JCTC Journal of Chemical Theory and Computation

Wavelet Transforms for Determining Time-Dependent Vibrational Frequencies

Asif Rahaman and Ralph A. Wheeler*

Department of Chemistry and Biochemistry, University of Oklahoma, 620 Parrington Oval, Room 208, Norman, Oklahoma 73019

Received April 17, 2005

Abstract: Fourier transforms of autocorrelation functions are typically used to calculate spectra and transport properties, but they require oscillations periodic in time (a stationary time series). The time-scale relation revealed by a wavelet transform (WT), on the other hand, gives a relation between time and pseudo-frequency that is used here to calculate vibrational frequencies for HONO and to detect the molecule's trans−cis conformational change in ab initio molecular dynamics simulations. Thus, the WT shows potential for giving new insights into nonstationary time series frequently encountered in chemistry and physics.

The Fourier transform (FT) is ubiquitous in chemistry for extracting the frequency components of a system's response to a time-dependent perturbation.[1,2] In computational chemistry, for example, spectra and transport properties are typically accessed by computing the FT of a time correlation function, $C(t)$:[3,4]

$$F(\omega) = \int C(t)\, e^{-i\omega t}\, dt \qquad (1)$$

where $e^{-i\omega t}$ is the Fourier basis function with frequency (ω) and time (t). Examples of important time correlation functions include the van Hove, dipole density, magnetization, and current−current correlation functions, whose FTs are related to the structure factor, infrared line shape, NMR line shape, and frequency-dependent conductivity, respectively.[3,4] Despite its enormous power, the traditional FT suffers several strict requirements, including the assumption of a stationary time series.[5,6] A powerful alternative to the traditional FT, the wavelet transform (WT), was recently developed and has potential for resolving frequency- *and* time-dependent properties[6,7] of nonstationary time series. Wavelet analysis was used to study energy transport in OCS,[8] for many-body dynamics[9] and intramolecular energy transfer.[10] The current contribution illustrates the method's power for resolving time-dependent spectral or transport properties by describing the use of WTs to calculate the vibrational spectrum of HONO in its most stable trans conformation and to detect HONO's trans−cis conformational change during ab initio molecular dynamics simulations.

Wavelet analysis has been used extensively in signal and image processing over the past decade[6,7] and has enjoyed increasing use in computational chemistry.[11−14] To understand the wavelet transform, consider replacing the Fourier basis function in eq 1 with a wavelet $\psi_{a,b}(t)$, where the WT can be written as

$$W(a,b) = \int C(t)\, \psi_{a,b}(t)\, dt \qquad (2)$$

The wavelet $\psi_{a,b}(t)$ can be defined by dilation and translation of a "mother wavelet" $\psi(t)$, that is,

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \qquad (a,b \in R;\ a \neq 0) \qquad (3)$$

where a is the scale (dilation) and b is the position (translation) parameter. In the discrete form, one can rewrite eq 3 as

$$\psi_{m,n}(t) = a_0^{-m/2} \psi(a_0^{-m} t - n b_0) \qquad (4)$$

where $a_0^m = a$ and $n b_0 a_0^m = b$ ($m,n \in Z$; $a_0 \neq 0$). Generally, $a_0 = 2$ and $b_0 = 1$ are used. Now, if one assumes a center frequency (ω_c) associated with a wavelet, then a pseudo-frequency (ω) can be obtained by using the relation

$$\omega = \omega_c / a \cdot \Delta \qquad (5)$$

where Δ is the sampling period and indicates the interval between two data points and a is a unitless scale parameter.[15] The lower scale (a) indicates a higher pseudo-frequency and vice versa. The frequency at the maximum of the mother wavelet's FT is the center frequency (ω_c). For applications envisioned here, pseudo-frequencies are found from the WT of an autocorrelation function.

Compared to the FT (eq 1), where sines and cosines are the analyzing functions, the WT (eq 2) uses a family of analyzing functions called wavelets, derived from the mother

* To whom correspondence should be addressed. E-mail: rawheller@chemdept.chem.ou.edu.

10.1021/ct050103d CCC: $30.25
Published on Web 08/10/2005

Table 1. Comparisons of Pseudo-Frequencies Calculated for Scales with Maximum Values of Wavelet Coefficients with Harmonic ab initio Frequencies Determined from Fourier Transformation and Experimental Frequencies[19]

pseudo-freqency (cm^{-1})	scale	ab initio frequency (cm^{-1})	FT frequency (cm^{-1})	exptl. IR frequency (cm^{-1})
		trans		
602	45	599	550, 610	544
796	34	790	750	596
1083	25	1099	990, 1050, 1070	790
1504	18	1514	1600	1263
2082	13	2043	2020	1700
3868	7	4082	3615, 3645, 4200	3591

wavelet $\psi(t)$ (eqs 3 and 4).[6,7] Wavelets are characterized by two parameters, the scale and position in time, so that pseudo-frequencies derived from the WT give a unique time-frequency relation for stationary *or* nonstationary time series that is inaccessible by FT. Here, the trans−cis isomerization of HONO renders the time series of atomic positions and velocities nonstationary, thereby producing two sets of vibrational frequencies, separated in time by a conformational change.

Figure 1a and b show the FT and WT, respectively, of a velocity autocorrelation function from a representative ab initio molecular dynamics trajectory of HONO. (The procedure for generating HONO trajectories is described in the references.[16]) In the Fourier spectrum, bands of frequencies appear near the frequency of each vibrational fundamental. In contrast, Figure 1b shows the wavelet coefficients, $W(a,b)$, for different values of scale, a, and position in time, b (several

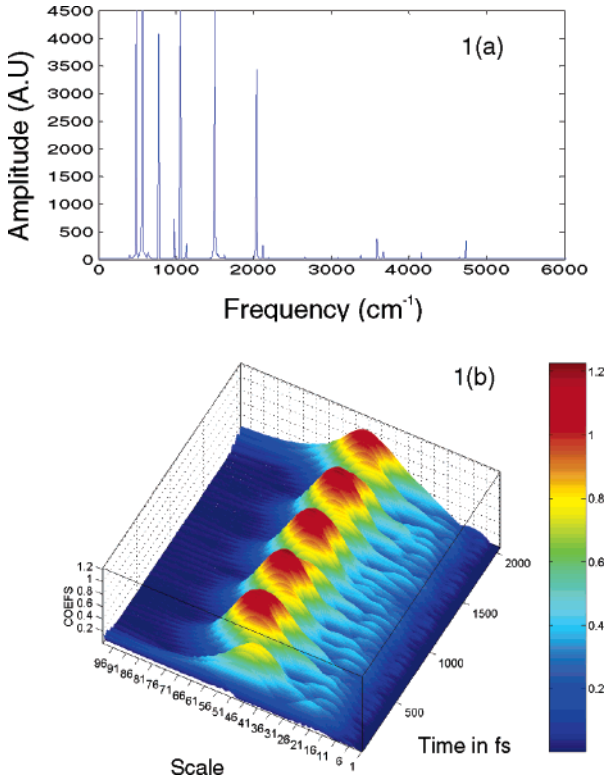


Figure 1. (a) Fourier transform of the velocity autocorrelation function $C(t)$ for a single *trans*- and *cis*-HONO molecule. (b) The plot of wavelet coefficients at different values of scale (a) and time obtained from the wavelet transform of $C(t)$ using the Morlet wavelet. The scale is unitless.

different mother wavelets were tested,[7,17,18] but all results reported here used the Morlet wavelet[17]). The scales for which the $W(a,b)$ values are the maximum were chosen, and the corresponding pseudo-frequencies were calculated. Table 1 lists the scales and corresponding pseudo-frequencies, compared with frequencies from ab initio quantum chemical calculations, the FT of the same correlation function, and experimental results.[19] Nearly all the pseudo-frequencies are in better agreement than Fourier frequencies with those frequencies determined by using ab initio quantum chemistry. The lower pseudo-frequencies are in better agreement than

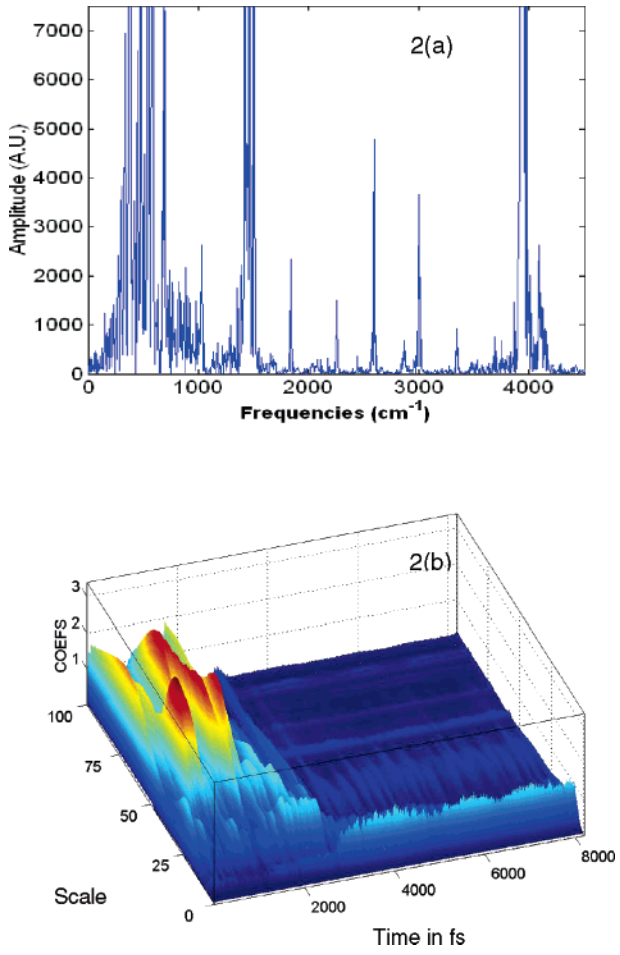


Figure 2. (a) Fourier transform of the velocity autocorrelation function $C(t)$ for the trans−cis isomerization of a single HONO molecule. (b) The plot of wavelet coefficients at different values of scale (a) and time obtained from the wavelet transform of $C(t)$ using the Morlet wavelet. The scale is unitless.

the higher pseudo-frequencies, a result that can be understood in terms of the uncertainty principle. At low scale, the time localization is more precise for WTs and, thereby, the uncertainty in pseudo-frequency increases. The high pseudo-frequencies corresponding to low scales, therefore, differ more from the ab initio quantum chemical results than the low pseudo-frequencies.

The velocity autocorrelation function for the trans−cis isomerization of a single HONO molecule was also generated from ab initio molecular dynamics simulations,[20] and its FT (Figure 2a) and WT (Figure 2b) were calculated. From the FT in Figure 2a, we cannot tell when the transition occurs from one conformation of HONO to another. The WT displayed in Figure 2b, however, shows discontinuities in wavelet coefficients along the scale at times corresponding to the conformational changes of HONO observed by monitoring the time history of the dihedral angle.

Acknowledgment. This study is supported by the National Science Foundation (Grant #CHE-0413965), NSF/NRAC supercomputer time allocation (No. MCA96-N019), and supercomputer time from the Oklahoma Supercomputing Center for Education and Research (OSCER).

References

(1) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U. K., 1989.

(2) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, California, 2000; Chapter 21.

(3) Berne, B. J.; Harp, G. D. *Adv. Chem. Phys.* **1970**, *17*, 63.

(4) Gordon, R. G. *Adv. Magn. Reson.* **1968**, *3*, 1.

(5) Percival, D. B.; Walden, A. T. *Spectral Analysis for Physical Applications*; Cambridge University Press: Cambridge, MA, 1993.

(6) Debnath, L. *Wavelet transforms and their applications*; Birkhauser: Boston, MA, 2002.

(7) Daubechies, I. *Ten lectures on wavelets*; SIAM press: Philadelphia, PA, 1992.

(8) Martens, C. C.; Davis, M. J.; Ezra, G. S. *Chem. Phys. Lett.* **1987**, *142*, 519.

(9) Martens, C. C. *Phys. Rev. A: At., Mol., Opt. Phys.* **1992**, *45*, 6914.

(10) Finney, L. M.; Borrmann, A.; Martens, C. C. *Chem. Phys. Lett.* **1993**, *214*, 159.

(11) Shao, X.-G.; Leung, A. K.-m.; Chau, F.-T. *Acc. Chem. Res.* **2003**, *36*, 276.

(12) Askar, A.; Cetin, E.; Rabitz, H. *J. Phys. Chem.* **1996**, *100*, 19165.

(13) Li, Z.; Borrmann, A.; Martens, C. C. *Chem.Phys. Lett.* **1993**, *214*, 362.

(14) Cho, A. E.; Doll, J. D.; Freeman, D. L. *J. Chem. Phys.* **2002**, *117*, 5971.

(15) Abry, P. *Ondelettes et turbulence. Multirésolutions, algorithmes de décomposition, invariance d'échelles*; Diderot Editeur: Paris, 1997.

(16) HONO simulations were started in the trans conformation with initial atomic velocities corresponding to $T = 300$ K. Constant energy, HF/6-31G(d,p), ab initio MD simulations were carried out with HONO defined as the quantum mechanical region using the CHARMM/GAMESS programs (Field, M. J.; Bash, P. A.; Karlus, M.; *J. Comput. Chem.* **1990**, *11*, 700−733). The integration time step for all simulations reported was 1 fs. Vibrational frequencies were calculated by using velocity correlation functions for both cis and trans conformations, but only vibrational frequencies for *trans*-HONO are reported here.

(17) Grossman, A.; Morlet, J. *SIAM J. Math.* **1984**, *15*, 723.

(18) Mallat, S. G. *IEEE Trans. Pattern Anal. Machine Learning* **1989**, *11*, 674.

(19) Chan, W.-t.; Pritchard, H. O. *Phys. Chem. Chem. Phys.* **2002**, *4*, 557.

(20) To investigate the cis−trans isomerization, CHARMM/GAMESS simulations were started with HONO at the transition state (with its dihedral angle at 0.0°) with initial atomic velocities corresponding to 100 K. Constant energy, HF/6-31G(d,p) simulations were performed for 10 ps.

CT050103D

JCTC Journal of Chemical Theory and Computation

Infrared Spectroscopy of *N*-Methylacetamide Revisited by ab Initio Molecular Dynamics Simulations

M. P. Gaigeot,*,† R. Vuilleumier,‡ M. Sprik,§ and D. Borgis†

Laboratoire de Modélisation des Systèmes Moléculaires Complexes, Université d'Evry val d'Essonne, Rue Père A. Jarland, 91025 Evry, France, Laboratoire de Physique Théorique des Liquides, Université Pierre et Marie Curie, Case Courrier 121, 4 place Jussieu, 75005 Paris, France, and Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Received February 14, 2005

Abstract: The density functional theory based molecular dynamics simulation method ("Car−Parrinello") was applied in a numerical study of the electronic properties, hydrogen bonding, and infrared spectroscopy of the trans and cis isomer of *N*-methylacetamide in aqueous solution. A detailed analysis of the electronic structure of the solvated molecules, in terms of localized Wannier functions and Born atomic charges, is presented. Two schemes for the computation of the solute infrared absorption spectrum are investigated: In the first method the spectrum is determined by Fourier transforming the time correlation function of the solute dipole as determined from the Wannier function analysis. The second method uses instead the molecular current−current correlation function computed from the Born charges and atomic velocities. The resulting spectral properties of *trans*- and *cis*-NMA are carefully compared to each other and to experimental results. We find that the two solvated isomers can be clearly distinguished by their infrared spectral profile in the 1000−2000 cm^{-1} range.

1. Introduction

Vibrational spectroscopy (infrared and Raman) can often provide information about molecular structures in the condensed phase which is complementary to or even hidden from diffraction methods. This is why the technique is commonly applied for the study of secondary structures of proteins[1−4] in solution or membranes. Structural motifs such as α-helices or β-sheets can be identified by characteristic changes in position and shape of linear infrared or Raman bands,[1−4] in particular the amide bands (mostly amide I, amide II, and amide III) are used as structural probes. Amide I gives the most sensitive response to changes in the secondary structure of proteins.[2] This is partly due to the fact that the amide I vibration appears in a spectral region with very little overlap with other internal modes of proteins and to the fact that its spectral intensity generally dominates the infrared spectrum. In addition to reflecting the molecular conformation,[2] the amide I band shape is also strongly affected by formation of intermolecular hydrogen bonds with the solvent.[2,5−7] Moreover, the frequency gap between the amide I and amide II bands represents some measure of the interaction between carbonyl and amide groups on the protein backbone. Consequently, the amide I mode is also used as a probe in nonlinear-IR-2D spectroscopy of small peptides.[8−11]

Assignment of vibrational modes is an important step in the structural and dynamical interpretation of the spectral fingerprints of conformation or solvation. Here is where modeling and computation can be most helpful, in particular when a full spectrum (including intensities) can be computed and directly compared to experimental data. As we have demonstrated in a recent study of the IR absorption of aqueous uracil[12] it is now possible to perform such a detailed calculation for small models of bio-organic molecules in finite temperature bulk solution with all water molecules treated at the same level of theory as the solute. This

* Corresponding author e-mail: gaigeot@ccr.jussieu.fr.
† Université d'Evry val d'Essonne.
‡ Université Pierre et Marie Curie.
§ University of Cambridge.

10.1021/ct050029z CCC: $30.25
Published on Web 07/07/2005

development is the result of the latest advances in density functional theory based molecular dynamics simulation ("Car−Parrinello"[13]), brought about by the implementation of the modern theory of polarization and the maximally localized Wannier functions. This opened up the field of first principle simulation to the study of infrared[14−16] and Raman[17] spectroscopy.

Following on our calculation of the infrared spectrum of aqueous uracil[12] (one of the RNA nucleic acid bases), we investigate in the present paper the infrared spectroscopy of another model bio-organic molecule, NMA, using similar ab initio molecular dynamics methods. Our main objective is again to assess whether Car−Parrinello simulations can be relied upon to give a proper account of the experimental spectrum of this elementary protein building block in its aqueous environment. Comparing the calculated infrared spectrum of gas-phase and aqueous-phase NMA, band shape modifications as well as band shifts will be correlated with the effect of the hydrogen bonding on internal motions of the molecule. Of special interest is the question of the spectroscopic characterization of the two distinct conformers. The change in structure between *trans*- and *cis*-NMA as well as an expected different organization of surrounding solvent molecules gives rise to different vibrational signatures recorded in infrared spectra. We have thus analyzed the infrared response to a change in conformation and compared the hydration properties of *trans*- and *cis*-NMA.

The paper also serves a more technical purpose, namely to present and compare two alternative ways of calculating infrared spectra of molecular solutes in a molecular solvent. (i) The first is the method applied in ref 12 which consists of calculating the individual molecular dipole moments (for the solute and each solvent molecule) by using localized Wannier functions.[12,14−16] (ii) The second approach, introduced here, is based on the calculation of dipole derivatives (or currents) through linear response theory. The total current is decomposed into solute and solvent contributions. Compared to the first scheme, based on partitioning of charge density, the second approach using the current representation is more fundamental leading to a rigorous separation of solute and solvent signal. However, requiring a coupled self-consistent perturbation calculation, the computational costs are also considerably higher.

Organization of the paper is as follows. Section 2 contains some relevant background information on NMA. Section 3 is a summary of the conditions of the Car−Parrinello molecular dynamics and the methods used in our work for the calculation of infrared spectra of solutes. In section 4 the structural properties of isolated and aqueous *trans*- and *cis*-NMA are described including details of the hydrogen bonding network between solute and solvent. The electronic properties of the solute are discussed too, in terms of localized Wannier functions, Born charges, and dipole moments. In section 6, we report our findings for the vibrational density of states (VDOS) of the various systems studied. Section 7 is devoted to the calculated infrared spectra. The comparison between the gas-phase and aqueous-phase spectral properties is discussed in detail. The last section concludes our study.

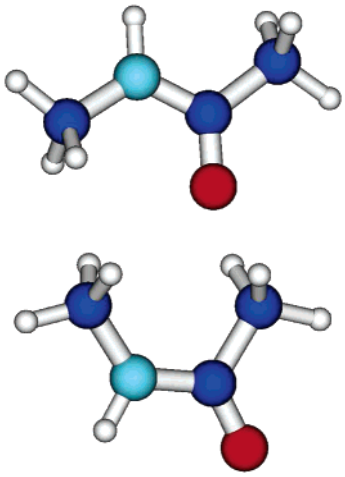


Figure 1. Schematic representation of NMA in the trans (top) and cis (bottom) conformation.

2. NMA as a Model Molecule

NMA (*N*-methylacetamide), a single amide "blocked" with methyl groups at both extremities, is the simplest building block model of the repeated peptide linkage of peptides and proteins. Figure 1 shows a schematic representation of the two conformers of the NMA molecule studied in this work, *trans*-NMA and *cis*-NMA. Structure and energetics of NMA have been the subject of numerous experimental and theoretical investigations.[18−26] Similarly the molecule has been a popular model system in infrared and Raman vibrational experiments either in the gas phase[7,27−29] or in solution.[7,30−34] Of particular concern in these studies were the possible relations between vibrational signatures of amide modes of this peptide model and those of secondary structures of peptides and proteins.

First theoretical studies were aimed at the understanding of electronic and geometrical properties of gas-phase NMA.[18−26] In these earlier studies, great emphasis was put on the electronic properties of the HNCO peptide linkage. Also the hydration structure of NMA has been investigated using computational methods, either ab initio geometry optimizations[19,20,35] with a few water molecules placed close to the main hydrophilic sites of NMA or applying classical molecular dynamics simulations of NMA surrounded by a few hydration shells. A variety of Hamiltonians were used, ranging from standard nonpolarizable force fields[36] to polarizable force fields[37] and mixed QM/MM dynamics.[23] Very recently, Martyna et al.[38] have studied the solvation structure of NMA using Car−Parrinello dynamics. As part of a careful classical molecular dynamics simulation of aqueous NMA involving a polarizable force field, Kato et al.[37] have performed a benchmark calculation of the infrared spectrum, which highlighted the importance of including polarization for an accurate description of the IR absorption. In a very recent work with ab initio geometry optimizations and harmonic vibrational analyses, Besley[35] has also shown the importance of incorporating a few layers of solvent around the solute (through a continuum model in that study) for a better description of the IR absorption patterns of aqueous NMA.

3. Computational Method

3.1. Ab Initio Molecular Dynamics. The DFT-based Car−Parrinello simulations performed in this work follow the general setup of previous ab initio molecular dynamics simulations. We used the Becke, Lee, Yang, and Parr (BLYP) gradient-corrected functional[39,40] for the exchange and correlation terms. The one-electron orbitals are expanded in a plane-wave basis set with a kinetic energy cutoff of 70 Ry restricted to the Γ point of the Brillouin zone. Medium soft norm-conserving pseudopotentials of the Martins-Trouillier type[41] are used. The core-valence interaction of C, N, and O is treated by *s* and *p* potentials with pseudization radii of 1.23, 1.12, and 1.05 au, respectively (taking the same radius for *s* and *p*), while H atoms are treated as an *s* potential with a 0.5 au radius. We should mention that the 70 Ry energy cutoff is mostly determined by the convergence of the H and O pseudopotentials. Energy expectations are calculated in reciprocal space using the Kleinman−Bylander transformation.[42]

Simulations were performed at constant volume using a fictitious electron mass of 500 au, a time step of 5 au (0.12 fs), and periodic boundary conditions. Gas-phase simulations were carried out in a cubic box of length 12 Å. This box size has been chosen after performing a series of wave function optimizations of NMA conformers in cubic boxes of increasing length. We found that, from 12 Å on, the electronic energy of the isolated molecule is converged within a criterion of 10^{-4} au, which ensures that periodic images of the molecule do not interact. Liquid simulations have been performed in a cubic box of length 12.07 Å, which is a compromise between spatial extent and duration of the MD runs. Starting configurations have been prepared using classical force field simulations, where we chose the precise number of water molecules in the simulation cell; this is done in order to fix the correct density of such a small simulated liquid system. We refer the reader to refs 12 and 43 for more details.

The cis/trans isomerization barrier of NMA in aqueous solution at room temperature is too high[38] to observe spontaneous isomerization on the molecular dynamics time scale. The infrared spectral features of *trans*-NMA and *cis*-NMA will thus be investigated by running two separate dynamics simulations each for a different conformation. In the *trans*-NMA model system the solute is surrounded by 50 water molecules and *cis*-NMA by 52 water molecules, which corresponds to two complete layers of solvent, and part of the third one, around the solute. These system dimensions are similar to the Car−Parrinello simulation of aqueous uracil performed previously.[12] The Car−Parrinello molecular dynamics simulations consist of two steps: An equilibration phase of 1 ps with a control of temperature through velocity rescaling, followed by data collection over trajectories of 9.35 and 6.35 ps, respectively, for aqueous *trans*-NMA and *cis*-NMA and 19.6 and 8 ps for gas-phase *trans*-NMA and *cis*-NMA. In that step, molecular dynamics are strictly microcanonical. The average ionic temperatures obtained along the simulations were 320 K (aqueous *trans*-NMA), 302 K (aqueous *cis*-NMA), 20 K (isolated *trans*-NMA), and 15 K (isolated *cis*-NMA). Hydrogen atoms were treated as classical particles with their true mass (1836 au). All simulations were carried out with version 3.7.1 of the CPMD ab initio molecular dynamics package.[44]

3.2. Computation of the Solute Spectrum. The first method applied here for the calculation of the infrared (IR) absorption coefficient $\alpha(\omega)$ makes use of the relation to the dipole time correlation function[45]

$$\alpha(\omega) = \frac{2\pi\beta\omega^2}{3n(\omega)cV} \times \int_{-\infty}^{\infty} dt \, \langle \boldsymbol{M}(t)\cdot\boldsymbol{M}(0)\rangle \exp(i\omega t) \quad (1)$$

where $\beta = 1/kT$, $n(\omega)$ is the refractive index, c is the speed of light in a vacuum, V is the volume, and $\mathbf{M}$ is the total dipole moment of the system, which is the sum of the ionic and electronic contribution.[15] The angular brackets indicate a statistical average. In this formula, we have taken into account a quantum correction factor (multiplying the classical line shape) of the form $\beta\hbar\omega/(1 - \exp(-\beta\hbar\omega))$, which was shown to give the most accurate results on calculated IR amplitudes.[12] For a complete discussion on quantum corrections, we refer the reader to ref 77. The IR spectra in this work are given as products $\alpha(\omega)n(\omega)$ expressed in cm^{-1} (decadic linear absorption coefficient) as a function of reciprocal wavenumber ω in cm^{-1}. The spectra have been smoothed with a window filtering applied in the time domain, i.e., each term of the correlation function $C(t)$ is multiplied by a Gaussian function $\exp(-0.5\sigma(t/t_{max})^2)$, where t_{max} is the length of the simulation, and σ is 10 for gas-phase simulations and 40 for liquid simulations. The periodic nature of an MD cell is a fundamental complication for the determination of the total dipole moment $\mathbf{M}$. This problem has been resolved by the modern theory of polarization[46,47] by describing polarization in terms of a Berry phase. This methodology was developed by Silvestrelli and Parrinello[14−16] in the context of Car−Parrinello molecular dynamics simulations.

When interested in the infrared spectrum of a solute immersed in liquid water, we face additional difficulties, namely separation of the absorption in solute and solvent contributions. Due to the limited statistics the subtraction method used in experiment is not an option for simulation. The approach applied previously in our calculation of the IR spectrum of aqueous uracil[12] is based on a decomposition of charge density using the maximally localized Wannier Functions scheme of Marzari and Vanderbilt.[48] This enabled us to write the total electronic dipole moment $\mathbf{M} = \mathbf{m}_{NMA} + \sum_W \mathbf{m}_W$ as the sum of the solute dipole moment and the sum of the solvent dipole moments. As a result, the total dipole correlation function is resolved in three contributions $C_{MM}(t) = C_{W/W}(t) + C_{W/NMA}(t) + C_{NMA/NMA}(t)$. The IR spectrum of solute NMA is calculated from only the Fourier transform of the self-correlation function $C_{NMA/NMA}(t)$. As discussed in ref 12, one disadvantage of this decomposition is that coupling with the surrounding solvent molecules is taken into account only through intermolecular polarization. The cross term $C_{W/NMA}(t)$, which includes the remaining IR contributions arising from correlations between the motion of solute and solvent, is ignored. In fact, this term is hard to compute. Opposite positive and negative contributions make it difficult to reach convergence within the short simulation

times of tens of picoseconds. Moreover while fully consistent in the sense that the molecular dipole moments rigorously sum to the cell dipole moment, the Wannier function method suffers from a certain arbitrariness in the definition of molecular dipole moment in solution.

An alternative approach to achieve a similar, but not necessarily equivalent, decomposition of the absorption signal is to return to the more fundamental current representation, involving the autocorrelation function of the total current, $\mathbf{j} = \dot{\mathbf{M}} = \mathrm{d}\mathbf{M}/\mathrm{d}t$:

$$n(\omega)\alpha(\omega) = \frac{2\pi}{3Vk_BT}\int_{-\infty}^{\infty}\langle \mathbf{j}(0)\mathbf{j}(t)\rangle e^{i\omega t}\mathrm{d}t \qquad (2)$$

Note again that in this formula we have taken into account a quantum correction factor of the form $\beta\hbar\omega/(1 - \exp(-\beta\hbar\omega))$. Since in the Born−Oppenheimer approximation the dipole moment of the system depends only on the atomic positions, one can make use of a chain rule to compute the current $\mathbf{j}(t)$ at time t

$$j^{\beta}(t) = \sum_{i,\alpha}\frac{\partial M^{\beta}}{\partial q_i^{\alpha}}(t)\frac{\mathrm{d}q_i^{\alpha}}{\mathrm{d}t}(t) = \sum_{i,\alpha}\frac{\partial M^{\beta}}{\partial q_i^{\alpha}}(t)v_i^{\alpha}(t) \qquad (3)$$

where $\mathbf{M}$ is the total dipole moment of the system, $\mathbf{q}_i$ is the position of atom i, and $\partial M^{\beta}/\partial q_i^{\alpha}$, $\alpha,\beta = x, y, z$, represent the components of the atomic polar tensor (APT) of atom i. The APT of atom i is thus defined as the derivatives of the dipole moment of the system with respect to infinitesimal displacements of atom i.[49,50] This tensor is directly related to intensities of infrared bands in the double harmonic approximation.[51] Numerous calculations of APT have been performed in the solid phase or in the gas phase.[49,50,52,53] Recently, a calculation of the APT of water molecules in liquid water has been performed.[54] For our system, the APT tensor was calculated for each atom in the solution using Density Functional Response Theory.[50] This was repeated every 1.2 fs over a ≈1 ps interval taken from the full trajectory of *trans*-NMA in solution. We have used the Putrino et al.[55] implementation of the linear response to an applied electric field in the CPMD code.[44] In this scheme, the macroscopic polarization of the periodically replicated cell is defined using the Berry phase approach of Resta.[56,57] The APT tensor for the whole system is then obtained from the Maxwell relation

$$\frac{\partial M^{\beta}}{\partial q_i^{\alpha}} = \frac{\partial F_i^{\alpha}}{\partial \epsilon^{\beta}} = \frac{\partial^2 E_{\mathrm{tot}}}{\partial \epsilon^{\beta}\partial q_i^{\alpha}} \qquad (4)$$

where ϵ is an applied uniform electric field and $\mathbf{F_i}$ is the force acting on particle i. This allows for the calculation of the $3N$ elements of the system APT tensor through only three linear response calculations ($\beta = x, y, z$).[50] Using the same methodology, we have computed the APT of *trans*-NMA in the gas phase at the equilibrium configuration. Relation of APT tensors to infrared intensities is well-known from quantum chemistry.[78−80]

To resolve the infrared spectrum of the whole system into solvent, solute and cross contributions, we separated the total current of the system into a solvent and a solute term

$$j^{\beta}(t) = j^{\beta}_{\mathrm{water}}(t) + j^{\beta}_{\mathrm{NMA}}(t) \qquad (5)$$

by restricting the sum in eq 3 to atoms of the solvent or the solute only. This natural decomposition of the current then provides a decomposition of the infrared spectrum which is different from the one obtained through the maximally localized Wannier orbitals. The infrared spectrum of NMA is now defined as

$$n(\omega)\alpha_{\mathrm{NMA}}(\omega) = \frac{2\pi}{3Vk_BT}\int_{-\infty}^{+\infty}\langle \mathbf{j}_{\mathrm{NMA}}(0)\mathbf{j}_{\mathrm{NMA}}(t)\rangle e^{i\omega t}\mathrm{d}t \qquad (6)$$

A major argument in favor of the current scheme is that any partitioning of the charge density is avoided. We can also take advantage of the calculation of the APT tensor of each atom of NMA in the gas phase and in the liquid phase to get the transverse effective charges, or Born charges, defined as

$$Z_i^B = \frac{1}{3}Tr\frac{\partial M^{\beta}}{\partial q_i^{\alpha}} \qquad (7)$$

which corresponds to the isotropic component of the APT of atom i. In the gas phase, this charge was shown to correlate very well with other charge definitions,[52] although the actual value was found to be rather sensitive to the approximation used for the density functional. We will discuss later the values of these charges for *trans*-NMA as well as their variations when going from the gas phase to the liquid.

4. Structural Analysis

4.1. Intramolecular Properties. Table 1 reports values of the main internal geometrical distances and angles of the *trans*-NMA and *cis*-NMA conformers obtained by averaging over the ab initio molecular dynamics trajectories, either in the gas or aqueous phase. Atoms are labeled according to the following convention (for molecular geometries see Figure 1): N is the amide nitrogen atom, C is the carbon of the carbonyl group, and C_C and C_N denote the two carbon atoms of the methyl groups located at the C-terminal and N-terminal part of the molecule. In the gas phase, it can be seen that bond lengths are very similar for both conformers. The largest difference is observed for the amide N−H bond which is increased by 0.003 Å in *cis*-NMA. Backbone distances, i.e., N−C, N−CH_3, and C−CH_3, are almost equal in both conformers, and we do not see any geometrical differences in C−H bonds whether the methyl group is located at the N-terminal or the C-terminal part of the molecule. In *cis*-NMA, the average distance between the carbonyl oxygen and amide hydrogen, i.e., 2.432 Å, is slightly too long to be considered as a proper intramolecular hydrogen bond.

Values of intramolecular distances obtained in our gas-phase ab initio molecular dynamics at low temperature are very similar to those obtained with high level ab initio geometry optimizations.[21,24,25] Depending on the ab initio method (HF, DFT, MP2) and on the basis set used,[19,21,24,25] the calculated C═O distance of *trans*-NMA lies between

Table 1. Geometrical Distances and Angles of *trans*-NMA and *cis*-NMA Averaged over the Car−Parrinello Molecular Dynamics Trajectory

	C═O	N−H	O···H	N−C	O−C−N−H	C−C−N−C
trans-NMA[a]	1.238 ± 0.006	1.018 ± 0.007	3.183 ± 0.011	1.376 ± 0.006	179.93 ± 4.62	180.05 ± 4.15
cis-NMA[a]	1.239 ± 0.005	1.021 ± 0.007	2.432 ± 0.017	1.378 ± 0.006	0.718 ± 3.351	−1.266 ± 5.845
trans-NMA[b]	1.268 ± 0.022	1.031 ± 0.032	3.186 ± 0.053	1.351 ± 0.028	173.98 ± 10.75	178.46 ± 10.70
cis-NMA[b]	1.272 ± 0.021	1.032 ± 0.023	2.423 ± 0.099	1.350 ± 0.024	−3.330 ± 10.540	−0.204 ± 12.896

	N−CH_3	C−CH_3	C−H (N-terminal)	C−H (C-terminal)
trans-NMA[a]	1.469 ± 0.006	1.529 ± 0.008	1.097 ± 0.003	1.098 ± 0.003
cis-NMA[a]	1.470 ± 0.005	1.527 ± 0.011	1.098 ± 0.003	1.097 ± 0.004
trans-NMA[b]	1.477 ± 0.031	1.520 ± 0.035	1.101 ± 0.021	1.098 ± 0.016
cis-NMA[b]	1.482 ± 0.039	1.514 ± 0.031	1.100 ± 0.023	1.100 ± 0.021

[a] Isolated molecule at ∼20 K. [b] Hydrated molecule at ∼300 K. Distances are given in angstroms, and angles in degrees.

1.221 and 1.237 Å, the N−C distance is comprised in the 1.364−1.383 Å interval, and the amide N−H distance lies in the 1.008−1.016 Å interval. The average distances obtained in our molecular dynamics of *trans*-NMA and *cis*-NMA are virtually equal to the ones calculated by DFT/BLYP geometry optimizations in ref 21, except for the N−C distance which appears slightly shorter (− ∼0.010 Å) in our dynamics. As usual with DFT calculations, it is found that distances obtained with the BLYP functional are overestimated compared to the B3LYP functional.[21] Comparing our results to experiment[58,59] for *trans*-NMA, we find that the C═O distance is overestimated by ∼0.010 Å, while the N−C distance is underestimated by ∼0.010 Å.

As can be anticipated, we find that C═O and N−H intramolecular distances are increased (almost equally by ∼0.030 Å) in aqueous solution, due to the formation of intermolecular hydrogen bonds between these groups and the surrounding water molecules. In aqueous solution, the backbone N−C distance is very much shortened (by ∼0.025 Å) for both NMA conformers. This is also observed for the backbone C−CH_3 distance in the trans (−0.009 Å) and cis (−0.013 Å) conformers. In contrast, the N−CH_3 bond length is increased from the gas to the liquid phase (+0.008 Å for *trans*-NMA and +0.012 Å for *cis*-NMA). Thus, in aqueous solution, we observe that the N−C and C−CH_3 backbone bonds are contracted, whereas the N−CH_3 bond is extended. As anticipated for methyl hydrophobic groups, C−H bond distances are not affected by the presence of solvent. The modifications of intramolecular distances observed here are consistent with previous geometry optimization calculations[20,26] with a few water molecules hydrogen-bonded to the N−H and C═O groups. The presence of two layers of solvent around NMA simulated in our work amplifies the variations observed with only a few solvent molecules. Overall, due to the higher temperature in the liquid simulation, the fluctuations around the mean distance or mean angle values displayed in Table 1 are higher in the liquid phase than in the gas phase.

In Table 1 we have also reported the average values of the OCNH and CCNC peptide bond dihedral angles obtained over the whole MD trajectory. Each NMA conformer did remain in its initial trans- or cis-orientation (OCNH angle), be it in the gas phase or in solution. We recall that gas-phase calculations have shown an energy difference in the range ∼2−3 kcal/mol between the trans and cis conformers.[21,23] According to the calculations of Martyna et al.[38] solvation enhances this barrier by ∼2−3 kcal/mol. We note that the OCNH dihedral angle of gas-phase *trans*-NMA obtained in our work is identical to that in ref 21, while we predict an angle of 0 degree for *cis*-NMA instead of 6.05 degrees.[21] We finally observe that the mean values of the OCNH dihedral angles of *trans*- and *cis*-NMA are decreased by a few degrees when going from the gas phase to the solution.

