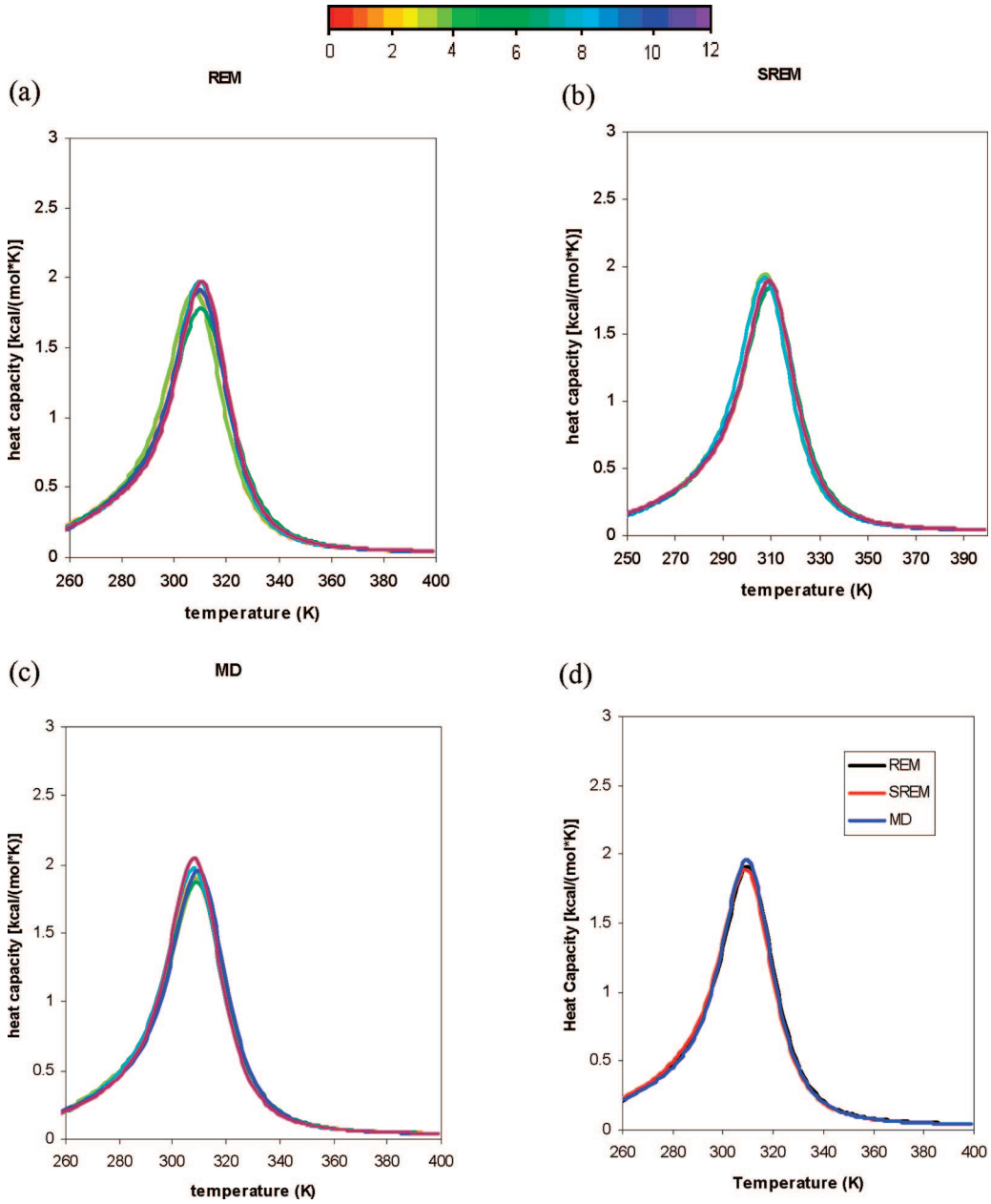


Figure 11. Plots of heat capacity using 2,000,000 consecutive MD steps per trajectory taken from (a) the REM run, (b) the SREM run, and (c) the canonical MD run of Ala_{10} with the temperature-dependent force field.[22] The curves in parts (a)-(c) are colored from red to blue according to the number of MD steps; the color codes are shown in the color bar with numbers indicating the number of millions of MD steps. (d) Three separate converged heat capacities obtained from the REM (black), SREM (red), and canonical MD (blue) simulations.

more complicated 1GAB system. As mentioned in section 2.4, accurate application of eq 5 would require the determination of a four-dimensional distribution of the UNRES energy components which share the same temperature dependence. Consequently, the advantage of minimizing the cost of communication between processors in the SREM algorithm is overwhelmed by the additional cost to obtain a converged distribution of energy components. Therefore, SREM is applicable only to our temperature-independent UNRES force field.

3.3. Parallel Performance of SREM. To check the performance of SREM, we compared it with REM under the same simulation conditions. During the simulation of 1E0L with a temperature-independent UNRES force field, REM and SREM were run for 20 million steps, and REM spent 10706 s while SREM used 9929 s to finish. During the simulation of polyalanine (Ala_{10}) with the temperature-dependent UNRES force field, REM and SREM were run for 20 million steps, and REM spent 101820 s while SREM used 99864 s to finish.

We also carried out a benchmark assessment of the SREM code using the Cray XT3 computer at the Pittsburgh Supercomputer Center. Weak scaling data, obtained by holding the per-node computational work constant (500,000 steps of MD) for SREM simulations (UNRES/SREM) using the Ala_{10} system, are presented. For comparison, a set of regular REM simulations (UNRES/REM) carried out using the Cray XT3 computer with synchronization for the same

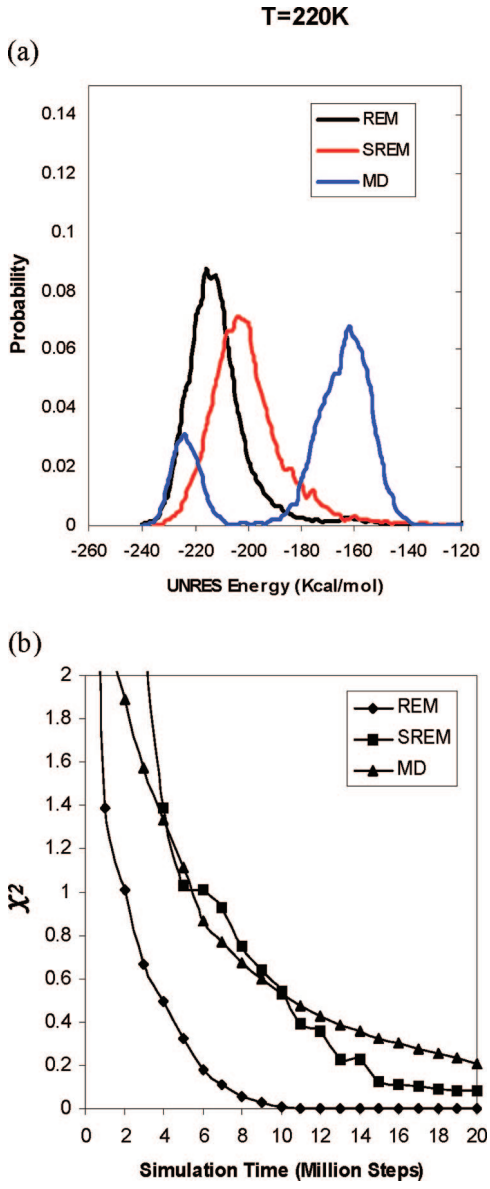


Figure 12. (a) The UNRES energy distribution for 1GAB at a temperature 220 K with the temperature-dependent force field.[22] The black, red, and blue curves show the UNRES energy distributions obtained from the REM, SREM, and canonical MD simulations for 20 million steps, respectively. (b) Convergence measure of the UNRES energy distributions for the REM (diamonds), SREM (squares), and canonical MD (triangles) simulations of 1GAB as a function of the simulation time. The χ^2 curves are plotted by using the converged UNRES energy distributions from the regular REM run for 20,000,000 steps as the reference, and the χ^2 values are averaged over all 20 temperatures.

system (Ala_{10}) is presented. Those data are given in Table 1, which shows that the time (i.e., setup time, total time, and nonsetup time) of the SREM simulations and REM simulations changes with a given number of processors. Nonsetup time in the table is the difference between the total time and the setup time. It is straightforward to compare the parallel performance of the SREM simulations and regular REM by plotting the nonsetup time of the simulations against the number of processors (see Figure 14). From the results (Table 1 and Figure 14), we observe that the nonsetup time of the SREM simulations increases slightly with increasing number of processors, whereas the nonsetup time of the REM simulations increases significantly with increasing number of processors. Speedup curves, calculated from the weak scaling data in Table 1 by applying eq 7, are shown in Figure 15

$$s(4n) = 4n\frac{t_{n-s}(4)}{t_{n-s}(4n)} \tag{7}$$

where $s(4n)$ and $t_{n\text{-}s}(4n)$ are the speed-up and the nonsetup time with $4n$ processors, respectively, given the same per-processor work load; the calculations were always carried out with a multiple of 4 processors.