4.2. Radial Pair Correlation Functions. As usual in simulations of liquids, determination of radial distribution functions (hereafter denoted rdfs) can provide a good first impression of hydrogen bonding among the solvent molecules and between the solute hydrophilic groups and the solvent. The relevant rdfs are reported in Figure 2, respectively, for NMA amide hydrogen/water oxygens (H−Ow, top left), NMA carbonyl oxygen/water hydrogens (O−Hw, top right), NMA methyl carbons/water oxygens (CH_3−Ow, bottom left), and water/water (Ow−Hw, Ow−Ow, and Hw−Hw, bottom right of Figure 2). In each figure, we compare results obtained for aqueous *trans*-NMA (plain lines) to those obtained for aqueous *cis*-NMA (dashed lines).

The broad picture of NMA hydration emerging from the Car−Parrinello molecular dynamics simulations follows what can be expected for this kind of organic molecule. H−Ow and O−Hw radial distribution functions display the general features of hydrogen bonding between hydrophilic sites, i.e., between N−H amide and water oxygens or between C═O carbonyl and water hydrogens. The first peak of the H−Ow rdfs is located at 2.0 and 1.9 Å for *trans*-NMA and *cis*-NMA, respectively, while the first minimum is found at 2.6 and 2.5 Å. The first peak of O−Hw rdfs appears at slightly shorter distances, 1.8 and 1.7 Å, whereas the position of the first minimum is at 2.5 Å for both conformers. These features are entirely consistent with the formation of strong hydrogen bonds between amide groups and water molecules, on the one hand, and between carbonyl groups and water molecules on the other hand. We note that H−Ow and O−Hw separations as measured by first maximum in rdfs are always 0.1 Å shorter for aqueous *cis*-NMA. By integrating rdfs up to the position of the first minimum, we find that the number of water molecules hydrogen bonded to the amide group is 0.87 for *trans*-NMA and 0.91 for *cis*-NMA, while for the

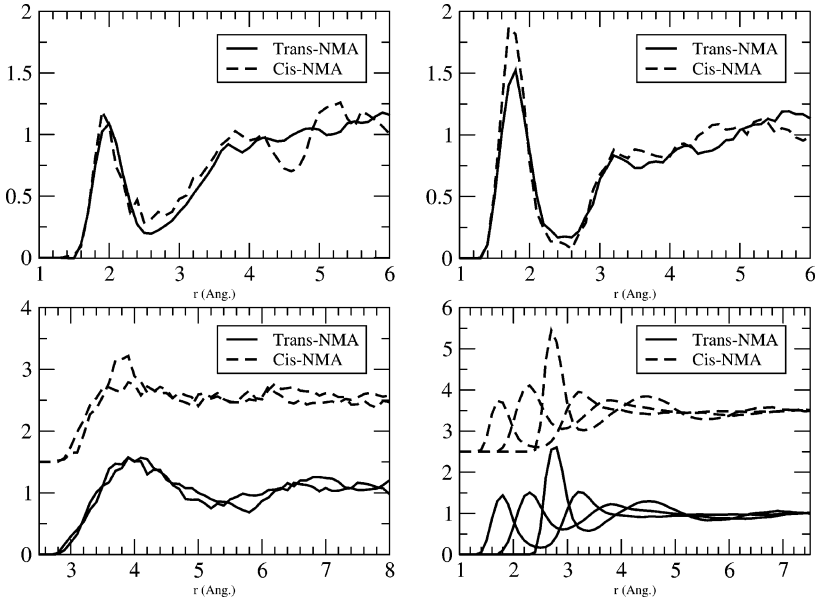


Figure 2. Solute−solvent and solvent−solvent radial distribution functions (RDF) calculated from Car−Parrinello molecular dynamics simulations of aqueous NMA. Conventions: H is NMA amide hydrogen atom, O is NMA carbonyl oxygen atom, C is one of the two NMA carbonyl carbon atoms, Ow and Hw are respectively the oxygen and hydrogen atoms of water molecules. RDF reported here: H−Ow (top left), O−Hw (top right), C−Ow (bottom left, with both carbonyl represented), Ow−Ow, Ow−Hw, Hw−Hw (bottom right). RDFs for *trans*-NMA are always in solid lines, and those for *cis*-NMA are in dashed lines. Distances are in angstroms.

carbonyl group the number is 2.0 for *trans*-NMA and 2.22 for *cis*-NMA. The total number of water molecules coordinated with the hydrophilic amide and carbonyl groups is thus slightly higher for *cis*-NMA than for *trans*-NMA. This results in a higher amplitude of the first O−Hw peak for *cis*-NMA (top right of Figure 2), whereas the amplitude of the first minimum is identical for both O−Hw rdfs. This difference suggests that water molecules are more tightly H-bonded to the C═O group of *cis*-NMA than to that of *trans*-NMA.

We observe a well-defined second maximum in the H−Ow rdf of aqueous *cis*-NMA, located around 4 Å with a finite width of ∼1 Å. Taking into account the characteristic distances displayed in our rdfs, this suggests that some water molecules belonging to the second hydration shell around the amide hydrogen of *cis*-NMA are also part of the first hydration shell of the carbonyl oxygen atom. This type of bridge formation is only possible for a cis conformer where the N−H and C═O are placed close to each other. This also implies that water molecules H-bonded to the amide group of *cis*-NMA are separated from the ones H-bonded to the carbonyl group, i.e., there are no water molecules shared by both groups in their first hydration shell. This situation is different from what would be expected in the geometry optimization (at 0 K) of *cis*-NMA with one water molecule, where the water molecule would be found to make a bridge so as to form a stable six-membered ring including the CO, CN, and NH bonds of *cis*-NMA. At this point, it should be reminded that the finite temperature (300 K) and the presence of a bulk solvent in our simulation lead to a completely different view. In particular, as will be shown below (and already found in our previous investigation[61]), solvent−solvent H-bonds are always energetically more favorable than solute−solvent H-bonds. At 300 K, water molecules bind to the solute, while maintaining H-bonds with the bulk water molecules so as to stabilize the collective water−water H-bond network.

First peak positions obtained in our ab initio dynamics of aqueous *trans*-NMA simulation are similar to the results of another Car−Parrinello MD study[38] and also of QM/MM calculations[23] as well as classical dynamics with a polarizable force field[37] or a QM-RISM calculation.[60] There are, however, minor discrepancies. QM/MM calculations give a first peak position for O−Hw at a slightly larger distance than in our work, i.e., 2.10 and 1.95 Å depending on the ab initio level. Furthermore, QM-RISM and classical O−Hw radial distribution functions of *trans*-NMA display a marked narrow second peak of low amplitude, located between 3.0 and 4.0 Å. This finding appears also in the QM/MM calculations at the Hartree−Fock level and a rather poor basis set. This feature is absent in our ab initio MD.

The main differences between our work and the previous studies come from the strength that can be attributed to the intermolecular H-bonds formed between solute and solvent. A rough estimate is given by the ratio of the amplitude of the first peak (denoted g^{max}) to the amplitude of the first minimum (denoted g^{min}) of the radial distribution functions. Using our data, this gives a H−Ow H-bond strength for aqueous *trans*-NMA of $g^{max}/g^{min} \sim 5$, which is a bit stronger compared to the $g^{max}/g^{min} \sim 4$ obtained using a polarizable force field and ∼3.6 with QM/MM (again with variations depending on the ab initio level of the QM part). The ab initio MD bond strength appears much stronger than in the QM-RISM calculation where a value of ∼2.0 is obtained. In contrast, our O−Hw H-bond strength is lower than the ones obtained in using QM/MM and MM methods ($g^{max}/g^{min} \sim 7.5$ versus ∼8.5 in QM/MM and ∼9.5 with polarizable

Table 2. Geometrical Distances between NMA Atoms and Wannier Centers Obtained by Averaging over the CPMD Trajectory of *trans*-NMA and *cis*-NMA

	trans-NMA[a]	*trans*-NMA[b]	*cis*-NMA[a]	*cis*-NMA[b]
N−WC N−C bond	0.475 ± 0.003	0.495 ± 0.032	0.459 ± 0.039	0.496 ± 0.033
N−WC N−C bond	0.476 ± 0.003	0.509 ± 0.030	0.480 ± 0.038	0.506 ± 0.031
N−WC N−H bond	0.607 ± 0.003	0.586 ± 0.013	0.602 ± 0.003	0.585 ± 0.011
N−WC N−CH3 bond	0.627 ± 0.002	0.626 ± 0.010	0.634 ± 0.002	0.626 ± 0.013
O−WC lone pair	0.314 ± 0.002	0.338 ± 0.014	0.318 ± 0.001	0.334 ± 0.011
O−WC lone pair	0.308 ± 0.002	0.323 ± 0.014	0.308 ± 0.001	0.342 ± 0.013
O−WC C═O bond	0.499 ± 0.002	0.469 ± 0.013	0.502 ± 0.002	0.463 ± 0.016
O−WC C═O bond	0.499 ± 0.002	0.463 ± 0.015	0.502 ± 0.002	0.467 ± 0.012
C_C−WC C_C−C bond	0.756 ± 0.004	0.764 ± 0.019	0.756 ± 0.006	0.762 ± 0.018
C_C−WC C_C−H bond	0.734 ± 0.009	0.721 ± 0.025	0.728 ± 0.004	0.718 ± 0.018
C_C−WC	0.726 ± 0.009	0.720 ± 0.021	0.722 ± 0.002	0.717 ± 0.020
C_C−WC	0.722 ± 0.003	0.721 ± 0.021	0.728 ± 0.003	0.718 ± 0.022
C_N−WC C_N−H bond	0.731 ± 0.009	0.721 ± 0.021	0.728 ± 0.003	0.727 ± 0.021
C_N−WC	0.728 ± 0.010	0.721 ± 0.021	0.722 ± 0.002	0.726 ± 0.016
C_N−WC	0.733 ± 0.008	0.721 ± 0.021	0.728 ± 0.003	0.728 ± 0.021

[a] Isolated molecule at ∼20 K. [b] Hydrated molecule at ∼300 K. Distances are given in angstroms.

force fields) but still well above the ∼6.4 determined in the QM-RISM work. Finally, we draw attention to the very good agreement between our results and Martyna's,[38] who used almost identical AIMD methods but a smaller model system consisting of 27 water molecules.

Also very instructive is the comparison of the H−Ow and O−Hw H-bond strength for *trans*-NMA and *cis*-NMA as implied by the structure of the ab initio MD rdfs. The H−Ow H-bonding is comparable in both systems (same values of g^{max} and g^{min}), whereas the O−Hw H-bond is stronger in *cis*-NMA ($g^{max}/g^{min} \sim 9$ versus ∼7.5). As previously emphasized, this can be traced back to water molecules more tightly H-binding to the carbonyl group of *cis*-NMA than to *trans*-NMA. This is also connected to the presence of a well-defined second peak in the H−Ow rdf of aqueous *cis*-NMA. Turning to the nonpolar sites in the molecule, Figure 2 (bottom left) shows all the rdfs involving the carbon atom of the CH_3 methyl groups. There is no indication of formation of hydrogen bonds between these groups and surrounding water molecules (first peaks located around 4.0 Å), and we do not see any distinction between C- or N-terminal methyls. The slightly higher amplitude of the peak observed for the N-terminal CH_3 of *cis*-NMA could be related to the more ordered second hydration shell around the N−H group as we have described previously.

Radial distribution functions among the solvent water molecules (Hw−Hw, Ow−Ow, and Hw−Ow) are reported in Figure 2 (bottom right). We find that each water molecule has 4.1 and 4.0 water neighbors on average for *cis*- and *trans*-NMA, which is similar, if perhaps somewhat smaller than the values in pure liquid water (≈4.3). Comparison of H-bond strengths estimated by the ratio g^{max}/g^{min} for all the computed rdfs (solute−solvent and solvent−solvent) leads to the following order: Hw−Ow > O−Hw > H−Ow, for both *trans*- or *cis*-NMA. Thus, H-bonds among the solvent molecules are always more beneficial than solute−solvent H-bonds. This result was already observed in the ab initio molecular dynamics of aqueous uracil.[61]

A comparison of the solvent−solvent rdfs of Figure 2 to pair distributions for the pure liquid can give a good indication of the extent to which the solvent is disturbed by the presence of the solute. The question is particularly relevant in view of our limited system dimensions. The rdfs of Figure 2 in fact appear very similar to the rdfs of (BLYP) homogeneous liquid water of refs 62−64 which, moreover, are in fair agreement with experiment. This suggests that solvent in our system essentially retains the structure of the pure liquid. However, the results of refs 62−64 have recently been questioned.[65] The rapid advances in computer power of recent years have made it possible to extend the length of the trajectories routinely to 20−30 ps and in exceptional cases even to longer times. Similarly, current AIMD simulations are carried using computationally more demanding implementations of the ab initio MD algorithm which approach the ideal Born−Oppenheimer limit closer than in the original simulations, using either smaller values of fictitious electron mass[65] or Born−Oppenheimer dynamics.[66] These new studies show the rdfs of BLYP water is significantly more structured than suggested by the first generation of MD simulations.[65] Even if the issue is far from settled,[81] the consensus seems now that a converged BLYP result at ambient conditions actually corresponds to the experimental structure of undercooled water at a substantially lower temperature (−20 °C).[66] The way this is manifested in practice is that the motion of the water (diffusion) slows down as the run progresses. Such a relaxation process was not observed in the present simulation of the model NMA solution. Note that the value of the fictitious electron mass of 500 au we have applied is significantly lower than the 800 au used in refs 62−64. Our trajectories, however, are still relatively short (∼10 ps). While it cannot be ruled out that further relaxation will take place on a longer time scale, the stability of our system for the duration of the simulation suggests that the system has settled in a (possibly) metastable state with a solvent structure and mobility in fair agreement with bulk liquid water.

5. Electronic Analysis

5.1. Wannier Functions and Born Charges. Table 2 reports the average positions of the Wannier centers (WC) obtained

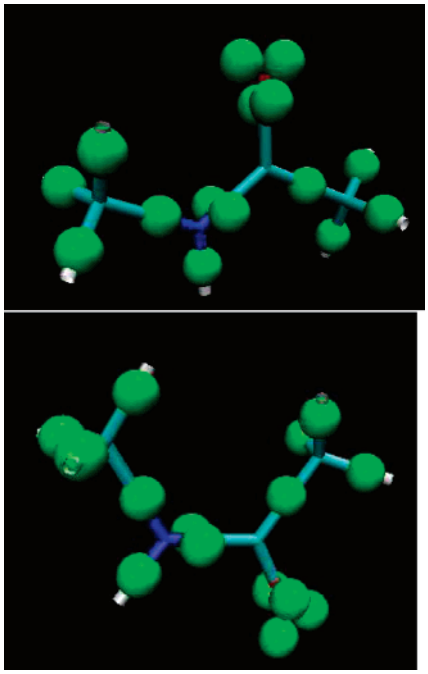


Figure 3. Representation of the positions of the Wannier centers (spheres) as calculated for *trans*-NMA (top) and *cis*-NMA (bottom). The Wannier center coordinates are given in Table 2 for the gas and aqueous phase.

from our gas- and aqueous-phase CPMD trajectories of *trans*- and *cis*-NMA. Wannier centers are represented in Figure 3 either by spheres centered on or near covalent bonds or by spheres that correspond to lone pairs of oxygen and nitrogen atoms. The same trends are observed for *trans*- and *cis*-NMA upon going from the gas to the aqueous phase. First, electrons along the N−H amide bond can be seen to be pulled toward the nitrogen atom in the liquid phase. This is expected from the formation of a hydrogen bond between N−H and surrounding water molecules. Similarly, we observe that the oxygen lone pairs are pushed out, also as a result of H-bonding, while electrons of the double bond C═O are displaced toward the oxygen atom. The polarization of the N−H and C═O bonds induced by the hydrogen bonds, is accompanied by a displacement of the electrons along the N−C bond away from the nitrogen atom, as indicated by the increase in nitrogen-Wannier centers distances (along that bond) in response to solvation. The electrons move closer to the middle of the bond suggesting that the N−C is stronger in the liquid, which is also consistent with the observed shrinking of the N−C bond length. This is a consequence of the nature of the peptide bond. Finally, we see that electrons along the $C-CH_3$ skeleton bond are pulled toward the carbonyl carbon in the liquid phase.

Born charges calculated according to eq 7 for the atoms of *trans*-NMA provide supplementary data to quantify and analyze the rearrangement of electrons, when going from the gas phase to the liquid phase. The results of this computation are listed in Table 3. Summing the charges over all atoms of the *trans*-NMA molecule, we verified that, as required by the sum rule for Born charges, the molecule is strictly charge neutral in the gas phase (absolute total charge less than 7×10^{-4} *e* due to numerical uncertainties). In

Table 3. Born Charges Calculated along the Trajectory of Gas-Phase *trans*-NMA[a] and Aqueous-Phase *trans*-NMA[b] (Units of Electrons)

atom type	gas phase[a]	aqueous phase[b]
C	1.049966	1.273998
O	−0.784995	−1.222748
N	−0.697676	−0.869836
H (amide)	0.161829	0.398282
C_C	−0.057427	−0.092539
C_N	0.339080	0.310198
H (C_C)	−0.011471	0.029868
H (C_C)	0.018505	0.036330
H (C_C)	0.020947	0.046038
H (C_N)	−0.003725	0.022462
H (C_N)	−0.039359	0.026497
H (C_N)	0.003645	0.018143

solution, however, *trans*-NMA acquires an average charge of −0.025 *e*. This relative small excess charge most likely originates in the dynamical contribution to the Born charge and should not be interpreted as a solvent to solute charge transfer. The dynamical charge measuring the response of the electronic charge density to displacements of the nuclei can lead to sometimes counterintuitive estimates of total charge. In particular in situations where the charge cloud lags behind, increasing the value of the dipole moment, the dynamical charge enhances the static contribution (for example the Born charge of a Na^+ ion in crystalline NaCl is +1.1 *e*).

Solvent induced changes of individual Born charges support the above view of electron flow based on the WC analysis. The C═O oxygen is made more negative by 0.44 *e*. The positive charge of the carbonyl carbon atom is increased by 0.22 *e*. On the N−H side, we observe that the hydrogen atom becomes more positive by 0.24 *e*, while the negative nitrogen atom is decreased by another 0.17 *e* in the aqueous phase compared to the gas phase. Alongside this polarization of the C═O and N−H bonds, most certainly arising from H-bonding with surrounding water molecules, it can be seen that the C═O end (including the C_CH_3 group) of *trans*-NMA gains 0.16 *e* in total which comes from the N−H group of the molecule. This charge transfer is again related to the special character of the peptide bond. Compared to atoms of the HNCO amide system, hydrogen atoms are globally chargeless, either in the gas phase or in the liquid phase.

The above picture can be rationalized with the help of the resonance model for the peptide bond: $C(=O)-NH \leftrightarrow C(-O^-)=NH^+$. In the gas phase, the neutral form is the most stable,[67] while in the aqueous phase we would expect the hydrogen bonding with the neighboring water molecules to stabilize the zwitterionic form with a more negatively charged oxygen (playing the role of an electron donor for the H-bond) and a positively charged N−H group (more ready to accept electrons from surrounding waters). This in turn is consistent with the observed charge transfer from the N−H end to the C═O end of *trans*-NMA (indicated by variations of the Born charges and WC displacements) and a strengthening of the N−C bond (shorter in the aqueous phase) and a weakening

Table 4

MD[a]	assignments[b]	expt[c]	expt[d]
3344	ν(N−H)		
2918, 2903, 2861	ν(C−H)		
1609	ν(C═O), δ(N−H) (amide I)	1707	1731, 1714
1458	δ(N-H), ν(N−C) (amide II)	1511	1499
1430	$\delta(C-H)_N$, ν(N−C)	1472	
1407	$\delta(C-H)_C$, ν(N−C)	1446	
1384	$\delta(C-H)_C$	1432	
1367	$\delta(C-H)_N$	1419	
1327	$\nu(C-CH_3)$, $\delta(C-H)_C$	1370	∼1420
1189	ν(N−C), δ(N−H) (amide III) + $\nu(C-CH_3)$, $\nu(N-CH_3)$	1266	1255
1098	$\nu(N-CH_3)$, $\delta(C-H)_N$	1168	
1015, 1005	ν(N−C), $\nu(N-CH_3)$, $\nu(C-CH_3)$, δ(N−H), CH_3 rockings	1089	

[a] Band positions (cm^{-1}) in the vibrational density of states of isolated *trans*-NMA computed at 20 K. [b] Assignments according to the method described in the Supporting Information. The notation of the bands are described in the text. [c] Experimental IR spectrum of *trans*-NMA in an argon matrix at 20 K.[27] [d] Experimental IR spectrum of *trans*-NMA in the gas phase at ∼100 K.[7]

of the C═O bond (longer in the liquid phase). This was already suggested by previous studies by Torii et al. for the lowest excited state of *trans*-NMA which has a clear zwitterionic form in the gas phase and, reversely to the ground state, a less pronounced zwitterionic form in aqueous phase.[67] The important consequences of these charge transfers on the IR intensities in solution will be examined in a later section.

5.2. Dipole Moments. Averaging the dipole moments of isolated *trans*-NMA and *cis*-NMA over the low temperature trajectories gives 3.99 D and 4.38 D, respectively. The dipole moment of *cis*-NMA is thus higher than the one of *trans*-NMA, an ordering also obtained in other theoretical works[23,24,60] using geometry optimizations of *trans*- and *cis*-NMA at various levels of calculations. The values obtained here are also very similar to those of refs 23, 24, and 60. We note that our 3.99 D value for *trans*-NMA is very close to the experimental value of 3.85 D measured in benzene.[68,69] In solution, the intramolecular charge transfers described in the previous section lead to a substantial increase of the dipole moments of both isomers. Adding the positions of the nuclei and centers of the maximally localized Wannier functions discussed in section 5.1 we find that solvent polarization enhances the average dipole moments to 6.96 D for *trans*-NMA and 7.33 D for *cis*-NMA. *cis*-NMA remains therefore the more polar conformer in solution. Also the dipole difference is, on average, conserved, i.e., 0.39 D in the gas phase versus 0.37 D in aqueous solution. Our assignment of the higher dipole in solution to *cis*-NMA is in agreement with the QM-RISM work of Wei et al.,[60] whereas the reverse ordering was found in a PCM-B3LYP model using the SCRF theory.[24] The absolute values of the aqueous dipole moments predicted by both calculations are considerably lower than the ab initio MD estimates. The values of ref 24 are especially low, almost equal to the values that we obtain in the gas phase. Our dipole moments are also ∼1.2−1.4 D higher compared to the ones computed in ref 60.

6. Vibrational Density of States

To understand our calculated infrared spectra in terms of vibrationally active infrared modes, we have computed the vibrational density of states (VDOS) of the isolated and aqueous *trans*- and *cis*-NMA molecules. The VDOS is obtained by Fourier transformation of the velocity autocorrelation function of the atoms. In addition we have decomposed the VDOS according to each atom type in order to get an interpretation of the vibrational bands in terms of individual atomic motions. Regarding the 1000−2000 cm^{-1} frequency domain which is of particular interest in experiments on peptides and proteins, and specifically related to the amide stretching and bending modes used for the characterization of helices and sheets structures, this simple decomposition proves sufficient to properly assign each vibrational peak of NMA to specific atomic motions. A complication is that the rocking motions of methyl groups have also a spectral signature in the same spectral domain. We have thus calculated also the Fourier transform of the correlation function of the three principal axis components of the methyl groups, to get more specific spectral signatures of rocking motions. The details of all this analysis are given in the Supporting Information. For each species, the final assignments of the different peaks are summarized in Tables 4−7. The overall spectrum in the 1000−2000 cm^{-1} region can be discussed in terms of coupled stretching (ν) and bending (δ) motions, more specifically ν(C═O), ν(N−C), and δ(N−H) for the central peptide region and $\nu(N-CH_3)$, $\delta(C-H)_N$, $\nu(C-CH_3)$, and $\delta(C-H)_C$ for the N-terminal and C-terminal regions, respectively. The systematic coupling of ν(C═O) and δ(N−H) gives rise to the so-called amide I mode, and the one between ν(N−C) and δ(N−H) results in amide II and amide III. The $\nu(N-CH_3)$ and $\nu(C-CH_3)$ stretches appear always slightly coupled to the C−H bends of the attached methyl groups, $\delta(C-H)_N$ and $\delta(C-H)_C$. A collective backbone motion involving all the backbone stretches, the N−H bend, and the rocking of the methyl groups is always observable at the lowest frequency part of the spectrum.

6.1. Isolated NMA. The vibrational frequencies and the corresponding assignments of gas-phase *trans*- and cis-NMA are reported in Tables 4 and 6. It appears that the two spectra have very similar features in the 1000−1700 cm^{-1} spectral domain: one single band related to amide I around 1600 cm^{-1}, a broad band between 1300 and 1500 cm^{-1} that

Table 5

MD[a]	assignments[b]	expt[c]
3259, 3200		
1983, 2920, 2861		
1535, 1522	amide I	1617[c1], 1620[c2], 1625[c3]
1483, 1466	amide II	1577[c1], 1580[c2], 1582[c3]
1450−1325	$\delta(C-H)_N$, $\delta(C-H)_C$	1416[c1], 1418[c2], 1378[c1], 1378[c2]
1233	amide III + $\nu(N-CH_3)$, $\delta(C-H)_N$, $\delta(C-H)_C$	1314[c1], 1315[c2]
1098	$\nu(N-CH_3)$, $\delta(N-H)$, $\delta(C-H)_N$	1167[c2]
1013	$\nu(N-C)$, $\nu(N-CH_3)$, $\nu(C-CH_3)$, $\delta(N-H)$, CH_3 rockings	1097[c2]

[a] Band positions (cm^{-1}) in the vibrational density of states of aqueous *trans*-NMA computed at 320 K. [b] Assignments according to the method described in the Supporting Information. The notation of the bands are described in the text. [c] Experimental IR spectrum of aquous NMA from refs 31 (c1), 32 (c2), and 7 (c3).

Table 6

MD[a]	assignments[b]	expt[c]
3327	$\nu(N-H)$	
2950, 2920, 2853	$\nu(C-H)$	
1606	amide I	
1428	$\delta(C-H)_N$, $\delta(C-H)_C$	1485
1403	$\delta(C-H)_N$, $\delta(C-H)_C$, $\nu(N-C)$, $\nu(N-CH_3)$	1454
1369	amide II	1432
1331	$\nu(N-C)$, $\nu(C-CH_3)$, $\delta(C-H)_C$	1387
1259	amide III	1325
1115, 1085	$\nu(N-CH_3)$, $\delta(C-H)_N$	
1015, 1005	$\nu(N-C)$, $\nu(N-CH_3)$, $\nu(C-CH_3)$, $\delta(N-H)$, CH_3 rockings	1075

[a] Band positions (cm^{-1}) in the vibrational density of states of isolated *cis*-NMA computed at 15 K. [b] Assignments according to the method described in the Supporting Information. The notation of the bands are described in the text. [c] Experimental IR spectrum of *cis*-NMA in an argon matrix at high temperature.[27]

Table 7

MD[a]	assignments[b]
3255, 3187	
2890	
1560, 1534	amide I
1466, 1433	$\delta(C-H)_C$, $\delta(C-H)_N$
1407, 1382	amide II
1365, 1339, 1328	amide III + $\nu(C-CH_3)$, $\delta(C-H)_C$, $\delta(C-H)_N$
1127	$\nu(N-CH_3)$, $\delta(C-H)_N$
1009	$\nu(N-C)$, $\nu(N-CH_3)$, $\nu(C-CH_3)$, $\delta(N-H)$, CH_3 rockings

[a] Band positions (cm^{-1}) in the vibrational density of states of aqueous *cis*-NMA computed at 302 K. [b] Assignments according to the method described in the Supporting Information. The notation of the bands are described in the text.

contains the most intense contribution to the whole VDOS (∼1330 cm^{-1}), and two bands identically located in both spectra in the 1000−1200 cm^{-1} spectral region. The amide I mode is therefore insensitive to a change in conformation (1609 cm^{-1} for *trans*- versus 1606 cm^{-1} for *cis*-NMA). The main differences between the two vibrational spectra come from the position of the amide II and amide III modes. Amide II in *trans*-NMA, located at 1458 cm^{-1}, is red-shifted by 89 cm^{-1} in *cis*-NMA (1369 cm^{-1}). On the other hand, the amide III band which appears as a single isolated peak in the 1100−1300 cm^{-1} region is blue-shifted by 70 cm^{-1} from the cis to the trans form.

There are four peaks related to methyl C−H bending in *trans*-NMA (∼1430 cm^{-1}, ∼1407, 1384, and 1367 cm^{-1}) and only two for *cis*-NMA (1428 and 1403 cm^{-1}). The mean frequencies of the two bands above 1400 cm^{-1} are identical for *trans*- and *cis*-NMA (with slight variations in their assignments). We also note that while we can clearly distinguish vibrational contributions from the methyl group located at the N-terminal or C-terminal position of *trans*-NMA, this is not the case for *cis*-NMA for which the signals of both methyl groups merge. The two vibrational modes peaked at 1384 and 1367 cm^{-1} in the *trans*-NMA spectrum, related to $\delta(C-H)$ bending, have disappeared in *cis*-NMA. In this region, *cis*-NMA exhibits only one band (1369 cm^{-1}) but related to amide II vibration. The backbone $\nu(C-CH_3)$ stretch coupled to $\delta(C-H)_C$ bend (∼1330 cm^{-1}) is not affected by the trans/cis conformation. Finally, the 1000−1100 cm^{-1} vibrational region is identical for both conformers, though we note that the broad band at 1098 cm^{-1} in the *trans*-NMA spectrum, related to the backbone $\nu(N-CH_3)$ stretch and $\delta(C-H)_N$ bend, is split into two bands located at 1085 and 1115 cm^{-1} for *cis*-NMA.

Much attention has been given in the literature to vibrational calculations for *trans*-NMA in the gas phase using standard ab initio geometry optimizations and frequencies determined at the harmonic level.[19−22,25,27,70,71] Very little work has been devoted to *cis*-NMA.[21] Detailed vibrational assignments of gas-phase *trans*-NMA have been reported by Wang et al.,[70] Mayne et al.,[71] Hagler et al.,[21] and Gerber et al.[22] (who went beyond the harmonic approximation). One interesting point revealed by Gerber's study is a downshift of up to 40 cm^{-1} of anharmonic frequencies with respect to their harmonic approximation. In these recent studies, no frequency scaling has been applied, contrary to older ab initio calculations.[19,20,25,27] A fundamental understanding of the vibrational modes was originally given in Miyazawa's study et al.,[72] on which most of the subsequent ab initio studies rely. Overall, frequencies calculated in our work using the BLYP functional are downshifted (by up to ∼100 cm^{-1} for amide I) compared to frequencies calculated with hybrid functionals (e.g. B3LYP) or with wave function based ab initio calculations. 5−10% underestimation of frequencies is typical for the BLYP functional. Although our calculated vibrational band positions show a systematic red-shift compared to the work of Gerber et al.,[22] we find good agreement with Gerber's anharmonic calculation for the frequency separations between amide II and amide III and between amide III and the backbone stretches (respectively, 269/264 cm^{-1} and 174/138 cm^{-1}). Our amide I−amide II

frequency gap (151 cm^{-1}) is too low in comparison to the result of ref 22 (204 cm^{-1}). The vibrational bands assignments proposed from our molecular dynamics are very similar to refs 21, 22, 70, and 71, except for the methyl motions which, we think, are not accounted for properly within a harmonic approximation. We find that the C−H bending motions of the C-terminal methyl group of *trans*-NMA are packed in the same spectral domain (1407−1384 cm^{-1}), whereas they are split apart at the N-terminal end. Hagler et al.[21] and Gerber et al.[22] find a systematic alternation of these modes in the same spectral domain. We remark that Hagler et al. have used the original assignments of Ataka et al.,[27] for which some ambiguities remain. Finally, in agreement with Hagler's work, we find that the amide I and amide II modes of *cis*-NMA are separated by ∼240 cm^{-1}; on the other hand, our amide II−amide III separation is equal to 110 cm^{-1}, to be compared to 185 cm^{-1} in their work.

6.2. Aqueous NMA. Again, for the detailed analysis of the VDOS of each isomer in solution we refer to the Supporting Information. Tables 5 and 7 list the position and our interpretation of the vibrational bands. A very noticeable effect of the solvent on the VDOS is that the amide I stretching motion undergoes a red-shift (87 cm^{-1} for *trans*-NMA and 46−72 cm^{-1} for *cis*-NMA), whereas the amide II is blue-shifted (∼8−25 cm^{-1} and 3−38 cm^{-1}, respectively). The red-shift of amide I is consistent with the formation of hydrogen bonds between the carbonyl group and the surrounding water molecules, as was emphasized in the structural analysis of solute−solvent H-bonding and in the charge flow analysis of section 5. The red-shift computed for *trans*-NMA is about twice larger than obtained for aqueous uracil[12] in our previous study. This more pronounced red-shift could be the signature of very strong hydrogen bonds formed between the NMA carbonyl group and the solvent. Similarly the g^{max}/g^{min} ratios obtained in the present simulations of *trans*- and *cis*-NMA are much larger than the ones obtained for aqueous uracil[61] (∼3.7−4.5 versus ∼7.5−9.3 for NMA). The blue-shift of amide II is also the result of H-bonding between the amide group of NMA and the surrounding water molecules and was also observed in the vibrational spectrum of aqueous uracil.[12]

The comparison of the VDOS of aqueous *trans*-NMA and *cis*-NMA reveals also several interesting features. The two bands around 1000 and 1100 cm^{-1} are identically located in both spectra. The 1300−1500 cm^{-1} spectral domain displays similar peaks for *trans*- and *cis*-NMA, although we have seen that their interpretation leads to very different assignments. The most obvious differences are the δ(C−H) bending mode and the amide II mode, the positions of which are roughly inverted in both spectra. The amide II bands of *trans*- and *cis*-NMA are shifted by ∼80−100 cm^{-1} with respect to each other, just as was observed in the gas phase. Finally, the amide I mode is ∼30 cm^{-1} blue-shifted from *cis*- to *trans*-NMA, i.e., 1560−1534 cm^{-1} versus 1535−1522 cm^{-1}. We recall that this mode was peaked at ∼1606 cm^{-1} in both gas-phase spectra.

Frequencies of *trans*-NMA hydrogen bonded with two water molecules via N−H and C═O groups have been calculated by Krimm et al.[34,73] at the Hartree−Fock level. Surprisingly (considering that our calculation fully takes into account two layers of solvent molecules around NMA instead of only two molecules and considering the difference in the level of calculations), the trends obtained here for the amide band separations are very comparable: We find ∼52−69 cm^{-1} between the amide I and amide II bands, while ∼50−70 cm^{-1} is found in refs 34 and 73, and our ∼250−230 cm^{-1} amide II/amide III gap compares very well with the 277 cm^{-1} gap of refs 34 and 73.

7. Infrared Spectroscopy

Unless stated otherwise, the infrared spectra discussed below are calculated using the dipole−dipole correlation function of the eq 1 of section 3. The alternative way of calculating infrared spectra through the current−current correlation function, eq 6 of section 3, will be discussed in section 7.3 for the particular case of aqueous *trans*-NMA.

7.1. Infrared Spectra of Gas-Phase NMA. We have reported in Figure 4 the infrared spectrum of isolated *trans*-NMA calculated from the quantum molecular dynamics trajectory, again in the 1000−2000 cm^{-1} spectral domain. The active infrared modes are identified by comparing to the vibrational density of states (Table 4). A schematic representation of vibrational assignments is presented at the top of Figure 5. Six infrared active bands are observed in the 1000−2000 cm^{-1} region. They are relatively narrow as expected in the gas phase. The most intense IR band located at 1609 cm^{-1} is related to the amide I mode. In the spectral domain between 1500 and 1300 cm^{-1}, we obtain four active infrared bands of much lower intensity: The amide II mode (1458 cm^{-1}), $\delta(C-H)_C$ (1407 cm^{-1}), and $\delta(C-H)_N$ (1367 cm^{-1}) bending, and the backbone stretch $\nu(C-CH_3)$ coupled to $\delta(C-H)_C$ (1327 cm^{-1}). As previously described in the VDOS analysis, the mirror mode on the N-terminal side, i.e., $\delta(C-H)_N/\nu(N-CH_3)$, should contribute to the modulation of the dipole moment of the molecule, thus giving rise to an active infrared band, but it does not. The more intense active band located at 1189 cm^{-1} is due to amide III. The intensity of this band remains nonetheless much smaller than the one of amide I. All other modes related to "pure" motions of the methyl groups (without any strong contributions from the skeleton atoms) do not show any infrared activity. Surprisingly, the collective backbone stretching deformation mode combined with methyl rocking, located at ∼1000 cm^{-1}, has no infrared activity either. The amplitude ratios obtained in our calculated bands are ∼10 for amide I−amide II, ∼5 for amide I−amide III, and ∼0.5 for amide II−amide III.

The result for the infra spectrum of gas-phase *cis*-NMA is given in Figure 4, and a schematic representation of the vibrational assignments appears at the top of Figure 6. We immediately see that this spectrum is totally different from the *trans*-NMA spectrum: in particular, the number of active bands is reduced from six to four, and the 420 cm^{-1} overall frequency range of the *trans*-NMA spectrum is reduced to ∼350 cm^{-1}. The three main infrared bands have respectively a very high (1606 cm^{-1}), high (1331 cm^{-1}), and medium intensity (1259 cm^{-1}). One additional band of very low intensity is located around 1400 cm^{-1}. The most intense active band is associated with the amide I mode. The

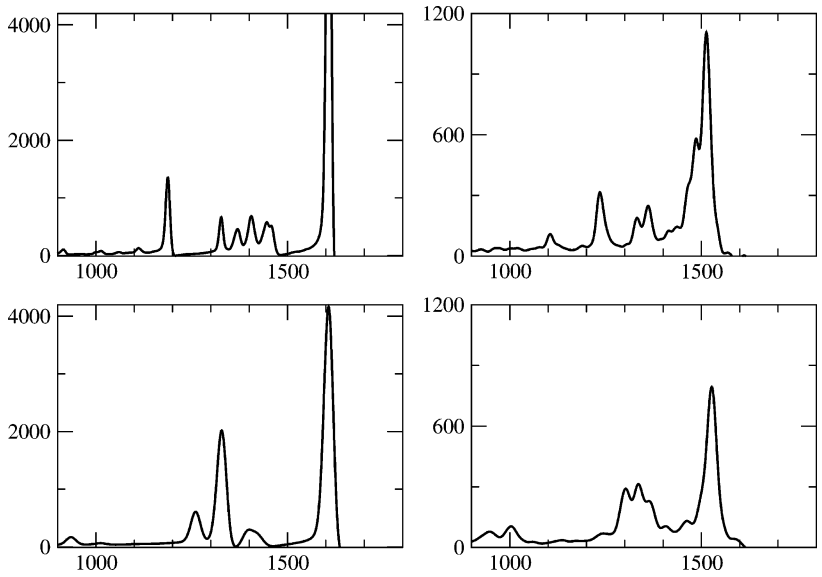


Figure 4. Calculated infrared spectra of gas-phase *trans*-NMA (top left), gas-phase *cis*-NMA (bottom left), aqueous *trans*-NMA (top right), and aqueous *cis*-NMA (bottom right). Spectra are reported in the 1000−1800 cm^{-1} wavenumber range, and IR amplitudes are reported in units of cm^{-1}.