We found that the increasing number of processors does not affect the efficiency of SREM but does decrease the efficiency of regular REM, which is shown in Figure 15. Therefore, we can conclude that the parallel performance for SREM is better than that for regular REM at this time.

4. Conclusions

We have implemented UNRES/SREM and tested it on three systems: 1E0L with the temperature-independent and Ala_{10} and 1GAB with the temperature-dependent force fields. By checking the convergence properties of the energy distributions and the thermodynamic properties of 1E0L, calculated from REM and SREM, we have demonstrated that SREM can reproduce the results of REM, while the canonical MD cannot, which means that the SREM algorithm is comparable to regular REM with the temperature-independent UNRES force field. For two simple systems studied, 1E0L and Ala_{10}, SREM turned out to be more efficient than REM. However, the data from section 3.3 indicate that the gain in wall-clock time is rather incremental than substantial. Moreover, SREM has the adaptive phase in which energy distributions are constructed, which can take a long time for a larger system.

With the temperature-dependent UNRES force field, SREM can be applicable to some simple systems such as Ala_{10}, but Ala_{10} is so simple that MD simulations can reproduce the same results as that of REM. To check the performance of SREM with our temperature-dependent UNRES force field, we used 1GAB as the target, but SREM failed to produce good results for this protein because of too-crude an estimation of the acceptance probability under the necessary assumption that the change of the effective energy can be neglected when moving from a current temperature to a neighboring temperature. SREM showed the same poor convergence behavior as canonical MD (Figures 12 and 13). One possible solution would be to use the distribution of groups of energy terms which share common temperature dependences instead of the distribution of energy. This modification would, however, kill the advantage of the SREM algorithm over the regular REM algorithm because of the enormous amount of time required to obtain a converged multidimensional distribution of energy components. Another possible solution is to create a database of $[E(T_{i\text{-}1}), E(T_i), E(T_{i+1})]$ triples instead of an energy distribution to sample an energy and translate it to the energy at a neighboring temperature; this is correct in principle, but

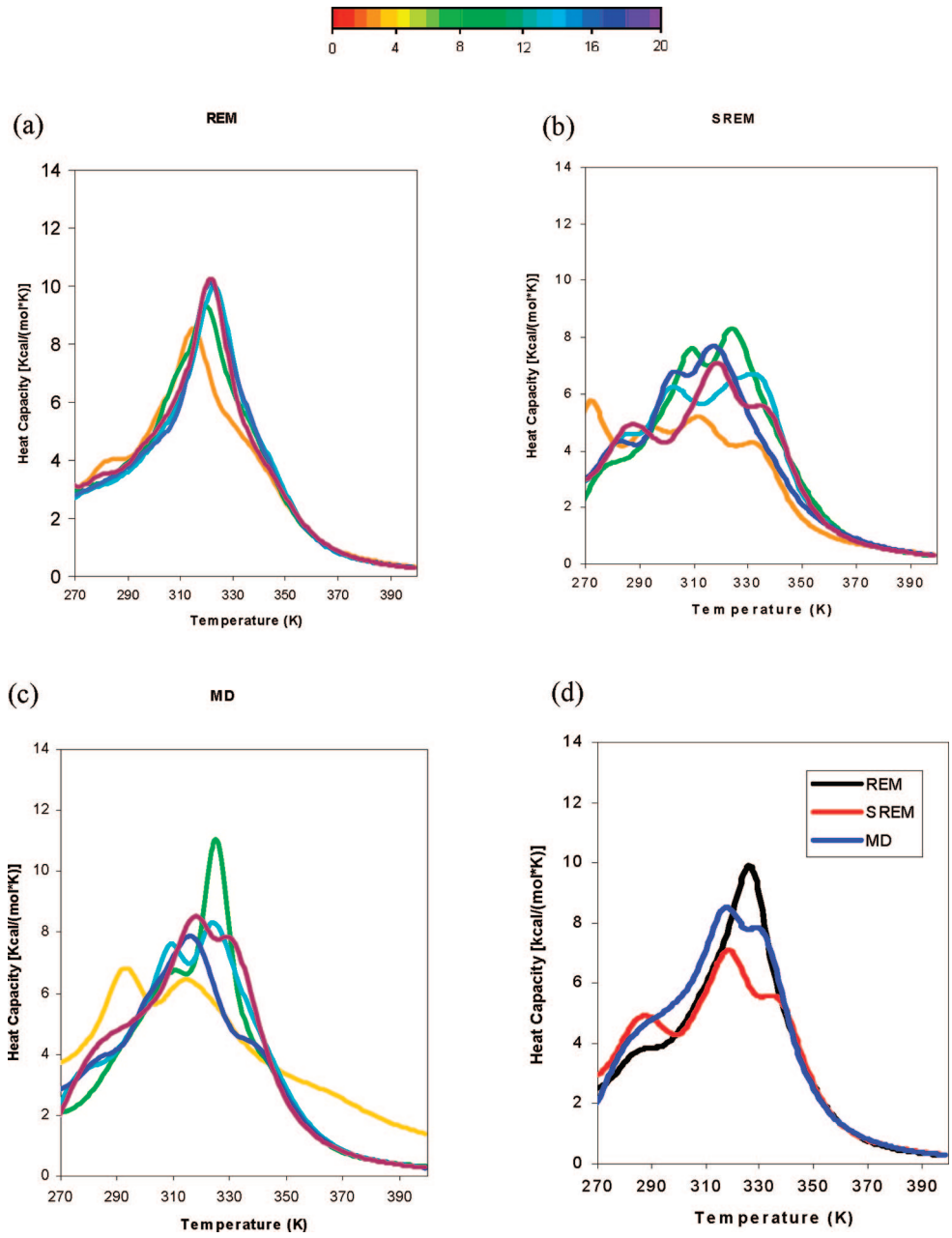


Figure 13. Plots of heat capacity using 4,000,000 consecutive MD steps per trajectory taken from (a) the REM run, (b) the SREM run, and (c) the canonical MD run of 1GAB with the temperature-dependent force field.[22] The curves in parts (a)-(c) are colored from red to blue according to the number of MD steps; the color codes are shown in the color bar with numbers indicating the number of millions of MD steps. (d) Three separated heat capacities obtained from the REM (black), SREM (red), and canonical MD (blue) simulations.

Table 1. Weak Scaling Data of UNRES REM and UNRES SREM for Ala_{10} Using the Cray XT3 Computer at the Pittsburgh Supercomputer Center

	UNRES/REM			UNRES/SREM		
no. of processors	setup time [s]	total time [s]	nonsetup time [s]	setup time [s]	total time [s]	nonsetup time [s]
4	0.14	67.15	67.01	0.16	88.57	88.41
8	0.17	66.75	66.58	0.17	88.35	88.18
16	0.24	68.11	67.87	0.24	87.51	87.27
32	0.40	67.52	67.12	0.47	88.17	87.70
64	0.61	68.77	68.16	0.66	89.85	89.19
128	1.07	69.75	68.68	1.15	88.73	87.58
256	2.05	73.87	71.82	2.29	89.28	86.99
512	3.94	78.74	74.80	4.20	94.99	90.79
1024	7.64	94.14	86.50	8.65	97.58	88.93

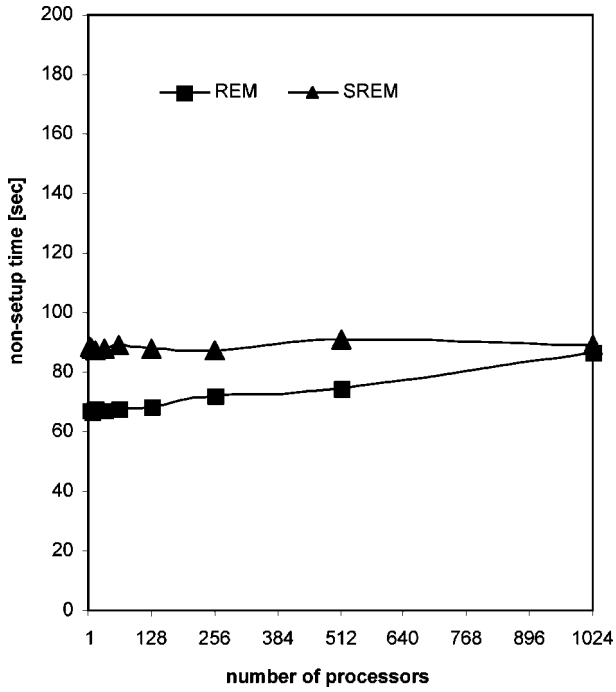


Figure 14. Plots of the weak scaling data, probed by holding the per-node computational work constant for Multiplexing Replica Exchange Molecular Dynamics (REM, squares) and Serial Replica Exchange Molecular Dynamics (SREM, triangles) codes using the Cray XT3 computer.

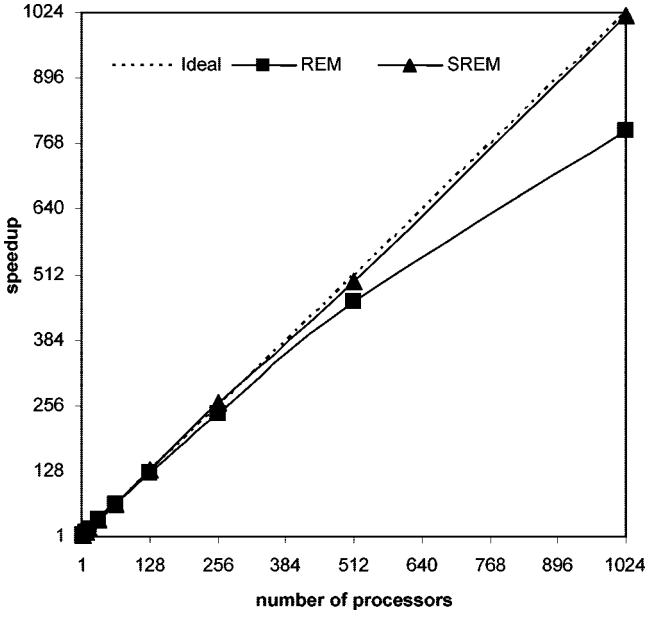


Figure 15. Speedup plots for Replica Exchange Molecular Dynamics (REM, squares) and Serial Replica Exchange Molecular Dynamics (SREM, triangles) codes using the Cray XT3 computer.

a problem here might be the large size of such a database. Assuming that 20 trajectories are run at different temperatures, and 500 most recent snapshots are saved from each (as in our study), giving 30,000 energy values to store, which is a reasonable size, this database will grow rapidly with the number of trajectories and number of snapshots to store, which is often necessary to treat larger systems. Therefore, SREM can be applied in a straightforward way only with our temperature-independent UNRES force field but not with our temperature-dependent one.

Acknowledgment. This work was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB05-41633), and the NIH John E. Fogarty International Center (TW7193). This research was conducted by using the resources of (a) our 800-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (b) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, (c) the John von Neumann Institute for Computing at the Central Institute for Applied Mathematics, Forschungszentrum Jülich, Germany, (d) our 45-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk, (e) the Informatics Center of the Metropolitan Academic Network (ICMAN) in Gdańsk, and (f) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) at the University of Warsaw.

References

(1) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619–1620.

(2) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1997**, *18*, 920–933.

(3) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.

(4) Geyer, C. J.; Thompson, E. A. *J. Am. Stat. Assoc.* **1995**, *90*, 909–920.

(5) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.

(6) Tesi, M. C.; van Rensburg, E. J. J.; Orlandini, E.; Whittington, S. G. *J. Stat. Phys.* **1996**, *82*, 155–181.

(7) Marinari, E.; Parisi, G.; Ruiz-Lorenzo, J. J. In *Spin Glasses and Random Fields*; Young, A. P., Ed.; World Scientific: Singapore, 1998; pp 58−98.

(8) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

(9) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(10) García, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345–354.

(11) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

(12) Rhee, Y. M.; Pandé, V. S. *Biophys. J.* **2003**, *84*, 775–786.

(13) Rodinger, T.; Howell, P. L.; Pomes, R. *J. Chem. Theory Comput.* **2006**, *2*, 725–731.

(14) Hagen, M.; Kim, B.; Liu, P.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2007**, *111*, 1416–1423.

(15) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1697–1714.

(16) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715–1731.

(17) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849–873.

(18) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 874–887.

(19) Liwo, A.; Kazmierkiewicz, R.; Czaplewski, C.; Groth, M.; Ołdziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. *J. Comput. Chem.* **1998**, *19*, 259–276.

(20) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323–2347.

(21) Liwo, A.; Arlukowicz, P.; Czaplewski, C.; Ołdziej, S.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1937–1942.

(22) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *111*, 260–285.

(23) Ołdziej, S.; Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **2003**, *107*, 8035–8046.

(24) Liwo, A.; Ołdziej, S.; Czaplewski, C.; Kozłowska, U.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 9421–9438.

(25) Ołdziej, S.; Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 16934–16949.

(26) Ołdziej, S.; Lagiewka, J.; Liwo, A.; Czaplewski, C.; Chinchio, M.; Nanias, M.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 16950–16959.

(27) Ołdziej, S.; Czaplewski, C.; Liwo, A.; Chinchio, M.; Nanias, M.; Vila, J. A.; Khalili, M.; Arnautova, Y. A.; Jagielska, A.; Makowski, M.; Schafroth, H. D.; Kazmierkiewicz, R.; Ripoll, D. R.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7547–7552.

(28) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13785–13797.

(29) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100–1120.

(30) Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *97*, 9412–9426.

(31) Chinchio, M.; Czaplewski, C.; Liwo, A.; Ołdziej, S.; Scheraga, H. A. *J. Chem. Theory Comput.* **2007**, *3*, 1236–1248.

(32) Robert, C. P. and Casella, G. Monte Carlo Integration. In *Monte Carlo Statistical Methods*; Springer: New York, 1999; pp 92−96

(33) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(34) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(35) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13798–13810.

(36) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362–2367.

(37) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637–649.

(38) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graph.* **1996**, *14*, 51–55.

(39) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. *Nat. Struct. Biol.* **2000**, *7*, 375–379.

(40) Johansson, M. U.; de Chateau, M.; Wikstrom, M.; Forsen, S.; Drakenberg, T.; Bjorck, L. *J. Mol. Biol.* **1997**, *266*, 859–865.

CT800063D

JCTC Journal of Chemical Theory and Computation

Cyameluric Acid as Anion-π Type Receptor for ClO_4^- and NO_3^-: π-Stacked and Edge-to-Face Structures

Dong Young Kim, N. Jiten Singh, and Kwang S. Kim*

Center for Superfunctional Materials, Department of Chemistry, Pohang University of Science and Technology, San 31, Hyojadong, Namgu, Pohang 790-784, Korea

Received April 9, 2008

Abstract: Based on the binding energies at high levels of ab initio theory including coupled cluster theory at the complete basis limit, we show that cyameluric acid ($C_6N_7O_3H_3$) is a potent receptor for ClO_4^- and NO_3^- anions through the anion-π interactions. In contrast, cyanuric acid ($C_3N_3O_3H_3$) binds Cl^-, NO_3^-, and ClO_4^- with the hydrogen bonding type structures, while their anion-π type structures show slightly weaker binding. Consequently, the cyameluric acid having the C_{3h} symmetric C_6N_7 nucleus with electron withdrawing oxygen atoms is a novel anion-π type receptor for trigonal-planar and tetrahedral anions. The structures of the cyameluric acid interacting with Cl^- and ClO_4^- are considered as the π stacking type. For the cyameluric acid interacting with NO_3^-, the π(edge) type complex is only slightly more favored over the π(stack) type in the gas phase, but the π(stack) type is likely to be as stable as the π(edge) type in the solvent phase.

Introduction

As π-rings are involved in diverse types of π-interactions[1] (π-π interactions,[2–7] H-π interactions,[8–10] cation-π interactions,[11,12] and anion-π interactions[13–15]), understanding these interactions are essential for studying protein structures,[8,16] nucleic acids structures,[17] carbon nanotube structures,[18] crystal packing,[19] molecular/biomolecular recognition,[20] and functional molecular/material systems.[21] Indeed, there have been numerous studies on π-interations.[22–24] However, as compared with other types of π-interactions, the anion-π interactions have been investigated very recently. Nevertheless, in the past few years, considerable progress has been made both theoretically and experimentally on the roles of anion-π interactions in supramolecular chemistry.[13–15] Recently, anion-π interactions have been extensively investigated for the host architecture and supramolecular self-assembly. In addition, anion-π interactions have been used for the design and synthesis of an anion host as anion recognition.[15] In this regard, we have been interested in designing a new type of anion receptors which utilize the anion-π interactions.