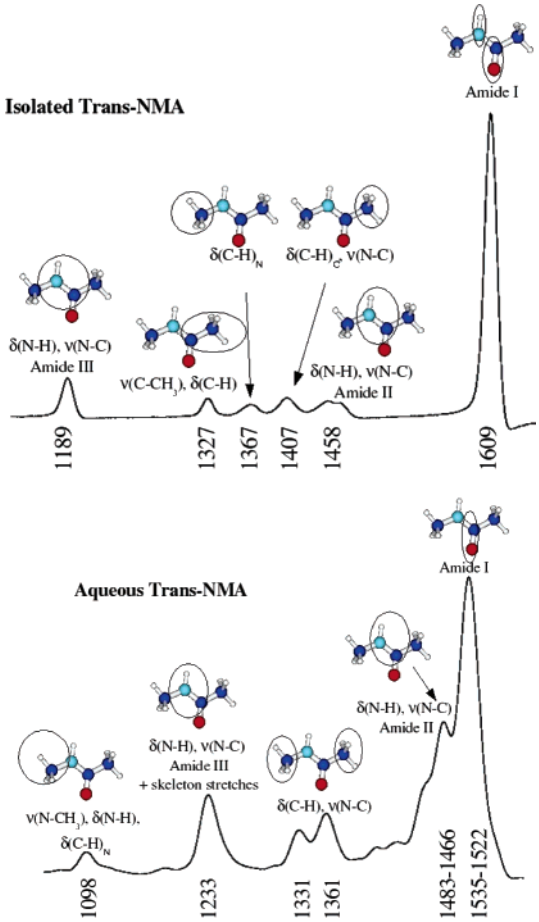


Figure 5. Schematic representation of the assignments of the calculated infrared spectrum of gas-phase (top) and aqueous-phase *trans*-NMA (bottom).

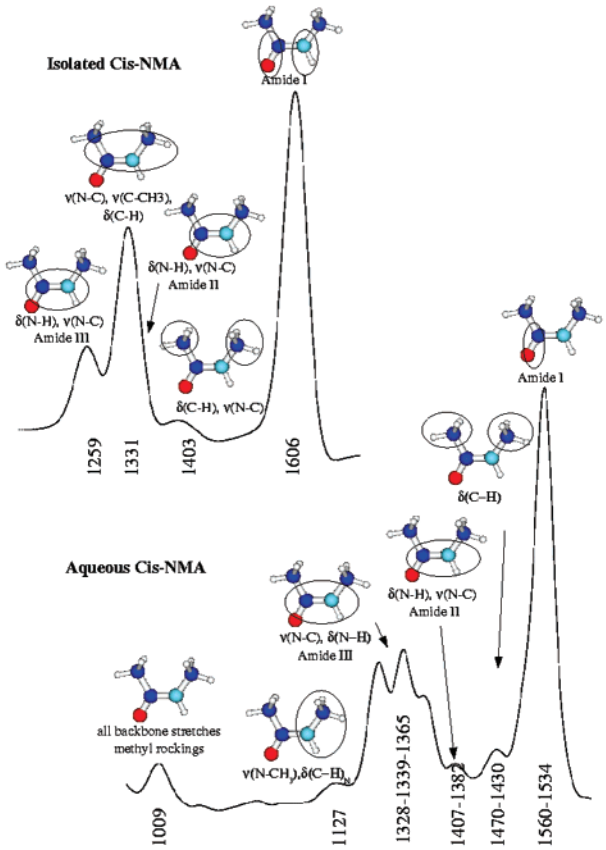


Figure 6. Schematic representation of the assignments of the calculated infrared spectrum of gas-phase (top) and aqueous-phase *cis*-NMA (bottom).

relatively broad band peaked at 1331 cm^{-1} and extending between ∼1300 cm^{-1} and ∼1400 cm^{-1} is composed of the $\nu(C-CH_3)/\delta(C-H)/\nu(N-C)$ mode identified previously at 1331 cm^{-1} in the VDOS, together with the amide II mode (1369 cm^{-1}) in the higher wavenumber part of the band. Thus, in our calculation, amide II does not appear in a single peak but is instead merged into a broader band which combines two vibrational modes. The infrared active mode located at 1259 cm^{-1} is due to amide III. The band of very weak intensity around 1400 cm^{-1} is due to $\delta(C-H)$ coupled

to ν(N−C) motions. Amplitude ratios between the peaks are ∼1.7 between amide I and the 1331 cm^{-1} band and ∼3.3 between amide I and amide III.

The amide I band is thus peaked in the two spectra of *trans*- and *cis*-NMA at the same frequency, and for both molecules, this band has the most intense activity of the whole spectrum. A second remark is that the IR spectrum of *trans*-NMA is composed of four distinct bands of rather low intensity in the 1500−1200 cm^{-1} domain (among which amide II) which are well separated from the more intense amide III band located at 1189 cm^{-1}. In the same domain, we observe only two distinct infrared bands, of respectively intense and medium intensities, in the spectrum of *cis*-NMA. A striking difference is the amide II mode which emerges as a single band in *trans*-NMA, whereas it is merged with the ν(C−CH$_3$) stretching band at 1331 cm^{-1} in *cis*-NMA. We note that the band with maximum at 1331 cm^{-1} in both spectra has a much higher activity in *cis*-NMA, presumably because of the merging with the amide II mode. The two δ(C−H) infrared bands of *trans*-NMA (1407 and 1367 cm^{-1}) are converted into one single peak in *cis*-NMA (∼1400 cm^{-1}) in which the N-terminal and C-terminal methyl groups jointly participate.

7.2. Infrared Spectra of Aqueous NMA. The infrared spectrum as computed for aqueous *trans*-NMA is shown in Figure 4. A schematic representation of vibrational assignments is added at the bottom of Figure 5. In the 1000−2000 cm^{-1} spectral domain the spectrum consists of six main peaks. The band broadly extending between 1450 and 1550 cm^{-1} has the highest intensity in the overall spectrum. Amide I and amide II modes are merged into this band, and amide II appears as a shoulder of high intensity. The infrared peaks of medium intensity located at 1361 and 1331 cm^{-1} come from δ(C−H) vibrations (combined with some ν(N−C) stretch). The peak of high intensity at 1233 cm^{-1} arises from absorption by amide III motion. Its high intensity is the consequence of a strong modulation of the dipole moment of the molecule through the N−C stretching and the combination with other skeleton stretching, evidenced in the VDOS analysis. The intensity of this band is about a third of that of the coupled amide I−amide II one. Finally, the band of low intensity at 1098 cm^{-1} is due to ν(N−CH$_3$)/δ-(C−H)$_N$/δ(N−H).

Turning to aqueous *cis*-NMA the calculated IR spectrum of this conformer in the 1000−2000 cm^{-1} region (Figures 4 and 6) is composed of two main bands positioned at ∼1526 cm^{-1} and around 1339 cm^{-1} and three peaks of lower intensities around 1450, 1400, and 1000 cm^{-1}. We can also distinguish a shoulder of very weak intensity around 1127 cm^{-1}. The very intense band peaking at 1526 cm^{-1} and roughly extending between 1400 and 1600 cm^{-1} is amide I. The intense band centered around 1339 cm^{-1} is resolved in three distinct maxima and is due to a mixing of amide III and ν(C−CH$_3$)/δ(C−H) (modes at 1328−1339−1365 cm^{-1} in the VDOS analysis of Table 7). This band combines on the high wavenumber side with the lower intensity amide II band (1382−1407 cm^{-1} in the VDOS). The shoulder of weak intensity at ∼1127 cm^{-1} is related to ν(N−CH$_3$)/δ(C−H)$_N$, while the collective backbone stretching and methyl rocking give rise to the active band of medium intensity located at ∼1000 cm^{-1}.

Thus, the band features displayed by the infrared spectra of aqueous *trans*- and *cis*-NMA are also quite different, as already observed in the gas phase. In particular, in the IR spectrum of aqueous *trans*-NMA the amide I and amide II modes are combined in a single very intense active band, while two intense bands appear in the spectrum of aqueous *cis*-NMA, i.e., amide I and the merged amide III−amide II. As a consequence, we can see that the 1300−1600 cm^{-1} spectral domain can be used as a very good spectral indicator of the liquid-phase trans versus cis conformation. The 1000−1300 cm^{-1} spectral domain also displays very different features for both conformers. The spectrum of *cis*-NMA appears flat in this region, whereas *trans*-NMA produces two bands, one at 1233 cm^{-1} corresponding to the amide III mode and the other at 1100 cm^{-1} for ν(N−CH$_3$)/δ(C−H)$_N$. On the other hand, the combination of skeleton stretching motions and methyl rocking around ∼1000 cm^{-1} has a substantial amplitude for *cis*-NMA but is almost absent in the spectrum of *trans*-NMA.

Comparing the infrared absorption in the gas and liquid phase, we can infer the following solvent effects on the infrared spectra of *trans*- and *cis*-NMA. Starting with *trans*-NMA, a striking consequence of the interaction with the solvent is the combination of a red-shift of the amide I and blue-shift of amide II band reducing the 150 cm^{-1} frequency gap observed in the gas phase to ∼40−50 cm^{-1} in solution. The result is the broad double band containing both modes. In both gas and liquid phases, apart from the very intense amide I band, the more intense band is due to the amide III mode, and solvation induces a blue-shift of 40 cm^{-1}. A further noticeable solvent effect is the gain of infrared activity of ν(N−CH$_3$)/δ(C−H)$_N$ at 1098 cm^{-1}. For *cis*-NMA, the main changes concern the two most intense bands in the gas-phase infrared spectrum, i.e., amide I at 1606 cm^{-1} and the merging of ν(C−CH$_3$), amide II, and amide III around 1200−1400 cm^{-1}. These bands carry comparable intensities in the gas phase and aqueous spectra. However, amide I is 80 cm^{-1} downshifted by the presence of the solvent, whereas amide III is blue-shifted by the same amount. Since ν(C−CH$_3$) and amide II keep basically the same position, it is the change of amide III which is mostly responsible for the band shape evolution. Note again the gain of intensity of the collective backbone motion at ∼1000 cm^{-1} in solution.

It is also instructive to relate the red- and blue-shifts of the amide bands to the charge flows and resonant states electronic picture described in section 5. The red-shift of amide I mode in the aqueous phase appears as being due to the weakening of the C═O bond which is the result of intermolecular hydrogen bonding and increased weight of the zwitterionic form. On the other hand, blue-shifts of the amide II and amide III bands can be related to the observed strengthening of N−C bond. Polarizations of both N−H and C═O bonds upon hydrogen bonding, which increase the Born charges of all these atoms (in absolute value), seems to play the most important role in the increase of IR intensities of the amide II and amide III bands, particularly with respect to the charges on N and C atoms. These

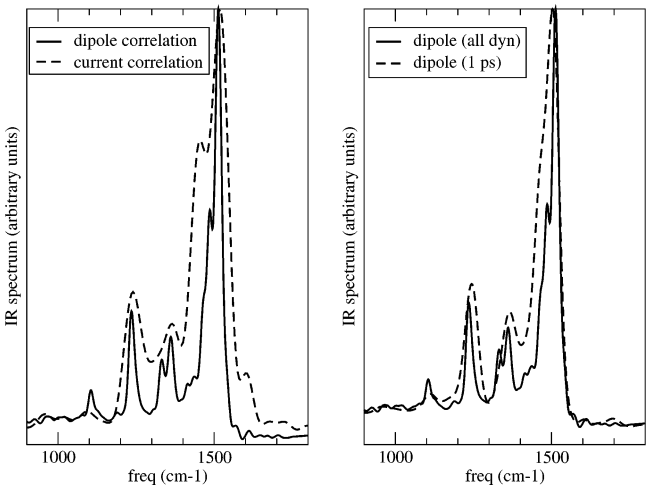


Figure 7. Left: comparison of the NMA infrared absorption spectrum calculated from the solute dipole−dipole correlation function (solid line) obtained over 9.35 ps of dynamics and from current−current correlation function (broken line) obtained over 1.0 ps of dynamics (see section 3). Right: NMA infrared spectrum calculated from the dipole−dipole correlation function over the 9.35 ps of dynamics (solid line) and over 1 ps of dynamics (broken line). See text for conclusions.

intensities however will be very sensitive to the charge transfer from the N-terminal to the C-terminal end, induced by the conjugated character of the N−C═O system, which would tend to lower the charges on N and C and increase (in absolute value) the charges of H and O. A proper account of the electronic structure of the peptide bond is thus needed to accurately reproduce the IR spectra of *cis*- and *trans*-NMA in solution. These observations will also have consequences on the interpretation of the peak shifts of the amide bands of peptides in different environments.

7.3. Dipole−Dipole versus Current−Current Correlation Functions. As explained in section 3, we have developed an alternative scheme for extracting the solute spectrum from the IR absorption of the full solution. This method is based on current−current autocorrelation functions (see eqs 2−6). The current is written in terms of atomic APT tensors calculated through linear response theory and atomic velocities. In Figure 7, we compare the IR spectrum of NMA computed in the dipole representation using Wannier centers to the result of the current representation based on atomic APT tensors. Note that the infrared spectrum determined from Wannier centers has been calculated over the whole ab initio trajectory, whereas only 1 ps of the total trajectory was used when the absorption was computed from current fluctuations. This limitation reflects the extra CPU cost of linear response theory for the calculation of the atomic APT tensors, as already mentioned in section 3. We thus expect that the statistics of the current spectrum will be better for the higher wavenumbers (∼above 1300−1400 cm^{-1}) than in the lower part of the spectrum (mostly below 1300 cm^{-1}). For the convenience of the discussion, the spectra of Figure 7 have been normalized to get the same amplitude of the amide I band. For a complete comparison, we have also shown in the same figure the infrared spectrum of NMA calculated with the dipole−dipole correlation function over a slice of 1 ps of the ab initio trajectory.

Despite the very short time span (1 ps) the IR spectrum determined from the current−current autocorrelation function reproduces most of the important features of the absorption. In particular, all amide bands are present, whereas the most intense amide I−amide II band is clearly not correctly obtained from the same 1 ps time interval when the dipole−dipole correlation function is used. This improved convergence is most likely an effect of the favorable statistics of velocities. Atomic velocities, in contrast to dipoles, are isotropic and fluctuate very quickly during the dynamics. Therefore, calculation of infrared spectra through current−current correlation functions can be done on shorter time scales of dynamics. This might be particularly important in the case of strong coupling between almost degenerate modes, such as for example the δ(O−H) bending mode of water and the amide I and amide II bands of NMA which both occupy the same ∼1600 cm^{-1} frequency band. The good agreement between the dipole and current based infrared spectra also validates our scheme for the computation of atomic APT tensors or, conversely, the method for separation of the total dipole moment in solute and solvent contributions.

8. Comparison to Experiment

Experimental infrared spectra are available only for *trans*-NMA, the most stable isomer in both the gas and liquid phase.

8.1. Infrared Spectra in the Gas Phase. Our calculated infrared spectrum of isolated *trans*-NMA can be compared to the experimental spectra obtained in an argon matrix[27] at 20 K and in the gas phase[7] at ∼100 K. The comparisons are displayed in Figure 8a. Note immediately that the two experimental results are in quantitative agreement, except obviously for substantially broader bands in the gas-phase experiment which was performed at higher temperature. Since the BLYP DFT functional is known to systematically underestimate the experimental frequencies, a conventional practice is to apply a constant multiplicative correction factor to the theoretical results. Here we have taken a factor of 1.064 which adjusts exactly the position of the amide III band to the experimental value of 1266 cm^{-1}. Furthermore the theoretical spectrum has also been normalized as to reproduce the correct intensity of this band. Doing so, we find that the theoretical amide I band is close to the correct position, but its intensity is overestimated. We ruled out that this overestimation be related to a nonthermalized amide I mode. Thermalization of all degrees of freedom can indeed be difficult to achieve within short gas-phase dynamics and can thus induce errors in calculated infrared intensities.[82] The intermediate region between 1300 and 1600 cm^{-1} is only approximately reproduced but with correct features. In light of our previous analysis of the ab initio spectrum, the experimental spectrum in that region can be understood as the succession of the ν(C−CH_3)/δ(C−H)$_C$ band at 1360 cm^{-1}, a series of two (gas phase) to four (matrix) low intensity mixed bands due to CH_3 bends, and the more intense amide II band. In the theoretical spectrum, the ν-(C−CH_3)/δ(C−H)$_C$ band is upshifted and slightly too weak,

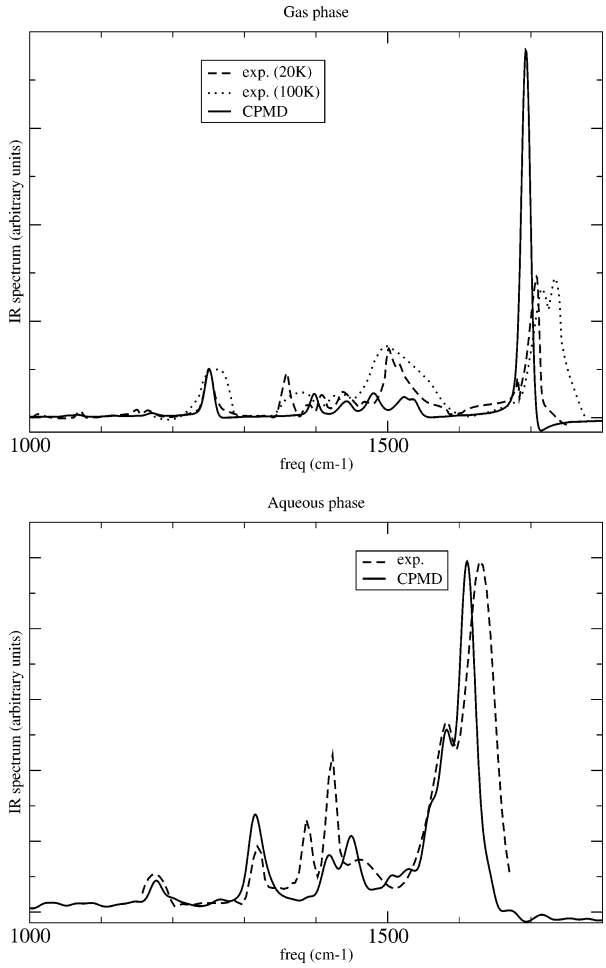


Figure 8. Comparison of the calculated and experimental infrared spectra of *trans*-NMA. Gas phase (top): Car−Parrinello calculation at 20 K (line), experiment in an argon matrix at 20 K[27] (dashed line), experiment in the gas phase at 100 K[7] (dotted line); we note that amplitude of the amide I band in ref 27 has been truncated so that the value is not known. Liquid phase (bottom): Car−Parrinello calculation at 320 K (solid line) and the experiment of refs 32 and 31 at 300 K (dashed line). See text section 8 for the explanation of adjustments made for the sake of comparison between simulation and experiment. See also text for conclusions.

and the amide II band is well positioned but its intensity is definitely too weak. Overall however, bearing in mind the known deficiency of the BLYP functional and the rather simple quantization procedure which was adopted and relies on a picture of coupled harmonic oscillators, it is fair to say that the performance of the ab initio simulations is rather satisfactory, even in the 1300−1600 cm^{-1} spectral range which is quite congested and involves a lot of coupled (anharmonic) motions.

8.2. Infrared Spectra in the Solution. Next we compare the calculated infrared spectrum of aqueous *trans*-NMA to experiments by Song et al.[32,31] and Keiderling et al.[7] All three experiments give very similar results (see Table 5), with subtle differences in band intensities. The experimental spectra presented in refs 31 and 32 have been recorded in the 1000−2000 cm^{-1} spectral domain, while the one of ref 7 is restricted to the 1200−2000 cm^{-1} domain. In Figure 8b, the data of Asher's et al.[31,32] have been used. Again, the computed frequencies have been uniformly scaled by a factor of 1.117 in order to align the amide III band with the experimental peak at ∼1318 cm^{-1}, and we have applied a normalization factor in order to reproduce the intensity for the amide I peak. It can be seen immediately that the overall agreement is very satisfactory with the correct number of bands and correct band shapes, but some discrepancies remain that are worth commenting in the light of our band assignments of Table 5 and Figure 5.

The broad and intense double band in the 1500−1700 cm^{-1} range due to the mixing of amide I and amide II is very well reproduced, except for a slightly lower amide I/amide II gap compared to experiment (54 instead of 60 cm^{-1}). The experimental spectrum exhibits three overlapping bands between 1350 and 1450 cm^{-1} and so does the computed one. In this triplet which was attributed to mostly δ(C−H) bending (coupled to some ν(N−C)), the two first bands appear with the correct relative intensity ratio, but they are slightly too weak and upshifted by ∼40 cm^{-1}. The third one, which in the experimental spectrum contributes to the triplet as a shoulder of lower intensity, is slightly shifted upward to become a shoulder of the amide II peak in the theoretical spectrum. Finally, the computed amide III band is found with a slightly too high intensity, and the lowest frequency band around 1170 cm^{-1}, attributed to ν(N−CH_3)/δ(C−H)$_N$/δ(N−H), is very well reproduced.

We conclude that the comparison of our calculations of the aqueous *trans*-NMA IR absorption to experiments gives overall excellent agreement in terms of the number of active bands, relative positions, and band shapes. The main discrepancy comes from the two twin bands related to δ(C−H) bending, for which one gets a spurious blue-shift and too weak intensities. Note that those two trends were also observed in the theoretical gas-phase spectrum. Comparison of our spectrum to the one simulated by Besley[35] using harmonic frequencies calculated over several thousand optimized structures (ab initio level) of NMA with three explicit water molecules embedded in a dielectric continuum, shows a better agreement to experiment. In particular, we obtain a better agreement on the amplitude of the amide II band, while the methyl bands lack any activity in ref 35. As in our calculated spectrum, the intensity of the amide III band is slightly overestimated in Besley's work.[35] We note that Besley's calculations[35] remain rather more cumbersome (several thousand conformations optimized) and expensive in comparison to our ab initio molecular dynamics. Apart from CH_3 bending, the comparison of our calculated infrared spectrum of aqueous *trans*-NMA with the one calculated by Kato et al.[37] using a classical polarizable force field shows that our quantum molecular dynamics simulations are doing much better. In Kato's work, the general features of the amide I−amide II double bands are well represented, though the amplitude of the amide II band is clearly underestimated. Nonetheless, all other infrared bands in the 1000−1500 cm^{-1} domain below this band are lacking activity. For example, the two peaks related to δ(C−H) bending have no infrared intensity, and the amide III band is barely discernible in the spectrum. Moreover, the weak band showing up in our

calculation and in experiments at ~1098 cm^{-1} is missing in Kato's calculation. Electronic polarization when treated at the ab initio level, as in Car−Parrinello calculations, gives thus a better insight into the infrared patterns. However, a number of problems remain to be resolved, such as underestimation and misplacement of CH_3 infrared bands.

9. Summary and Conclusions

In this paper, we have studied the characteristic vibrational features of *N*-methylacetamide, a simple but essential building block of peptides and proteins, using Car−Parrinello ab initio molecular dynamics simulations. The infrared spectra of *trans*-NMA and *cis*-NMA in the gas and aqueous phase were analyzed and compared in detail. The key computational tool for the calculation of molecular dipoles is the maximally localized Wannier functions. As demonstrated in a previous study of aqueous uracil,[12] and confirmed here, the Wannier functions approach leads to accurate infrared spectra which can be directly compared to experiments. Although band positions are systematically red-shifted (mainly due to our use of the BLYP functional), the main infrared vibrational features, including band shapes and relative bands gaps, can be satisfactorily reproduced. The focus in the present NMA study was on the amide bands. Our results for the relative position and amplitude of the amide I and amide II are in very good agreement with experiments. Amide III also ends up in the right position relative to amide I and II but with a slightly overestimated intensity. It is thus fair to say that our methodology is able to reproduce correctly all amide features in the 1000−2000 cm^{-1}, which is indeed an essential prerequisite to the future study of the vibrational spectroscopy of larger polypeptides and proteins. The amplitudes of methyl groups δ(C−H) bands were underestimated in our calculations. Since C−H/water interactions are more sensitive to dispersion than to electrostatics forces, this deficiency could be related to the lack of a proper dispersion term in DFT calculations.

To compare our calculated infrared spectra to the experiments (gas and liquid phase), we made use of two different scaling factors that adjusted the position of the calculated amide III band (1.064 and 1.117, respectively, for the gas phase and the solution). This point deserves some more comments. There is no real reason as to why the scaling factor between gas phase and solution should be rigorously the same. This scaling factor depends on frequency and thus could change in a condensed phase environment. What could also be different in solution is an enhanced inertia (giving rise to frequency red-shifts) due to the fictitious electron mass used in the Car−Parrinello molecular dynamics scheme. As we have shown in our previous paper,[12] the fictitious electron mass can contribute to the underestimation of the frequencies (up to 40−50 cm^{-1} for the study[12]). Added to the well-known frequency red-shifts due to the use of the BLYP functional, this leads to an increased underestimation of the frequency positions.

An important technical issue that concerned us in this paper was the decomposition of the total spectrum into solute and solvent components. We have used two alternative schemes to achieve such a separation. One way, already mentioned above, is based on the resolution of the cell dipole moment in molecular dipoles computed from maximally localized Wannier functions and the subsequent calculation of the infrared spectrum by Fourier transformation of the solute dipole−dipole correlation function. This strategy is the same as adopted in our previous work.[12] An alternative route, applied here for the first time for a solute in a solvent, involves the representation of the IR signal in terms of a current correlation function and the separation of the current into atomic components. The atomic charges used in the expression of the current were obtained from atomic polar tensors (APT's) calculated using linear response methods. We showed that both definitions yield very similar results. The relative merits of the two methods deserve some comment. A clear drawback of the current based scheme is that the computation of APT tensors is expensive. However, the definition of molecular dipole moments is not unique. The definition of molecular currents is relatively straightforward and not subject to the same arbitrariness as the dipoles. Moreover, such an analysis seems to yield meaningful results with relatively short dynamics, at least shorter than those required with the Wannier decomposition.

Our results show how the activity of the infrared bands of the two NMA isomers is affected by the liquid phase. We were also able to relate these shifts of the amide bands to the solute−solvent intermolecular hydrogen bond network. The study of the Wannier centers and, in the case of *trans*-NMA, of the Born charges were found to provide a useful physical picture of the charge rearrangement in response to solvation. The electrons flows observed in the HNCO peptide link are consistent with an increased weight of the zwitterionic form in a two-state resonance model. This relates the solvent-induced spectral changes to the weakening of the C═O bond and strengthening of the N−C bond. We believe that these observations will be useful for further modeling of the IR spectra (amide bands) of larger peptides in different environment and/or conformations using molecular force fields.

Our simulations also emphasize the importance of the amide III band among the infrared signatures. In particular, the blue-shift of this band in going from the gas to the liquid phase can lead to specific spectral patterns in the IR spectra, as observed for aqueous *cis*-NMA. Interpretation of this band is the least understood of amide bands of peptides and proteins[74−76] and is thought to be highly dependent on the Φ/Ψ angles. We hope that further Car−Parrinello dynamics on more realistic models of peptides will contribute to give a better understanding.

Finally, by comparing the infrared spectra of *trans*-NMA and *cis*-NMA in water determined at the ab initio level (in the DFT sense), we predicted that the two isomers can be clearly distinguished by their IR signatures. In particular the amide I and amide II bands display characteristic differences, both in the gas and aqueous phase. We hope that our findings for NMA can be confirmed experimentally. We also hope that a similar approach can be extended in the future to various helix or β-sheetlike conformations of polypeptides, with the ultimate goal of contributing to the understanding

of how vibrational spectroscopy can help to discriminate protein structures.

Acknowledgment. The authors thank CINES (Montpellier, France) and IDRIS (Orsay, France) for generous access to their computational facilities. The linear response calculations were performed using the HPCx facilities at Daresbury Laboratories as part of the UKCP grant. M.P.G. acknowledges support from Genopole-France through the action 'ATIGE' Action Thématique Incitative de Génopole.

Supporting Information Available: Analysis of the vibrational density of states (VDOS) of *trans*- and *cis*-NMA inferred from the Car−Parrinello trajectories of the isomers in both the gas- and liquid-phase is summarized in Figures 1−4. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Krimm, S.; Bandekar, J. *Adv. Prot. Chem.* **1986**, *38*, 181.

(2) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379.

(3) *Biological Applications of Raman Spectroscopy;* Spiro, T., Ed.; Wiley-Interscience: New York, 1987; Vol. I.

(4) *Infrared Spectroscopy of Biomolecules*; Mantsch, H., Chapman, D., Eds.; Wiley-Liss: New York, 1996.

(5) Kato, K.; Matsui, T.; Tanaka, S. *Appl. Spectrosc.* **1987**, *41*, 861.

(6) Pancoska, P.; Kubelka, J.; Keiderling, T. A. *Appl. Spectrosc.* **1999**, *53*, 655.

(7) Kubelka, J.; Keiderling, T. A. *J. Phys. Chem. A* **2001**, *105*, 10922.

(8) Gnanakaran, S.; Hochstrasser, R. M. *J. Am. Chem. Soc.* **2001**, *123*, 12886.

(9) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123.

(10) Woutersen, S.; Hamm, P. *J. Phys. Chem. B* **2000**, *104*, 11316.

(11) Moran, A. M.; Dreyer, J.; Mukamel, S. *J. Chem. Phys.* **2003**, *118*, 1347.

(12) Gaigeot, M.-P.; Sprik, M. *J. Phys. Chem. B* **2003**, *107*, 10344.

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

(14) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 3572.

(15) Silvestrelli, P. L.; Bernasconi, M.; Parrinello, M. *Chem. Phys. Lett.* **1997**, *277*, 478.

(16) Bernasconi, M.; Silvestrelli, P. L.; Parrinello, M. *Phys. Rev. Lett.* **1998**, *81*, 1235.

(17) Putrino, A.; Sebastiani, D.; Parrinello, M. *J. Chem. Phys.* **2000**, *113*, 7102.

(18) Polavarapu, P. L.; Deng, Z.; Ewig, C. S. *J. Phys. Chem.* **1994**, *98*, 9919.

(19) Mirkin, N. G.; Krimm, S. *J. Am. Chem. Soc.* **1991**, *113*, 9742.

(20) Mirkin, N. G.; Krimm, S. *J. Mol. Struct. (THEOCHEM)* **1991**, *236*, 97.

(21) Cuevas, G.; Renugopalakrishnan, V.; Madrid, G.; Hagler, A. T. *Phys. Chem. Chem. Phys.* **2002**, *4*, 1490.

(22) Gregurik, S. K.; Chaban, G. M.; Gerber, R. B. *J. Phys. Chem. A* **2002**, *106*, 8696.

(23) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182.

(24) Garcia-Martinez, A.; Teso-Vilar, E.; Garcia-Fraile, A.; Martinez-Ruiz, P. *J. Phys. Chem. A* **2002**, *106*, 4942.

(25) Nandini, G.; Sathyanarayana, D. N. *J. Mol. Struct. (THEOCHEM)* **2002**, *579*, 1.

(26) Han, W. G.; Suhai, S. *J. Phys. Chem.* **1996**, *100*, 3942.

(27) Ataka, S.; Takeuchi, H.; Tasumi, M. *J. Mol. Struct.* **1984**, *113*, 147.

(28) Fillaux, F.; De Loze, C. *J. Chim. Phys.* **1976**, *73*, 1004.

(29) Grenie, Y.; Avignon, M.; Garrigou-Lagrange, C. *J. Mol. Struct.* **1975**, *24*, 293.

(30) Chen, X. G.; Schweitzer-Stenner, R.; Krimm, S.; Mirkin, N. G.; Asher, S. A. *J. Am. Chem. Soc.* **1994**, *116*, 11141.

(31) Song, S. S.; Asher, S. A.; Krimm, S.; Bandekar, J. *J. Am. Chem. Soc.* **1988**, *110*, 8547.

(32) Song, S. S.; Asher, S. A.; Krimm, S.; Shaw, K. D. *J. Am. Chem. Soc.* **1991**, *113*, 1155.

(33) Herrebout, W. A.; Clou, K.; Desseyn, H. O. *J. Phys. Chem. A* **2001**, *105*, 4865.

(34) Chen, X. G.; Schweitzer-Stenner, R.; Asher, S. A.; Mirkin, N. G.; Krimm, S. *J. Phys. Chem.* **1995**, *99*, 3074.

(35) Besley, N. A. *J. Phys. Chem. A* **2004**, *108*, 10794.

(36) MacKerrel, A. D.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.

(37) Iuchi, S.; Morita, A.; Kato, S. *J. Phys. Chem. B* **2002**, *106*, 3466.

(38) Mantz, Y. A.; Gerard, H.; Iftimie, R.; Martyna, G. J. *J. Am. Chem. Soc.* **2004**, *126*, 4080.

(39) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098.

(40) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(41) Trouillier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993.

(42) Kleinman, L.; Bylander, D. M. *Phys. Rev. Lett.* **1982**, *48*, 1425.

(43) Vuilleumier, R.; Sprik, M. *J. Chem. Phys.* **2001**, *115*, 3454.

(44) Hutter, J. et al. *CPMD, version 3.7.1*. IBM Research Division, IBM Corp and Max Planck Institute Stuttgart.

(45) McQuarrie, D. A. *Statistical Mechanics*; Harper-Collins Publishers: New York, 1976.

(46) King-Smith, R. D.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 1651.

(47) Vanderbilt, D.; King-Smith, R. D. *Phys. Rev. B* **1993**, *48*, 4442.

(48) Marzari, N.; Vanderbilt, D. *Phys. Rev. B* **1997**, *56*, 12847.

(49) Cioslowski, J. *J. Am. Chem. Soc.* **1989**, *111*, 8333.

(50) Gonze, X. *Phys. Rev. B* **1997**, *55*, 10337.

(51) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.

(52) De Proft, F.; Martin, J. M. L.; Geerlings, P. *Chem. Phys. Lett.* **1996**, *250*, 393.

(53) Pasquarello, A.; Car, R. *Phys. Rev. Lett.* **1997**, *79*, 1766.

(54) Pasquarello, A.; Resta, R. *Phys. Rev. B* **2003**, *68*, 174302.

(55) Putrino, A.; Sebastiani, D.; Parrinello, M. *J. Chem. Phys.* **2000**, *113*, 7102.

(56) Resta, R. *Phys. Rev. Lett.* **1998**, *80*, 1800.

(57) Silvestrelli, P. L.; Parrinello, M. *Phys. Rev. Lett.* **1999**, *82*, 3308.

(58) Blondel, A.; Karplus, M. J. *J. Comput. Chem.* **1996**, *17*, 1132.

(59) Kitano, M.; Fukuyama, T.; Kuchitsu, K. *Bull. Chem. Soc. Jpn.* **1973**, *46*, 384.

(60) Du, Q.; Wei, D. *J. Phys. Chem. B* **2003**, *107*, 13463.

(61) Gaigeot, M.-P.; Sprik, M. *J. Phys. Chem. B* **2004**, *108*, 7458.

(62) Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **1995**, *103*, 150.

(63) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys.* **1999**, *111*, 3572.

(64) Izvekov, I.; Voth, G. A. *J. Chem. Phys.* **2002**, *116*, 10372.

(65) Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *120*, 300.

(66) VandeVondele, J.; Mohamed, F.; Krack, M.; Hutter, J.; Sprik, M.; Parrinello, M. *J. Chem. Phys.* **2005**, *122*, 014515.

(67) Torii, H.; Tatsumi, T.; Tasumi, M. *J. Raman Spec.* **1998**, *29*, 537.

(68) Rodrigo, M. M.; Tarazona, M. P.; Saiz, E. *J. Phys. Chem.* **1986**, *90*, 2236.

(69) Pralat, K.; Jadzyn, J.; Balanicka, S. *J. Phys. Chem.* **1983**, *87*, 1385.

(70) Wang, Y.; Purrello, R.; Jordan, T.; Spiro, T. G. *J. Am. Chem. Soc.* **1991**, *113*, 6359.

(71) Mayne, L. C.; Hudson, B. *J. Phys. Chem.* **1991**, *95*, 2962.

(72) Miyazawa, T.; Shimanouchi, T.; Mizushima, S. I. *J. Chem. Phys.* **1956**, *24*, 408.

(73) Mirkin, N. G.; Krimm, S. *J. Mol. Struct.* **1996**, *377*, 219.

(74) Mikhonin, A. V.; Ahmed, Z.; Ianoul, A.; Asher, S. A. *J. Phys. Chem. B* **2004**, *108*, 19020.

(75) Mirkin, N. G.; Krimm, S. *J. Phys. Chem. A* **2002**, *106*, 3391.

(76) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K.; Mroz, P. A.; Kozlowski, P. M. *J. Phys. Chem. B* **2002**, *106*, 4294.

(77) Ramirez, R.; Lopez-Ciudad, T.; Kumar, P.; Marx, D. *J. Chem. Phys.* **2004**, *121*, 3973.

(78) Person, W. B.; Newton, J. H. *J. Chem. Phys.* **1974**, *61*, 1040.

(79) Person, W. B. In *Vibrational Intensities In Infrared and Raman Spectroscopy;* Person, W. B., Zerbi, G., Eds.; Elsevier: 1982.

(80) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 6142.

(81) Kuo, I.-F. W.; Mundy, C.; McGrath, M. J.; Siepmann, J. I.; VandeVondele, J.; Sprik, M.; Hutter, J.; Chen, B.; Klein, M. L.; Mohamed, F.; Krack, M.; Parrinello, M. *J. Phys. Chem. B* **2004**, *108*, 12990.

(82) Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, *121*, 12233.

CT050029Z

JCTC Journal of Chemical Theory and Computation

Exploring the Essential Dynamics of B-DNA

Alberto Pérez,[†,‡] José Ramón Blas,[†] Manuel Rueda,[†,§] Jose María López-Bes,[‡] Xavier de la Cruz,[†,||] and Modesto Orozco*[,†,§,⊥]

Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain, Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda Diagonal 643, Barcelona 08028, Spain, Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain, Institució Catalana per la Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, 08018 Barcelona, Spain, and Structure and Modeling Node, Instituto Nacional de Bioinformática, Spain

Received March 2, 2005

Abstract: The essential dynamics of different normal and chemically modified DNA duplexes pertaining to the B family have been extensively explored from molecular dynamics simulations using powerful data mining techniques. Some of them, which are presented here for the first time, might become standard, powerful tools to characterize the dynamical behavior of complex biomolecular structures such as nucleic acids. Their potential impact is illustrated by examining the extended trajectories sampled for the set of DNA duplexes considered in this study, which allows us to discuss the degree of conservation of the natural flexibility pattern of the different DNAs, which in specific cases contain severe chemical modifications.