Cyclic C_3N_3 nucleus systems are known to be a remarkable unit for supramolecular assembly.[25] Because of the partially positive charge on C atoms of cyclic C_3N_3 type molecule with D_{3h} symmetry, one might expect that the cyclic C_3N_3 nucleus systems could interact favorably with trigonal-planar anions. In this regard, it would be interesting to investigate if cyclic C_6N_7 nucleus systems would be a template for the construction of tripodal and cylindrophane molecular systems for the recognition of trigonal-planar and tetrahedral anions. C_6N_7 nucleus systems are also well-known as the building blocks of the diamondlike graphitic carbon(IV) nitrides[26] and as the burn-rate suppressants for solid rocket propellants.[27]

Toward this direction, using high level ab initio calculations, we here investigate how the keto forms of cyanuric acid (**1**: $C_3N_3O_3H_3$) with the C_3N_3 nucleus and the cyameluric acid (**2**: $C_6N_7O_3H_3$)[28,29] with the C_6N_7 nucleus (in Figure 1) interact with a halide anion (Cl^-), a trigonal-planar anion (NO_3^-), and a tetrahedral anion (ClO_4^-). Indeed, we find that **2** is a novel anion receptor moiety which binds trigonal-planar or tetrahedral anions through the anion-π interaction. Cyanuric acid and cyameluric acid have five and seventeen tautomeric forms, respectively. In the gas phase, **1** and **2** are the most stable keto forms among the tautomers of cyanuric acid and cyameluric acid, respectively, and these keto forms are likely to remain stable in water and methanol.[28] Recently, it is experimentally noted that the keto form of cyameluric acid is favored in the crystal.[29] Furthermore,

* Corresponding author e-mail: kim@postech.ac.kr.

10.1021/ct8001255 CCC: $40.75
Published on Web 06/28/2008

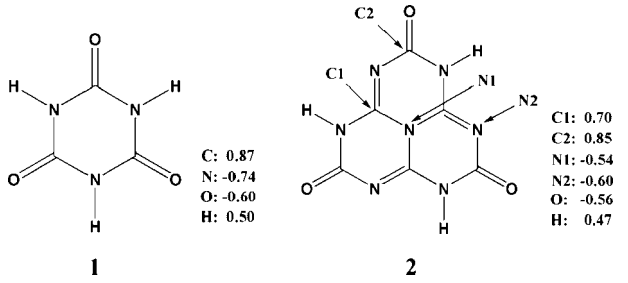


Figure 1. NBO charges (au) of cyanuric acids (**1**: $C_3N_3O_3H_3$) and cyameluric acids (**2**: $C_6N_7O_3H_3$).

in consideration of the futuristic synthetic functionalization of the hydrogen atoms with the functional groups that could enhance the binding affinity of the trigonal-planar and tetrahedral oxyanions, we have chosen **1** and **2**. This will enable us to understand the strength of the anion-π interaction in the anion complexed with the tripodal based receptors consisting of cyameluric acid ring template.

Calculation Methods

We performed ab initio calculations using Gaussian 03 and Molpro programs.[30] The initial structures were optimized by using Møller−Plesset second-order perturbation theory (MP2) with the 6-31+G* basis set and then optimized at the MP2 level of theory with the aug-cc-pVDZ (abbreviated as aVDZ) basis set. To confirm their minimum energy structures, the frequency analysis was done at the MP2/aVDZ level of theory. Then, the low lying energy structures were fully optimized on the potential surfaces that are corrected for basis set superposition error (BSSE) using the counterpoise (CP) method of Boys and Bernardi[31] for the complexes of **1**. However, these BSSE corrected geometries were obtained by optimizing only the vertical distances of the anion from the ring plane with the fixed monomer geometries using the CP method for the **2**-anion complexes. The effects of the inclusion of higher level of correlation energies were examined by performing the coupled cluster calculations with single and double excitations (CCSD) and the CCSD with perturbative triple excitations [CCSD(T)] using the aVDZ basis set on the BSSE corrected MP2/aVDZ optimized geometries. Then, the MP2/aug-cc-pVTZ (abbreviated as aVTZ) calculations were carried out, and the complete basis set (CBS) limit interaction energies at the MP2 level were obtained based on the extrapolation method exploiting that the electron correlation energy is proportional to N^{-3} for the aug-cc-pV*N*Z basis set.[32,33] For **1** interacting with Cl^-, the CCSD(T)/aVTZ calculations were performed, and the CBS limit interaction energies at the CCSD(T) level [CCSD(T)/CBS] were evaluated. These CCSD(T)/CBS binding energies (which were evaluated by extrapolating the aVDZ and aVTZ values) are found to be very close to the CCSD(T)/CBS* binding energies which were evaluated from the MP2/CBS binding energies simply by adding the difference between the MP2/aVDZ and CCSD(T)/aVDZ binding energies (without the CCSD(T)/ aVTZ value).[32] Owing to the large size of NO_3^- and ClO_4^-, these interaction energies with **1** are reported with the CCSD(T)/CBS* binding energies. Using the MP2/aVDZ zero point energies (ZPE) and thermal energies, the ZPE-corrected interaction energies (ΔE_0), enthalpies (ΔH_{298} at 1 atm), and free energies (ΔG_{298} at 1 atm) at 298 K were evaluated. The natural bond orbital (NBO) charges were obtained at the MP2/aVDZ level. In the case of **2** interacting with anions, we evaluated the MP2/CBS and CCSD/CBS* values (but not the CCSD(T)/CBS* values due to their large size). To evaluate the interaction energies in the condensed phase, single point energy calculations were performed using self-consistent reaction field (SCRF) theory with the isodensity surface polarized continuum model (IPCM)[34] at the MP2/aVDZ level. The solvent dielectric constant used for water is 78.

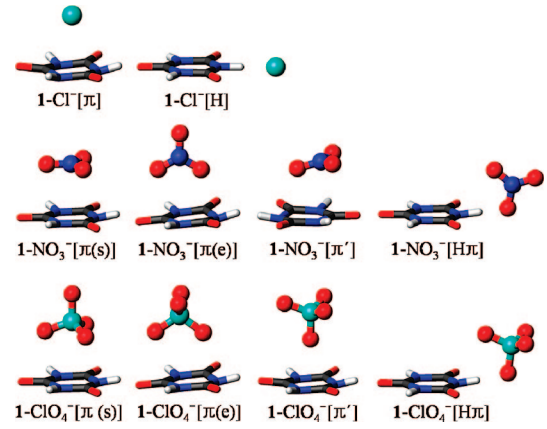


Figure 2. Structures of the 1-Cl^-[π, H], 1-NO_3^-[π(s: stack), π(e: edge), π', Hπ], and 1-ClO_4^-[π(s), π(e), π', Hπ] complexes.

To understand the nature of the interaction energies, we analyzed the energy components [E_{es}: electrostatic energy, E_{ind*}: effective induction energy including the induction-induced exchange energy and the contribution of third and higher order on the uncorrelated level ($\Delta E_{ind*} = \Delta E_{ind} + \Delta E_{exch\text{-}ind} + \Delta E_{int,HF}$), E_{disp*}: effective dispersion energy including the dispersion-induced exchange energy ($\Delta E_{disp*} = \Delta E_{disp} + \Delta E_{exch\text{-}disp}$), and E_{exch*}: effective exchange repulsion energy with the induction-induced and dispersion-induced exchange energies excluded ($\Delta E_{exch*} = \Delta E_{exch} - \Delta E_{exch\text{-}ind} - \Delta E_{exch\text{-}disp}$)]. The symmetry adapted perturbation theory (SAPT)[35] combined with density functional theory (DFT) for monomer properties, called DFT-SAPT by Hesselmann et al.,[36] is known to yield reliable energy components with the asymptotically corrected PBE0 (PBE0AC) exchange-correlation (xc) functional. Using the density-fitting (DF) approximation, SAPT calculations can be successfully performed on medium sized systems. In the present study, the PBE0AC xc functional with the ALDA xc kernel was used. In the DF-DFT-SAPT calculations, a purely local ALDA xc kernel was used for hybrid xc functionals.[37] The aVDZ set was used as the atomic basis set, and the cc-pVTZ MP2-fitting set was employed for the DF approximations.