Introduction

Though the first molecular dynamics (MD) simulation of a nucleic acid (NA) was done in the 1980s,[1,2] technical problems limited the systematic use of this powerful simulation technique to explore NAs in solution until the middle of the 1990s. The implementation of methods to treat long-range electrostatic forces, such as the Particle Mesh Ewald method,[3] and the development of more accurate force fields[4−8] are responsible for the huge explosion of MD simulations of NAs during the past decade. Current state-of-the-art simulations cover several nanoseconds of fully solvated 10−14-mer duplexes,[9−13] and simulations approaching the 0.1 μs range have been published by different groups (for example, see refs 14−16), showing that reasonable convergence can be obtained for many structural and enegetic properties in the sub-microsecond simulation time, at least for normal B-type DNA duplexes. Very recently, the Ascona B-DNA Consortium (ABC) published its first conclusions of B-DNA structure and flexibility obtained from the analysis of all of the 136 unique tetranucleotide sequences of DNA duplexes.[17] Clearly, the field of MD simulations of NAs has reached maturity and well-defined standards for simulation conditions and protocols are being established. Unfortunately, the capacity to run longer and more stable trajectories has not been accompanied by a similar increase in our ability to extract all the information included in such trajectories, and a large amount of useful information present in the trajectory is escaping our analysis.

The study of MD trajectories has traditionally pursued the verification of whether the trajectory was converged and the characterization of the average structural and energetic details of the molecular system.[9−13] Convergence has been typically checked by looking at the time evolution of the positional root-mean-square deviation (rmsd), or alternative global geometrical properties. The average structural features of molecules have been analyzed with a variety of geometrical

* Corresponding author e-mail: modesto@mmb.pcb.ub.es.
† Institut de Recerca Biomèdica.
‡ Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona.
§ Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona.
|| Institució Catalana per la Recerca i Estudis Avançats (ICREA).
⊥ Instituto Nacional de Bioinformática.

10.1021/ct050051s CCC: $30.25
Published on Web 06/30/2005

measurements, such as the helical parameters for NAs. Interaction properties have been explored by integrating the solvent population around the molecule, or by computing interaction potentials. Dynamical properties of the solvent distribution around the solute can be characterized by parameters such as the maximum or average residence time.[14−18] Recently, the ensemble of structures collected in the MD simulation has been used to estimate energetic properties by combining the average intramolecular energy with estimates of the solvation free energy derived from continuum[12,13,19−23] or discrete[24,25] linear-response calculations. This strategy has been very powerful in determining the relative stability of different DNA (or RNA) helical conformations,[12,23,26] but some caution is needed when used to discriminate between states very close in energy.

Convergence problems precluded, for many years, a careful analysis on the dynamical properties of macromolecules. However, it seems that even short (2−20 ns) simulations, which can be obtained quite easily with current computational resources, can be enough to capture the essential dynamics of small polymers such as B-DNA duplexes,[14−16] which opens the possibility of using MD as a method to reproduce NA flexibility in physiological conditions. In this paper, we want to examine this point in depth. For this purpose, we will first discuss several approaches to describe the dynamical behavior of NAs, and particularly new measurement techniques will be presented. Second, the potential impact of these techniques will be used to analyze the essential deformation pathways of DNA duplexes by analyzing a local library of MD trajectories collected for normal, mutated, and chemically modified DNA duplexes.

Methodological Approach

The dynamic analysis of trajectories is a typical data mining problem, where the information of interest is buried in many gigabytes of noise. Particularly, the trajectory contains a large amount of fast and irrelevant movements, which hide the soft deformation modes that are important to understand the flexibility of DNA. Most approaches to extract information from the trajectory are based on the construction of covariance matrices. For DNA simulations, such matrices can be built considering all or part of the system and using three different coordinate models: (i) Cartesian coordinates ($\mathbf{C}_x$), (ii) mass-weigthed Cartesian coordinates ($\mathbf{C}_m$), and (iii) helical coordinates ($\mathbf{C}_h$). We will summarize, in the following, how to process these covariance matrices to derive dynamic information on NAs.

Essential Dynamics. Following the principal component analysis (PCA) method, diagonalization of the Cartesian covariance matrix ($\mathbf{C}_x$) yields a set of $3N - 6$ eigenvectors (where N is the number of atoms used to define $\mathbf{C}_x$) and their associated eigenvalues. The eigenvectors display a series of interesting properties: (i) they explain all the variance of the system, (ii) they are orthogonal, and (iii) the fraction of the total variance of the trajectory explained by a given eigenvector is given by the magnitude of its eigenvalue. Thus, the diagonalization of $\mathbf{C}_x$ allows us to characterize the most important deformation modes and to evaluate their relative weight in the global flexibility of the molecule.[12,13,27−32] PCA is especially powerful in the study of DNA trajectories, since this biopolymer has quite simple dynamics (compared to proteins), which can be represented by a small number of "essential movements" (see Results section).

The analysis of the eigenvalues obtained by diagonalization of $\mathbf{C}_x$ provides very direct information on the flexibility of DNA. For example, they can be used to derive harmonic force constants associated with the essential deformations of DNA (see eq 1; refs 12,13,30−32).

$$K_i = k_B T/\lambda_i \tag{1}$$

where k_B is Boltzmann's constant (in kcal/mol K), T is the absolute temperature, and λ_i is the eigenvalue associated with the essential movement i (in Å^2).

Note that once the force constant is known, the deformation energy along the essential mode i can be easily determined from eq 2.

$$E_i = \frac{K_i}{2}(\Delta X_i)^2 \tag{2}$$

where ΔX_i is a Cartesian deformation along the eigenvector i.

Assuming that the dynamical behavior of two molecules can be described by means of a limited set of t-harmonic oscillators, the accessible volume for a system at temperature T can be defined as shown in see eq 3a (see refs 31−33). Note that the practical use of the configurational volume (Vol) as a descriptor of flexibility faces a serious problem: Vol increases rapidly with the number of eigenvectors until a maximum is found; the introduction of additional eigenvectors (with $\lambda < 1$) decreases the volume, and a value of zero is reached when t becomes close to m. This behavior reflects the fact that the real dimensionality of the DNA is not equal to the total number of eigenvectors and that the stiffer eigenvectors do not contribute significantly to an explanation of trajectory variance. Several approaches can be used to overcome the intrinsic limitations of the volume: (i) the definition of an arbitrary number of important eigenvectors (t) for the systems studied, (ii) the use of an ad-hoc shifting value (see eq 3b) as, in fact, is performed in the quasiharmonic analysis of entropy (see below), and (iii) the use of the number of eigenvectors for which the maximum of Vol is obtained as an estimate of the dimensionality of the configurational space. Note that the dimensionality will roughly correspond to the order number of the eigenvector whose eigenvalue is closer to 1 (see eq 3a) in a given unit measuring system.

$$\mathrm{Vol} = \prod_{i=1}^{t}(\lambda_i)^{1/2} \tag{3a}$$

$$\mathrm{Vol}_g = \prod_{i=1}^{t}(1 + \lambda_i)^{1/2} \tag{3b}$$

where, here, t is the number of eigenvalues considered.

The selection of a common threshold for the definition of Vol might be acceptable when very similar molecules are compared, but for this study, it is not recommended. The

generalized volume shown in eq 3b provides information somehow redundant to that derived from quasiharmonic entropy measures (see below). Thus, in this paper, we decide to use the dimensionality as a measure of the configurational volume accessible to the DNA. We should note that the concept of dimensionality is close to physical intuition, since it can be interpreted as the limit above which the addition of additional dimensions does not improve our representation of the systems. For example, a sheet of paper is well represented by just two dimension, and the third one is irrelevant [i.e., the three-dimensional volume of a sheet of paper is close to zero, whereas its two-dimensional volume (surface) is not]. Note, however, that the value of the dimensionality depends on the units used to measure variance. In our case, we consider that the first irrelevant mode is that along which, at temperature T, the expected displacement of the molecule is less than 1 Å (a classical heteroatom−hydrogen distance).

The analysis of the eigenvectors projected on the Cartesian space gives very useful information on the nature of the essential movements of the DNA. Unfortunately, manipulation of this information is, again, difficult because of its high dimensionality, which forces us to focus on a reduced set of "important" n eigenvectors ($\{\mu_i\}_0$), defined as those needed to explain a certain degree of variance (for example, 80 or 90%, which, for normal DNA dodecamers, implies around 10 modes). Then, we can quantify the similarity in the pattern of deformation sampled in two trajectories (A and B) by comparing the corresponding sets of eigenvectors ($\{\mu\}_{i0}^{A}$ and $\{\mu_i\}_0^{B}$). This can be done by using eq 4,[12,13,30−32,34−36] which implicitly assumes that all the n eigenvectors have the same weight in defining the variance of the trajectory.

$$\gamma_{AB} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}(\nu_i^{A}\nu_j^{B})^2 \tag{4}$$

where n is the minimum number of eigenvectors needed to define the set of "important" eigenvectors $\{\mu_{i}\}_0^{A}$ and $\{\mu_i\}_0^{B}$ and where ν_i^X stands for the i-unitary eigenvector of trajectory X (i.e., $\nu_i^X = \mu_i^X/|\mu_i^X|$). Note that γ_{AB} takes values from 0 (null similarity) to 1 (identical essential movements).

The absolute similarity index γ_{AB} is useful to verify[12,13,30−32,34−36] the convergence of a single trajectory by defining A and B as two distant portions of the same trajectory. Indeed, it can be modified (see eq 5) to determine how well the essential dynamics of a given molecule follow a given conformational transition.

$$\gamma_A^R = \sum_{i=1}^{n}(\nu_i^{A}R)^2 \tag{5}$$

where R is the unitary transition vector defining the transition between two conformations.

The most powerful use of γ_{AB}, however, is as a measure of the similarity between the essential dynamics of two molecules for which Cartesian covariance matrices of the same dimensionality can be created (this might imply, in some cases, a restriction of the number of atoms considered in the calculation of $\mathbf{C}_x$). As noted elsewhere,[12,13] eq 4 assumes that (i) trajectories are very long and (ii) a small number of eigenvectors are included in $\{\mu_i\}_0^{A}$ and $\{\mu_i\}_0^{B}$. The second requirement is not a real problem for NAs (see above), but clearly, current MD simulations are too short and part of the dissimilarities between the trajectories of two molecules can be simply due to the limited configurational sampling. Thus, it is often convenient to use relative similarity indexes (see eqs 6 or 7), where the noise in the original trajectories is largely canceled by normalizing cross similarities ($X-Y$) with self similarities ($X-X$ and $Y-Y$).

$$\kappa_{AB} = 2\frac{\gamma_{AB}}{(\gamma_{AA}^T + \gamma_{BB}^T)} \tag{6}$$

$$\kappa_{AB} = \frac{\gamma_{AB}}{(\gamma_{AA}^T\,\gamma_{BB}^T)^{1/2}} \tag{7}$$

where γ_{XX}^T is the absolute self-similarity index (eq 4) for trajectory X obtained by comparing the first and second halves of the trajectory.

As noted above, eqs 4−7 implicitly assume that all eigenvectors in the "important space" have the same weight in the dynamical behavior of the molecule. This means that, for the similarity index, there is a good overlap between the first eigenvectors (which might explain 40% of the variance) that is equivalent with that of higher modes, which might explain no variance. Furthermore, the method is insensitive to permutations in the ordering of the eigenvectors, which is quite counterintuitive since the same similarity index will be found if the first eigenvector of trajectory A matches the first of trajectory B, or if the matching happens between eigenvectors 1(A) and 1000(B).

Clearly, the use of a small set of "important eigenvectors" in the comparison ($n \leq 10$; see above) reduces but does not eliminate this intrinsic problem of indexes γ and κ, since cases can exist where the ordering of eigenvectors or the associated eigenvalues are quite different in the two trajectories. Here, we propose a new strategy to solve this problem. The method starts with the assumption that the molecule moves sampling states defined by a common displacement (Δx) along the different eigenvectors. Thus, the weight of each eigenvector in defining the flexibility space will be given by its Boltzman factor computed from the harmonic energy penalty given by eq 2 (see Figure 1 and eq 9). Within this approach, the relative importance of two eigenvectors (i and j) in defining the flexibility space is given by the ratio between their respective Boltzmann's factors (see eq 9). Accordingly, the relative importance of an eigenvector in the set of essential movements is given by eq 9, where the index z runs under either the "important" ($z = n$) or entire ($z = m$) set of eigenvectors.

$$w_i = \frac{e^{-(\Delta x)^2/\lambda_i}}{\sum_{j=1}^{z} e^{-(\Delta x)^2/\lambda_j}} \tag{8}$$

$$W_{ij} = \frac{e^{-(\Delta x)^2/\lambda_i}}{e^{-(\Delta x)^2/\lambda_j}} \tag{9}$$

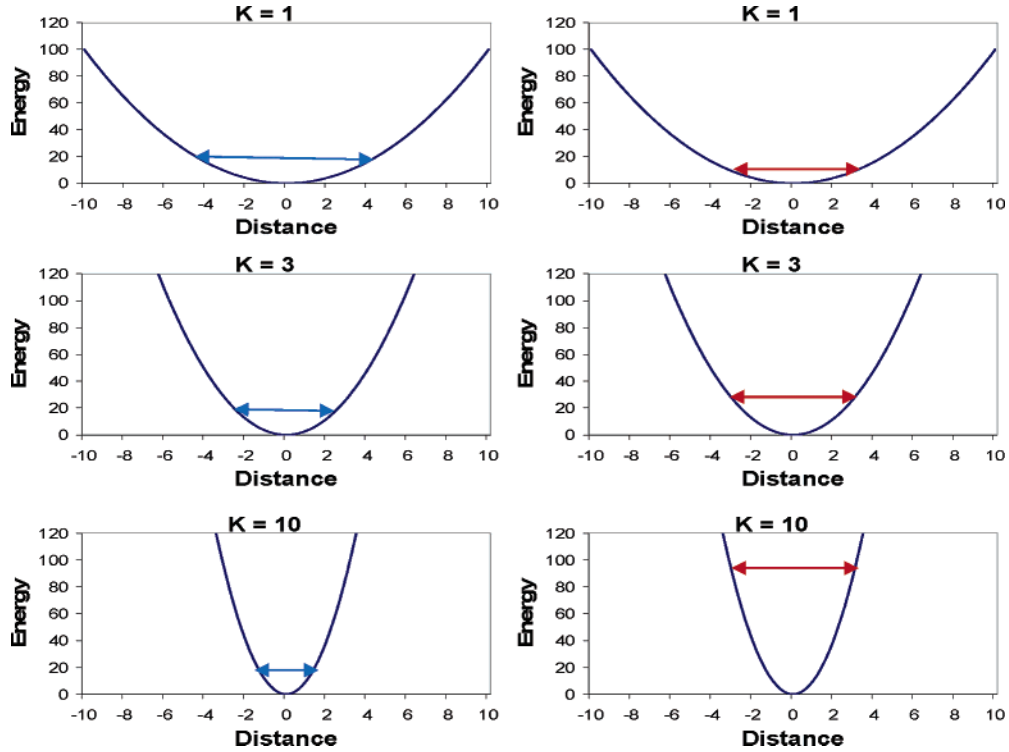


Figure 1. Scheme of the movements along three eigenvectors (softer to stiffer from top to bottom). Panels at the left indicate the sampling obtained for a common energy (equipartition principle). Panels at the right indicate the different energy levels needed to sample each node for a common displacement.

When comparing two sets of eigenvectors $\{\mu_i\}^A$ and $\{\mu_i\}^B$, the contribution to the global similarity of two individual eigenvectors μ_i^A and μ_j^B is assumed to be equal to the product of the relative probabilities of each eigenvector. Then, the extension of this concept to all (or the "important" set of) eigenvectors and the subsequent normalization considering the energy distribution of the sets of eigenventors- $\{\mu_i\}^A$ and $\{\mu_i\}^B$ yields to eq 10, which is equivalent to eq 4, but sensitive to the relative importance of the different eigenvectors in explaining trajectory variance.

$$\xi_{AB} = \frac{2\sum_{i=1}^{i=z}\sum_{j=1}^{j=z}\left\{(\nu_i^A\nu_j^B)\dfrac{\exp\left[-\dfrac{(\Delta x)^2}{\lambda_i^A}-\dfrac{(\Delta x)^2}{\lambda_j^B}\right]}{\sum_{i=1}^{i=z}\exp\left[-\dfrac{(\Delta x)^2}{\lambda_i^A}\right]\sum_{j=1}^{j=z}\exp\left[-\dfrac{(\Delta x)^2}{\lambda_j^B}\right]}\right\}^2}{\sum_{i=1}^{i=z}\left\{\dfrac{\exp\left[-2\dfrac{(\Delta x)^2}{\lambda_i^A}\right]}{\left(\sum_{i=1}^{i=z}\exp\left[-\dfrac{(\Delta x)^2}{\lambda_i^A}\right]\right)^2}\right\}^2+\sum_{j=1}^{j=z}\left\{\dfrac{\exp\left[-2\dfrac{(\Delta x)^2}{\lambda_j^B}\right]}{\left(\sum_{j=1}^{j=z}\exp\left[-\dfrac{(\Delta x)^2}{\lambda_j^B}\right]\right)^2}\right\}^2} \quad (10)$$

where λ_i is the eigenvalue (in Å^2) associated with eigenvector μ_i, whose unitary vector is ν_i. The sum can be extended to all ($z = m$) or the "important set" ($z = n$) of the eigenvectors.

By analogy to the κ_{AB} index, the new absolute index ξ_{AB} can be manipulated to obtain relative similarity measures (δ_{AB}; see eq 11). The derivation of transition indexes equivalent to eq 5 is also straightforward.

$$\delta_{AB} = 2\frac{\xi_{AB}}{(\xi_{AA}^T + \xi_{BB}^T)} \quad (11)$$

Note that a key issue in the calculation of eq 10 is the selection of the common displacement (Δx). Large values of Δx will make it so that only the first modes will have importance in the comparison (see eqs 8 and 9), whereas small values will increase the weight in the comparison of higher essential modes, some of them with small impact in the molecule variance [note that for $\Delta x = 0$, eqs 10 and 11 converge to eqs 4 and 6, which means that all the modes (in the important space) will have the same impact in defining the similarity index]. As a compromise, the common distortion level (Δx) can be defined (i) by determining the number of eigenvectors (n') required to capture a given degree (for example 90%) of the variance and (ii) determine the smallest Δx for which the contribution of the $n' + 1$ mode to the similarity index is less than 1%. This process guarantees that no weight will be given to irrelevant modes and that, within the "important space", all eigenvectors will have some impact (depending on their relative Boltzman's factors) in the determination of the similarity function.

The different behavior of the two similarity indexes can be illustrated in the comparison of the essential movements of the central eight-mer portion of two duplexes of different lengths but identical sequence: $d(A)_{11}$ and $d(A)_{15}$ (trajectories

Table 1. DNA Duplexes Studied by MD Simulations[a]

code	sequence
trajectory 1	*d*(GATTAATTAATTAATC)
trajectory 2	*d*(GATTAATT***A***ATTAATC)
trajectory 3	*d*(GGCCGGCCGGCCGGCC)
trajectory 4	*d*(GGCCGGCC***G***GCCGGCC)
trajectory 5	*d*(AAAAAAAAAAAAAAAA)
trajectory 6	*d*(CGCGAATTCGCG)
trajectory 7	*d*(CTTTTC***F***TTCTT)
trajectory 8	*d*(CTTTTCTTTCTT)
trajectory 9	*d*(CTTTTC***T***TTCTT)
trajectory 10	*d*(CTTTTC***T***TTCTT)
trajectory 11	*d*(AAAAAAAAAAAA)
trajectory 12	*d*(GAAGGAGGAGA)
trajectory 13	*d*(CCAAGCTTGG)
trajectory 14	*d*(CCAA***G***CTTGG)

[a] Bases in bold correspond to special cases: trajectories 2 and 4 have *S*-methylphosphonate substituting the charged group after the marked base; 7, 8, and 9 contain apolar surrogates of bases (T, Q, and Z respectively; see ref 51). Finally, trajectory 14 contains thymine paired to the marked guanine.

Table 2. Example of the Behavior of Absolute Similarity Indexes (γ and ξ; See Eqs 5 and 10) in the Comparison of the Essential Dynamics of the Eight-mer Portion of Trajectories 5 and 11 in Normal Conditions and Forcing Some Eigenvector Rotations (See Text for Details)

	γ	ξ
normal	0.77	0.87
1↔10 rotation (traj11)	0.77	0.63
1↔10 rotation (traj5)	0.77	0.64
1↔10 rotation (both)	0.77	0.62
1↔456 rotation (both)	0.66	0.53
10↔456 rotation (both)	0.66	0.86

5 and 11, see Table 1). By defining an important space of 10 eigenvectors, absolute similarity indexes obtained from the indexes γ (eq 5) and ξ (eq 10) are large, which shows the similarity in the nature of the movement of both duplexes (see Table 2). When eigenvectors are permuted in the essential space, that is, by interchanging the position of vectors 1 and 10, the γ index remains unaltered whereas the ξ index detects the smaller similarity between the two trajectories (see Table 2). The substitution of the first eigenvector by an irrelevant one (taken here as the 456th eigenvector) decreases the γ index at a level similar to that obtained when such a replacement affects the 10th eigenvector despite the difference in importance of these two eigenvectors (see Table 2). On the contrary, the new ξ index properly captures the difference in impact of the substitution of the 1st and 10th eigenvectors (see Table 2). In summary, although, for normal cases, the behavior of both similarity indexes is similar, the new similarity ξ index is more powerful in analyzing anomalous situations.

As noted above, the new indexes (10 and 11) are dependent on the common displacement (Δx). However, in practical use, for the series of the trajectories considered here, where no large permutation of eigenvectors exists, the dependence of the new indexes on Δx is very small, as seen for selected cases in Table 3. Note that, in general, the

Table 3. Absolute Similarity Index (ξ; Eq 10) between Trajectories 1−4 for Different Values of the Common Displacement (Δx)[a]

	traj1	traj2	traj3	traj4
traj1		0.73	0.65	0.69
		0.75	0.67	0.71
		0.82	0.73	0.77
		0.83	0.74	0.77
traj2			0.63	0.65
			0.66	0.68
			0.72	0.74
			0.74	0.75
traj3				0.69
				0.70
				0.74
				0.72

[a] Values in each cell correspond, top to bottom, to Δx= 0.0, 1.0, 5.0, and 15.0 Å.

similarity indexes increase when the displacements do until large Δx values are reached (for which only the first mode contributes to the similarity), indicating that, for the series of trajectories studied, the first eigenvectors are the better conserved (see Results section).

Entropy Calculation. The fluctuations in the root-mean square deviation or mass-weighted root-mean-square deviation (rmsd or rmsd_w) along the trajectory have been traditionally used as an indicator of the flexibility of a molecule. The rmsd is defined as the average deviation of a conformation with respect to a reference structure (see eq 12).

$$\text{rmsd}_k = [1/N \sum_{l=1}^{3N}(x_{kl} - x_l)^2]^{1/2} \tag{12}$$

where x_{kl} stands for the l coordinate in structure k and x_l is the value of the l coordinate in the reference structure.

The average deviation ($\langle \text{rmsd}^2 \rangle_t$) can be defined in terms of the time-averaged position of each Cartesian or mass-weigthed coordinate ($\langle x_l \rangle_t$; see eq 13). It can be easily shown that the average deviation is equal to the variance of the trajectory and the deviation between the average and the reference structures (see eq 14). Note that the variance of the trajectory equals the sum of the eigenvalues, which for most covariance matrices is closely related to the product of the eigenvalues, which as shown below is closely related to the entropy of the system. The connection between rmsd fluctuation and entropy that has been implicitly assumed in many MD studies is then clear.

$$\langle \text{rmsd}^2 \rangle_t = \frac{1}{Z}\sum_{k=1}^{Z}\left[\frac{1}{N}\sum_{l=1}^{3N}(x_{kl} - x_l)^2\right] = \frac{1}{N}\sum_{l=1}^{3N}\left[\frac{1}{Z}\sum_{k=1}^{Z}(x_{kl} - \langle x_l \rangle_t - x_l + \langle x_l \rangle_t)^2\right] \tag{13}$$

$$\langle \text{rmsd}^2 \rangle_t = \text{rmsd}_{\text{ref}}{}^2 + \frac{1}{N}\sum_{l=1}^{3N}\frac{1}{Z}\sum_{k=1}^{Z}(x_{kl} - \langle x_l \rangle_t)^2 = \text{rmsd}_{\text{ref}}{}^2 + \frac{1}{N}\sum_{l=1}^{3N}\text{var}_l \tag{14}$$

where Z stands for the number of snapshots collected during the trajectory.

The conformational freedom of a molecule is then qualitatively captured in the oscillation of the mean square deviation, in the generalized configurational volume, and in the dimensionality of the space (see Methods section). However, the best thermodynamic property to describe the conformational freedom is entropy. In principle, for any system, entropy can be computed using basic thermodynamic principles (see eq 15). However, in general, the direct use of eq 15 is not advisable since the definition of microstates is difficult and somehow arbitrary for DNA and convergence problems might be severe.

$$S \propto \sum_k P_k \log_2 P_k \tag{15}$$

where P_k is the probability of microstate k and the sum extends to all the possible conformational space.

More useful in DNA simulations is the evaluation of molecular entropy from the harmonic oscillator model using either Schlitter's (ref 37; see eq 16) or Andricioaei and Karplus' (ref 38; see eq 17) methods. The two approaches rely on the same principles: the quantum oscillator and the concept of "generalized configurational volume". Both require the diagonalization of the mass-weighted covariance matrix to obtain the frequencies (derived directly from the eigenvalues) associated with the essential deformations. Note that the sum of the eigenvalues of a $\mathbf{C}_x$ (or $\mathbf{C}_w$) in displacement (or mass-weighted displacement) units equals the variance, connecting, then, the concepts of entropy and rmsd fluctuations (see above).

In practice, Schlitter's and Andricioaei and Karplus' methods provide very close results and share the same intrinsic shortcomings derived from (i) the assumption that the molecule moves from the equilibrium geometry according to harmonic modes and (ii) the time dependence of the entropy estimate. Thus, both methods can be used in trajectories near the equilibrium but not in those that sample irreversible conformational transitions. The time dependence of both methods is intrinsic to any simulation-based estimate of entropy, since as the length of the trajectory increases the number of microstates visited and, then, the entropy become larger. The time-dependence problem can be alleviated by using an exponential correction formula developed by Harris et al. (see eq 18 and ref 39), which using partial entropy estimates obtained for different simulation time windows (snapshots always collected every 1 ps) allows us to estimate the entropy at infinite simulation time.

$$S \approx 0.5k\sum_i \ln\left(1 + \frac{e^2}{\alpha^2}\right) \tag{16}$$

$$S = k\sum_i \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \tag{17}$$

where $\alpha_i = \hbar\omega_i/kT$, ω being the eigenvalues obtained by diagonalization of the mass-weighted covariance matrix ($\mathbf{C}_w$), and the sum extends to all the nontrivial vibrations.

$$S(t) = S_\infty - \frac{\alpha}{t^\beta} \tag{18}$$

where α and β are fitted parameters and t is the simulation time (in picoseconds) used to obtain the entropy estimate.

Helical Stiffness. Standard double helical NAs such as physiological DNA are often represented by means of helical coordinates,[40−42] which implies a moderate loss of information, but offers two major advantages: (i) it is very close to chemical intuition, and (ii) it dramatically simplifies the definition of the DNA conformational space. Thus, for canonical base pairs, the conformational space of DNA can be reasonably represented by only $3(K - 1)$ rotational and $3(K - 1)$ translational degrees of freedom, where K is the number of base pair steps in the duplex. For large and regular polymers, the coordinate systems can be even more simplified by using a reduced set of global polymer parameters (like global twist, bending, or stretch; see ref 43).

Covariance matrices defined in the helical reference system ($\mathbf{C}_h$) can be manipulated to determine the flexibility of DNA with respect to perturbations along helical coordinates. As shown by Olson and co-workers[44,45] and Lankas and co-workers,[43,46] this can be obtained by inversion of $\mathbf{C}_h$, which yields (see eq 19) a stiffness matrix $\mathbf{F} = F(K_{ij})$, whose diagonal elements represent contributions due to deformations arising from pure helical variables, whereas off-diagonal components account for coupling terms. Once the stiffness matrix is known, the mesoscopic calculation of the elastic energy of a DNA helix can be easily done even for very large DNA fragments (see eq 20).

$$\mathbf{F} = kT\mathbf{C}_h^{-1} \tag{19}$$

$$E = \sum_a \frac{K_a}{2}(X_a - X_a^0)^2 + \sum_{a,b} \frac{K_{ab}}{2}(X_a - X_a^0)(X_b - X_b^0) \tag{20}$$

where X_a and X_b stand for two different helical coordinates ($a \neq b$), K_a stands for a diagonal force constant (for variable a), and K_{ab} represents out-of-diagonal force constants.

Note that the analyses of the stiffness matrix and of the elastic energy provide complete information on the global isotropic and anisotropic deformability of any NA, provided its geometry can be well-represented in the helical space.[12,13,30−32,43−46]

Computational Details. Trajectories for 14 B-DNA duplexes of length 10−16-mer (see Table 1) were obtained from isothermal−isobaric simulations ($T = 298$ K; $P = 1$ atm). The duplexes considered here involve different sequences and, in some cases, include point alterations, such as mismatched pairs, unusual bases, or the substitution of a phosphate by a neutral phosphonate. The PARM-98 AMBER force field[8] was used in conjunction with the TIP3P water model[47] to describe standard nucleotides and water. When needed, force-field parameters were developed to describe unusual bases using the RESP/HF-6-31G(d)[48] and PARMQMD[49] methodologies. SHAKE[50] was used in all simulations to constrain all bonds at equilibrium lengths, which

Table 4. Percentage of Variance Explained for Different Trajectories (Central Eight-mer) When Different Number of Eigenvectors Are Considered[a]

trajectory	1	5	10	50	100
traj1	25,1	63,2	78,2	95,9	98,5
traj2	28,0	61,4	76,6	95,4	98,3
traj3	21,9	60,2	76,6	96,1	98,4
traj4	30,0	66,2	79,4	96,5	98,6
traj5	31,7	66,8	79,3	95,9	98,5
traj6	16,2	52,4	70,0	93,8	97,5
traj7	24,5	64,3	76,9	95,1	98,1
traj8	31,0	66,7	79,0	95,5	98,3
traj9	21,7	57,0	72,7	94,4	97,8
traj10	20,1	57,6	73,4	94,9	98,0
traj11	23,9	62,4	74,4	94,6	97,9
traj12	18,0	56,5	73,6	94,8	98,0
traj13	21,9	58,7	74,0	94,8	97,9
traj14	17,5	57,8	72,9	94,6	97,8
average	23.7 ± 4.8	60.8 ± 4.8	75.5 ± 2.8	95.2 ± 0.7	98.1 ± 0.3
traj5[b]	29.7	57.3	72.5	94.5	97.7

[a] The average values with standard deviations are also shown. [b] This value corresponds to trajectory 5 considering the central 13-mer.

allowed us to use a 2 fs time step for integration of Newton's equations. All studied systems were neutralized by adding a suitable amount of Na^+ and hydrated by immersion in suitable rectangular boxes of TIP3P waters.[47] The PME strategy[3] and periodic boundary conditions were used to account for long-range effects. Some of the simulations were partially or totally obtained from our local trajectory library, whereas others were especially performed for this paper (see Table 1). All the trajectories were extended for at least 5 ns, and the trajectory comprised between the 1st and 5th nanosecond was used for analysis.

Some analyses were performed considering the entire helices, whereas others considered only the common eight-mer portion. Since the sequences were different, we were often forced to use simplified covariance matrices to facilitate comparison between different duplexes (purines were represented by atoms N9, C8, H8, N7, C5, C6, N6/O6, N1, C2, N3, and C4 and pyrimidines by N1, C6, H6, C5, C4, N4/O4, N3, C2, and O2; i.e., 20 atoms × base dimer). For other studies (for example, to study conformational transitions), a further reduced coordinate system was used by taking only four atoms for each base (C1′−C4−N9−C8 for purines and C1′−C6−N1−C2 for pyrimidines). Finally, a last set of analyses (for example, to compare essential movements between trajectories) was performed considering only backbone atoms (up to N1 or N9). In all cases, comparisons were performed always considering covariance matrices of the same dimensionality.

Results and Discussion

Essential Dynamics. Around 100 eigenvectors are able to explain nearly all the variance of the eight-mer DNA duplexes irrespective of its composition and of the presence of chemical alterations in the structure (see Table 4). Interestingly, not many more eigenvectors are necessary to explain all the variance of longer duplexes (see Table 4). Thus, 100 eigenvectors explain between 97.9% (traj11) and 98.5% (traj5) of the variance in $d(A)_8$, and the same number of eigenvectors represent 97.7% of the variance of $d(A)_{13}$.

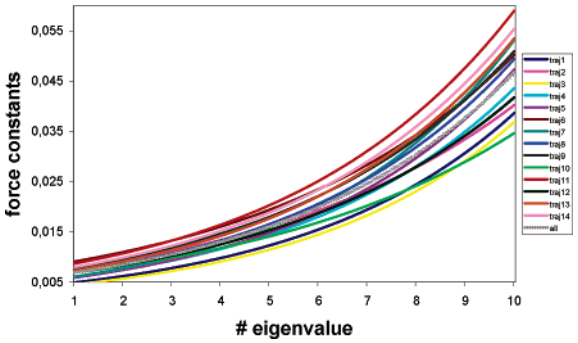


Figure 2. Force constants (in kcal/mol Å^2) associated with the first 10 essential deformations of different DNA duplexes. Line in gray labeled "all" corresponds to the exponential fit [$K_i = 0.005\,572 \exp(0.2126i)$] performed considering data for all the duplexes. Force constants were determined considering only the central eight-mer portion of the different DNA duplexes.

This suggests that the dynamics of the DNA is quite simple (see below). For qualitative comparison, when the same analysis is performed with a set of trajectories taken from our local database of small protein trajectories (PDB entries 1AA2, 1AA3, 1ACF, 1ACX, 1AG4, 1ARK, 1ATA, 1B40, 1BM8, 1CDZ, 1CK2, 1EXG, 1FOW, and 1HY8), the average variance explained by 100 eigenvectors is 88%.

The softer deformation modes have associated elastic force constants below 10 cal/mol Å^2 for all the central eight-mer segments (see Figure 2), and even for the 10th mode, force constants are, in general, below 50 cal/mol Å^2. This implies very low frequencies in the range of 13−44 (1st−10th) cm^{-1} for the studied systems, values that are similar to those obtained for proteins (range between 9 and 36 cm^{-1}) and to those associated with weak rotations in small molecules. The low force constants associated with the essential modes indicate that thermal energy can introduce important fluctuations in the DNA structure and that the DNA duplex is an extremely flexible entity irrespective of the sequence or the presence of chemical alterations. Thus, the rmsd's between the snapshots that have explored more distant regions along

Table 5. Dimensionality (see Methods Section) of the Configurational Space of the Different Trajectories[a]

trajectory	dimensionality
traj1	48
traj2	44
traj3	45
traj4	42
traj5	45
traj6	42
traj7	50
traj8	46
traj9	45
traj10	48
traj11	46
traj12	46
traj13	42
traj14	42

[a] In all cases, values are computed considering the central eight-mer fragments.

the first eigenvector are in the range 3.5−4.6 Å (these structures are selected from the values of the projection of the trajectory along the corresponding first eigenvector), a quite large value considering that they are computed for the central eight-mer and considering the deformation only along one eigenvector.

The increase in the value of the elastic constants with the eigenvectors follows well an exponential law (see Figure 1). In fact, the force constants (in kcal/mol Å^2) associated with the ith eigenvector can be reasonably determined (at least for eight-mer duplexes) using the equation $K_i = 0.005\,572 \exp(0.2126i)$, obtained by analyzing all the force constants of the central eight-mer. Trajectories leading to force constants far from these consensus values must be analyzed with special care. Also interestingly, the different lines in Figure 2 are very close for the fist eigenvector, whereas they diverge as higher modes are considered. This indicates that (in absolute terms) force constants associated with the first modes are less dependent on the nature of the oligonucleotides than those associated with higher modes.

The average dimensionality of the configurational space of the different eight-mer duplexes considered here is 45 ± 3 (see Table 5), which suggests that the configurational space of eight-mer DNAs can be represented in a small space of 42−50 dimensions corresponding to the important essential movements (see also Table 4). This small number sharply contrasts with the thousands of degrees of freedom of the studied duplexes and confirms the relative simplicity of the accessible configurational space of the DNA duplex. The difference in dimensionality due to changes in sequence, presence, or chemical alterations and sequence environments is rather small (see Table 5) and does not seem to be clearly related to the chemical nature of the duplex.

As expected, the dimensionality of the system increases with the length of the duplex (see Figure 3). An analysis of the data shows that, at least within the interval studied, we can predict the dimensionality of the system from the linear equation dim = 4.71 × length + 8.9 ($r^2 > 0.999$). This relationship suggests that, on average, each base pair adds only around five more dimensions to the DNA configurational space. These empirical equations can then be used to estimate the complexity of accessible space of large DNA duplexes, defining then suitable simulation protocols, since simulations that are able to explore a space of small dimensionality will not be large enough to reproduce that of larger spaces well.

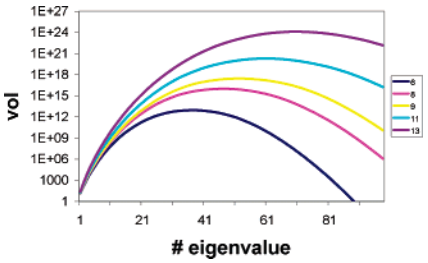


Figure 3. Variation in the configurational volume (Vol in Å^t; where t is the number of eigenvalues, logarithmic scale) for a common sequence [d(A)$_n$] with the length of the duplex and the number of eigenvalues.

There is a close similarity in the nature of movements sampled by the eight-mer duplexes in the different trajectories. Absolute similarity indexes (γ, eq 4 and ξ, eq 10) range typically between 0.6 and 0.8. The ξ values (eq 10) are generally larger than the γ ones (eq 4), showing that the similarity is greater for the most important deformation modes (see Table 6), whereas it decreases as higher modes are considered. Clearly, the new index seems to be more powerful than traditional Hess metrics[36] to properly weight the importance of low-frequency modes to draw the essential dynamics of DNA.