Results and Discussion

Figure 2 shows optimized structures of **1** interacting with Cl^-, NO_3^-, and ClO_4^-. Those anions can interact with **1** in either the hydrogen bonding type [H-type] or the anion-π type [π-type]. For **1**-Cl^-, the covalent bonding type of

Table 1. Intermolecular Distances (R in Å) and Interaction Energies (ΔE_e in kcal/mol) for the **1**-Anion Complexes[a]

	MP2				CCSD		CCSD(T)	
	R	aVDZ	aVTZ	CBS	aVDZ	CBS*	aVDZ	CBS*
Cl^-								
π	2.94	−18.00	−18.91	**−19.29**	−17.57	**−18.86**	−18.16	**−19.45**
H	1.84	−21.74	−22.92	**−23.41**	−19.61	**−21.28**	−20.94	**−22.62**
NO_3^-								
π(s)	2.90	−15.14	−16.63	**−17.25**	−13.61	**−15.71**	−14.94	**−17.05**
π(e)	2.56	−16.72	−17.33	**−17.59**	−16.32	**−17.20**	−17.13	**−18.01**
π'	2.64	−15.92	−16.63	**−16.92**	−14.74	**−15.75**	−15.87	**−16.88**
Hπ	1.55	−21.56	−22.20	**−22.47**	−20.16	**−21.07**	−21.38	**−22.29**
ClO_4^-								
π(s)	2.78	−14.02	−14.59	**−14.84**	−12.02	**−12.84**	−13.37	**−14.19**
π(e)	2.55	−13.05	−13.95	**−14.32**	−11.96	**−13.24**	−13.02	**−14.29**
π'	2.51	−11.54	−12.22	**−12.51**				
Hπ	1.73	−15.05	−15.60	**−15.84**	−13.95	**−14.73**	−15.00	**−15.78**

[a] The BSSE-corrected MP2/CBS and CCSD(T)/CBS energies were estimated based on the extrapolation method exploiting that the electron correlation energy is proportional to N^{-3} for the aV*N*Z basis set. The CCSD(T)/CBS* energies were estimated using [$E_{CCSD/CBS} = E_{MP2/CBS} + (E_{CCSD/aVDZ} - E_{MP2/aVDZ})$]. R is either the vertical distance from the ion to the ring plane of **1** (for π complexes) or the shortest H-bond distance from the ion to the N atom of **1** (for H complexes). In the case of **1**−Cl^-, we also calculated the CCSD(T)/aVTZ energies (−18.94/−22.30 kcal/mol for π-type/H-type). By using these aVTZ values, the extrapolated CCSD(T)/CBS energies for the π-type/H-type are −19.27/−22.87 kcal/mol, which are very close to the CCSD(T)/CBS* energies (−19.45/−22.62 kcal/mol). Structures *NO_3^-:π'* and *ClO_4^-:π*(e) are transition states, and structure *ClO_4^-:π'* is a saddle point with three imaginary frequencies.

Table 2. Thermodynamic Quantities (in kcal/mol at the CCSD(T)/CBS* Level) for the **1**-Anion Complexes

	Cl^-		NO_3^-			ClO_4^-	
	π	H	π(s)	π(e)	Hπ	π(s)	Hπ
ΔE_0	−18.95	−23.64	−16.27	−17.18	−22.40	−13.46	−15.12
ΔH_{298}	−19.15	−23.81	−15.92	−16.77	−22.01	−13.03	−14.62
ΔG_{298}	−12.15	−18.13	−4.99	−7.52	−13.54	−3.45	−7.26

complex (where an anion attacks a partially positive carbon atom of π system, resulting in transformation from the sp^2 to sp^3 hybrid orbital) is not observed. For the complexes of **1** with NO_3^- and ClO_4^-, the H-type is combined with the π-type [Hπ-type]. In the case of π-type, three different isomers [π(stack), π(edge), and π' types] are observed. **1**-NO_3^-[π'] and **1**-ClO_4^-[π(edge)] complexes with one small imaginary frequencies would be the transition states of the Hπ-type. **1**-ClO_4^-[π'] has three imaginary frequencies. Other optimized geometries are the local or global minima geometries. Similar structural and energetic studies for the case of s-triazine with NO_3^- and BF_4^- have been done by other groups.[38] We have listed the CBS limit interaction energies (ΔE_e) at the MP2/CCSD/CCSD(T) level in Table 1. The ΔE_0, ΔH_{298} and ΔG_{298} are obtained from the CCSD(T)/CBS* binding energies with the MP2/aVDZ ZPE and thermal energies corrections (Table 2). In tables and figures, π(stack)/π(edge) will be simply denoted as π(s)/π(e) for brevity.

At the MP2/CBS* level of theory, the π(edge)-type complex is 0.3 kcal/mol more stable than the π(stack)-type for **1**-NO_3^-, while the π(stack)-type is 0.5 kcal/mol more stable than the π(edge)-type for **1**-ClO_4^-. For **1**-ClO_4^-, however, the π(edge)-type is 0.1 kcal/mol more stable than the π(stack)-type at the CCSD(T)/CBS* level of theory. For the complexes of **1** with Cl^-, NO_3^-, and ClO_4^-, the interaction energies of the H(π)-type complexes are lower than the most stable π-type complexes by 3.2, 4.2, and 1.5 kcal/mol at the CCSD(T)/CBS* level, respectively. For Cl^-, the interaction energy for the H-type/π-type is −19.27/−22.87 kcal/mol at the CCSD(T)/CBS level and −19.45/−22.62 kcal/mol at the CCSD(T)/CBS* level, showing that both CBS and CBS* values are almost the same. In Table 1, the MP2/CBS interaction energies are close to the CCSD(T)/CBS* values, while the CCSD/CBS* values are slightly underestimated but consistently lower than the CCSD(T)/CBS* values. In this regard, in the case of **2** interacting with the anions, we report the CCSD/CBS* energies.

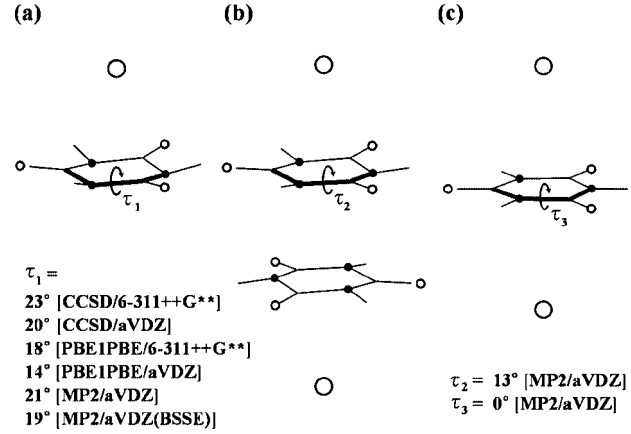


Figure 3. The C−N−C−N torsion angles (τ: dihedral angles) in the optimized geometries of **1**-Cl^-(a), Cl^--**1**-**1**-Cl^-(b), and Cl^--**1**-Cl^-(c).

For **1**-Cl^-/NO_3^-/ClO_4^- complexes, $\Delta\Delta E_0$ [ΔE_0(π-type)−ΔE_0(H(π)-type)] is 4.7/5.2/1.7 kcal/mol, and $\Delta\Delta G_{298}$ [ΔG_{298}(π-type)−ΔG_{298}(H(π)-type)] is 6.0/6.0/3.8 kcal/mol, respectively. Therefore, for all the three cases, the H(π) types are much more stable.

The MP2/aVDZ optimized geometry of **1**-Cl^-[π] shows that the ring plane is significantly distorted with the C−N−C−N torsion angles of 21° in Figure 3a, which is in good agreement with the value (20°) at the CCSD/aVDZ level. This ring distortion is due to the partial change in the hybrdization from sp^2 to sp^3 of both peripheral carbon and nitrogen atoms. Since **1** does not show strong electron delocalization in contrast to the aromatic ring systems, **1** can

Table 3. DFT-SAPT (Density Fitting) Interaction Energy Components (kcal/mol) for the **1**-Anion Complexes[a]

	Cl^-		NO_3^-			ClO_4^-	
	π	H	π(s)	π(e)	Hπ	π(s)	Hπ
ΔE_{int}	−17.40	−21.22	−13.85	−15.61	−20.02	−11.77	−14.07
ΔE_{deform}	2.91	5.10	1.99	2.46	4.60	1.83	2.31
ΔE_{sapt}	−20.31	−26.32	−15.84	−18.07	−24.62	−13.60	−16.37
ΔE_{es}	−28.33	−32.61	−17.09	−21.99	−33.06	−19.48	−19.41
ΔE_{ind^*}	−8.07	−24.55	−5.90	−7.33	−20.83	−6.20	−10.95
ΔE_{disp^*}	−6.97	−6.60	−8.57	−7.63	−6.99	−9.96	−5.73
ΔE_{exch^*}	23.07	37.45	15.72	18.89	36.25	22.04	19.72