As expected, the best similarity is found when two similar sequences are compared (see, for example, similarity indexes for trajectories 8, 9, and 10 or for 5 and 11), and the worst are found when duplexes of different sequences and containing chemical alterations are compared (see Table 6). When the entire set of trajectories is combined into a single one, a consensus covariance matrix can be generated ($\mathbf{C}_x^{\text{cons}}$), containing not only data on the variability of DNA conformation due to natural deformations in the Cartesian space but also information on the deformation of the DNA structure due to change in the sequence or the presence of unusual modifications in the chemical structure. The eigenvectors obtained upon diagonalization of $\mathbf{C}_x^{\text{cons}}$ provide, then, information on both types of deformations (sequence-derived and due to thermal fluctuations). It is worth noting that these two types of conformational changes could, in principle, be very different, but a comparison of eigenvectors obtained from $\mathbf{C}_x^{\text{cons}}$ with those obtained by diagonalization of the different individual covariance matrices ($\mathbf{C}_x$) yields very large similarity indexes (in the range 0.7−0.8; see Table 3). This finding strongly supports the idea that there is an essential deformation pattern for the DNA, which can explain 70−80% of the normal flexibility of any DNA duplex. Furthermore, our results also suggest that the conformational changes in the DNA induced by alterations in its covalent structure happen, in most cases, along the natural deformation modes of the duplex.[31]

Entropy calculations. Entropies determined from Schlitter's[37] and Andricioaei and Karplus'[38] methods provide a

Table 6. Absolute Similarity Indexes (Indexes γ, Eq 5, above the Diagonal in Roman Font) and Indexes ξ, below the Diagonal in Italics) between the Eigenvectors Associated the Central Eight-mer Portion of the Different Duplexes[a]

	traj1	traj2	traj3	traj4	traj5	traj6	traj7	traj8	traj9	traj10	traj11	traj12	traj13	traj14	**All**
traj1		0.73	0.65	0.69	0.62	0.55	0.68	0.66	0.69	0.67	0.66	0.64	0.57	0.56	**0.72**
traj2	*0.84*		0.63	0.65	0.59	0.55	0.63	0.65	0.62	0.56	0.61	0.59	0.55	0.59	**0.68**
traj3	*0.76*	*0.75*		0.69	0.62	0.62	0.60	0.63	0.60	0.57	0.63	0.65	0.61	0.54	**0.69**
traj4	*0.78*	*0.76*	*0.74*		0.59	0.52	0.58	0.63	0.67	0.59	0.60	0.61	0.52	0.55	**0.65**
traj5	*0.73*	*0.70*	*0.73*	*0.74*		0.61	0.65	0.64	0.63	0.62	0.78	0.68	0.60	0.59	**0.69**
traj6	*0.63*	*0.65*	*0.68*	*0.63*	*0.70*		0.61	0.66	0.56	0.57	0.59	0.64	0.79	0.67	**0.68**
traj7	*0.76*	*0.74*	*0.69*	*0.71*	*0.77*	*0.70*		0.82	0.70	0.73	0.65	0.65	0.62	0.58	**0.79**
traj8	*0.77*	*0.73*	*0.72*	*0.76*	*0.79*	*0.72*	*0.89*		0.70	0.73	0.65	0.65	0.61	0.59	**0.80**
traj9	*0.79*	*0.75*	*0.69*	*0.74*	*0.74*	*0.66*	*0.79*	*0.78*		0.74	0.66	0.65	0.55	0.54	**0.72**
traj10	*0.76*	*0.68*	*0.68*	*0.70*	*0.73*	*0.65*	*0.81*	*0.79*	*0.80*		0.63	0.64	0.56	0.51	**0.71**
traj11	*0.77*	*0.75*	*0.73*	*0.77*	*0.87*	*0.69*	*0.79*	*0.79*	*0.76*	*0.74*		0.70	0.60	0.57	**0.69**
traj12	*0.72*	*0.67*	*0.73*	*0.74*	*0.79*	*0.68*	*0.72*	*0.74*	*0.71*	*0.72*	*0.82*		0.63	0.59	**0.72**
traj13	*0.65*	*0.64*	*0.65*	*0.66*	*0.73*	*0.85*	*0.70*	*0.72*	*0.68*	*0.66*	*0.71*	*0.70*		0.70	**0.66**
traj14	*0.69*	*0.68*	*0.65*	*0.71*	*0.75*	*0.76*	*0.74*	*0.76*	*0.70*	*0.66*	*0.74*	*0.71*	*0.78*		**0.63**
ALL	**0.81**	**0.77**	**0.79**	**0.75**	**0.78**	**0.72**	**0.83**	**0.84**	**0.79**	**0.81**	**0.80**	**0.79**	**0.71**	**0.75**	

[a] Similarities of each trajectory with respect to a "consensus DNA trajectory" obtained by combining the 14 individual trajectories are also displayed, in bold.

Table 7. Intramolecular Entropies (in kcal/mol K) Computed for the Different Trajectories from 3.5 ns Samplings[a]

name	Schlitter	Andreociaei–Karplus	rmsd fluctuations
trajectory 1	1.57/*1.73*	1.45	1.23
trajectory 2	1.54/*1.68*	1.42	1.03
trajectory 3	1.58/*1.79*	1.46	1.31
trajectory 4	1.51/*1.69*	1.39	1.15
trajectory 5	1.56/*1.72*	1.43	1.32
trajectory 6	1.51/*1.63*	1.39	1.07
trajectory 7	1.53/*1.71*	1.41	1.53
trajectory 8	1.57/*1.74*	1.43	1.59
trajectory 9	1.55/*1.70*	1.42	1.24
trajectory 10	1.61/*1.81*	1.48	1.56
trajectory 11	1.55/*1.71*	1.43	1.26
trajectory 12	1.57/*1.74*	1.44	1.35
trajectory 13	1.53/*1.69*	1.41	1.33
trajectory 14	1.52/*1.65*	1.39	1.29

[a] Values in italics are those obtained by extrapolating to infinite simulation time. Total rmsd fluctuations ($\langle \mathrm{rmsd}^2 \rangle_t - \mathrm{rmsd}_{\mathrm{ref}}^2$; see eqs 13–14) from the average are in Å^2. The central eight-mer portions of all duplexes were considered in the calculations, where the base pairs were represented by a common reference system comprising 20 atoms.

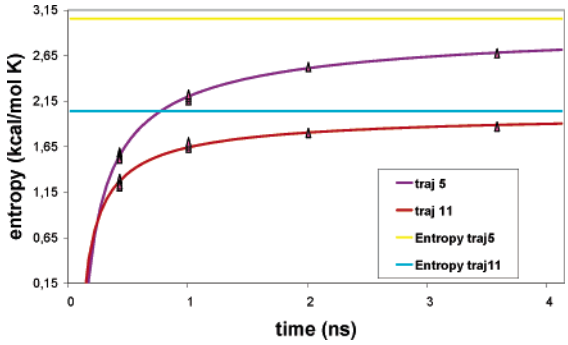


Figure 4. Change in (Schlitter) entropy with simulation time (snapshots collected every picosecond) for two duplexes of different lengths but the same composition: traj 5 is $d(\mathrm{A})_{15}$ and traj 11 is $d(\mathrm{A})_{11}$. Individual values used to fit the exponential profiles are shown, as well as the estimate of entropy at infinite simulation time (labeled as "Entropy traj*x*").

quantitative measure of the conformational freedom of a molecule moving in a quasi-harmonic regime. Andricioaei and Karplus' method provides systematically smaller values than the Schlitter approach, but in relative terms, both approaches give identical results (see Table 7). As noted previously (see Methods section), the entropy strongly depends on the extent of sampling used to create the mass-weighted covariance matrix. Not surprisingly, such dependence becomes more evident (see Figure 4) as the length of the duplexes increases, which warns against the quantitative value of entropy estimates of large duplexes obtained from 1 to 10 ns simulation times (note the difference between the last entropy estimate and the entropy at infinitive simulation time in Figure 4). In fact, a 5 ns simulation time is probably not enough to reproduce the real intramolecular entropy of even the shorter eight-mer duplexes.

Overall, entropies are only moderately dependent on the eight-mer duplex considered (a range of less than 10% variation in the total entropy, with a standard deviation in S_∞ of only 2%), confirming that global flexibility is mostly determined by the general polymeric properties of the duplex, and changes in sequence, presence of chemical alterations, or other moderate changes do not have a dramatic influence on the DNA entropy. Interestingly, the entropy differences between the DNA duplexes analyzed here are evident only for long simulation times, when the structure is allowed to explore less populated regions of the configurational space (Supporting Information, Figure S1). The fitting of entropy estimates at different simulation times to eq 18 yields similar α and β values for all the duplexes, and consensus values ($\alpha = 24.3 \pm 3.2$, $\beta = 0.62 \pm 0.03$; see eq 18) can be used to compute entropies at infinite simulation time from a unique estimate at 3500 ps with a reasonable error (rmsd = 0.02 kcal/mol K) for a variety of DNA sequences. Simulations yielding α and β values far from these consensus ones (for the central eight-mer) should be taken with caution, since

Table 8. Diagonal Stiffness Constants (Translations in kcal/mol Å^2 and Rotations in Kcal/Mol Degree^2) Associated with Harmonic Deformation of Local Helical Parameters for Some Selected Base Pair Steps[a]

steps	shift	slide	rise	tilt	roll	twist
AG/CT	1.36 (0.50)	2.14 (0.45)	6.34 (1.10)	0.029 (0.008)	0.024 (0.002)	0.018 (0.008)
GA/TC	1.54 (0.31)	2.04 (0.13)	8.30 (0.97)	0.040 (0.004)	0.025 (0.001)	0.032 (0.007)
GC/GC	2.07 (0.65)	3.61 (0.77)	10.46 (0.96)	0.045 (0.003)	0.030 (0.007)	0.041 (0.008)
GG/CC	1.19 (0.25)	1.78 (0.13)	8.37 (0.55)	0.042 (0.003)	0.024 (0.000)	0.030 (0.004)
TT/AA	1.66 (0.50)	2.72 (0.59)	8.31 (0.81)	0.035 (0.007)	0.025 (0.002)	0.032 (0.008)

[a] Values were averaged for all the base pair steps of a given type in the entire set of duplexes (the standard deviation is given in parentheses). Only values for which at least 10 examples were found are reported in this table.

they might be capturing conformational transitions moving out of the harmonic regime.

Helical Stiffness. The helical elasticity of DNA has a moderate dependence on the local sequence (20−50% in force constants; see Table 8), with some steps being systematically more rigid than others. However, the data in Table 8 show that the general helical elastic properties of the DNA are marked by its general polymeric structure rather than by its sequence. When stiffness constants for a given (5′-XY-3′) step are computed in different chemical environments (for example, in different positions of the same duplex or in different duplexes), no negligible differences are found, as noted in the standard deviations associated with averages in Table 8. The same range of variation is evident when present force constants are compared with those obtained by Lankas and co-workers[43,46] using sequences different from those considered here. Overall, this indicates that the molecular environment surrounding the base pair step can play a major role in determining the local rigidity of DNA, suggesting then that fragments longer than two base pairs are needed to capture the deformability of DNA. Data in Table 8, combined with an analysis of flexibility in Cartesian space (see above), clearly show that the concept of flexibility is rather diffuse and should be linked to the nature of the perturbation introduced to the system. Some duplexes might be more flexible in terms of twist deformations but more rigid for bending distortions. In any case, large flexibility in the local helical space might yield, depending on the sequence, small flexibility in the Cartesian space. Caution is then necessary when playing with very anisotropic concepts such as flexibility or rigidity.

Conclusions

(1) MD trajectories contain a large amount of information on the flexibility of DNA duplexes, but hidden in gigabytes of noise. Data mining methods such as those discussed here can be used for the routine analysis of nanosecond-scale MD simulations.

(2) DNA is a very flexible entity, whose essential deformation movements are associated with very small force constants. However, despite this large flexibility, the pattern of DNA deformability is not very complex and can be described by a relatively small set of movements.

(3) The pattern of essential movements of DNA is largely conserved irrespective of length, sequence, or chemical alterations. It is possible to obtain a reduced set of consensus essential movements of DNA duplexes that are able to explain a large amount of variance in the MD trajectories. These movements are also informative in understanding how the DNA can change its structure as a result of small covalent changes.

(4) For a given length, the level of disorder and flexibility in the Cartesian space of the different DNAs studied is quite similar. However, different DNAs can react in different ways when they are perturbed along specific helical coordinates, indicating that concepts such as flexibility or rigidity are not universal, but must be tightly related to the coordinate frame used to measure the conformational space and on the nature of the deformation introduced to the structure.

Acknowledgment. We thank Prof. Luque and two anonymous reviewers for their valuable comments. This work has been supported by the Spanish Ministry of Education and Science (BIO2002-06848), the BBVA, and La Caixa Foundation. We also thank the support of the Instituto Nacional de Bioinformatica (INB—Genoma España) and the Centre de Supercomputacio de Catalunya (CESCA).

Supporting Information Available: Projection of some of the trajectories in the first eigenvectors (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Levitt, M. *Cold Spring Harbor Symp. Quant. Biol.* **1983**, *47* (Pt 1), 251−62.

(2) Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. *J. Biomol. Struct. Dyn.* **1983**, *1*, 231.

(3) Darden, T.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98* (12), 10089−92.

(4) Cornell, W. D.; Cieplak, P.; Baily, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−97.

(5) MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117* (48), 11946−75.

(6) Foloppe, N.; Mackerell, A. D. *J. Comput. Chem.* **2000**, *21* (2), 86−104.

(7) Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16* (3), 487−509.

(8) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.

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

(10) Cheatham, T. E.; Kollman, P. A. *Annu. Rev. Struct. Dyn.* **2000**, *51*, 435−471.

(11) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350−357.

(12) Orozco, M.; Pérez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350−364.

(13) Orozco, M.; Rueda, M.; Blas, J. R.; Cubero, E.; Luque, F. J.; Laughton, C. A. *Encyclopedia of Computational Chemistry*; Wiley: New York, 2004 (published online April 15, 2004; doi: 10.1002/0470845015.cn0080).

(14) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (41), 14771−5.

(15) Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2004**, *87* (2), 800−11.

(16) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32* (14), 4269−80.

(17) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. *Biophys. J.* **2004**, *87* (6), 3799−813.

(18) McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2000**, *304*, 803−820.

(19) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211.

(20) Srinavasanm, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc*. **1998**, *120*, 9401−9409.

(21) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127−9.

(22) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246* (1, 2), 122−9.

(23) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem*. **2002**, *23* (14), 1297−304.

(24) Morreale, A.; de la Cruz, X.; Meyer, T.; Gelpi, J. L.; Luque, F. J.; Orozco, M. *Proteins* **2004**, *57* (3), 458−67.

(25) Morreale, A.; de la Cruz, X.; Meyer, T.; Gelpi, J. L.; Luque, F. J.; Orozco, M. *Proteins* **2005**, *58* (1), 101−9.

(26) Cubero, E.; Abrescia, N. G. A.; Subirana, J. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 14603−14612.

(27) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121* (25), 5981−5991.

(28) Amadei, A.; Linssen, B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412−425.

(29) Wlodek, S. T.; Clark, T. W.; Scott, L. R.; McCammon, J. A. *J. Am. Chem. Soc.* **1997**, *119*, 9513−9522.

(30) Noy, A.; Perez, A.; Lankas, F.; Luque, F., J.; Orozco, M. *J. Mol. Biol.* **2004**, *343* (3), 627−38.

(31) Perez, A.; Noy, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2004**, *32* (20), 6144−51.

(32) Noy, A.; Pérez, A.; Márquez, M.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc*. **2005**, *127*, 4910−20.

(33) Go, M.; Go, N. *Biopolymers* **1976**, *15* (6), 1119−27.

(34) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125* (26), 8007−14.

(35) Rueda, M.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc*. **2005**, in press.

(36) Hess, B. *Phys. Rev.* **2000**, *62*, 8438.

(37) Schliter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.

(38) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289.

(39) Harris, S.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658.

(40) Lu, X. J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300* (4), 819−40.

(41) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31* (17), 5108−21.

(42) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn*. **1989**, *6* (4), 655−67.

(43) Lankas, F.; Sponer, J.; Hobza, P.; Langowski, J. *J. Mol. Biol.* **2000**, *299*, 695−709.

(44) Olson, W. K.; Zhurkin, V. B. *Curr. Opin. Struct. Biol*. **2000**, *10*, 286−97.

(45) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163−8.

(46) Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E., III. *Biophys. J.* **2003**, *85* (5), 2872−83.

(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(48) Bayly, C. E.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

(49) Aleman, C.; Canela, E. I.; Franco, R.; Orozco, M. *J. Comput. Chem.* **1991**, *12*, 664−674.

(50) Rickaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1997**, *23*, 327−341.

(51) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323−2324.

CT050051S

JCTC Journal of Chemical Theory and Computation

Theoretical Conformational Analysis for Neurotransmitters in the Gas Phase and in Aqueous Solution. Serotonin

Giuliano Alagona,*,† Caterina Ghio,† and Peter I. Nagy*,‡

Molecular Modeling Lab, CNR-IPCF, Institute for Physical Chemistry Processes, Via Moruzzi 1, I-56124 Pisa, Italy, and Department of Medicinal and Biological Chemistry and Center for Drug Design and Development, The University of Toledo, Toledo, Ohio 43606-3390

Received April 5, 2005

Abstract: Conformational analyses have been performed for protonated serotonin in the gas phase, aqueous solution, and in the binding cavity of a 5-HT2A receptor model. DFT geometry optimizations have been performed in the gas phase at the B3LYP/6-31G* levels. Optimized calculations up to the B3LYP/6-311++G** level find two low-energy gauche conformations separated by 8−10 kcal/mol barriers from a trans conformation with relative energy of about 6 kcal/mol. In aqueous solution as concluded from IEF-PCM/B3LYP/6-31G* and IEF-PCM/MP2/6-31G*//IEF-PCM/B3LYP/6-31G* continuum solvent calculations as well as Monte Carlo free energy perturbation simulations with explicit solvent molecules, those barriers decrease to 2−7 kcal/mol, while the two gauche and one trans conformers are within a 3 kcal/mol relative free energy range. The solute is strongly hydrated by about three water molecules around the $-NH_3^+$ group and by one water molecule for each of the pyrrole and phenolic hydrogen atoms. Docking studies of the protonated ligand predicted both gauche and trans ligand conformers to favorably interact with the 5-HT2A receptor in its hypothesized binding cavity. The theoretical studies confirm the experimental results regarding strong interactions with the Asp155 and Ser159 residues (TM helix III) and the interactions of the indole ring with Phe, Trp, and Tyr side chains in TM V, VI, and VII helices within a 24 kcal/mol range for the relative interaction energies.

I. Introduction

The present study is a continuation of our interest toward the in-solution structure of biogen amine neurotransmitters. After former studies for histamine,[1] dopamine,[2] and norepinephrine[3] by us and others, the present target of our investigation is serotonin (5-hydroxytryptamine, 5-HT). In CNS, serotonin has a neurotransmitter function on different 5-HT receptor subtypes, influencing numerous physiological processes such as sleep, cognition, motor activity, temperature regulation, nociception, appetite, sexual behavior, etc. In periphery, serotonin plays role as a regulator of smooth muscle function in the cardiovascular and gastrointestinal systems and also in platelet aggregation.

The neutral, gas-phase serotonin has been theoretically studied recently by Van Mourik and Emson.[4] The biogen amine neurotransmitters above are primarily protonated, however, in aqueous solution at the physiological tissue pH = 7.4,[3a] and according to a widely accepted concept,[5] this cationic form is the acting one as an agonist for the corresponding G-protein coupled receptor (GPCR).[6] GPCRs are transmembrane receptors comprising a densely packed seven-helix system, and using the recently determined rhodopsin receptor[7] as a template for their model, the binding sites are located about 11 Å away from their extracellular surfaces, down along the main axes of the helices.

* Corresponding authors phone: +39-050-3152450; fax: +39-050-3152442; e-mail: G.Alagona@ipcf.cnr.it (G.A.) and phone: (419)530-1945; fax: (419)530-1909; e-mail: pnagy@utnet.utoledo.edu (P.I.N.).

† CNR-IPCF, Institute for Physical Chemistry Processes.

‡ The University of Toledo.

10.1021/ct050088c CCC: $30.25
Published on Web 07/20/2005

One of the most important questions is the following: how does the cationic neurotransmitter reach the binding cavity from the synaptic cleft, which may be essentially characterized as an aqueous solution. This mechanism has not been explored yet. Accordingly, we address two questions in the present study. The focus of our interest is to find the prevalent conformers of serotonin in aqueous solution mimicking the extracellular environment in the synaptic cleft. The low-energy serotonin conformers, as starting models for the bound form for the agonist of the 5-HT2A serotonin receptor, will be utilized in simple molecular mechanics docking studies. In these latter studies our goal is to explore whether the agonist has to undergo remarkable conformational changes as compared to its favored in-solution structure, to fit optimally to the amino acid side chains in the binding cavity.

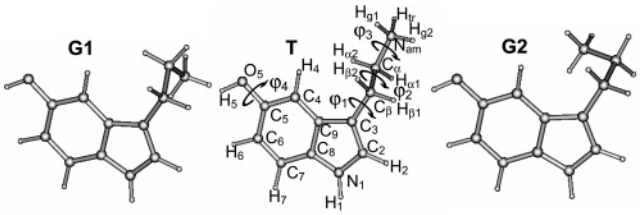
For histamine, dopamine, and norepinephrine as well as for serotonin the protonation is a reversible process and depends on the chemical character of the environment. The protonated state of these molecules must be thermodynamically stable in the aqueous extracellular environment at the physiological tissue pH (see above). Considering, however, the narrow pore in the seven-helix transmembrane structure of the GPCRs, it is an essential question whether the protonated form is maintained throughout the traverse process from the receptor surface to the binding cavity at a depth of about 11 Å. Since there is no space for accommodating a large number of water molecules in the pore, it is also possible that the proton is lost throughout the traverse process and is regained in the binding cavity. In fact, our computer modeling for the complex of the norepinephrine ligand with the β-adrenergic receptor indicates that a few water molecules can be accommodated in the binding cavity even in the presence of the agonist.[3a]

Common in the binding of these neurotransmitters is that there is an ion-pair interaction with the protonated nitrogen of the agonist and the negatively charged Asp 3.32 side chain (using the Ballesteros−Weinstein numbering scheme[8] for the receptor residues) in transmembrane III. This interaction anchors the ligand, and if there are at least two water molecules in the neighborhood or the local dielectric constant is larger than nine, the environment helps maintain the ion-pair structure.[9] On this basis, a conformational study of protonated serotonin seems to be justified, and the present investigation explores the relevant part of the in-solution conformational space by utilizing continuum dielectric and Monte Carlo methods in combination with ab initio and density functional theory calculations.

II. Methods and Calculations

Atom numbering and the three main conformations, G1, G2 gauche, and T trans, for protonated serotonin are shown in Scheme 1. The minimum energy OH arrangement has been determined from the flexible scan of the $H_5O_5C_5C_6$ (φ_4) dihedral angle at the DFT level by applying the B3LYP functional[10] using the 6-31G* basis set[11] for G1, G2, and T in the gas-phase. Single point MP2 calculations[12] have been carried out for comparison. A potential energy surface (PES) was calculated changing the $C_2C_3C_\beta C_\alpha$ (φ_1) and the $C_3C_\beta C_\alpha N$ (φ_2) torsion angles in 30° increments at the same level and

Scheme 1. G1, T, and G2 Protonated Serotonin



optimizing the rest of the internal coordinates. Three local minima and three transition states (TS), verified by frequency analysis, were identified on the ($0° \leq \varphi_1 \leq 180°$) PES. Since the rotation of the serotonin side chain is possible both clockwise and counterclockwise, there is a mirror image for each structure with equal energy. The PES reflects this character upon 360° rotation of both the φ_1 and φ_2 torsion angles and possesses C_2 symmetry, accordingly. The relative energies for the local minima and TS calculated through gas-phase optimizations at the B3LYP/6-31G* and B3LYP/6-311++G** levels have been compared. The calculations were carried out using the Gaussian 03 software[13] running in Pisa.

Using thermal corrections, the relative molar internal free energies for the selected conformers, ΔG(int), were calculated at $T = 310$ K and $p = 1$ atm, using the rigid rotator-harmonic oscillator approximation.[14] Accordingly, the molar ΔG(int) was calculated as

$$\Delta G(\text{int}) = \Delta E + \Delta G_{\text{th}}(T) = \Delta E + \Delta\text{ZPE} + \Delta(H(T) - \text{ZPE}) - T\Delta S(T) \quad (1)$$

Here ΔE and ΔG_{th}(T) are the quantum chemically calculated internal energy difference and the so-called thermal correction, respectively. Contributions to the latter, ZPE, $H(T)$, and $S(T)$, are the zero point energy, the vibrational enthalpy, and the total entropy at T (= 310 K), respectively. Since the proposed scaling factor for the B3LYP/6-31G* ΔZPE correction[15] (0.9804) is very close to 1, no scaling factor was applied.

In-solution calculations started with B3LYP/6-31G* geometry optimizations at the integral-equation-formalism (IEF)-PCM level. In the IEF-PCM framework,[16] the solvent is represented by a continuous medium characterized by its dielectric permittivity, where the solute is embedded inside a cavity of proper shape created around it. The cavity, built according to the actual geometric structure of the solute, univocally defines the closed surface that separates solute from solvent, used to formulate the basic electrostatic equations characterizing the solute−solvent (su−sv) system. Bondi radii[17] have been used throughout, except for the united atom description of the CH (R=1.9 Å) and CH_2/CH_3 (R=2.0 Å) groups, applying a 1.2 scaling factor. In addition to computational efficiency, a prominent advantage of the IEF-PCM approach with respect to standard PCM[18] is that errors, such as those linked to the partition of the continuous apparent surface charge into discrete elements and to the electronic charge escaped outside the cavity, are reduced at least by an order of magnitude, because of the use of operators related to the electric potential instead of the electric field.[19]

Starting from gas-phase optimized geometries, structures corresponding to energy minima and TSs were optimized in aqueous solution. In addition to these structures, six further structures, called midpoints henceforth, were optimized, while the $C_2C_3C_\beta C_\alpha$ and $C_3C_\beta C_\alpha N$ torsion angles were kept fixed at values taken as averages of those related to the two closest stationary points. [Only in the case of midpoint 1, because of the shape of the PES, it was necessary to optimize φ_1 at φ_2 = average of the corresponding value for G1 and TS1 structures.] Then all other internal coordinates were relaxed. Employing this strategy, a route connecting the extreme points of the in-solution PES was selected. Since we are interested in the relative free energies only for points G1, TS1, G2, TS2, T, and TS3, the calculated values are independent of the selected paths to connect the six points mentioned. The IEF-PCM relative free energies were calculated as

$$\Delta G_{tot} = \Delta E^{s}_{int} + 1/2\Delta E_{susv} + \Delta G_{drc} \quad (2)$$

The first two energy terms were obtained by the iterative solution of the Schrödinger equation for a solute immersed in a polarizable continuum dielectric.

$$E^{s}_{int} = [<\Psi_s|H^\circ|\Psi_s>] \quad (3a)$$

$$E_{susv} = [<\Psi_s|V_R^{sol}|\Psi_s>] \quad (3b)$$

H° is the solute Hamiltonian, V_R^{sol} is the solvent reaction field generated by the fully polarized solute in solution, and Ψ_s is the converged wave function of the solute. The G_{drc} term stands for the nonelectrostatic, dispersion-repulsion-cavity part of the free energy of the solute in solution. $C_3C_\beta C_\alpha N$ rotational potential curves in solution have been calculated at different levels and with different basis sets for the serotonin cation. Local energy minima and transition states in solution were verified by frequency analysis. Vibrational frequencies at the IEF-PCM/B3LYP/6-31G* level have been computed.

By utilizing the IEF-PCM/B3LYP/6-31G* optimized molecular structures, the procedure we proposed recently[20] has been followed: electrostatic potential fitted CHELPG[21] and RESP[22] charges were derived, using the IEF-PCM/B3LYP/6-31G* wave function, for the 26 atoms of the solute in 12 conformations. These atomic charges were applied in Monte Carlo (MC) simulations and in docking studies when a model receptor was considered. Relative conformational solvation free energies were obtained by using the free energy perturbation method (FEP)[23a] as implemented in Monte Carlo simulations.[23b] Calculations were carried out by the use of the BOSS 4.2 software.[24]

MC simulations[25] for the aqueous solution of a protonated neurotransmitter was described in our earlier study for norepinephrine.[3a] Therefore, only the main points and the important differences as compared to the present modeling will be described here. MC simulations for the serotonin cation were performed in NpT (isobaric−isothermal) ensembles at T = 310 K and p = 1 atm. A water box, including 485 TIP4P water molecules[26] and a single solute with or without a chloride counterion, was considered for the aqueous solution model. Solvation free energy changes were calculated for rotation about the $C_\beta-C_\alpha$ axis. Geometries of the reference structures with fixed changes of ∼30° in the reaction path torsion angle were determined from IEF-PCM/B3LYP/6-31G* calculations relaxing all other internal coordinates. The interaction energy of the solution elements was calculated using the all-atom 12-6-1 OPLS-AA pair potential.[27] Steric OPLS parameters were taken from the program library, while potential derived charges were obtained as stated above.

Three different MC simulations in which the $-NH_3^+$ group, as a unit, was rotated by 360° about the $C_\beta-C_\alpha$ axis were carried out. The CHELPG charge set was applied for the $SerH^+\cdots Cl^-$ ion pair; the RESP fitted atomic charges were used for the single-solute serotonin cation ($SerH^+$) and the $SerH^+\cdots Cl^-$ ion pair. Application of a counterion is in place because the whole system (solute and its environment) must be neutral. In physiological conditions, the liquid environment can be characterized as a saline solution of 0.15 mol/dm^3 concentration. By simply assuming that the salt is NaCl and/or KCl, the concentration corresponds approximately to 1 Cl^-:360 water molecules. Our system is slightly more dilute with 1 Cl^-:485 water molecules.

In the original ion pair approximation with the CHELPG charges, the two ions were allowed to move independently. With a freely moving chloride ion, however, increments to the free energy were calculated with such a large noise that the standard deviation was regularly close to the free energy increment value itself, even for torsion angle changes as small as 0.8° and considering 5000K and 6000K configurations in the equilibration and averaging phases, respectively. These studies, however, pointed out that MC simulations predict a largely separated cation···anion system. On this basis, the $C_3\cdots Cl$ distance was kept fixed at 12 Å, meeting the requirement that Δcoord < edge/2. Nonetheless, the serotonin cation could rotate about a random axis through atom C_3, thus the relative orientation of the ion pair was not rigid. The solute−solvent and solvent−solvent cutoff radii were set to 12 and 9.75 Å, respectively, whereas the solute−solute atom−atom interactions were calculated at any distance of the ion pair.

With the RESP charges, the FEP calculations were repeated for the $SerH^+\cdots Cl^-$ system. This second series became necessary because the net free energy change with the CHELPG set was about 0.8 kcal/mol after completing a 360° rotation of the $-NH_3^+$ group. This error is known as the hysteresis of the calculation, which should ideally vanish, and, in fact, it was within the calculated standard deviation of the closed perturbation circle for dopamine[2d] and norepinephrine.[3a] The third MC simulation utilized the RESP atomic charge set for the "cation only" model. Long-range electrostatic corrections were calculated by means of the IEF-PCM procedure. Since the solute−solvent cutoff was set to 12 Å, solute−solvent interactions beyond this limit were calculated as corrections. Spheres with radii of 12 Å were formed around each solute atom, and their overlapping volume created the cavity to be used in IEF-PCM/B3LYP/6-31G* single-point calculations. By comparing the results obtained with the two charge sets, the effect of the selected

partial charge model can be investigated. Analogously, from the RESP calculations for the single cation and the ion pair, the counterion effect on relative solvation free energies for the different conformers can be evaluated.

Since the derived charge sets reflect the in-solution geometries and the polarization of the solute by the solvent, the relative internal energy was calculated also from the polarized solute model, thus from IEF-PCM/B3LYP/6-31G* calculations, and the relative total free energy of the protonated serotonin conformers on the basis of Monte Carlo simulations is

$$\Delta G_{tot} = \Delta E^{s}_{int}(\mathrm{PCM}) + \Delta G_{solv}(\mathrm{MC}) \qquad (4)$$

Conformational free energy differences (without the frequency dependent corrections), calculated according to eq 4, have been compared for the three MC models. The structure and some peculiarities of the first hydration shell of the different conformers have been characterized by utilizing radial distribution functions (rdfs) and pair-energy distribution functions (pedfs).

For investigating the interaction of the $SerH^+$ ligand and a model for the 5-HT2A receptor, flexible docking after a short molecular dynamics study was performed. A recently developed muscarinic-1 receptor model has been exploited,[28] based on the rhodopsin structure[7] as a template. On the basis of the correspondence of the residues for the muscarinic-1 and 5-HT2A receptors,[6b] 16−24 amino acids were switched in the middle third of each of the seven transmembrane α-helices. These altogether 150 amino acid side chains may surround the binding cavity or are at the corresponding depth of each helix. Based on studies by Roth et al.[29] and Almaula et al.,[30] the $-NH_3^+$ group was oriented toward the Asp 3.32 (residue 155) and Ser 3.36 (residue 159) side chains of the hypothesized binding cavity in the starting arrangement. The indole part of serotonin pointed toward different aromatic side chains including Phe 5.44, 5.47, 5.48, 6.51, 6.52, and 7.38 (240, 243, 244, 339, 340, and 365), Trp 6.48 and 7.40 (336 and 367), and Tyr 7.43 (370). As mentioned above, each gauche and trans $SerH^+$ conformers have a mirror image isomer with equal internal energy. Since the cavity has, however, a given stereochemistry, the elements of the image pairs would interact with the receptor with different energies. For this reason, the mirror image pairs of protonated serotonin in the G1, G2, and T conformations were placed into the cavity. This resulted in small differences for the interaction of the Asp 3.32 carboxylate and the cationic head of the ligand, but the indole part interacted with different residues from TM V, VI, and VII. The docking studies were performed by the aid of the Sybyl 6.91 software[31] and using the Tripos force field. For the receptor, the all-atom Kollman charge set of Sybyl was utilized, whereas the solute charges were accepted from our in-solution fitted CHELPG set. After removing the poor interactions for the receptor···ligand complex in a pre-energy minimization, a 10 ps long gas-phase molecular dynamics calculation was carried out with each ligand conformations. The Sybyl molecular dynamics module was used in a gas-phase N*VT* ensemble at $T = 310$ K. Solvent effects were implicitly considered in the distance-dependent dielectric approximation of the form $\epsilon = 4r$. The nonbonded cutoff was set to 8 Å. The 10 ps simulation length was supposed to be long enough for a possible change of the ligand conformation and rotation of involved receptor side chains into a more preferable position without destroying the basic structure of the receptor model. The energies of the structures with the smallest potential energy in the 5−10 ps interval as well as the structures obtained after 10 ps MD simulations were minimized. The interaction energy (EI) between the receptor and the ligand was calculated for these minimized structures as

$$\mathrm{EI} = E(\mathrm{add}) - E(\mathrm{rec,add}) - E(\mathrm{lig,add}) \qquad (5)$$

where E(rec,add) and E(lig,add) are the energies of receptor and ligand, respectively, in the optimized adduct. The ligand distortion energy, E(lig,dist), was calculated as E(lig,add) − E(lig,opt), where E(lig,opt) derived from a molecular mechanics optimization with the Tripos force field and CHELPG charges from the present study. Structures and relative energies of the isolated ligands were also considered with the MMFF94 force field.[32] Taking the receptor structure from the adduct after 10 ps MD and minimization, the optimized ligand was placed into the binding cavity at O•• •N distances of 4, 5, and 7 Å for the Asp155 carboxylate oxygen and the $SerH^+$ protonated nitrogen. The indole site was oriented toward the Phe, Trp, and Tyr side chains in TM V, VI, and VII. By performing docking studies from these starting arrangements, 18 further receptor−ligand binding modes were identified. In the absence of the knowledge of the absolute energy minimum of the receptor (whose determination is practically impossible), relative interaction energies and ligand distortion energies may provide qualitative information about the favorable binding mode of protonated serotonin.

III. Results and Discussion

A. Gas-Phase Results. A systematic conformational analysis of all degrees of freedom of protonated serotonin has been carried out in vacuo at the B3LYP/6-31G* level. To explore the conformational space of the indole OH group, the flexible scan of the $H_5O_5C_5C_6$ (φ_4) dihedral angle has been carried out for G1, G2, and T (Scheme 1). Two conformational minima have been found at $\varphi_4 \approx 0°$ and 180° (Figure 1) with the lower barrier joining them about $\varphi_4 \approx 90°$. The local minimum about $\varphi_4 \approx 180°$ is invariably higher than that about $\varphi_4 \approx 0°$ (by 1.2, 1.8, and 2.2 kcal/mol for G1, T, and G2, respectively). Thus all the subsequent calculations took $\varphi_4 \approx 0°$ as a starting conformation.

In accord with our earlier studies on neurotransmitters, the potential energy surface (displayed in Figure 2) was calculated in the gas phase by changing the $C_2C_3C_\beta C_\alpha$ (φ_1) and $C_3C_\beta C_\alpha N$ (φ_2) torsion angles at the B3LYP/6-31G* level. Three local minima, G1, G2, and T, were identified with the TS1, TS2, and TS3 transitions states connecting two neighboring minima when the φ_2 torsion angle changes between 0° and 360° and $0° \le \varphi_1 \le 180°$.

The B3LYP/6-31G* energies for the G1 and T local energy minima relative to the G2 conformer are 0.8 and 6.0 kcal/mol, respectively (Table 1). High barriers, ranging from

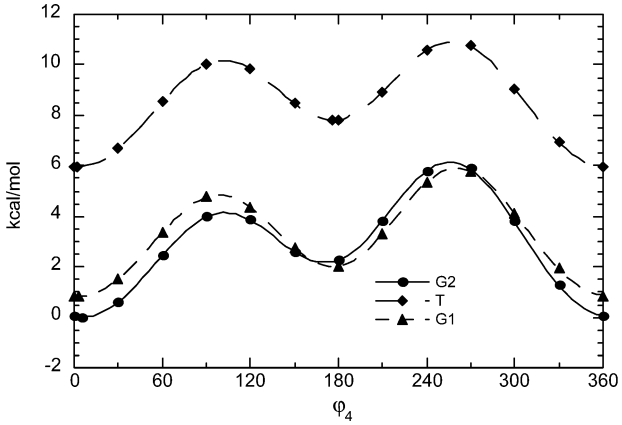


Figure 1. B3LYP/6-31G* potential energy profiles in vacuo for the rotation about $H_5O_5C_5C_6$ (φ_4) for the three main conformers.

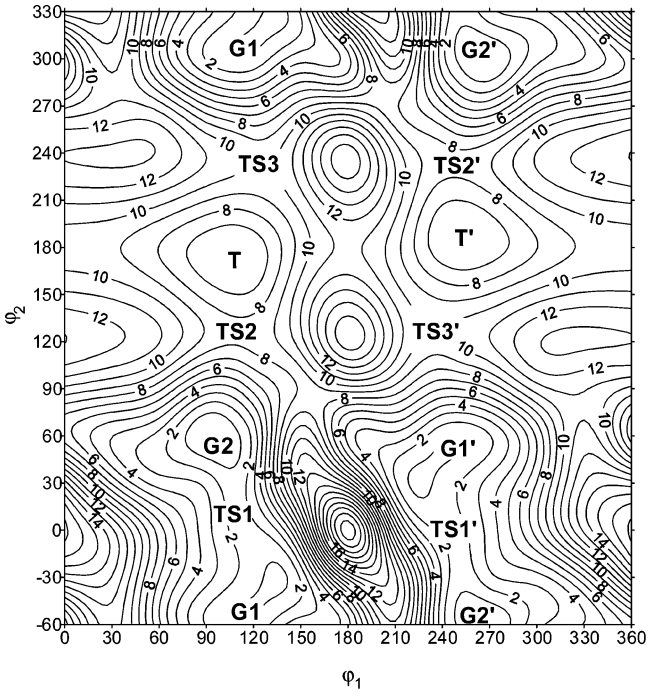


Figure 2. B3LYP/6-31G* potential energy surface in vacuo for the rotation about $C_2C_3C_\beta C_\alpha$ (φ_1) and $C_3C_\beta C_\alpha N$ (φ_2) with the main stationary points indicated.

Table 1. B3LYP Relative Energies (kcal/mol) at Optimized Geometries for Protonated Serotonin in the Gas Phase

	6-31G*	6-311++G**
G2	0[a]	0[b]
TS2 (G2→T)	8.25	7.68
T	5.97	5.62
TS3 (T→G1)	10.15	9.18
G1	0.84	0.66
TS1 (G1→G2)	1.65	1.94

[a] Reference energy (hartrees): −573.389277. [b] Reference energy (hartrees): −573.553883.

9.5 (φ_1=0°, φ_2=180°) to 11.2 kcal/mol (φ_1 = 180°, φ_2 = 180°), hinder the rotation about the C_3-C_β axis. Conversely, a favorable interconversion path with rotation about the $C_\beta-C_\alpha$ axis lies along the $\varphi_1 \approx \pm 120°$ regions.[2] (Hereafter only the path in the $\varphi_1 \approx 120°$ region has been considered, unless

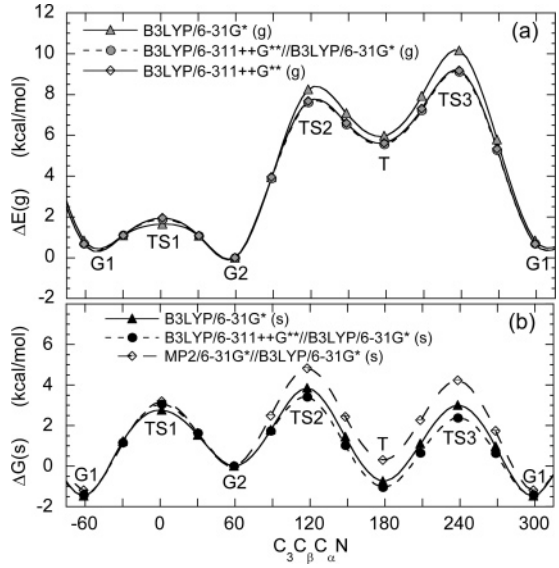


Figure 3. Profiles (kcal/mol) for the rotation about the $C_\beta-C_\alpha$ axis in protonated serotonin of the potential energy in the gas phase (a) (g, empty markers) and of the IEF-PCM free energy in aqueous solution (b) (s, solid markers). Solid lines = optimized structures; dashed lines = single point calculations at the B3LYP/6-311++G** or MP2/6-31G* level on the B3LYP/6-31G* structures.

otherwise specified (stationary points in the $\varphi_1 \approx -120°$ region are marked with a prime).) The TS1 transition state (ΔE = 1.6 kcal/mol) separates G1 and G2, with barriers of 0.8 kcal/mol for the G1 to G2 transformation. This transition state is, however, much lower in energy than T (ΔE = 6.0 kcal/mol). The TS2 and TS3 transition state energies between T/G2 and T/G1 are 8.3 and 10.2 kcal/mol, respectively, providing barriers of 2.3 and 4.2 kcal/mol for the T to G2 and T to G1 rotations, respectively. Finally, the G1 to T barrier turns out to be 9.3 kcal/mol, as can be derived from Table 1 and Figure 3a.

From the point of view of a conformational analysis, determination of the potential energy as a function of the torsion angles is the most critical task. Due to the importance of diffuse functions when calculating relative energies with DFT,[33] B3LYP/6-311++G** optimizations for the local energy minima and TS have been carried out starting from the corresponding B3LYP/6-31G* structures. Figure 3a shows the energy profiles as a function of the $\varphi_2 = C_3C_\beta C_\alpha N$ torsion angle at the B3LYP/6-31G*, B3LYP/6-311++G**// B3LYP/6-31G*, and B3LYP/6-311++G** levels in vacuo. The rotational curves, with a very similar 3-fold maximum−minimum pattern separated by about 120°, show limited variations in the barrier heights among the various descriptions. For the TS3 barrier at φ_2 = 240°, the values in kcal/mol with the 6-31G*/6-311++G** basis sets are 9.3/8.5 (G1 to T) and 4.2/3.6 (T to G1), respectively. For the TS2 barrier at φ_2 = 120°, the relevant values in kcal/mol are 2.3/2.1 (T to G2) and 8.3/7.7 (G2 to T), respectively. Finally, the TS1 barrier in kcal/mol at φ_2 = 0° (or 360°) is 1.6/1.9 (G2 to G1) and 0.8/1.3 (G1 to G2) at the two levels (Table 1).

The basis set effect is almost insignificant on the geometries, whereas relative energies change by 0.2−1.0 kcal/mol, primarily for the TS structures. The larger basis set

generally decreases the relative energies with the exception of the TS1 conformer energy. Nonetheless, Table 1 suggests unequivocally that the G2 structure is the lowest energy conformer of the six ones, and the other gauche structure is higher in energy by less than 1 kcal/mol. The trans, extended side-chain T structure has a relative energy of about 6 kcal/mol in the gas phase. The TS1 barrier separating G1 and G2 is relatively small, less than 2 kcal/mol. TS2 and TS3 barriers separating the T structure from any gauche conformers are, however, as high as 8 kcal/mol (Figure 3a).

The optimized values of the $C_2C_3C_\beta C_\alpha$ and $C_3C_\beta C_\alpha N$ torsion angles calculated at the two levels (reported in Table S1 in the Supporting Information) differ moderately, and the values are close to the classical torsion angles in a staggered or an eclipsed ethane conformer. Variations of the corresponding torsion angles at the different levels are generally less than 5°.

B3LYP/6-31G* optimizations were followed by frequency analysis, which confirmed the extreme point character of the six structures on the PES. Some selected normal frequencies are provided in the Supporting Information (Table S2), where the calculated IR spectra for cationic serotonin in the gas phase are also available (Figure S1). Similarly to our previous results for protonated gas-phase norepinephrine,[3a] the lowest three frequencies correspond to the $C_2C_3C_\beta C_\alpha$ torsion, the $C_3C_\beta C_\alpha$ bending, and the $C_3C_\beta C_\alpha N$ torsion, respectively. Although we calculated norepinephrine frequencies at the HF/6-31G* level, the B3LYP/6-31G* values are similar, and mainly the separation of the calculated frequencies resemble in the two calculations. This similarity suggests that low frequencies and their distribution are close to each other in the protonated, 2(aromatic-ring)-substituted ethylamines.

B. In-Solution Results. *1. Continuum Solvent.* In the aforementioned norepinephrine study[3a] we faced two problems when in-solution calculations were performed. We used gas-phase optimized geometries and utilized atomic charge parameters derived from these calculations for in-solution Monte Carlo simulations. Both problems are overcome in the present theoretical approach, which requires, however, much larger computational effort.

IEF-PCM/B3LYP/6-31G* geometry optimizations in aqueous solution led to small changes in the torsion angles as compared to the gas-phase structure. Bond lengths and angles vary also upon solvation, but the energy effect concomitant to a bond increase or shortening is considerably larger than a change of a few degrees in the torsion angles. In a recent study[20a] we found that the geometry distortion upon solvation, calculated at the PCM/B3LYP/6-31G* level, causes an increase in the internal energy (calculated in the gas phase at the in-solution optimal geometry) up to a few tenths of a kcal/mol for small molecules. This energy change is not negligible in the case of a conformational equilibrium for components in nearly equal concentration.

In-solution geometry optimizations are more important, however, to derive relevant net atomic charges to be applied in the 12-6-1 effective pair potential utilized in Monte Carlo simulations. As discussed earlier,[20] the BOSS program calculates the interaction energy of the solution elements by the implication that the atomic charge parameters represent a polarized electron distribution existing in the solution. The charge parameters for the TIP4P water model were developed accordingly.[26] In prior applications, atomic charges for the solute were derived based on a fit of the gas-phase HF/6-31G* electrostatic potential by Orozco et al.[34] The argument supporting the use of this procedure was that the HF/6-31G* basis set overestimates the gas-phase dipole moments by about 20%, so a fit to the corresponding molecular electrostatic potential could implicitly account for the polarized charge distribution of the solute in solution. This procedure, however, may not work well for every kind of solute molecules and does not account for the different charge distributions in different solvents. Conversely, the present approach is applicable for any molecule in any solvent.

The relative free energies for the different protonated serotonin conformers, as calculated based on IEF-PCM/B3LYP/6-31G* geometry optimizations as well as from IEF-PCM/B3LYP/6-311++G** and IEF-PCM/MP2/6-31G* single point calculations at the IEF-PCM/B3LYP/6-31G* optimized geometries, are compared in Figure 3b.

The gas-phase and in-solution curves in Figure 3 differ substantially. Although both types of curves show three maxima and minima, the barrier heights and stabilities of the local minima are quite different. In the gas phase, G2 is the most stable structure, and the T conformer corresponds to a high-energy local minimum. The barrier heights corresponding to the TSs occurring at $C_3C_\beta C_\alpha N$ values of around 120° and 240° are 8−10 kcal/mol. In contrast, IEF-PCM/B3LYP/6-31G* calculations predict the G1 structure to be the lowest free energy one, while T is higher in free energy only by 0.7 kcal/mol. The G2 relative free energy is 1.5 kcal/mol. The G1 to G2 barrier, TS1, increased to 4.3 kcal/mol, on the contrary the G2 to T (TS2) and T to G1 (TS3) barriers decreased to 3.8 and 3.7 kcal/mol, respectively. Overall, the in-solution relative conformational free energy curve shows a 3-fold potential with three similar energy minima in a 1.5 kcal/mol range, and the barriers are in a 3−4.5 kcal/mol range. Similar conclusions are obtained from the analyses of the other two in-solution potentials from single point calculations. The MP2/6-31G*//B3LYP/6-31G* results in the $60° < \varphi_2 < 300°$ region turn out to be somewhat less favorable than the B3LYP ones, probably because of the decrease of the dispersion interaction for the extended T form, not accounted for by B3LYP. A stabilization of about 1 kcal/mol for the trans conformer relative to a gauche one for 2-phenylethylamines seems to be a general feature at the B3LYP/6-31G* level in comparison to the MP2/6-31G* level, as we found either for protonated dopamine[2d] and norepinephrine[3a] or for neutral and zwitterionic tyramine (2-(4OH-phenyl)ethylamine) and dopamine in a recent study.[2g]

Using the IEF-PCM/B3LYP/6-31G* frequencies, the IR spectra of protonated serotonin have been calculated for the G1, G2, and T conformations (see Figure S2 in the Supporting Information). Some selected frequencies are provided in Table S3 in the Supporting Information. Comparison of the normal frequencies in aqueous solution with those in vacuo (Table S2) shows a remarkable lowering of the $C_3C_\beta C_\alpha N$ torsion frequency for the G2 conformer. From

Table 2. Unscaled Free Energy Correction Terms (kcal/mol) at the B3LYP/6-31G* Level Relative to the G2 Conformer in Vacuo and in Solution at $T = 310$ K and $p = 1$ atm

vacuo	ΔZPE	$\Delta(H(T)-\text{ZPE})$	$-T\Delta S_{tot}(T)$	$\Delta G_{th}(T)$
TS2	−0.26	−0.31	0.55	−0.02
T	−0.19	0.24	−0.86	−0.81
TS3	−0.29	−0.26	0.02	−0.53
G1	−0.13	0.04	−0.30	−0.39
TS1	−0.05	−0.38	1.16	0.73
water	ΔZPE	$\Delta(H(T)-\text{ZPE})$	$-T\Delta S_{tot}(T)$	$\Delta G_{th}(T)$
TS2	0.02	−0.49	1.29	0.83
T	0.13	0.03	0.12	0.28
TS3	0.08	−0.49	1.27	0.86
G1	0.17	−0.09	0.46	0.54
TS1	0.07	−0.51	1.48	1.04

Figure 3, the barriers are high enough to consider the motion as a vibration instead of a hindered rotation. The above finding has an important effect on the calculated relative thermal free energy contribution, $\Delta G_{th}(T)$, as defined in eq 1 (Table 2). Since the molecular weight is constant, and the moments of inertia change generally only a little for the substituted ethylamine conformers,[3a] the main contribution to $\Delta G_{th}(T)$ comes from the $T\Delta S_{vibr}(T)$ term for the local energy minimum structures. Within this term, the largest contributions come from the lowest frequency vibrations, which are, however, the least accurate and hence also contribute the largest error to the entropy. Comparison of the in vacuo and in-water $\Delta G_{th}(T)$ values confirm our previous finding. Whereas the in vacuo thermal corrections relative to the G2 conformers are negative values (with the exception of the +0.7 kcal/mol for the TS1 transition state), all $\Delta G_{th}(T)$ values become positive in solution.

The curves in Figure 3 do not contain thermal corrections. The use of the rigid rotator-harmonic oscillator model must be approximate even in the gas phase and probably even less adequate for solutes. Considering, however, that $\Delta G_{th}(T)$ refers to conformers of the same species, a remarkable cancellation error may be expected. By considering these correction terms, the G2-G1 and the T-G2 free energy separation decreases by 0.5 and 0.3 kcal/mol, respectively, providing total free energies for the G2 and T forms relative to the G1 form of 1.0 and 0.4 kcal/mol, respectively.

2. Monte Carlo Simulations. Two different charge sets, RESP and CHELPG, were used in MC simulations for each rotamer as stated in the Calculation section. They are visually compared in Figure S3, while a detailed numerical summary is provided in Tables S4 and S5 in the Supporting Information. The general shape of the two charge-alteration curves is similar. The largest difference in the *shape* is on the left-hand side of the plots referring to the C_3 to C_7 atoms, all belonging to the aromatic ring system (Scheme 1). RESP assigns to the C_2, C_7, and C_6 atoms nearly equal atomic charges (~ -0.2 charge units). For these atoms the CHELPG charges are within $\sim$0.3 charge units. In contrast, the RESP change amounts to $\sim$0.1 units for C_9 and C_8, which are almost constant from the CHELPG fit. Trends are similar for the C_6 to H_{g1} atoms. The CHELPG charges differ considerably for the amine nitrogen, N_{am}, and for the g2 gauche proton in the $-NH_3^+$ group as a function of the conformation. The conformation dependence for these atoms is smaller at the RESP level. Considering the numerical values for specific atoms, the largest difference was found for the pyrrole nitrogen, N_1, and C_5 to N_{am} atoms. Although the O_5 charge is similar in both sets, the C_5-O bond is more polarized with the CHELPG fit than with the RESP one.

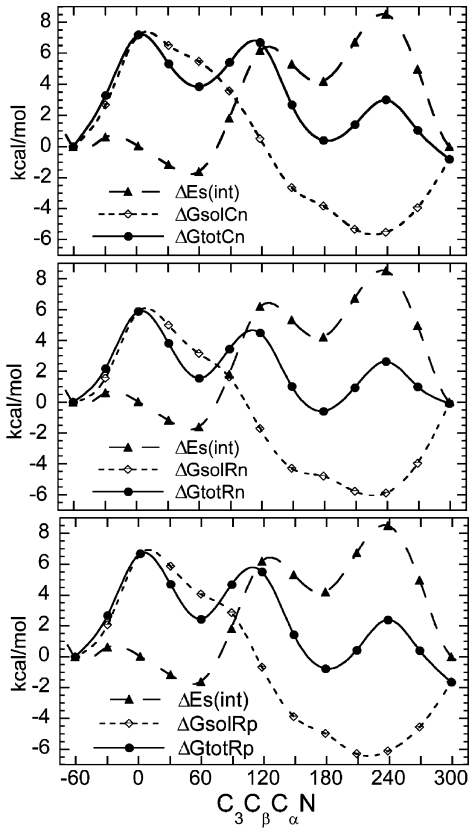


Figure 4. Conformational free energy differences compared using the three MC models. ΔE s(int) stands for ΔE^{s}_{int} (IEF-PCM/B3LYP/6-31G*) of eq 3a, ΔG_{sol}Cn (top) and ΔG_{sol}Rn (middle) stand for ΔG_{solv} in the $SerH^+\cdots Cl^-$ system with the CHELPG and RESP charges, respectively, while ΔG_{sol}Rp (bottom) stands for the ΔG_{solv} values in the $SerH^+$ system with the RESP charges. For ΔG_{tot}, the sum of ΔE^{s}_{int} (IEF-PCM) and ΔG_{solv}, the same notation as for ΔG_{solv} is used.

The largest difference in polarity was found for the $C_\alpha-N_{am}$ bond. The RESP polarity is $\sim$0.3 units, whereas the CHELPG value amounts to 0.7 units. Overall, the CHELPG as compared to the RESP fit results in a more polar charge distribution for protonated serotonin. This finding is consistent with the calculated ΔG_{solv} values in Figure 4.

Figure 4a−c shows the internal energy of the polarized solute, ΔE^{s}_{int} (IEF-PCM/B3LYP/6-31G*), ΔG_{solv}, and the total relative free energy in solution, ΔG_{tot}, computed as a summation of ΔE^{s}_{int} (IEF-PCM) and ΔG_{solv} for the three MC models studied. The ΔE^{s}_{int} (IEF-PCM) curve is similar to the in vacuo, nonsymmetric, 3-fold B3LYP/6-31G* potential in Figure 3a. The polarization of the solute in solution leads, however, to a deeper minimum at φ_2 ($C_3C_\beta C_\alpha N$ about 60° (G2 conformer), and the relative internal energy for the T conformer (φ_2 about 180°) is

Table 3. Total Relative Free Energies for the Protonated Serotonin Conformers Optimized at the IEF-PCM/B3LYP/6-31G* Level in Aqueous Solution

	continuum model (IEF-PCM) B3LYP/6-31G*	Monte Carlo		
		CHELPG	RESP	RESP (cation)
G1	0	0	0	0
TS1	4.2	7.2 ± 0.2	5.9 ± 0.1	6.7 ± 0.1
G2	1.5	3.9 ± 0.3	1.5 ± 0.3	2.4 ± 0.2
TS2	5.3	6.7 ± 0.4	4.5 ± 0.4	5.5 ± 0.2
T	0.7	0.4 ± 0.4	−0.6 ± 0.4	−0.8 ± 0.3
TS3	4.5	3.0 ± 0.4	2.6 ± 0.4	2.4 ± 0.3
	continuum model (IEF-PCM) MP2/6-31G*//B3LYP/6-31G*	Monte Carlo		
		CHELPG	RESP	RESP (cation)
G1	0	0	0	0
TS1	4.4	6.9 ± 0.2	5.6 ± 0.1	6.4 ± 0.1
G2	1.2	3.4 ± 0.3	1.1 ± 0.3	2.0 ± 0.2
TS2	6.0	7.3 ± 0.4	5.1 ± 0.4	6.1 ± 0.2
T	1.5	1.0 ± 0.4	0 ± 0.4	−0.2 ± 0.3
TS3	5.4	3.8 ± 0.4	3.4 ± 0.4	3.2 ± 0.3

reduced from 6.0 to 4.2 kcal/mol. The two TS occurring at φ_2 values of about 120° (TS2) and 240° (TS3) are less high by ∼2 kcal/mol each.

The ΔG_{solv} curves are similar in shape but different in amplitudes. Each of them have a maximum at $\varphi_2 = 0°$ (TS1) with relative values of 7.2, 5.9, and 6.7 kcal/mol, respectively, for a, b, and c simulations. The minima are in the $\varphi_2 = 210-240°$ range with −5.5, −5.9, and −6.3 kcal/mol, respectively, for a, b, and c. Thus ΔE^{s}_{int} (IEF-PCM) and ΔG_{solv} curves run oppositely. Slightly negative relative internal energies are combined with largely positive ΔG_{solv} values in the φ_2 range of −60° to 60°. In contrast, large positive ΔE^{s}_{int} (IEF-PCM) contributions are combined with strongly negative ΔG_{solv} terms for $\varphi_2 = 180°$ to 300°. It is important to remark that the solvation curve in Monte Carlo simulations is not sensitive to internal energy changes, thus it does not show a periodicity for local energy minimum and TS structures. Rather it depends on the relatively buried or solvation-exposed character of the $-NH_3^+$ group. Accordingly, ΔG_{solv} is positive in the $C_3C_\beta C_\alpha N$ torsion angle range of −60° to 60°, where the protonated group leans above the ring system and is the most negative in the $C_3C_\beta C_\alpha N = 180-240°$ range with an extended, nearly trans side-chain conformation. The solvation term reminds of a sine curve, irrespective of the parametrization or whether the chloride counterion has been included or not.

As a result, the $\Delta G_{tot} = \Delta E^{s}_{int}$(IEF-PCM) + ΔG_{solv}(MC) total relative free energy curve corresponds to a moderately nonsymmetric 3-fold potential in the entire 360° rotation range. The overall conclusion is that the three Monte Carlo solvation models as well as the IEF-PCM/B3LYP/6-31G* continuum solvent model are qualitatively equivalent but provide results differing in fine details. These details are important, however, in calculating the equilibrium conformer composition. Table 3 summarizes the relative free energies without considering thermal corrections. In addition, in the bottom part of the table, results based on IEF-PCM/MP2/6-31G*//IEF-PCM/B3LYP/6-31G* calculations are reported. With respect to the remarkable shift of the relative free energy for the T conformer at the two levels, the MP2 results deserve special attention.

All calculations agree that the G2 gauche conformer is of high relative energy and can be present in the equilibrium composition with only a small fraction. If ΔE^{s}_{int} is calculated on the basis of the IEF-PCM/B3LYP/6-31G* values, the corresponding continuum solvent results for the total relative free energies and the CHELPG MC predicts the G1 gauche conformer to be the most stable, but the relative CHELPG value of 0.4 ± 0.4 for the T trans conformation allows an almost equal G1 and T population. The RESP parametrization favors the T form both with and without considering the counterion. If the relative thermal corrections from Table 2 are also considered, all Monte Carlo models predict the trans conformation as the prevailing one in aqueous solution, whereas the IEF-PCM model still prefer the G1 form over the T conformation by 0.5 kcal/mol.

A different composition is predicted, however, on the basis of IEF-PCM/MP2/6-31G*//IEF-PCM/B3LYP/6-31G* calculations. At that level, the relative free energies predict a 1.5 kcal/mol destabilization of the T form as compared to the 0.7 kcal/mol gap obtained from the B3LYP/6-31G* results. If ΔE^{s}_{int} is calculated using the MP2 internal energies of the in-solution polarized solute, the subsequent CHELPG/MC predicts the prevalence of the G1 form in solution, in accord with the IEF-PCM calculations. The mean ΔG_{tot} value of T calculated with the RESP parametrization for the ion pair allows a nearly equal population for the G1 and T conformers in aqueous solution. The inclusion of thermal corrections leads to ΔG_{tot} values of 1.2, 0.7 ± 0.4, and −0.3 ± 0.4 kcal/mol for the T conformer at the IEF-PCM, CHELPG/MC, and RESP/MC levels, respectively. The RESP(cation) ΔG_{tot} value with thermal corrections for T is −0.5 ± 0.3 kcal/mol, favoring the T conformer over G1. The solution structure analysis (see next section) has pointed out, however, that the solvation of the $SerH^+$ cation only may exaggerate the stabilization of the trans conformer relative to the gauche $SerH^+$ conformations.

In summary, the theoretical models are sensitive to the

selected method (continuum or explicit solvent), to the theoretical level in continuum calculations (MP2 or B3LYP), and to the charge parametrization within the MC simulations. The predicted relative free energies for the conformers allow a sensitive equilibrium between the extended trans and the more globular gauche conformers in aqueous solution. In the absence of experimental results, the theoretical prediction cannot be controlled at present. The significant shift in the trans relative energy at the MP2/6-31G* vs B3LYP/6-31G* levels for 2-phenylethylamines supports, however, the MP2/6-31G* internal energies in comparison with the B3LYP/6-31G* values.[35] By calculating a trans−gauche energy separation of about 1 kcal/mol at the MP2/6-31G* level for neutral tyramine, Melandri and Maris[38] found a good accord between the theoretical and experimental compositions derived on the basis of the free-jet microwave spectrum. In contrast, B3LYP/6-31G* geometry optimizations lead to about equal energies for these conformers, whereas the conformer composition calculated on the basis of MP2/6-31G*//B3LYP/6-31G* results agree fairly well with the experimental one.[2g]

A comparison with earlier calculations shows that the conformational preference for 2-substituted ethylamines in aqueous solutions at pH = 7 is a resultant of a subtle balance between internal and solvation free energies. Histamine, dopamine, norepinephrine, and serotonin are protonated at their amine groups at least 92% at pH = 7.4.[3a] Protonated histamine prefers a gauche side chain ($C_{ring}C_\beta C_\alpha N$) conformation with an intramolecular N−H···N hydrogen bond as predicted theoretically[1c] and found experimentally.[1g] On the basis of Solmajer et al.[39] experimental study, the protonated dopamine favors a $C_{ring}C_\beta C_\alpha N$ gauche conformation at pH = 7, whereas these authors found the $C_{ring}C_\beta C_\alpha N$ trans arrangement as the most preferred conformation under similar conditions for norepinephrine. Our previous theoretical results for those systems were at least in qualitative accord with the experimental conformer populations.[2d,3a] All the above results correspond to free energy differences of a few tenths of a kcal/mol for the two most stable conformers. The present sophisticated approach slightly favoring the trans conformer in aqueous solution is in accord with the previous findings.

3. Solution Structure. The $O_{w(ater)}\cdots H(N_{am(monium)})$ radial distribution functions (rdfs) for the trans and gauche hydrogens of the $-NH_3^+$ group in the G1 conformation are compared in Figure 5. As was mentioned in the previous section, the $-NH_3^+$ group leans above the pyrrole ring in the G1 conformer. Since the $-NH_3^+$ group has a staggered conformation relative to the H_2C_α group, the trans hydrogen in the NH_3^+ group is still exposed to favorable solvation, whereas the gauche hydrogens point toward the ring system. This difference is clearly seen on the corresponding rdfs. The g(R) peak for $H_{t(rans)}$ is about twice as high as for the $H_{g(auche)}$ atom. The difference means remarkably more localized water solvent molecules around the trans hydrogen than the gauche one. The maximum of the peak occurs at a shorter distance for the trans than for the gauche hydrogen. This means that the favorable $O\cdots H_t$ separation is shorter than the $O\cdots H_g$ distance, which can also be explained by

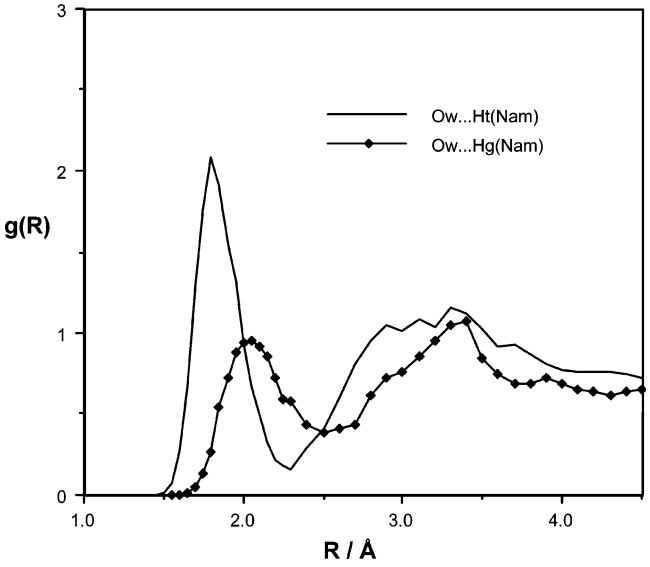


Figure 5. $O_w-H(N_{am})$ radial distribution functions for the trans and the gauche hydrogens of the $-NH_3^+$ group as calculated for the G1 conformer of the $SerH^+$ cation in the $SerH^+\cdots Cl^-$ system. RESP charge parametrization.

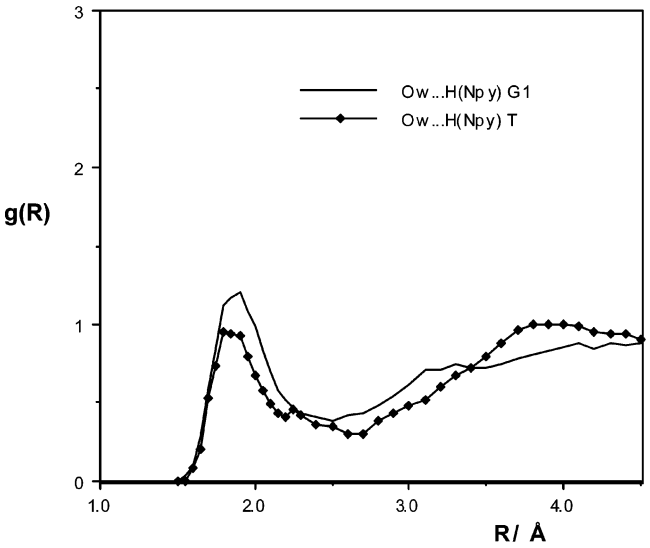


Figure 6. $O_w-H(N_{py})$ radial distribution functions for the G1 and T conformers of the $SerH^+$ cation in the $SerH^+\cdots Cl^-$ system. RESP charge parametrization.

steric effects. The rdf represents, however, only a short-range effect, because the curves run very close to each other for $R > 3.4$ Å. The difference in the hydration of gauche and trans $-NH_3^+$ hydrogens is considerable only for G conformations. In the T structure, the rdfs (not shown) have peak g(R) values in the range of 1.36−1.63 at R = 1.80−1.85 Å.

The hydration of the pyrrole hydrogen is less sensitive to the side-chain conformation (Figure 6). In fact, the $O_w\cdots H(N_{py(rrole)})$ rdf has a higher g(R) peak for the G1 than for the T conformer. A possible explanation is that water molecules are strongly bound to the protonated side chain, which leans above the pyrrole ring in the G1 conformation. As a secondary effect, these water molecules contribute to the hydration of the pyrrole hydrogen. Nonetheless, the peak values are in the range of 1.0−1.2, only slightly exceeding the value of g(R) = 1. On the basis of the definition of radial

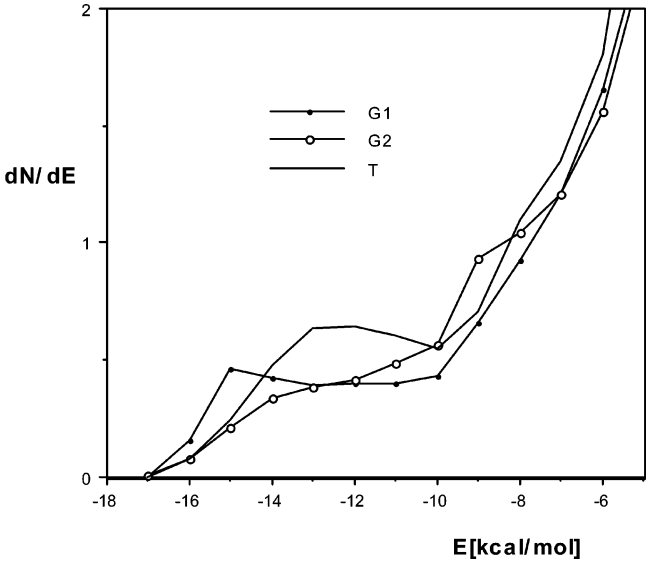


Figure 7. Pair-energy distribution functions for the G1, G2, and T conformers of the $SerH^+$ cation in the absence of the counterion. RESP charge parametrization.

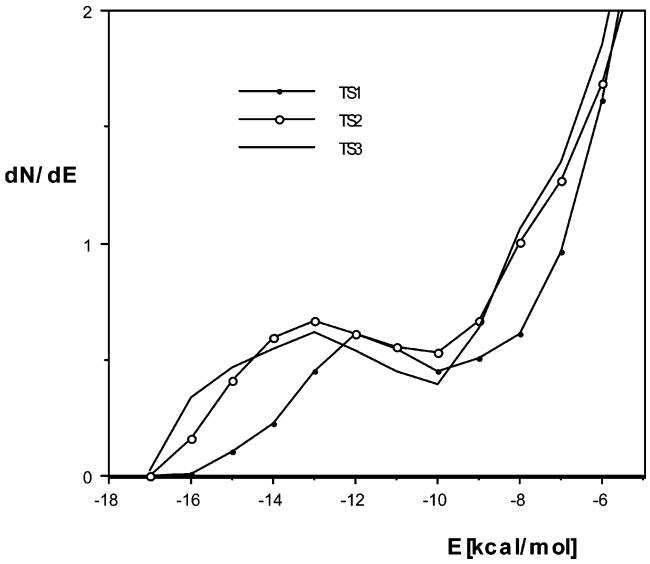


Figure 8. Pair-energy distribution functions for the TS1, TS2, and TS3 transition states of the $SerH^+$ cation in the absence of the counterion. RESP charge parametrization.

distribution functions, this means that the probability of finding a water oxygen at a distance of about 1.9 Å from the pyrrole hydrogen is up the 20% larger than finding a water molecule anywhere in the bulk solvent.

Figures 7 and 8 compare the pair-energy distribution functions for the $SerH^+$ cation (thus for the system without the chloride counterion) calculated with the RESP charges. The pair energy distribution functions (pedfs), dN/dE as a function of E, give the number of water molecules (N) in energy interaction of $E \pm \Delta E$ with the solute, where ΔE was set to 0.5 kcal/mol in the present study. The pedfs in Figure 7 refer to the G1, G2, and T structures with local internal energy minima. Their general trend is essentially similar: all of them start at $E = -17$ kcal/mol. The G1 curve rises more rapidly and is twice as high as the other two at $E = -15$ kcal/mol. It means that the strongest water−solute

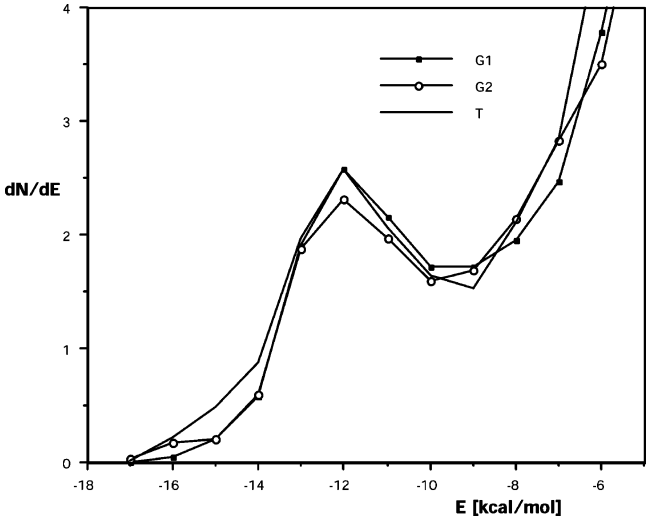


Figure 9. Pair-energy distribution functions for the G1, G2, and T conformers of the $SerH^+$ cation in the $SerH^+\cdots Cl^-$ system. RESP charge parametrization.

interactions are established for a gauche $C_3C_\beta C_\alpha N$ side-chain conformation. However, the G1 pedf does not increase further on until $E = -9$ kca/mol. In contrast, the pedf is highest for the T conformer in the E range of -14 to -10 kcal/mol, which means that T is solvated by the largest number of water molecules with interaction energy of -14 to -10 kcal/mol. Accordingly, this energy range has the largest contribution to the integral of the corresponding pedf, providing the largest N_{HB} value for the trans conformer in a forthcoming table. For less negative E values every pedf rises steeply.

It is interesting to observe that the pedfs for TS structures (Figure 8) do not substantially differ from those for local internal energy minima conformers. Pedfs for TS start also at $E = -17$ kcal/mol and reach a slight maximum in the -17 to -10 kcal/mol interaction range, and after this breaking point the curves sharply increase. Both in Figures 7 and 8, the dN/dE values are around 0.5 in the range of $E = -14$ to -10 kcal/mol. These similarities in the pedf values for TS and local internal energy minimum structures support the explanation given in the previous section that the solvent effect is more determined by the molecular geometry than by the internal energy. A very implicit effect still must be there: the electrostatic solute−solvent interaction energy at a given geometry depends only on atomic charges, and atomic charges were derived by fitting the molecular electrostatic potential, which, however, depends on the internal structure of the solute.

Pedfs for the $SerH^+\cdots Cl^-$ solute system are compared in Figure 9. These pedfs for G1, G2, and T conformers are completely different from those in Figure 7. Although they also start at $E = -17$ kcal/mol, all pedfs have a sharp peak at $E = -12$ kcal/mol and a minimum in the -10 to -9 kcal/mol range. The basic difference of pedfs calculated in the presence or absence of the chloride ion indicates the importance of considering a counterion in the modeling process. Based only on the existence of such a difference, one still could not make a choice. Since, however, under

Table 4. Coordination Numbers (n_{Coo}) and Numbers of Hydrogen Bonds (N_{HB}) in the Presence and Absence of a Chloride Counterion[a]

		$SerH^+\cdots Cl^-$	$SerH^+$
		G1	
n_{Coo}	Cl/H	7.5	
	$H(N_{am})/O^b$	1.1	0.9
	H(O)/O	0.9	1.1
	$H(N_p)/O$	1.1	1.0
N_{HB}		10.9 ($E_{max} = -9$)	2.7 ($E_{max} = -10$)
		G2	
n_{Coo}	Cl/H	7.6	
	$H(N_{am})/O^b$	1.0	1.0
	H(O)/O	1.0	1.0
	$H(N_p)/O$	1.0	1.0
N_{HB}		10.4 ($E_{max} = -9$)	2.5 ($E_{max} = -10$)
		T	
n_{Coo}	Cl/H	7.5	
	$H(N_{am})/O^b$	1.1	1.3
	H(O)/O	1.0	1.0
	$H(N_p)/O$	1.0	1.0
N_{HB}		11.4 ($E_{max} = -9$)	3.2 ($E_{max} = -10$)

[a] For the sake of comparison, all values are relevant to MC simulations with RESP charges; E_{max} in kcal/mol. [b] Average value.

real circumstances the overall system must be neutral, the inclusion of a counterion would provide a more realistic model.