[a] The DFT-SAPT values without density fitting are also calculated. In the case of **1**−Cl^-, ΔE_{int} = −17.86/−21.74, ΔE_{sapt} = −20.77/−26.84, ΔE_{es} = −28.32/−32.61, ΔE_{ind^*} = −8.06/−24.53, ΔE_{disp^*} = −7.44/−7.13, and ΔE_{exch^*} = 23.05/37.43 kcal/mol. ΔE_{deform}: CCSD(T)/aVDZ deformation energy (ΔE_{int} = ΔE_{sapt} + ΔE_{deform}).

be easily distorted when interacting with a highly electronegative anion. However, in the crystal structures of cyanuric acid nucleus that shows favorable anion-π interactions with anions, the planarity of the ring is well restored. For example, the crystal structure of Fe-pseudocryptand [CCDC No. 623175] in which the cyanuric acid nucleus interacts with Cl^- through the anion-π interactions[39] shows the C−N−C−N torsion angle of 8° (see the Supporting Information). Here, the cyanuric acid ring plane interacts with Cl^- in one side, with another cyanuric acid ring plane in the other side through the π-π interactions, forming an anion-π-π-anion complex. Another crystal structure of the cyanuric acid with Cl^- is found to have the Cl^--**1**-Cl^- complex with the C−N−C−N dihedral angle of 0°.[40] In the MP2 calculated geometries of the Cl^--**1**−**1**-Cl^- complex and the Cl^--**1**-Cl^- complex, the C−N−C−N torsion angle is 13° and 0°, respectively, which is in good agreement with the crystal structures. Therefore, the difference between the calculated geometry of the isolated **1**-Cl^- complex and the X-ray strcuture in the crystal form arises from the fact that the X-ray structure is for the condensed phase, while the calculated structure is for the isolated molecule in the gas phase.

Table 3 shows the DFT-SAPT interaction energy components. It should be noted that ΔE_{sapt} is the interaction energy between individual monomers on the complex geometries. Upon the complexation with an anion, there are some deformations in geometries of the π-rings and anions. At the CCSD(T)/aVDZ level, ΔE_{deform} for the H-type/Hπ-type of **1**-Cl^-/NO_3^- complex is 5.1/4.6 kcal/mol. What distinguishes the interactions between the H(π)-type and the π-type is the magnitude of the induction energies and dispersion energies. In the case of the H(π)-type, the induction contribution (38/34/30% for Cl^-/NO_3^-/ClO_4^-) to the total attractive interaction energies (ΔE_{es} + ΔE_{ind^*} + ΔE_{disp^*}) is much larger than that for the π-type (19/20/17%), while the dispersion contribution (10/11/16%) is much smaller than that for the π-type (16/21/28%). The electrostatic energy components of both the H(π)-type (51/54/54% for Cl^-/NO_3^-/ClO_4^-) and the π-type (65/60/55%) are large. Owing to the large diffuse nature of the electron cloud in the anionic systems, the magnitude of the exchange repulsion energies in these anion complexes, which is similar to the magnitude of the electrostatic interaction energy, is substantially larger than that in the case of the complexes involving cations.[12f]

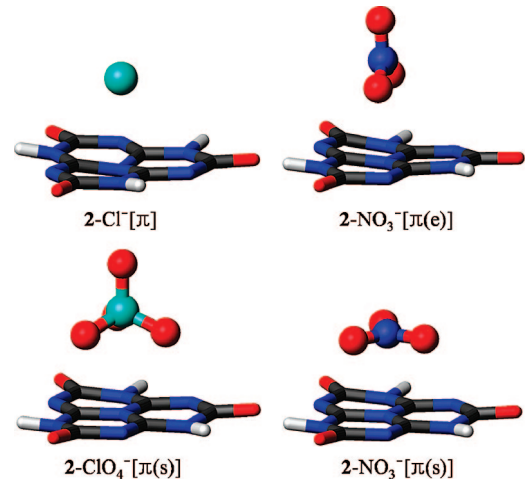


Figure 4. Structures of the **2**-Cl^-[π], **2**-NO_3^-[π(e), π(s)], and **2**-ClO_4^-[π(s)] complexes.

Table 4. Intermolecular Distance (R in Å), Interaction Energies (ΔE_e in kcal/mol), and SAPT(DF-DFT(PBE0AC)/aVDZ) Interaction Energy Components (in kcal/mol) for the **2**-Anion Complexes

	Cl^-	NO_3^-		ClO_4^-
	π	π(s)	π(e)	π(s)
R[a]	2.94	2.74	2.54	2.72
MP2/aVDZ	−32.19	−29.84	−30.74	−28.08
MP2/ aVTZ	−33.44	−31.11	−31.68	−29.40
MP2/CBS	**−33.96**	−31.64	−32.08	−29.96
CCSD/aVDZ	−30.94	−27.23	−29.58	−25.81
CCSD/CBS*	**−32.71**	−29.30	−30.92	−27.69
ΔE_{int}	−31.38	−27.39	−29.14	−25.69
ΔE_{def}	2.59	2.01	2.46	2.15
ΔE_{sapt}	−33.97	−29.40	−31.60	−27.84
ΔE_{es}	−44.16	−32.99	−36.90	−30.17
ΔE_{ind^*}	−12.66	−10.52	−11.86	−9.05
ΔE_{disp^*}	−9.64	−13.46	−11.07	−12.85
ΔE_{exch^*}	32.49	27.57	28.23	24.24
$\Delta E_{ipcm(water)}$[b]	−11.94	−15.89	−15.30	−18.14

[a] R is the shortest vertical distance from an anion atom (O) to the ring plane. [b] Interaction energies in water (dielectric constant: ϵ = 78) are estimated using SCRF/IPCM(MP2/aVDZ).

Since the structure of **1** with D_{3h} symmetry has a highly positive charge on each C atom (0.87 au) due to the neighboring N atoms (with larger electronegativity than the C atom) and the electron withdrawing substituents of =O (Figure 1), it favors interacting with trigonal-planar and tetrahedral anions. Similarly, we note that the structure of **2** with C_{3h} symmetry also has highly positive charge on each C1 carbon atom (0.70 au) surrounded by three N atoms including the centroid N atom (−0.54 au) and on each C2 carbon atom (0.85 au) surrounded by two N atoms and one =O atom. Therefore, we can expect that **2** would be a novel anion receptor recognizing trigonal-planar and tetrahedral anions.

Figure 4 shows the optimized structures of **2**-complexes interacting with Cl^-, NO_3^-, and ClO_4^-, and Table 4 lists the CBS interaction energies and the SAPT interaction energy components.

For the **2**-Cl^-/NO_3^-/ClO_4^- complexes, the interaction energies of the π-type complexes are ~30 kcal/mol, which are greater than those of the charged H-bonding complexes. Any H-types of the **2**-NO_3^-/ClO_4^- complexes and any covalent types of the **2**-Cl^-/NO_3^-/ClO_4^- complexes merged to the π-type complexes during the geometry optimization. The H-type of the **2**-Cl^- complex, though it exists, is much less stable than the π-type complex. As compared with the π(stack)-type of **1**, in the π(stack)-type of **2** the electrostatic energy increases almost twice (from −17 to −33 kcal/mol) for the binding with NO_3^-, while it increases only 1.5 times (from −28/−19 to −44/−30 kcal/mol) for the binding with Cl^-/ClO_4^-. It is because the positive N on the centroid of **2** favorably interacts with the negative N on the center of NO_3^-. Nevertheless, in the case of the **2**-NO_3^- complex, the π(edge)-type (i.e., edge-to-face type) is slightly more stable than the π(stack)-type by ~1 kcal/mol, because the π(edge)-type (wherein the highly positively charged N atom of NO_3^- stays away from the highly positive C atoms on the rings of **2**) has the larger electrostatic energy gain by ~4 kcal/mol (−33.0 vs −36.9 kcal/mol in ΔE_{es} for the edge vs stack form). However, both π(edge)-type and the π(stack)-type are nearly isoenergetic, so that the structure can be changed depending on the solvation and packing effects. The π(edge)-type which has a larger electrostatic energy gain is expected to be slightly less stabilized than the π(stack)-type in the condensed phase, so that the π(stack)-type would be easily observable in crystals. Indeed, the SCRF/IPCM(MP2/aVDZ) calculations show that in water the π(stack)-type is ~0.6 kcal/mol more stable than the π(edge)-type (Table 4). Thus, all the structures of the **2**-Cl^-/NO_3^-/ClO_4^- complexes would have the anion-π interactions in the π-stacked form in solvents and crystals, while for the **2**-NO_3^- complex the π(edge)-type is also compatible. What is more important is that the interaction energy of **2** with Cl^-/NO_3^-/ClO_4^- in the gas phase changes from −32/−31/−28 kcal/mol (at the MP2/aVDZ level) to −12/−16/−18 kcal/mol in water. Thus, ClO_4^- and NO_3^-, which have smaller binding energies than Cl^- in the gas phase, have larger binding energies in water, indicating that **2** is a selective receptor for ClO_4^- and NO_3^- in solution.