The above conclusion may be, however, only conceptual and may affect only fine details of the simulations. Some solution structure characteristic numbers, compared in Table 4, show small differences for the solute solvation in the presence or absence of the counterion. The chloride by water hydrogen coordination number (Cl/H) hardly changes upon conformational variations in the solute (its calculated value is stably 7.5−7.6).

Only small changes have been computed in the average $H(N_{am})/O_w$ ($H(N_{am})/O$), H(phenolic)/O_w (H(O)/O), and $H(N_{py})/O_w$ ($H(N_p)/O$) coordination numbers for G1 and G2 solute conformers. All these values are 1.0 ± 0.1. The only "large" difference, 0.2 units, was calculated for the $H(N_{am})/O$ coordination number in the T conformation. Since the $C_3\cdots Cl$ distance was set to 12 Å, the distance between the $SerH^+$ and the anion is fairly large in most orientations of the side chain. Nonetheless, the extended side chain, thus the T conformer, can feel the effect of the counterion at a great degree in some arrangements. The decrease of the $H(N_{am})/O$ value of 1.3 for $SerH^+$ to 1.1 in the $SerH^+\cdots Cl^-$ system suggests that the two ionic sites compete in the strong localization of water molecules. In the absence of a chloride ion, $SerH^+$ could localize a larger number of water molecules with its extended side chain. The competition is not obvious for G1 and G2 conformers, where the side chain leans above the ring system in such a position that the hydration is less favorable even in the absence of the counterion.

Numbers of strong hydrogen bonds, N_{HB}, were calculated by the integration of the pedfs up to the first minima or until the ends of the plateaus. The upper energy limits are indicated in parentheses in Table 4. If $E_{max} = -9$ kcal/mol is allowed for the $SerH^+$ systems as well, the computed numbers are 3.3, 3.4, and 3.9 for G1, G2, and T conformers, respectively. Subtracting these values from the N_{HB} values for $SerH^+\cdots Cl^-$, nearly the corresponding Cl/H are obtained. Thus, the large difference in the N_{HB} numbers is related to the hydrogen bonds connected to the chloride ion. Since this difference is either equal to the Cl/H coordination numbers within the rounding error (G1 and T) or slightly less than the Cl/H value, one can interpret the result in this way: all hydrogens in the first shell (about 7.5) are H-bonded to the chloride with the G1 and T conformations, or about 7 hydrogen bonds are there with the G2 solute rotamer. The number of hydrogen bonds according to this calculation is just 3.3−3.4 units for G conformers, and N_{HB} is slightly larger for the more favorably hydrated T conformer. Using the more severe integration limit of $E = -10$ kcal/mol, 2.5−2.7 H-bonds are obtained for G conformers and 3.2 for the T form. The difference in the number of the hydrogen bonds for G and T conformers is maintained in the two calculations.

The N_{HB} numbers in Table 4 were calculated by an integration till $E = -9$ kcal/mol at most. The comparison of the values for the two models indicates the chloride as hydrated by 7−8 water molecules (almost all of them) H-bonded to it. There are about three *very* strong hydrogen bonds ($E = -17$ to -9 kcal/mol) to the protonated serotonin solute, whose primary location sites have been identified here, in accord with previous studies,[1c,2d,3a] in the first hydration shell of the $-NH_3^+$ group. There are, however, at least two additional hydrogen-bonding sites in the molecule: the phenolic (aromatic) OH and the pyrrole NH site. Both sites were studied by us previously: the phenolic $OH\cdots O_w$[40] and pyrrole $NH\cdots O_w$[41] hydrogen bond energies were found to be less favorable than those involving the ammonium group (i.e. less negative than about −9 kcal/mol). The pedfs are monotonically increasing after this limit, while no separate and resolved maximum−minimum character can be seen for these interactions, because solute−water interaction energies with water molecules in the second hydration shell appear also in the indicated energy range. Nonetheless, the clear shoulder for G2 and the less obvious one for T (see Figure 7), located in the energy range −10 to −7 kcal/mol, are assigned to the nonionic $H\cdots O_w$ interactions including the two mentioned sites.

C. Docking Studies. The 2-(aromatic ring) substituted ethylamine neurotransmitters (histamine, dopamine, norepinephrine, epinephrine, and serotonin) may have their biochemical specificities simply due to the chemical difference in the aromatic ring, ring or side-chain substituents (norepinephrine and epinephrine have a 2-OH group, as well), different protonation abilities of the amine group (this group is an N-methyl secondary amine group for epinephrine) or due to the differences in the favorable conformation(s) of the molecules. This latter may also be a consequence of the chemical character of the 2-substituent.

In Figure 10, our optimized structures for dopamine and norepinephrine (from previous studies) and for serotonin are superimposed in the three main side-chain conformations. In the top part of the figure, the catechol ring and the phenol substructures of the 5-OH indole system are superimposed, while in the bottom part the side chains. Whereas the

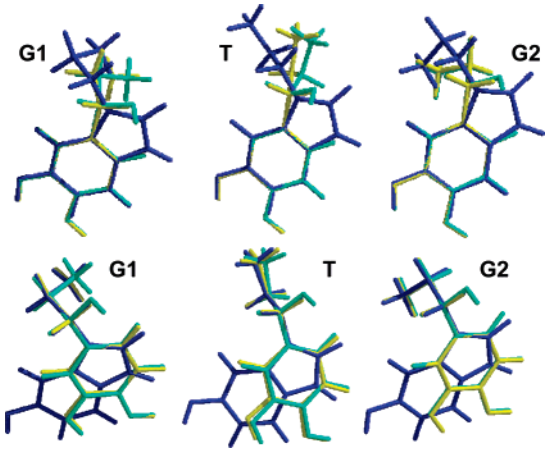


Figure 10. Optimized structures for protonated serotonin (blue) in the three main side-chain conformations, superimposed to those for dopamine (yellow) and norepinephrine (cyan) derived from our previous studies, on: (top) catechol rings and the phenol substructure of 5-OH indole; (bottom) $C_2C_3C_\beta C_\alpha$.

dopamine and norepinephrine molecules, differing only in the 2-OH group, are in both cases very well superimposable, superimposition of serotonin to these molecules fails. The superimposition is better for the side-chain fit, indicating that the three molecules have similar stable conformations with G1, T, and G2 side-chain arrangements. Thus we conclude that the neurotransmitter specificity for these molecules lies in the difference of the chemical structure: additional 2-OH for the dopamine−norepinephrine pair, catechol versus the 5-OH indole ring for the dopamine−serotonin pair. This difference may not be confined to the direct difference in the interactions with the receptor side chains. The chemical difference may have an implicit effect, as well: the bioactive conformation may (and probably does) differ from the local free energy minimum structures, and the readiness for a distortion to the bioactive conformation may also depend on the chemical structure.

The following question may, however, emerge: is our complicated combined DFT/ab initio/MC approach necessary at all for determining the in-solution relevant conformations and relative free energies. For a comparison, molecular mechanics geometry optimizations and energy calculations have also been performed using three values for the dielectric constant, ϵ (4r, 1r, and 1, respectively), shown in Table 5. The basic conclusion considering the values from Tables 1−3 is that the relative T and G2 free energies considerably decrease and increase, respectively, as compared to the gas-phase internal energies. In molecular mechanics optimizations, the three different forms for ϵ are decisive, because even with an NB cutoff = 8.0 Å almost all nonbonded intramolecular interactions were considered in the $SerH^+$ cation. The $\epsilon = 4r$ functional form mimics the solvent effects to a greater extent than the other two approximations. Since the CHELPG charges were derived by means of the in-solution wave function, this charge set should be most logically combined with the $\epsilon = 4r$ parameter. The derived relative energies (first set) show small variations for T and G2 with the Tripos force field; therefore, these MM calculations cannot reproduce the trend in the relative conformational free energy in aqueous solution. The MMFF94 relative energy for T is in fair agreement with the DFT/MC value. In this case, however, the MMFF94 charges were used, which did not depend on the conformation and sometimes differ from the CHELPG charges very remarkably. For example, the CHELPG charges for N_{am} varied between −0.21 and −0.35, whereas the MMFF94 value is −0.85. Not even the sign agrees for $N_{pyrrole}$ (the CHELPG values varied in the range −0.40 to −0.42 as compared to an MMFF94 charge of +0.03).

The charge sensitivity of the relative free energy in solution clearly appears from the difference in the DFT/MC and MMFF94 predicted G2 values. In contrast, the MMFF94 parametrization reproduces very well the B3LYP/6-31G* gas-phase relative energies.

The torsional angles can be predicted with a deviation of 10° at most with both MM methods. Changes in the ϵ function lead to a remarkable shift of the calculated key torsional angles. By comparing the first sets with the IEF-PCM/B3LYP/6-31G* values and the third set with the B3LYP/6-31G* gas-phase optimized dihedral angles, the trends are reproduced. Overall, while the conformer geometries can be fairly determined for protonated serotonin, the relative free energies have not been reproduced with either MM force fields and parametrizations.

For examining the serotonin···5-HT2A receptor interactions, six protonated serotonin conformers were considered in the present modeling. Since the optimized stable conformers take geometries without a symmetry plane, all these conformers exist in pairs of mirror images with equal energies. The receptor, however, has a set stereochemistry, thus binding energies of the elements of the mirror-image conformers must be in general different due to different geometric fits to the receptor binding site. Two modes of the ligand binding in conformations of G1 and G2′ are shown in Figure 11. Calculated geometric parameters as well as interaction and distortion energies are summarized in Table 6.

Throughout the successive energy minimization, molecular dynamics step and repeated energy minimization of the ligand···receptor complexes, the ligands generally saved their original G1, G2, and T as well as the corresponding mirror image characters even in the bound complex (Table 6). In the altogether 29 different cases, however, remarkable G1′ to T′, T′ to G2′, G2′ to G1′, and G2′ to T′ conformational changes were noticed, as well.

In a molecular mechanics study only comparisons among distinct conformations of the same system have physical meaning. Thus, the properly calculated differences may have physical relevance. In Table 6, all EI values calculated on the basis of eq 5 are negative, indicating a stabilizing interaction. Those values, however, depend on the parametrization; therefore, a more realistic prediction can be made from the ΔEI values. By using this procedure, the relative binding energies of the different conformers can be obtained, allowing assignment of the more likely binding form. The most negative EI interaction energies are in the range of −35 to −43 kcal/mol with different ligand conformations. In its strongest binding mode, each conformer forms at least one

Table 5. Relative Energies/Free Energies (kcal/mol) and Torsion Angles (deg) for Protonated Serotonin from Different Approaches

	relative energies or free energies		
	G1	T	G2
ΔE(B3LYP/6-31G*), gas phase	0.	5.1	−0.8
ΔG(IEF-PCM/B3LYP+MC/CHELPG charges), solution	0.	0.4± 0.4	3.9 ± 0.3
ΔE(Tripos force-field, CHELPG charges)[a]	0.	−0.4, −2.2, −8.0	−0.2, −2.1, −7.0
ΔE(MMFF94 force-field and charges)[a]	0.	0.9, 2.9, 5.3	−0.1, −0.3, −1.0

	torsion angles		
	G1	T	G2
B3LYP/6-31G*, gas phase			
$C_3C_\beta C_\alpha N$	−52.1	173.2	57.7
$C_2C_3C_\beta C_\alpha$	113.1	107.0	94.3
IEF-PCM/B3LYP/6-31G*, solution			
$C_3C_\beta C_\alpha N$	−61.0	178.5	59.4
$C_2C_3C_\beta C_\alpha$	105.7	102.3	84.5
Tripos force-field, CHELPG charges[a]			
$C_3C_\beta C_\alpha N$	−58.4, −53.0, −43.5	179.5, 179.2, 177.1	61.2, 58.5, 57.7
$C_2C_3C_\beta C_\alpha$	110.5, 113.8, 122.8	107.9, 108.9, 111.1	90.4, 91.7, 95.4
MMFF94 force-field and charges[a]			
$C_3C_\beta C_\alpha N$	−66.4, −58.9, −49.8	180.1, 180.2, 179.0	67.3, 61.0, 57.8
$C_2C_3C_\beta C_\alpha$	104.0, 107.6, 116.1	96.6, 96.4, 102.5	80.8, 80.6, 85.7

[a] The three consecutive values provided for MM calculations were obtained in energy minimizations with parameters for the dielectric function, ϵ, and the nonbonded cutoff, NB as (ϵ=4r, NB=8.0 Å), (ϵ=1r, NB=12.0 Å), and (ϵ=1, NB=12.0 Å), respectively.

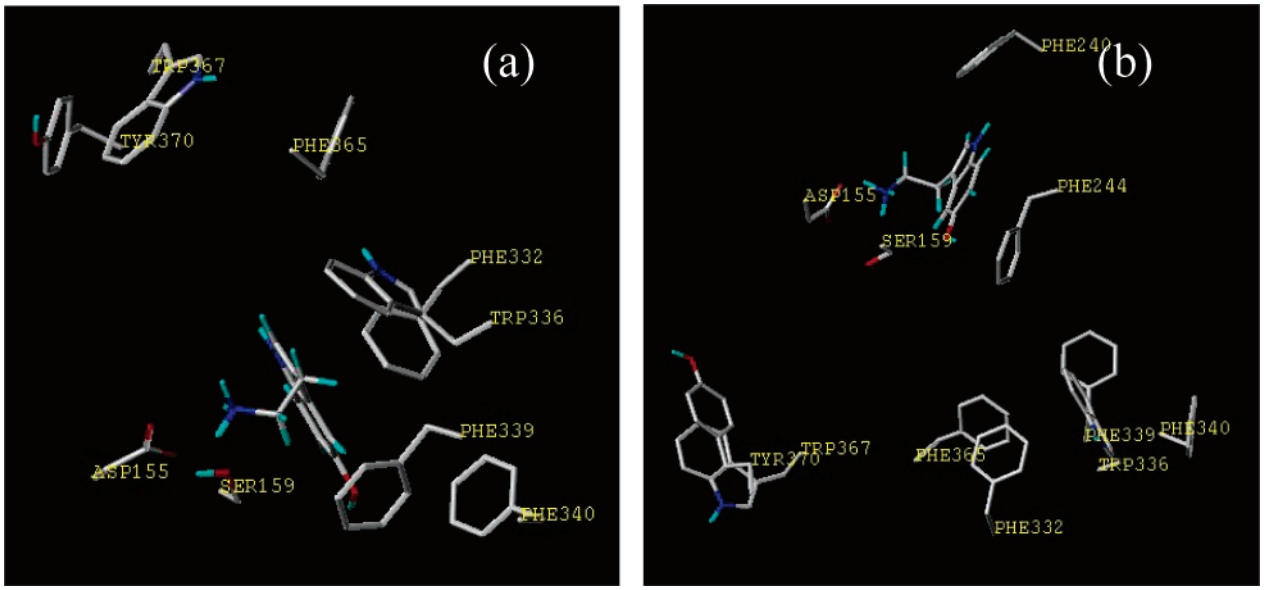


Figure 11. Receptor-bound forms of the ligand in the G1 (a) and G2′ (b) conformations.

strong $-N_{am}H^+\cdots{}^-OOC-$(Asp155) hydrogen bond with O···H distance of 1.6−2.0 Å. In the absence of such a bond, the EI values are remarkably less negative. These poses were found from starting arrangements involving $N_{am}\cdots{}^-OOC-$(Asp155) = 5−7 Å. Unless the $-N_{am}H^+\cdots{}^-OOC$ hydrogen bond was reestablished, the binding energy decreased (EI increased). The EI values form a range of 24 kcal/mol, indicating that each conformer has stronger and weaker binding modes.

Since in the strongest binding modes the different conformers of the $SerH^+$ ligand always form the $NH_3{}^+\cdots{}^-OOC$ hydrogen bond, the differences are attributed to the presence or absence of remarkable van der Waals interactions between the amino acid side chains and aromatics rings. Such an interaction was accepted as meaningful if at least one C···C distance between the indole ring of the ligand and the aromatic ring of the protein was less then 4 Å. In their most stable interactions with the receptor, the T′, G2′, and G2 conformers exhibited van der Waals interactions with amino acid residues Phe 240, 244 (TM V), Phe 339 (TM VI) and Trp 336, Phe 339, 340 (TM VI), respectively. In some binding modes of the G2′ and G1′ conformers (EI = −35.8 and −34.0 kcal/mol, respectively), the ligand interacts with the Phe365 and Tyr370 side chains, respectively, from TM VII. In weaker binding modes, van der Waals interactions

Table 6. Molecular Mechanics Energy (kcal/mol) and Geometry Terms Related to the Serotonin···5-HT2A Receptor-Model Interactions

	−EI	E(lig,dist)	ΔEI
G1	23.2−35.4	0.2−5.5	7.7−19.9
G1′	29.4−34.9	0.3−2.6	8.2−13.7
G2	29.9−41.0	0.5−1.5	2.1−13.2
G2′	32.8−41.4	1.0−2.0	1.7−10.3
T	29.3−37.4	0.3−3.3	5.7−13.8
T′	19.3−43.1	0.7−4.4	0.0−23.8

	torsion angles (deg)			
	ligand in adduct		optimized ligand	
	$C_2C_3C_\beta C_\alpha$	$C_3C_\beta C_\alpha N$	$C_2C_3C_\beta C_\alpha$	$C_3C_\beta C_\alpha N$
G1	95−173	−54 − (−72)	110.5	−58.4
G1′	−112 − (−129)	50−56	−110.5	58.4
G2	80−98	52−64	90.4	61.2
G2′	−88 − (−99)	−57 − (−82)	−90.4	−61.2
T	102−134	171−194	107.9	179.5
T′	−62 − (−143)	166−203	−107.9	−179.5

were noticed with Phe243 (TM V) and Trp367 (TM VII) residues, as well.

Two binding modes of the ligand in the G1 and G2′ conformations, establishing two medium-strong interactions (EI = −35.4 and −39.5 kcal/mol, respectively) with the receptor, are shown in Figure 11a,b. The cationic head of the ligand points toward the carboxylate group of Asp 3.32 (Asp155) in both cases. This interaction anchors the ligand and allows a motion of the rest of the molecule with a basically fixed location for the $-NH_3^+$ group, which is maintained in its place also by the interaction with Ser 3.36 (Ser159) found in experimental studies.[30] The indole ring interacts (Figure 11a) with several aromatic rings in the neighborhood, including Phe339, Phe340, Trp336, and Phe332 (TM VI), but it is far away from the aromatic part of Phe 7.38 (Phe365), Trp 7.40 (Trp367), and Tyr 7.43 (Tyr370). As mentioned above, interactions with these residues were found with some other ligand conformations. All these interactions were found important based on experimental studies by Roth et al.[29] In Figure 11b, as anticipated above, the side chain is still interacting with Asp and Ser, but the indole ring has moved to a different location because of its almost opposite φ_1 value: now the closest aromatic rings in the neighborhood are Phe244 and Phe240 in TM V.

As a summary of the docking studies, the rhodopsin-based 5-HT2A receptor model accounts for all the important ligand−receptor interactions found experimentally. The calculations provide an about 24 kcal/mol energy range for different binding modes, allowing G1, G2, and T conformations and their mirror images for the ligand. The binding mode and ligand conformation producing the most negative interaction energy do not necessarily correspond to the biologically active ligand conformation and the receptor activating binding mode. The diverse interaction energies and ligand conformations obtained in the docking study indicate, however, that neither steric hindrance nor activation problems should be expected when a protonated serotonin cation interacts with the 5-HT2A receptor in the binding cavity.

IV. Conclusions

Protonated serotonin takes two stable $C_{ring}C_\beta C_\alpha N$ gauche and one trans conformations in the gas phase. Transformation of the two gauche conformers by rotation about the $C_\beta-C_\alpha$ axis is hindered by barriers of about 2 kcal/mol, as calculated at the DFT level of theory (B3LYP) when the 6-31G* and 6-311++G** basis sets were applied. The gauche conformers are separated in energy by about 1 kcal/mol. The trans form is higher in energy by 6 kcal/mol than the most stable G2 conformer and is separated from the gauche structures by barriers of about 8 kcal/mol.

The in-solution conformational analysis was performed at the IEF-PCM level and by considering explicit solvent molecules within Monte Carlo simulations. Optimized geometries and internal energies were obtained at the IEF-PCM/B3LYP/6-31G* level. All calculations agree that two gauche and one trans side-chain conformations, at torsional angles close to those in the gas phase, represent the local minimum energy structures even in solution. A remarkable difference compared to the gas phase is, however, that the G1 to G2 barrier increased to 4 kcal/mol, whereas the T conformer is separated from the gauche forms by 4 kcal/mol. If an estimate for the relative thermal corrections is also considered, then the IEF-PCM/B3LYP/6-31G* calculations favor a gauche conformer, whereas the Monte Carlo simulations predict the trans form as the prevailing conformation. By considering, however the IEF-PCM/MP2/6-31G*//IEF-PCM/B3LYP/6-31G* results and the derived relative internal energies for the conformers, most calculations predict a nearly equal population for the G1 and T conformers in aqueous solution. Although any calculation predicts a conformational equilibrium in aqueous solution with detectable fractions for all three conformers, the authors prefer the MP2//B3LYP results because of the capacity of this theoretical level to predict conformer populations in fairly good agreement with the experimental results for a representative set of the 2-phenylethylamine family.

The solute is strongly hydrated by about three water molecules around the $-NH_3^+$ group with −9 to −17 kcal/mol interaction energy. Two water molecules, one for each, are expected to hydrate the pyrrole and the phenolic hydrogen atoms, as concluded from the analysis of the radial distribution functions.

Docking studies of the protonated ligand predicted both gauche and trans ligand conformers to favorably interact with the 5-HT2A receptor in its hypothesized binding cavity. The theoretical studies confirm the experimental results regarding strong interactions with the Asp155 and Ser159 residues (TM helix III) and the interactions of the indole ring with Phe, Trp, and Tyr side chains in TM V, VI, and VII helices. Neither steric hindrance nor activation problems should be expected when the protonated serotonin conformers interact with the 5-HT2A receptor with relative interaction energies up to about 24 kcal/mol.

Acknowledgment. The authors are indebted to Professor Jorgensen for permission to use the BOSS 4.2 program.

Supporting Information Available: B3LYP/6-31G* IR spectra in vacuo (Figure S1) and in solution at the IEF-PCM/B3LYP/6-31G* level (Figure S2) for G1, T, and G2 protonated serotonin, B3LYP optimized torsion angles for protonated serotonin in the gas phase (Table S1), selected wavenumbers in vacuo at the B3LYP/6-31G* level for protonated serotonin (Table S2), selected wavenumbers at the IEF-PCM/B3LYP/6-31G* level for protonated serotonin in aqueous solution (Table S3), and IEF-PCM/B3LYP/6-31G* CHELPG and RESP charges in aqueous solution (Figure S3) for the conformers considered (Tables S4 and S5, respectively). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Vogelsander, B.; Godfrey, P. D.; Brown, R. D. *J. Am. Chem. Soc.* **1991**, *113*, 7864−7869. (b) Worth, G. A.; Richards, W. G. *J. Am. Chem. Soc.* **1994**, *116*, 239−250. (c) Nagy, P. I.; Durant, G. J.; Hoss, W. P.; Smith, D. A. *J. Am. Chem. Soc.* **1994**, *116*, 4898−4909. (d) Karpinska, G.; Dobrowolski, J. C.; Mazurek. A. P. *J. Mol. Struct. (THEOCHEM)* **1996**, *369*, 137−144. (e) Godfrey, P. D.; Brown, R. D. *J. Am. Chem. Soc.* **1998**, *120*, 10724−10732. (f) Kovalainen, J. T.; Christiaans, J. A. M.; Ropponen, R.; Poso, A.; Perakyla, M.; Vepsalainen, J.; Laatikainen, R.; Gynther, J. *J. Am. Chem. Soc.* **2000**, *122*, 6989−6996. (g) Kraszni, M.; Kokosi, J.; Noszal, B. *J. Chem. Soc. Perkin 2*, **2002**, 914−917. (h) Ramirez, F. J.; Tunon, I.; Collado, J. A.; Silla E. *J. Am. Chem. Soc.* **2003**, *125*, 2328−2340. (i) Raczynska, E. D.; Darowska, M.; Cyranski, M. K.; Makowski, M.; Rudka, T.; Gal, J. F.; Maria, P. C. *J. Phys. Org. Chem.* **2003**, *16*, 783−796.

(2) (a) Urban, J. J.; Cramer, C. J.; Famini, G. R. *J. Am. Chem. Soc.* **1992**, *114*, 8226−8231. (b) Alagona, G.; Ghio, C. *Chem. Phys*. **1996**, *204*, 239−249. (c) Urban, J. J.; Cronin, C. W.; Roberts, R. R. Famini, G. R. *J. Am. Chem. Soc.* **1997**, *119*, 12292−12299. (d) Nagy, P. I.; Alagona, G.; Ghio, C. *J. Am. Chem. Soc.* **1999**, *121*, 4804−4815. (e) Aliste, M. P.; Cassels, B. K. *J. Chem. Soc., Perkin 2* **2001**, 906−915. (f) Alagona, G.; Ghio, C. *Int. J. Quantum Chem*. **2002**, *90*, 641−656. (g) Nagy, P. I.; Völgyi, G.; Takács-Novák, K. *Mol. Phys*. **2005**, *103*, 1589−1601.

(3) (a) Nagy, P. I.; Alagona, G.; Ghio, C.; Takács-Novák, K. *J. Am. Chem. Soc.* **2003**, *125*, 2770−2785. (b) Snoek, L. C.; Van Mourik, T.; Simons, J. P. *Mol. Phys*. **2003**, *101*, 1239−1248. (c) Snoek, L. C.; Van Mourik, T.; Carcabal, P.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4519−4526.

(4) Van Mourik, T.; Emson, L. E. V. *Phys. Chem. Chem. Phys*. **2002**, *4*, 5863−5871.

(5) (a) Goodman & Gilman's: *The Pharmacological Basis of Therapeutics*, 9th ed.; Hardman, J. G., Limbird, L. E., Molinoff, P. B., Goodman Gilman, A., Eds.; McGraw-Hill: New York, 1996. (b) Foye's: *Principles of Medicinal Chemistry*, 5th ed.; Williams, D. A., Lemke, T. L., Eds.; Lippincott, Williams & Wilkins: Baltimore, MD, 2002.

(6) (a) Tota, M. R.; Candelore, M. R.; Dixon, R. A. F.; Strader, C. D. *Trends Pharmacol. Sci.* **1991**, *12*, 4−6. (b) Trumpp-Kallmeyer, S.; Hoflack, J.; Bruinvels, A.; Hibert, M. *J. Med. Chem.* **1992**, *35*, 3448−3462. (c) Strader, C. D.; Fong, T. M.; Tota, M. R.; Underwood, D. *Annu. Rev. Biochem. Soc.* **1994**, *63*, 101−132. (d) Lu, Z.-L.; Saldanha, J. W.; Hulme, E. C. *Trends Pharmacol. Sci*. **2002**, *23*, 140−146.

(7) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739−745.

(8) Ballesteros, J. A.; Weinstein, H. *Methods Neurosci.* **1995**, *25*, 366−428.

(9) Liljefors, T.; Norrby, P.-O. *J. Am. Chem. Soc.* **1997**, *119*, 1052−1058.

(10) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789. (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(11) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(12) (a) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622. (b) Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem.* **1976**, *10s*, 1−19. (c) Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys*. **1980**, *72*, 4244−4245. (d) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968−5975.

(13) 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.

(14) McQuarrie, D. A. *Statistical Mechanics*; University Science Book: Sausalito, CA, 2000.

(15) Wong, M. W. *Chem. Phys. Lett.* **1996**, *256*, 391−399.

(16) (a) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys*. **1997**, *107*, 3032−3041. (b) Cancès, E.; Mennucci, B. *J. Chem. Phys*. **1998**, *109*, 249−259. (c) Cancès, E.; Mennucci, B. *J. Chem. Phys.* **1998**, *109*, 260−266.

(17) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.

(18) (a) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094. (b) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210−3221.

(19) Amovilli, C.; Barone, V.; Cammi, R.; Cancès, E.; Cossi, M.; Mennucci, B.; Pomelli, C. S.; Tomasi, J. *Adv. Quantum Chem.* **1998**, *32*, 227−261.

(20) (a) Alagona, G.; Ghio, C.; Nagy, P. I. *Int. J. Quantum Chem.* **2004**, *99*, 161−178. (b) Nagy, P. I.; Takács-Novák, K. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2838−2848. (c) Nagy, P. I. *J. Phys. Chem. B* **2004**, *108*, 11105−11117.

(21) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.

(22) (a) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280. (b) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc*. **1993**, *115*, 9620−9631.

(23) (a) Zwanzig, R. W. *J. Chem. Phys*. **1954**, *22*, 1420−1426. (b) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050−3054.

(24) Jorgensen, W. L. *BOSS version 4.2*; Yale University: New Haven, CT, 2000.

(25) (a) Jorgensen, W. L.; Madura, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 1407−1413. (b) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 1489−1496. (c) Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174−2182. (d) Jorgensen, W. L.; Briggs, J. M.; Contreras, M. L. *J. Phys. Chem.* **1990**, *94*, 1683−1686.

(26) (a) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935. (b) Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381−1392.

(27) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(28) Rajeswaran, W. G.; Cao, Y.; Huang, X. P.; Wroblewski, M. E.; Colclough, T.; Lee, S.; Liu, F.; Nagy, P. I.; Ellis, J.; Levine, B. A.; Nocka, K. H.; Messer, W. S., Jr. *J. Med. Chem.* **2001**, *44*, 4563−4576.

(29) Roth, B. L.; Shoham, M.; Choudhary, M. S.; Khan, N. *Mol. Pharmacol.* **1997**, *52*, 259−266.

(30) Almaula, N.; Ebersole, B. J.; Zhang, D.; Weinstein, H.; Sealfon, S. C. *J. Biol. Chem.* **1996**, *271*, 14672−14675.

(31) Sybyl 6.91, Tripos Inc., St. Louis, MO, 2003.

(32) (a) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490−519, 520−552, 553−586, 616−641. (b) Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, *17*, 587−615.

(33) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384−1388.

(34) (a) Orozco, M.; Jorgensen, W. L.; Luque, F. J. *J. Comput. Chem.* **1993**, *14*, 1498−1503. (b) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J. Comput. Chem.* **1993**, *14*, 1240−1249.

(35) One of the referees of this article raised the problems of MP2 "used with a small basis set", without hydrogen polarization, and of intramolecular BSSE that can be mistaken for dispersion. Concerning 6-31G* (a medium-size basis set), for protonated norepinephrine in the gas phase the MP2/6-31G* optimized structures and stabilities were fairly consistent with those obtained at the MP2/6-311++G** level.[3a] Furthermore those latter results were almost exactly superimposed to the minima of the MP2/6-311++G**//HF/6-31G*curve. An analogous trend was obtained by Nielsen et al.[36] for a related substituted ethylammonium in water, where extending the basis set from 6-31+G* to 6-311+G** did not lead to any significant difference. In that case, however, it was necessary to use diffuse functions to describe the anionic head of the zwitterion. Concerning BSSE, actually MP2 is known to be more prone than HF to intermolecular BSSE at least, error that remains almost constant despite basis set enlargement, as observed also in polyatomic cation−water interactions.[37]

(36) Nielsen, P. A.; Norrby, P.-O.; Liljefors, T.; Rega, N.; Barone V. *J. Am. Chem. Soc.* **2000**, *122*, 3151−3155.

(37) Alagona, G.; Biagi, A.; Ghio, C. *Mol. Eng.* **1992**, *2*, 137−152.

(38) Melandri, S.; Maris, A. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2863−2866.

(39) Solmajer, P.; Kocjan, D.; Solmajer, T. *Z. Naturforsch.* **1983**, *38c*, 758−762.

(40) Nagy, P. I.; Dunn, W. J., III.; Alagona, G.; Ghio, C. *J. Phys. Chem.* **1993**, *97*, 4628−4642.

(41) Nagy, P. I.; Durant, G. J.; Smith, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 2912−2922.

CT050088C

JCTC Journal of Chemical Theory and Computation

Extension of the PDDG/PM3 Semiempirical Molecular Orbital Method to Sulfur, Silicon, and Phosphorus

Ivan Tubert-Brohman, Cristiano Ruch Werneck Guimarães, and William L. Jorgensen*

Department of Chemistry, Yale University, 225 Prospect St., New Haven, Connecticut 06520-8107

Received February 11, 2005

Abstract: The PDDG/PM3 semiempirical molecular orbital method has been parametrized for molecules, ions, and complexes containing sulfur; the mean absolute error (MAE) for heats of formation, ΔH_f, of 6.4 kcal/mol is 35−40% smaller than those for PM3, AM1, and MNDO/d. For completeness, parametrization was also carried out for silicon and phosphorus. For 144 silicon-containing molecules, the ΔH_f MAE for PDDG/PM3, PM3, and AM1 is 11−12 kcal/mol, whereas MNDO/d yields 9.4 kcal/mol. For the limited set of 43 phosphorus-containing molecules, MNDO/d also yields the best results followed by PDDG/PM3, AM1, and PM3. The benefits of the d orbitals in MNDO/d for hypervalent compounds are apparent for silicon and phosphorus, whereas they are masked in the larger dataset for sulfur by large errors for branched compounds. Overall, for 1480 molecules, ions, and complexes containing the elements H, C, N, O, F, Si, P, S, Cl, Br, and I, the MAEs in kcal/mol for ΔH_f are 6.5 (PDDG/PM3), 8.7 (PM3), 10.3 (MNDO/d), 10.8 (AM1), and 19.8 (MNDO).

Introduction

Semiempirical methods based on the neglect of diatomic differential overlap (NDDO)[1] approximation, such as MNDO,[2] AM1,[3] PM3,[4] and MNDO/d,[5] occupy an important place in computational chemistry because of their speed and excellent scaling with increasing system size. Even as constant advances in computer resources permit application of ab initio and DFT methods to ever-larger systems of chemical interest, semiempirical methods allow the exploration of new frontiers such as full quantum mechanical calculations for proteins[6] or long Monte Carlo and molecular dynamics simulations of reactions in solution and in enzymes by means of coupled quantum and molecular mechanics (QM/MM).[7−16]

In recent articles, a new NDDO-based method, PDDG/PM3, was introduced.[17,18] It is derived from the PM3 method by the addition of small pairwise distance-directed Gaussians to the core repulsion function. The method was initially parametrized for the basic organic elements C, H, N, and O[17] and later extended to the halogens F, Cl, Br, and I.[18] The use of the PDDG function in conjunction with extensive parametrization using large datasets resulted in a reduction of about 30% in the mean absolute error (MAE) for heats of formation in comparison to PM3. Several systematic errors such as for homologation and branching were overcome, and improvements for activation barriers for S_N2 reactions involving halogens were also obtained. The PDDG/PM3 method yields heats of formation and isomerization energies that are more accurate than those obtained from B3LYP/6-31G* calculations, and for some classes of compounds, chemical accuracy is approached; for example, the MAE is 1.17 kcal/mol for alkanes. It has also been used successfully in QM/MM studies of nucleophilic aromatic substitution (S_NAr) and S_N2 reactions in solution.[7,8]

In this paper, the PDDG/PM3 method is extended to sulfur, silicon, and phosphorus. With these additions, the complete set of "organic elements" is available for calculations. This will allow the application of the PDDG/PM3 method to a wide variety of molecules of chemical and biological interest. Although it is expected that methods that do not include d orbitals will have difficulties with hypervalent compounds, the question is how far can one go with an sp basis set and the PDDG approach?

* Corresponding author email: william.jorgensen@yale.edu.

10.1021/ct0500287 CCC: $30.25
Published on Web 06/25/2005

PDDG Formalism

The rationale behind the PDDG formalism has been discussed.[17] The key difference between the PDDG/PM3 method and its predecessor is the addition of pairwise distance-directed Gaussian terms (PDDG) to the core repulsion function, as given in eq 1.

$$\mathrm{PDDG}(A,B) = \frac{1}{n_A + n_B}\{\sum_{i=1}^{2}\sum_{j=1}^{2}(n_A P_{A_i} + n_B P_{B_j}) \exp[-10\ \text{Å}^{-2}(R_{AB} - D_{A_i} - D_{B_j})^2]\} \quad (1)$$

This equation is an empirical correction to the core repulsion between atoms *A* and *B*, which are separated by a distance R_{AB}; each element requires four parameters, P_{A1}, P_{A2}, D_{A1}, and D_{A2}. The function is weighted using n_A and n_B, which are the number of valence electrons for atoms *A* and *B*, respectively. For each pair of atoms, there are four Gaussians, which depend only on atomic parameters. The Gaussians are small compared to those used in the original AM1 and PM3 core repulsion functions and work by addressing systematic errors associated with bonds and functional groups.[17,18]

Another difference between the traditional NDDO methods and the PDDG method is the way in which the molecular energy obtained from the self-consistent field calculation is converted to a heat of formation at 298 K. In both cases, the difference between the heat of formation of each atom and its electronic energy (*eisol*) is added to the molecular energy. In the traditional method, *eisol* is obtained as a derived parameter by calculating the energy of an isolated atom with a restricted single-determinant wave function using the semiempirical formalism and parameter set;[19] for the PDDG methods, *eisol* is treated as an optimizable parameter, obtained from a through-origin linear regression so that the conversion from molecular energy to heat of formation gives as small an error as possible. A similar approach has been applied in recent work by Winget and Clark using density functional theory (DFT).[20]

Parameter Optimization

To extend the PDDG/PM3 method to sulfur, silicon, and phosphorus, a similar procedure to the one used for the halogens was followed: the U_{ss}, U_{pp}, β_s, β_p, ζ_s, ζ_p, and α MNDO parameters, as well as the PM3 and PDDG/PM3 Gaussian pre-exponential and distance terms, were optimized by a combination of gradient-based methods (Fletcher−Powell) and simulated annealing.[18] The available reference data had very few molecules involving more than one of the elements discussed in this paper, so little coupling between different elements' parameters was expected, which allowed each element to be optimized separately. A total of 527 reference values were used in the error function during the optimization of the three elements, as detailed in Table 1. The error function that was minimized was the weighted sum of the square deviation between the calculated and reference values, as discussed previously.[18] Most of the optimizations were done with fixed PM3 geometries, adding gradients to the error function to ensure that the geometric minimum for each molecule did not stray far from the PM3 minimum. Only the final optimization stage involved fully flexible geometries; in this stage, the gradients were not included in the error function, since they are almost zero.

Table 1. Composition of the Training Sets[a]

data type	S	Si	P
heat of formation	81	47	23
ionization potential	28	15	13
dipole moment	16	12	7
bond length	78	96	16
bond angle	54	33	8

[a] For detailed data and references, see the Supporting Information.