Conclusions

Using high levels of ab initio theory including coupled cluster theory at the complete basis limit, we have investigated the possibility of designing anion receptors which strongly bind trigonal-planar and tetrahedral anions. The cyanuric acid ($C_3N_3O_3H_3$) binds Cl^-, NO_3^-, and ClO_4^- as H(π)-type structures, while its anion-π type structures are less stable. Thus, we studied the cases of the cyameluric acid ($C_6N_7O_3H_3$) having the C_{3h} symmetric C_6N_7 nucleus with electron withdrawing O atoms and Cl^-, NO_3^-, and ClO_4^- anions having π-type structures with the binding energies of ~30 kcal/mol in the gas phase. Although Cl^- has larger binding energies than NO_3^- and ClO_4^- in the gas phase, this trend is reversed in water. The cyameluric acid is a novel anion-π type receptor for ClO_4^- and NO_3^-, and consequently, it would be used as a template for building tripodal and cylindrophane receptors for the selective recognition of trigonal-planar and tetrahedral anions toward the design of novel anion receptors.[41]

Acknowledgment. This work was supported by GR-L(KICOS) and BK21(KRF).

Supporting Information Available: XYZ coordinates of the MP2/aug-cc-pVDZ BSSE-corrected optimized geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Hobza, P.; Selzle, H. L.; Schlag, E. W. *Chem. Rev.* **1994**, *94*, 1767. (b) Brutschy, B. *Chem. Rev.* **2000**, *100*, 3891. (c) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *Chem. Rev.* **2000**, *100*, 4145.

(2) (a) Janda, K. C.; Hemminger, J. C.; Winn, J. S.; Novick, S. E.; Harris, S. J.; Klemperer, W. *J. Chem. Phys.* **1975**, *63*, 1419. (b) Steed, J. M.; Dixon, T. A.; Klemperer, W. *J. Chem. Phys.* **1979**, *70*, 4940.

(3) (a) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525. (b) Hunter, C. A. *Chem. Soc. Rev.* **1994**, *23*, 101.

(4) (a) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1990**, *93*, 5893. (b) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1993**, *97*, 3937. (c) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500. (d) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790. (e) Špirko, V.; Engkvist, O.; Soldán, P.; Selzle, H. L.; Schlagm, E. W.; Hobza, P. *J. Chem. Phys.* **1999**, *111*, 572. (f) Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802. (g) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142.

(5) (a) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887. (b) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*, 8377. (c) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200. (d) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690. (e) Ringer, A. L.; Sinnokrot, M. O.; Lively, R. P.; Sherrill, C. D. *Chem. Eur. J.* **2006**, *12*, 3821. (f) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.

(6) (a) Tsuzuki, S.; Uchimaru, T.; Sugawara, K. -I.; Mikami, M. *J. Chem. Phys.* **2002**, *117*, 11216. (b) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104. (c) Tsuzuki, S.; Uchimaru, T.; Mikami, M. *J. Phys. Chem. A* **2006**, *110*, 2027. (d) Grimme, S. *Chem. Eur. J.* **2004**, *10*, 3423. (e) Piacenza, M.; Grimme, S. *J. Am. Chem. Soc.* **2005**, *127*, 14841.

(7) (a) Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, *111*, 3446. (b) Kim, K. S.; Suh, S. B.; Kim, J. C.; Hong, B. H.; Lee, E. C.; Yun, S.; Tarakeshwar, P.; Lee, J. Y.; Kim, Y.; Ihm, H.; Kim, H. G.; Lee, J. W.; Kim, J. K.; Lee, H. M.; Kim, D.; Cui, C.; Youn, S. J.; Chung, H. Y.; Choi, H. S.; Lee, C.-W.; Cho, S. J.; Jeong, S.; Cho, J.-H. *J. Am. Chem. Soc.* **2002**, *124*, 14268.

(8) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23.

(9) (a) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1990**, *93*, 5893. (b) Vaupel, S.; Brutschy, B.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2006**, *128*, 5416. (c) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27. (d) Paliwal, S.; Geib, S.; Wilcox, C. S. *J. Am. Chem. Soc.* **1994**, *116*, 4497. (e) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768.

(10) (a) Lee, E. C.; Hong, B. H.; Lee, J. Y.; Kim, J. C.; Kim, D.; Kim, Y.; Tarakeshwar, P.; Kim, K. S. *J. Am. Chem. Soc.* **2005**, *127*, 4530. (b) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *J. Am. Chem. Soc.* **2001**, *123*, 3323. (c) Hong, B. H.; Lee, J. Y.; Cho, S. J.; Yun, S.; Kim, K. S. *J. Org. Chem.* **1999**, *64*, 5661. (d) Tarakeshwar, P.; Lee, S. J.; Lee, J. Y.; Kim, K. S. *J. Chem. Phys.* **1998**, *108*, 7217. (e) Ren, R.; Jin, Y.; Kim, K. S.; Kim, D. H. *J. Biomol. Struct. Dyn.* **1997**, *15*, 401.

(11) Meot-Ner (Mautner) , M.; Deakyne, C. A. *J. Am. Chem. Soc.* **1985**, *107*, 469.

(12) (a) Dougherty, D. A.; Stauffer, D. *Science* **1990**, *250*, 1558. (b) Kumpf, R. A.; Dougherty, D. A. *Science* **1993**, *261*, 1708. (c) Kim, K. S.; Lee, J. Y.; Lee, S. J.; Ha, T.-K.; Kim, D. H. *J. Am. Chem. Soc.* **1994**, *116*, 7399. (d) Lee, J. Y.; Lee, S. J.; Choi, H. S.; Cho, S. J.; Kim, K. S.; Ha, T. K. *Chem. Phys. Lett.* **1995**, *232*, 67. (e) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303. (f) Kim, D.; Hu, S.; Tarakeshwar, P.; Kim, K. S.; Lisy, J. M. *J. Phys. Chem. A* **2003**, *107*, 1228.

(13) Schneider, H. J. *Angew. Chem.* **1991**, *103*, 1419.

(14) (a) Alkorta, I.; Rozas, I.; Elguero, J. *J. Org. Chem.* **1997**, *62*, 4687. (b) Mascal, M.; Armstrong, A.; Bartberger, M. *J. Am. Chem. Soc.* **2002**, *124*, 6274. (c) Alkorta, I.; Rozas, I.; Elguero, J. *J. Am. Chem. Soc.* **2002**, *124*, 8593. (d) Mascal, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 2890. (e) Quiñonero, D.; Garau, C.; Rotger, C.; Frontera, A.; Ballester, P.; Costa, A.; Deyà, P. M. *Angew. Chem., Int. Ed.* **2002**, *41*, 3389. (f) Frontera, A.; Saczewski, F.; Gdaniec, M.; Dziemidowicz-Borys, E.; Kurland, A.; Deya, P. M.; Quiñonero, D.; Garau, C. *Chem.-Eur. J.* **2005**, *11*, 6560. (g) Quiñonero, D.; Garau, C.; Frontera, A.; Ballester, P.; Costa, A.; Deyà, P. M. *J. Phys. Chem. A* **2005**, *109*, 4632. (h) Kim, D.; Tarakeshwar, P.; Kim, K. S. *J. Phys. Chem. A* **2004**, *108*, 1250. (i) Kim, D.; Lee, E. C.; Kim, K. S.; Tarakeshwar, P. *J. Phys. Chem. A* **2007**, *111*, 7980. (j) Berryman, O. B.; Bryantsev, V. S.; Stay, D. P.; Johnson, D. W.; Hay, B. P. *J. Am. Chem. Soc.* **2007**, *129*, 48. (k) Hermida-Ramón, J. M.; Estevéz, C. M. *Chem.-Eur. J.* **2007**, *13*, 4743. (l) Bryantsev, V. S.; Hay, B. P. *J. Am. Chem. Soc.* **2005**, *127*, 8282. (m) Gamez, P.; Mooibroek, T. J.; Teat, S. J.; Reedijk, J. *Acc. Chem. Res.* **2007**, *40*, 435.

(15) (a) Demeshko, S.; Dechert, S.; Meyer, F. *J. Am. Chem. Soc.* **2004**, *126*, 4508. (b) Hoog, P. de.; Gamez, P.; Mutikainen, I.; Turpeinen, U.; Reedijk, J. *Angew. Chem., Int. Ed.* **2004**, *43*, 5815. (c) Rosokha, Y. S.; Lindeman, S. V.; Rosokha, S. V.; Kochi, J. K. *Angew. Chem., Int. Ed.* **2004**, *43*, 4650. (d) Schottel, B. L.; Chifotides, H. T.; Shatruk, M.; Chouai, A.; Pérez, L. M.; Bacsa, J.; Dunbar, K. R. *J. Am. Chem. Soc.* **2006**, *128*, 5895. (e) Mascal, M.; Yakovlev, I.; Nikitin, E. B.; Fettinger, J. C. *Angew. Chem., Int. Ed.* **2006**, *45*, 2890. (f) Tanaka, D.; Masaoka, S.; Horike, S.; Furukawa, S.; Mizuno, M.; Endo, K.; Kitagawa, S. *Angew. Chem., Int. Ed.* **2006**, *45*, 4628. (g) Gorteau, V.; Bollot, G.; Mareda, J.; Matile, S. *Org. Biomol. Chem.* **2007**, *5*, 3000. (h) Schneider, H.; Vogelhuber, K. M.; Schinle, F.; Weber, M. *J. Am. Chem. Soc.* **2007**, *129*, 13022. (i) Schottel, B. L.; Chifotides, H. T.; Dunbar, K. R. *Chem. Soc. Rev.* **2008**, *37*, 68.

(16) Gazit, E. *FASEB J.* **2002**, *16*, 77.

(17) (a) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608. (b) Jurečka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985. (c) Sponer, J.; Jurečka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. *Chem. Eur. J.* **2006**, *12*, 2854. (d) Sedlak, R.; Jurečka, P.; Hobza, P. *J. Chem. Phys.* **2007**, *127*, 075104.

(18) Hong, B. H.; Small, J. P.; Purewal, M. S.; Mullokandov, A.; Sfeir, M. Y.; Wang, F.; Lee, J. Y.; Heinz, T. F.; Brus, L. E.; Kim, P.; Kim, K. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14155.

(19) (a) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210. (b) Hong, B. H.; Lee, J. Y.; Lee, C.-W.; Kim, J. C.; Bae, S. C.; Kim, K. S. *J. Am. Chem. Soc.* **2001**, *123*, 10748. (c) Hong, B. H.; Bae, S. C.; Lee, C.-W.; Jeong, S.; Kim, K. S. *Science* **2001**, *294*, 348. (d) Mooibroek, T. J.; Teat, S. J.; Massera, C.; Gamez, P.; Reedijk, J. *Cryst. Growth Des.* **2006**, *6*, 1569.

(20) (a) Oh, K. S.; Lee, C.-W.; Choi, H. S.; Lee, S. J.; Kim, K. S. *Org. Lett.* **2000**, *2*, 2679. (b) Yun, S.; Kim, Y.-O.; Kim, D.; Kim, H. G.; lhm, H.; Kim, J. K.; Lee, C.-W.; Lee, W. J.; Yoon, J.; Oh, K. S.; Yoon, J.; Park, S.-M.; Kim, K. S. *Org. Lett.* **2003**, *5*, 471.

(21) (a) Singh, N. J.; Lee, H. M.; Suh, S. B.; Kim, K. S. *Pure Appl. Chem.* **2007**, *79*, 1057. (b) Singh, N. J.; Lee, E. C.; Choi, Y. C.; Lee, H. M.; Kim, K. S. *Bull. Chem. Soc. Jpn.* **2007**, *80*, 1437. (c) Singh, N. J.; Lee, H. M.; Hwang, I.-C.; Kim, K. S. *Supramol. Chem.* **2007**, *19*, 321.

(22) Ruhl, E.; Bisling, P. G. F.; Brutschy, B.; Baumgartel, H. *Chem. Phys. Lett.* **1986**, *126*, 232.

(23) (a) Adamovic, I.; Li, H.; Lamm, M. H.; Gordon, M. S. *J. Phys. Chem. A* **2006**, *110*, 519. (b) Mishra, B. K.; Sathyamurthy, N. *J. Phys. Chem. A* **2005**, *109*, 6.

(24) (ah) Sato, T.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2005**, *123*, 104307. (b) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 4209. (c) Piacenza, M.; Grimme, S. *Chem. Phys. Chem.* **2005**, *6*, 1554. (d) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345.

(25) (a) Mooibroek, T. J.; Gamez, P. *Inorg. Chim. Acta* **2007**, *360*, 381. (b) Gamez, P.; Reedijk, J. *Eur. J. Inorg. Chem.* **2006**, 29.

(26) (a) Liu, A. Y.; Cohen, M. L. *Science* **1989**, *245*, 841. (b) Souto, S.; Pickholz, M.; Santos, M. C. Dos.; Alvarez, F. *Phys. Rev. B* **1998**, *57*, 2536.

(27) Williams, G. K.; Palopoli, S. F.; Brill, T. B. *Combust. Flame* **1994**, *98*, 197.

(28) (a) Liang, X.; Pu, X.; Zhou, H.; Wong, N.-B.; Tian, A. *THEOCHEM* **2007**, *816*, 125. (b) Liang, X.; Zheng, W.; Wong, N.-B.; Shu, Y.; Tian, A. *THEOCHEM* **2005**, *732*, 127. (c) Alkorta, I.; Jagerovic, N.; Elguero, J. *ARKIVOC* **2004**, 130.

(29) (a) Sattler, A.; Schnick, W. *Z. Anorg. Allg. Chem.* **2006**, *632*, 1518. (b) Wager, J.; El-Gamel, N. E. A.; Kroke, E. *Z. Naturforsch.* **2006**, *61*, 975. (c) El-Gamel, N. E. A.; Seyfarth, L.; Wagler, J.; Ehrenberg, H.; Schwarz, M.; Senker, J.; Kroke, E. *Chem. Eur. J.* **2007**, *13*, 1158.

(30) (a) 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. (b) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO*, a package of *ab initio* programs, version 2006.1.

(31) (a) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553. (b) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.

(32) (a) Helgaker, T.; Ruden, T. A.; Jorgensen, P.; Olsen, J.; Klopper, W. *J. Phys. Org. Chem.* **2004**, *17*, 913. (b) Lee, H. M.; Kim, D.; Singh, N. J.; Kolaski, M.; Kim, K. S. *J. Chem. Phys.* **2007**, *127*, 164311.

(33) Min, S. K.; LEE, E. C.; Lee, H. M.; Kim, D. Y.; Kim, D. *J. Comput. Chem.* **2007**, DOI 10.1002/jcc.20880.

(34) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098.

(35) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.

(36) (a) Jansen, G.; Hesselmann, A. *J. Phys. Chem. A* **2001**, *105*, 646. (b) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464. (c) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *362*, 319. (d) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778. (e) Hesselmann, A.; Jansen, G. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010.

(37) (a) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103. (b) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Am. Chem. Soc.* **2006**, *128*, 11730. (c) Fiethen, A.; Jansen, G.; Hesselmann, A.; Schütz, M. *J. Am. Chem. Soc.* **2008**, *130*, 1802.

(38) (a) Maheswari, P. U.; Modec, B.; Pevec, A.; Kozlevčar, B.; Massera, C.; Gamez, P.; Reedijk, J. *Inorg. Chem.* **2006**, *45*, 6637. (b) Garau, C.; Quiñonero, D.; Frontera, A.; Escudero, D.; Ballester, P.; Costa, A.; Deyà, P. M. *Chem. Phys. Chem.* **2007**, *438*, 104.

(39) Nabeshima, T.; Masubuchi, S.; Taguchi, N.; Akine, S.; Saiki, T.; Sato, S. *Tetrahedron Lett.* **2007**, *48*, 1595.

(40) Wang, Y.; Wei, B.; Wang, Q. *J. Crystallogr. Spectrosc. Res.* **1990**, *20*, 79.

(41) (a) Mizuno, T.; Wei, W. H.; Eller, L. R.; Sessler, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 1134. (b) Yoon, J.; Kim, S. K.; Singh, N. J.; Kim, K. S. *Chem. Soc. Rev.* **2006**, *35*, 355. (c) Chellappan, K.; Singh, N. J.; Hwang, I.-C.; Lee, J. W.; Kim, K. S. *Angew. Chem., Int. Ed.* **2005**, *44*, 2899. (d) Angew. Chem. 2005, *117*, 2959. (e) Ihm, H.; Yun, S.; Kim, H. G.; Kim, J. K.; Kim, K. S. *Org. Lett.* **2002**, *4*, 2897. (f) Gale, P. A. *Chem. Soc. Rev.* **2008**, *199*, 81.

CT8001255