The prior PDDG/PM3 parametrizations were performed as local optimizations that tried to stay close to the original PM3 parameters.[17,18] However, preliminary results for sulfur showed that the local approach could provide little improvement over PM3, so a more global optimization was undertaken. As a first step, to perturb the PM3 parameters reasonably far from their original values, 384 (3×2^7) initial parameter sets were generated by crossover between the U_{ss}, U_{pp}, β_s, β_p, ζ_s, ζ_p, and α parameters of PM3, MNDO, and AM1. These parameter sets were optimized in the usual way while using the PM3 Hamiltonian and local optimization for the PM3 Gaussians; the PDDG Gaussians were not added yet. The six best results were chosen, and an exhaustive crossover was performed between them, resulting in 1920 (15×2^7) parameter sets. After optimizing the latter, the best set was chosen, and 256 (4^4) different combinations of PM3 Gaussian parameters were tried in a grid-like fashion. After optimization of these, 256 combinations of PDDG Gaussians were added to the best result. The best parameter set from the last step was then subjected to 3000 steps of stochastic search; at each step, a random "kick" was applied to the parameters, which were then optimized using the same gradient-based algorithm. The best parameter set obtained so far was subjected to a flexible geometry optimization using simulated annealing, resulting in the final parameter set for sulfur. Suffice it to say that this protocol was not planned in advance in its entirety but was the result of much trial and error. However, it resulted in remarkable improvement over the simple local optimization, as discussed below.

Similar procedures were tried for silicon and phosphorus, with a few variations. The MNDO method was no longer included in the crossover part, as it was found to not be helpful; also, the stochastic search was not performed for these elements. For silicon, it was found that the results for the global optimization were not better than those from the simple local optimization, whereas for phosphorus, gains were made over PM3. The buildup procedure of parametrization for C, H, N, and O first; then the halogens; and then S, Si, and P clearly leads to the largest errors for the last elements. Simultaneous optimization for all elements is desirable but logistically taxing.

Results and Discussion

The optimized and dependent parameters for the PDDG/PM3 method are shown in Table 2 along with the original PM3 parameters for comparison. Some PDDG/PM3 parameters

Table 2. Optimized PDDG/PM3 Parameters for S, Si, and P, along with the Standard PM3 Parameters[a]

	PDDG/PM3			PM3		
	S	Si	P	S	Si	P
U_{ss}	−43.906 366	−26.332 522	−37.882 113	−49.895 371	−26.763 483	−40.413 096
U_{pp}	−43.461 348	−22.602 540	−30.312 979	−44.392 583	−22.813 635	−29.593 052
β_s	−2.953 912	−3.376 445	−12.676 297	−8.827 465	−2.862 145	−12.615 879
β_p	−8.507 779	−3.620 969	−7.093 318	−8.091 415	−3.933 148	−4.160 040
ζ_s	1.012 002	1.586 389	2.395 882	1.891 185	1.635 075	2.0175 63
ζ_p	1.876 999	1.485 958	1.742 213	1.658 972	1.313 088	1.504 732
a	2.539 751	2.215 157	2.005 294	2.269 706	2.135 809	1.940 534
eisol	−166.336 554	−66.839 000	−117.212 854	−183.453 740	−67.788 214	−117.959 174
DD	1.006 989	1.310 515	0.893 978	1.121 431	1.314 455	1.064 495
QQ	0.891 487	1.126 089	0.960 457	1.008 649	1.274 340	1.112 039
ρ_0[b]	1.517 625	2.695 556	1.743 870	1.517 625	2.695 556	1.743 870
ρ_1[b]	0.711 672	1.630 757	1.050 851	0.748 602	1.633 605	1.160 242
ρ_2[b]	0.754 336	0.949 200	1.208 907	0.814 668	1.025 130	1.339 579
a_1	−0.330 692	−0.071 314	−0.398 055	−0.399 191	−0.390 600	−0.611 421
b_1	6.000 000	6.000 000	1.997 272	6.000 669	6.000 054	1.997 272
c_1	0.823 837	0.237 995	0.950 073	0.962 123	0.632 262	0.794 624
a_2	0.024 171	0.089 451	−0.079 653	−0.054 899	0.057 259	−0.093 935
b_2	6.000 000	6.000 000	1.998 360	6.001 845	6.007 183	1.998 360
c_2	2.017 756	1.897 728	2.336 959	1.579 944	2.019 987	1.910 677
P_{A1}	0.120 434	−0.091 928	0.462 741			
P_{A2}	−0.002 663	−0.040 753	−0.020 444			
D_{A1}	0.672 870	1.163 190	0.714 296			
D_{A2}	2.032 340	2.190 526	2.041 209			

[a] Units are (eV) U_{ss}, U_{pp}, β_s, β_p, *eisol*, a_1, a_2, P_{A_1}, P_{A_2}; (au) ζ_s, ζ_p; (Bohr) *DD, QQ*, ρ_0, ρ_1, ρ_2; (Å) c_1, c_2, D_{A_1}, D_{A_2}; (Å^{-1}) α, b_1, b_2. [b] For use in MOPAC 6, $\rho_0 = 0.5/AM$, $\rho_1 = 0.5/AD$, $\rho_2 = 0.5/AQ$.

for sulfur, such as U_{ss}, ζ_s, and β_s, are quite different from the PM3 parameters. This is a result of the global optimization method along with the differences in the training set. However, the PDDG/PM3 parameters are still more similar to PM3 than to AM1 values, while the magnitude of the differences between PDDG/PM3 and PM3 parameters is similar to the differences between PM3 and AM1. The differences in the parameters for silicon and phosphorus are generally smaller than those for sulfur.

Sulfur Thermochemistry. The performances of PDDG/PM3 and the other common semiempirical methods for the calculation of standard heats of formation are summarized in Table 3. Full details are provided for all individual compounds in the Supporting Information. The MAE for the entire set of 249 molecules is 6.4 kcal/mol for PDDG/PM3, which is 35−40% smaller than the MAE for the older methods. This large improvement was observed for most classes of compounds, but it could only be achieved after the extensive global optimization discussed above. Initial tests with local optimizations resulted in only ca. 15% improvements with respect to PM3. It is remarkable that the errors for hypervalent molecules with PDDG/PM3 and its sp basis set are not large and that notably improved performance for classes such as sulfones is achieved, while simultaneously improving the performance for normal valence functional groups. The MAEs for thiols, sulfides, and disulfides are decreased by 63%, 13%, and 76%, respectively, when compared to PM3. Compounds with both sulfur and halogens were the only notable class where PDDG/PM3 did not yield improvement over PM3 or MNDO/d. The problem cases here are predominantly small fluorine-containing species including SF_2, FSSF, the transition state for the $SH^- + CH_3F$ S_N2 reaction, and the $HFHS^+$ complex. If more data were available for larger molecules, the benefits of PDDG for homologation and branching could be expected to dominate for this class too.

Table 3. MAEs for Heats of Formation for Sulfur Compounds (kcal/mol)

	N	PDDG	MNDO	AM1	PM3	MNDO/d
all	249	6.4	41.2	10.6	10.5	10.0
training	81	7.1	34.2	9.9	10.6	10.0
test	168	6.0	44.6	10.9	10.4	10.1
halides	15	13.7	56.5	14.5	6.6	5.1
sulfoxides	7	5.4	43.2	3.8	6.1	5.6
sulfones	36	5.9	143.8	18.9	18.1	10.4
sulfates	5	1.7	162.9	10.5	5.1	5.9
sulfites	4	8.9	54.1	22.0	11.2	7.6
thiols	29	1.4	5.5	4.4	3.8	7.3
sulfides	38	3.9	9.5	5.9	4.5	6.4
disulfides	14	1.8	9.7	5.1	7.6	9.5
aromatics	8	1.3	4.7	2.6	4.5	1.8
thioamides	9	7.4	14.6	11.8	22.2	27.5
dithiocarbamates	5	4.6	19.2	9.1	17.8	34.6
thioesters	10	5.4	11.6	6.5	7.0	7.8
thiocarbonates	7	4.6	11.8	5.6	15.6	8.2
anions	7	5.3	5.7	5.3	7.2	3.5
cations	7	11.3	19.2	18.0	23.0	11.0
transition structs.	4	10.6	26.7	26.2	14.9	29.9
complexes	18	13.0	41.6	15.0	17.5	19.6
others	16	10.2	10.2	9.1	8.4	6.6

Table 4. Heat of Formation Results (kcal/mol) for Selected Molecules that Highlight Systematic Problems

			Deviation (calc-exp)				
formula	name	exp	PDDG	PM3	AM1	MNDO	MNDO/d
CH_4S	methanethiol	−5.4	0.2	−0.1	1.1	−1.9	1.1
$C_4H_{10}S$	1-butanethiol	−21.1	2.2	1.6	−3.1	−1.8	1.8
$C_{10}H_{22}S$	1-decanethiol	−50.7	−1.3	−1.3	−14.6	−0.5	3.1
$C_4H_{10}S$	isobutyl thiol	−23.1	0.3	4.3	1.4	4.2	8.5
$C_5H_{12}S$	isopentyl thiol	−27.4	−0.5	2.4	−1.1	4.4	8.0
$C_5H_{12}S$	neopentyl thiol	−30.8	−0.7	7.7	7.0	15.3	20.4
$C_4H_{10}S$	*sec*-butyl thiol	−23.0	1.3	4.4	1.0	2.9	7.3
$C_4H_{10}S$	*tert*-butyl thiol	−26.0	1.0	8.1	7.3	10.9	14.9
C_2H_6S	dimethyl sulfide	−8.9	−5.3	−2.1	−0.4	−8.2	0.4
$C_4H_{10}S$	diethyl sulfide	−20.0	−0.7	2.8	−1.8	−9.0	0.5
$C_8H_{18}S$	diisobutyl sulfide	−42.9	−2.4	5.7	−0.9	3.1	14.7
$C_6H_{14}S$	diisopropyl sulfide	−33.9	−0.8	7.8	3.2	0.6	11.0
$C_8H_{18}S$	di-*tert*-butyl sulfide	−45.1	−6.1	12.0	11.3	21.5	31.6
$C_2H_6S_2$	dimethyl disulfide	−5.8	−3.5	1.0	1.6	−9.0	−0.7
$C_2H_6S_3$	dimethyl trisulfide	−3.0	−2.0	−3.9	−1.7	−10.2	−2.4
$C_8H_{18}S_2$	dibutyl disulfide	−37.9	−1.2	3.5	−6.8	−7.9	3.0
$C_8H_{18}S_2$	di-*tert*-butyl disulfide	−47.8	0.8	20.1	16.0	21.0	34.5
$C_2H_6O_2S$	dimethyl sulfone	−89.2	1.7	12.9	18.9	142.9	3.6
$C_4H_{10}O_2S$	diethyl sulfone	−102.6	9.6	21.7	21.4	143.2	5.6
$C_8H_{18}O_2S$	dibutyl sulfone	−121.8	7.7	19.5	13.3	143.6	7.1
$C_8H_{18}O_2S$	diisobutyl sulfone	−128.0	7.9	27.6	24.4	160.2	25.0
$C_8H_{18}O_2S$	di-*tert*-butyl sulfone	−130.6	8.0	38.2	41.8	178.9	49.3

The latter point becomes apparent in Table 4, which highlights some of the systematic errors found with the different semiempirical methods. All methods except PDDG/PM3 have notable problems with homologation or branching or both. In particular, AM1 yields errors proportional to the number of methylene groups that results, for example, in a deviation of 14.6 kcal/mol for decanethiol. All of the methods except PDDG/PM3 tend to overestimate the heats of formation of branched compounds. The problem is particularly large for both MNDO and MNDO/d; the latter method yields deviations of 31.6 kcal/mol for di-*tert*-butyl sulfide and 49.3 kcal/mol for di-*tert*-butyl sulfone. Although a major part of the branching errors may be attributed to the hydrocarbon parameters, sulfur parameters must also play a role, because MNDO/d, while having the same parameters as MNDO for the first-row elements, tends to have larger branching errors. All of the methods, including PDDG/PM3, overestimate the heat of formation of sulfones to varying degrees. MNDO/d might be expected to be the best method for sulfones because of its inclusion of *d* functions; however, its advantages for small sulfones are minor, and the branching problems for larger sulfones lead to large errors. PDDG/PM3 still overestimates the heat of formation of sulfones, but when all the sulfones in the dataset are considered, the mean signed error is +2.7 kcal/mol versus about +18 kcal/mol for AM1 and PM3.

The 18 complexes in the dataset are considered further in Table 5, which focuses on the computed enthalpies of complexation, that is, the difference between the heats of formation of the complex and the separated molecules. PDDG/PM3 gives the lowest MAEs for neutral as well as positively and negatively charged complexes. The overall MAE of 3.8 kcal/mol from PDDG/PM3 shows large improvement over the alternative semiempirical methods. The improvements here are clearly not from better treatment of branching because the molecules are all small. Perhaps surprisingly, the original MNDO method gives the second best results for all classes; it does not show the serious overestimate of binding for positively charged complexes as from AM1 or PM3. Thus, its problems with the heats of formation of hypervalent molecules such as SO_2 and dimethyl sulfoxide mostly cancel out when calculating interaction enthalpies.

A new method based on AM1, called AM1*, has been reported recently.[26] This method includes d orbitals for phosphorus, sulfur, and chlorine as well as a modified core-repulsion function with two-element parameters. The standard AM1 parameters and formalism are used for hydrogen and the first-row elements. The intersection of the dataset used in the AM1* study with the dataset used here has 82 sulfur-containing molecules and results in the following MAEs in kcal/mol: MNDO/d (6.0), PM3 (7.2), PDDG/PM3 (7.3), AM1* (8.0), AM1 (10.1), and B3LYP/6-311+G(2d,p)//LYP/6-31+G(d) (11.8). The dataset contains few large or branched molecules, so MNDO/d and PM3 perform relatively well. There is no clear benefit for the AM1* methodology, whereas the B3LYP results are the poorest. As discussed previously,[17,18] PDDG/PM3 consistently outperforms the much slower DFT methods for calculating heats of formation owing, presumably, to some systematic errors with the DFT methods.

The lower errors for heats of formation with PDDG/PM3 carry over to lower errors for isomerization enthalpies, as illustrated for several series in Table 6. PDDG/PM3 some-

Table 5. Interaction Enthalpies of Sulfur-Containing Complexes

complex	ref	PDDG	MNDO	AM1	PM3	MNDO/d
$NH_3\cdots SO_2$	−4.49[a]	−4.26	−2.93	−13.37	−16.55	−2.74
$MeNH_2\cdots SO_2$	−6.70[a]	−3.41	−2.48	−10.63	−10.69	−1.94
$Me_2NH\cdots SO_2$	−7.93[a]	−3.42	−2.18	−10.00	−1.05	−1.53
$NMe_3\cdots SO_2$	−10.65[a]	−2.98	−1.46	−8.12	−0.89	−0.92
$NH_3\cdots SO_2\cdots NH_3$	−3.00[a]	−7.60	−5.43	−23.34	−2.31	−4.68
$MeNH_2\cdots SO_2\cdots MeNH_2$	−4.28[a]	−6.06	−4.60	−18.25	−1.87	−3.83
$CH_3SH\cdots F^-$	−38.04[b]	−34.73	−60.55	−103.16	−65.37	−61.76
$CH_3F\cdots HS^-$	−12.54[b]	−9.78	−5.35	−7.77	−5.37	−4.30
$CH_3SH\cdots H_2O$[c]	−1.16[d]	−3.56	−0.52	−1.90	−1.02	−0.57
$CH_3SH\cdots H_2O$[e]	−2.21[d]	−3.42	−0.62	−2.37	−2.43	−0.51
$CH_3SCH_3\cdots H_2O$	−2.88[d]	−4.32	−0.69	−2.44	−2.45	−0.25
$DMSO\cdots H_2O$	−6.56[d]	−8.15	−2.20	−5.61	−6.17	−1.90
$EtSH\cdots CH_3S^-$	−5.40[f]	−12.39	−11.45	−9.90	−14.37	−3.32
NH_3-HS^+	−111.45[g]	−104.00	−101.22	−138.30	−141.19	−70.1
H_2O-HS^+	−72.51[g]	−64.71	−69.00	−92.11	−89.45	−42.4
H_2S-HS^+	−93.36[g]	−93.88	−84.32	−124.81	−122.12	−86.61
$HF-HS^+$	−33.17[g]	−24.90	−48.45	−102.27	−64.40	−39.39
$HCl-HS^+$	−60.87[g]	−58.60	−49.50	−84.69	−90.34	−36.89
MAE (negative or neutral)		3.21	5.23	9.88	6.19	5.26
MAE (positive)		5.26	9.89	34.16	27.23	21.68
MAE (all, kcal/mol)		3.78	6.52	16.62	12.03	9.82

[a] MP3/6-31+G(2d,p) values at 298 K.[21] [b] CCSD(T)/aug-cc-pVTZ ZPE corrected with MP2 frequencies.[22] [c] Water as hydrogen-bond acceptor. [d] MP2(fc)/6-31++G(2d(X+),p) ZPE corrected.[23] [e] Water as hydrogen-bond donor. [f] MP2/6-31+G** ZPE corrected.[24] [g] G2 values at 0 K.[25]

Table 6. Heat of Formation Results (kcal/mol) for Some Isomeric Series

	exptl.	PDDG	MNDO	AM1	PM3	MNDO/d
$CH_3CH_2SCH_3$	−14.2	−17.4	−23.0	−15.6	−14.1	−14.0
$CH_3CH_2CH_2SH$	−16.2	−15.3	−18.2	−17.4	−14.1	−14.7
$(CH_3)_2CHSH$	−18.2	−17.1	−16.2	−16.0	−14.4	−12.5
$HSCH_2CH_2CH_2SH$	−7.1	−6.2	−11.0	−9.8	−3.8	−4.0
$HSCH_2CHSHCH_3$	−7.1	−7.1	−8.3	−9.5	−5.0	−0.4
$CH_3CH_2SSCH_3$	−11.8	−13.7	−20.6	−10.7	−8.8	−11.5
$CH_3CH_2SCH_2CH_3$	−20.0	−20.7	−29.0	−21.8	−17.2	−19.5
$CH_3CH_2CH_2CH_2SH$	−21.1	−18.9	−22.9	−24.2	−19.5	−19.4
$(CH_3)_2CHSCH_3$	−21.4	−24.8	−25.3	−20.1	−18.4	−15.9
$CH_3CH_2CHSHCH_3$	−23.0	−21.7	−20.0	−21.0	−18.6	−15.7
$(CH_3)_2CHCH_2SH$	−23.1	−22.8	−18.9	−21.6	−18.8	−14.6
$(CH_3)_3CSH$	−26.0	−25.0	−15.1	−18.7	−17.9	−11.1
$HSCH_2CH_2CH_2CH_2SH$	−12.0	−11.5	−16.0	−17.1	−9.7	−8.9
$CH_3CH_2SSCH_2CH_3$	−17.9	−18.0	−26.5	−17.3	−12.7	−16.4
$(CH_3)_2CHSSCH_3$	−18.9	−19.5	−23.4	−15.2	−11.9	−13.6

what overestimates the stability of sulfides relative to thiols but does particularly well for dithiols relative to disulfides and for branching isomers.

Silicon Thermochemistry. Table 7 summarizes the results for heats of formation for various classes of silicon-containing compounds. Considering all 144 molecules in the dataset, MNDO/d provides the smallest MAE, and the differences between PM3, AM1, MNDO/d, and PDDG/PM3 are small. There are some classes of compounds for which MNDO/d is particularly advantageous, specifically, divalent silicon molecules, ions, and complexes. PDDG/PM3 gives results that are generally similar to those of PM3 or AM1 for most classes, though it does particularly well for silyl ethers. All of the methods have difficulties with the compounds containing halogens and silicon; however, most of these are very small polyhalosilanes.

Phosphorus Thermochemistry. Reliable thermochemical data for organophosphorus compounds has historically been scarce,[27] which makes the development of semiempirical methods difficult. The present dataset consists of only 43 molecules, taken mostly from the paper reporting the parametrization of MNDO/d.[5] The dataset is both too small and poorly representative of key functional groups containing phosphorus. Thus, the parametrizations should be considered tentative and used with caution. The summary of results in Table 8 should be viewed in this light. PDDG/PM3 gives a MAE for heats of formation similar to that of AM1 and shows a 17% smaller error than PM3. The best method for this dataset is MNDO/d, with a MAE of 12.7 kcal/mol. Because the dataset contains mostly small molecules, the advantage of d orbitals apparently outweighs the homologation and branching problems of MNDO. The MAE for

Table 7. MAEs for Heats of Formation for Silicon Compounds (kcal/mol)

	N	PDDG	MNDO	AM1	PM3	MNDO/d
all	144	11.9	16.2	11.7	11.1	9.4
training	56	10.3	13.0	9.1	9.8	7.4
test	88	12.9	18.2	13.3	11.9	10.7
CHSi	40	7.6	8.2	7.5	6.4	8.3
CHONSi	5	4.4	7.2	4.5	13.5	2.7
halides	31	15.2	26.5	13.5	9.8	15.0
ethers	9	6.9	11.1	7.8	8.6	9.6
anions	4	17.0	21.7	14.4	9.2	6.7
cations	10	13.8	21.0	8.5	24.6	7.3
radicals	14	11.9	14.1	16.8	9.5	7.8
complexes	9	21.7	14.0	20.6	28.5	11.9
divalent	11	14.6	19.0	12.1	12.2	5.2
others	9	13.0	14.3	11.8	6.3	6.8

Table 8. MAEs for Heats of Formation for Phosphorus Compounds (kcal/mol)

	N	PDDG	MNDO	AM1	PM3	MNDO/d
all	43	17.9	44.8	18.2	21.5	12.7
training	21	11.3	47.7	14.5	18.8	9.0
test	22	24.1	42.1	21.7	24.0	16.3
low valence	1	9.0	24.8	36.7	27.6	7.9
normal valence	25	11.0	20.9	11.6	16.3	9.4
hypervalent	14	31.0	94.6	28.3	30.9	18.5
halides	8	20.0	56.5	20.8	14.7	14.4
radicals	3	16.8	18.4	19.3	19.1	15.8
phosphines	4	10.9	14.9	8.8	7.0	4.8
phosphates	3	34.6	80.3	6.3	34.4	15.3
phosphites	4	9.4	54.7	15.6	42.5	9.2

phosphorus halides increases for PDDG/PM3 when compared to PM3, as for sulfur and silicon halides. This trend suggests that the problem may be the halogen parameters. Though they give significantly improved results for CHNOX molecules, where X is a halogen, the halogen parameters are not optimal for molecules with second-row atoms.

For comparison with the new AM1* method, the results for the intersection of the AM1* dataset for phosphorus and the dataset used in this work were analyzed. For the 33 common molecules, the MAEs (kcal/mol) are MNDO/d (8.6), AM1 (13.1), PDDG/PM3 (13.8), PM3 (17.4), AM1* (17.6), and B3LYP/6-311+G(2d,p)//B3LYP/6-31+G(d) (12.6). MNDO/d, again, gives the best result; PDDG/PM3, AM1, and the DFT method give similar MAEs near 13 kcal/mol, and AM1* and PM3 yield MAEs of 17−18 kcal/mol.

Another recent development is the optimization by Lopez and York of the AM1/d method to treat nucleophilic attack on biological phosphates.[28] The method was optimized using a set of B3LYP reference values for phosphates, phosphoranes, and transition structures. Because of the different scope of the AM1/d and the PDDG/PM3 parametrizations and the lack of overlap between the datasets used, making a fair comparison is difficult; it is to be expected that the AM1/d method will perform significantly better for the types of systems for which it was parametrized, whereas PDDG/PM3 has a wider scope in terms of functional groups included in the parametrization.

Table 9. MAEs for Ionization Potentials (IP), Dipole Moments, and Geometries

	element	*N*	PDDG	MNDO	AM1	PM3
IP (eV)	S	32	0.66	0.71	0.48	0.35
	Si	15	1.08	0.86	0.63	1.21
	P	13	0.42	1.27	0.66	0.61
dipole (D)	S	28	0.54	0.44	0.43	0.46
	Si	26	0.92	0.84	0.45	0.72
	P	12	0.77	0.83	0.79	0.58
bond length (Å)	S	78	0.06	0.13	0.07	0.08
	Si	96	0.09	0.13	0.06	0.11
	P	16	0.12	0.08	0.08	0.06
angle (deg)	S	54	5.50	9.46	6.74	7.42
	Si	33	3.25	2.47	2.39	2.75
	P	8	3.36	3.39	3.76	1.91

Table 10. MAEs for Heats of Formation (kcal/mol) for All Molecules Studied

		standard NDDO methods				PDDG	
	N	MNDO	AM1	PM3	MNDO/d	MNDO	PM3
CHNO[a]	622	8.4	6.7	4.4	8.4	5.2	3.2
halogens[b]	422	14.0	11.1	8.1	13.4	6.6	5.6
S	249	41.2	10.6	10.5	10.0		6.4
Si	144	16.2	11.7	11.1	9.4		11.9
P	43	44.8	18.2	21.5	12.7		17.9
all	1480	19.8	10.8	8.7	10.3		6.5

[a] Reference 17. [b] Reference 18.

Other Observables. The parametrization of the PDDG/PM3 method has focused on improving heats of formation and, therefore, enthalpies of reaction in general while retaining accuracy comparable to prior NDDO-based methods for other observables such as geometries, ionization potentials, and dipole moments. Table 9 shows the MAEs for these properties, while the complete results are, again, reported in the Supporting Information. The performances of the various methods are similar overall, with no striking problems.

Summary

The parametrization of PDDG/PM3 for sulfur showed the benefits of a global optimization over simpler local optimizations and the value of the availability of a large amount of experimental data for training and testing. Through global optimization, a method was obtained that yields MAEs for heats of formation that are 40% smaller than those from the alternatives, including semiempirical methods that employ d orbitals and DFT at the B3LYP/6-311+G(2d,p)//B3LYP/6-31+G(d) level. Extension to silicon and phosphorus was also carried out, though it was hampered by limited experimental data.

PDDG/PM3 can now be used to treat molecules with any combination of the elements C, H, N, O, F, Cl, Br, I, S, Si, and P. Table 10 summarizes the results for all molecules, ions, and complexes that have been treated so far. The PDDG/PM3 method is the most accurate semiempirical method available for calculating heats of formation with a MAE of 6.5 kcal/mol for the 1480 molecules in our full dataset; PM3 is the next best with an MAE of 8.7 kcal/mol.

Particularly striking improvements have been obtained for molecules containing C, H, N, O, S, and the halogens. Moreover, the PDDG/PM3 method can be easily implemented into existing software such as MOPAC 6.[29]

Acknowledgment. Gratitude is expressed to the National Science Foundation (CHE-013996) and to the National Institute of Health (GM032136) for support of this work.

Supporting Information Available: An Excel table with the experimental and calculated heats of formation, ionization potentials, and dipole moments for the molecules in this study. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.

(2) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4917.

(3) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 7, 3902−3909.

(4) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221−264.

(5) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616−626.

(6) Gogonea, V.; Suarez, D.; Van der Vaart, A.; Merz, K. M., Jr. *Curr. Opin. Struct. Biol.* **2001**, *11*, 217−223.

(7) Acevedo, O.; Jorgensen, W. L. *Org. Lett.* **2004**, *6*, 2881−2884.

(8) Vayner, G.; Houk, K. N.; Jorgensen, W. L.; Brauman, J. I. *J. Am. Chem. Soc.* **2004**, *126*, 9054−9058.

(9) Guimarães, C. R. W.; Repasky, M., P.; Chandrasekhar, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2003**, *125*, 6892−6899.

(10) Guimarães, C. R. W.; Udier-Blagović, M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2005**, *127*, 3577−3588.

(11) Chandrasekhar, J.; Shariffskul, S.; Jorgensen, W. L. *J. Phys. Chem. B* **2002**, *106*, 8078−8085.

(12) Ruiz-Pernia, J. J.; Silla, E.; Tunon, I.; Marti, S.; Moliner, V. *J. Phys. Chem. B* **2004**, *108*, 8427−8433.

(13) Ranaghan, K. E.; Mulholland, A. J. *Chem. Commun.* **2004**, 1238−1239.

(14) Devi-Kesavan, L. S.; Garcia-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2003**, *109*, 133−139.

(15) Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2003**, *125*, 1532−1540.

(16) Kaminski, G. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787−1796.

(17) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.

(18) Tubert-Brohman, I.; Guimarães, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 138−150.

(19) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−105.

(20) Winget, P.; Clark, T. *J. Comput. Chem.* **2004**, *25*, 725−733.

(21) Wong, M. W.; Wiberg, K. B. *J. Am. Chem. Soc.* **1992**, *114*, 7527−7535.

(22) Gonzales, J. M.; Cox, R. S., III; Brown, S. T.; Allen, W. D.; Schaefer, H. F., III. *J. Phys. Chem.* **2001**, *105*, 11327.

(23) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem. A* **1998**, *102*, 3782−3797.

(24) Gronert, S.; Lee, J. M. *J. Org. Chem.* **1995**, *60*, 6731−6736.

(25) Solling, T. I.; Radom, L. *Chem.—Eur. J.* **2001**, *7*, 1516−1524.

(26) Winget, P.; Horn, A. H. C.; Selcuki, C.; Martin, B.; Clark, T. *J. Mol. Model.* **2003**, *9*, 408−414.

(27) Pilcher, G. In *The Chemistry of Organophosphorus Compounds*; Hartley, F. R., Ed.; John Wiley & Sons: New York, 1990; Vol. 1, pp 127−136.

(28) Lopez, X.; York, D. M. *Theor. Chem. Acc*. **2003**, *109*, 149−159.

(29) Directions, code, and parameter files for implementation of the PDDG methods in *MOPAC* are freely available for download at http://www.jorgensenresearch.com. The PDDG methods are also available in the *BOSS* program; see http://www.cemcomco.com.

CT0500287

JCTC Journal of Chemical Theory and Computation

Rationalization of the $\pi-\sigma$ (Anti)aromaticity in All Metal Molecular Clusters

Ayan Datta and Swapan K. Pati*

Theoretical Sciences Unit and Chemistry and Physics of Materials Unit, Jawaharlal Nehru Center for Advanced Scientific Research, Jakkur P. O, Bangalore-560064, India

Received May 20, 2005

Abstract: A $\sigma-\pi$ separation analysis of the energies in Al_4Li_4 reveals that the system is more π-antiaromatic than the σ-aromaticity in it. This is true also for C_4H_4 and Ga_4Li_4. Unlike C_4H_4 that has a very large component of π-antiaromaticity, for these all-metal clusters, these energy scales are comparable though π-antiaromaticity is the major driving force for the distortion of the molecules from the square (σ-aromatic) structure to the rectangular (π-antiaromatic) architecture. For the dianion $Al_4Li_4^{2-}$, the σ-equalization prevails over the π-distortion in Al_4Li_4, and for the dication $Al_4Li_4^{2+}$, π-equalization is the driving force for the square symmetric structure.

The past decade has witnessed rapid progress in the new chemistry of small metal clusters of Al, Si, and Ga facilitated by computational strategy, synthesis, and characterization.[1−3] These metal clusters have a close resemblance with the cyclic organic π-conjugated molecules in structure as well as properties. Al_4Li_4 and their anions, $Al_4Li_3^-$ or Al_4^{4-}, follow the properties of cyclobutadiene (C_4H_4).[4] They also undergo interaction with transition metals such as Fe and Ni to form complexes of the type (Al_4Li_4)-$Fe(CO)_3$ and sandwich complexes of the type $(Al_4Li_4)_2Ni$ thereby resembling C_4H_4 in its role as a ligand.[5] Al_4^{4-} also satisfies the simple Huckel criteria for antiaromaticity as it possesses 4π electrons in its frontier π-orbitals. Despite the similarities between these all-metal systems and their organic counterpart (C_4H_4), serious doubts have recently been raised whether these systems are aromatic or antiaromatic, based on the π-electron scheme alone.[6,7] The σ-backbone appears to be quite important and based on a nucleus independent chemical shift (NICS), Schleyer and co-workers suggested that these clusters are net aromatic. But Boldyrev, Wang, and co-workers concluded that $Li_3Al_4^-$ and Li_4Al_4 clusters are net antiaromatic. Santos and co-workers agreed with the net antiaromaticity of these clusters on the basis of electron localization function (ELF) analysis.[8]

In this letter we show that the confusion associated with these clusters can be settled through a simple $\sigma-\pi$ separation analysis which provides an unambiguous answer to all such questions. The different results obtained by the previous workers arise primarily because of the indirect methods used to characterize aromaticity/ antiaromaticity. We stress that the σ-backbone is quite an important component in the structural features of almost all molecular systems. Even in C_6H_6, it has been found that the σ-backbone is responsible for the symmetric D_{6h} structure and the π-electrons actually tend to distort the symmetric structure.[9,10]

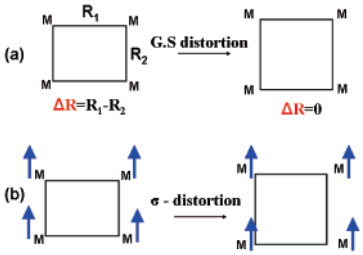
We have considered a variety of molecular systems, Al_4Li_4, $Al_4Li_4^{2-}$, Ga_4Li_4, and Al_4^{2-}, and compared with C_4H_4 and similar organic analogues at each step of our $\sigma-\pi$ analysis. These systems have either 4π, 6π, or 2π electrons in their frontier orbitals and provide a diverse set for studying aromaticity or antiaromaticity. All the geometries were optimized at the B3LYP/6-311G++ (d, p) level[11,12] (see the Supporting Information file for structures and energies). The ground-state geometry for both Al_4Li_4 and Ga_4Li_4 have a planar rectangular structure for the ring with the Li ions occupying positions so as to maintain a C_{2h} architecture. The bond-length alteration (BLA) for Al_4Li_4 and Ga_4Li_4 are 0.12 and 0.16 Å, respectively. Note that the same for C_4H_4 is 0.2 Å. The fact that the BLA for Ga_4Li_4 is more than that in Al_4Li_4 suggests that Ga_4Li_4 is more antiaromatic than Al_4Li_4, in analogy with C_4H_4, and a $\sigma-\pi$ separation should be ideal to quantify such a statement.

We distort the geometry optimized structures by ΔR (where ΔR is the difference between the long M−M and short M−M bond in the M_4 ring) so that that the distortion keeps the sum of two adjacent M−M bonds constant

* Corresponding author e-mail: pati@jncasr.ac.in.

10.1021/ct0501351 CCC: $30.25
Published on Web 06/23/2005

Scheme 1. (a) Distortion Mode for the M_4 Rings (M=C, Al, Ga) in the Ground State[a] and (b) Distortion in the σ-Electrons Involving the Distortion in a High-Spin Configuration



[a] Li atoms not shown for the sake of clarity.

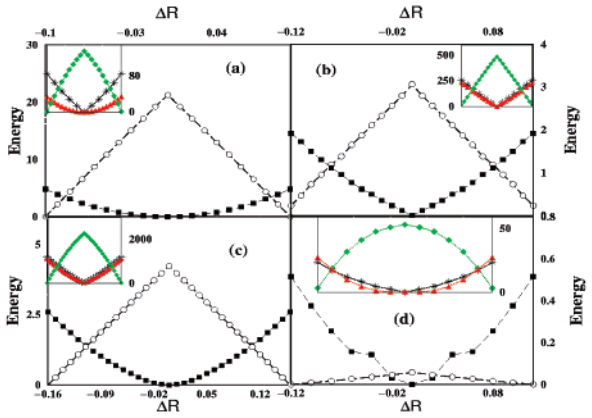


Figure 1. Variation of the σ-energy (square) and the π-energy (circles), both in kcal/mol as a function of the distortion axis, ΔR for (a) C_4H_4, (b) Al_4Li_4, (c) Ga_4Li_4, and (d) $Al_4Li_4^{2-}$ derived from $(Al_4Li_4)Fe(CO)_3$. The insets show V_{core} (green), V_{ee} (black), and V_{nn} (red) components in the ground-state structures. All the energies are scaled to make the most stable geometry zero in energy and positive values in the energy-axis correspond to destabilization.

[Scheme 1 (a)]. The energy associated with the distortion is partitioned into σ and π components as $\Delta E_\pi = \Delta E_{GS} - \Delta E_\sigma$. One of the simplest methods to get the contribution associated with a distortion only along the σ-backbone is to freeze the π-electrons with all-parallel spins. The σ-backbone for a M_4 ring with 4π electrons can be modeled as M_4^{4-} with a high spin configuration ($S = 2$) with all the 4π electrons being parallel [Scheme 1(b)]. Similarly, for the 2π electron sytems such as $C_4H_4^{2+}$ and Al_4^{2-}, $S = 1$ corresponds to the high spin state. For the 6πe $Al_4Li_4^{2-}$ however, there are only four π-orbitals and thus a high spin configuration with $S = 3$ is not feasible, rather two parallel spins with $S = 1$ state in $Al_4Li_4^{2-}$ corresponds to the high spin state. We thus define, $\Delta E_\sigma = \Delta E_{HS}$. Such an analysis gives a very clear picture of the nature of interactions in the system and has been extensively used in the literature for various organic molecules.[13,14] For the high spin systems, we perform UB3LYP calculations at the same basis set level with annihilation of the first spin-contaminant.

In Figure 1, we plot the σ-energy and the π-energy as a function of the distortion parameter, ΔR. In the inset, the core energy, V_{core} (sum of kinetic energy and nuclear-electron (ne) interactions), electron−electron interactions (V_{ee}), and the nuclear−nuclear interactions (V_{nn}) are plotted. For all the systems, we find that the π-electrons have a general tendency of forming a distorted structure (π-energy is most stable at large ΔR), while the σ-framework opposes the distortion and tends to equalize the bonds. The final structure and thus the aromatic/antiaromatic features will crucially depend on the predomination of either of the forces. In Figure 1(a), the result for the well-known C_4H_4 system is shown. The instability associated with the σ-backbone distortion is very little (4 kcal/mol for $\Delta R = 0.1$), while the stability for π-distortion is quite substantial (22 kcal/mol for $\Delta R = 0.1$), clearly overwhelming the tendency for σ-backbone equalization. Thus the C_4H_4 has a rectangular structure and is overall π-antiaromatic with a minor σ-aromatic component. Both V_{ee} and V_{nn} are destabilized with distortion, while the V_{core} component is stabilized. We have further analyzed that it is the V_{ne} term in the V_{core} that favors the distorted structure. This is easy to understand as the V_{ne} component is associated with the electron−lattice interactions and leads to the Jahn−Teller stabilization in the distorted structure. However components such as V_{ee} and V_{nn} stabilizes the $\Delta R = 0$ structure associated with the delocalized π-electrons (for nonzero ΔR, the electron density is localized in shorter bonds).

For the all-metal system however, the $\sigma-\pi$ separation energy plays a crucial role. For example, in Al_4Li_4, the distortion in the σ-framework leads to a destabilization of 2.5 kcal/mol, while the π-framework gains energy of 3.5 kcal/mol (Figure 1(b)). The ground-state energy is thus stabilized by the distortion along the ring. Accordingly thus the Al_4Li_4 is π-antiaromatic though the σ-aromatic component is also substantial. However overall Al_4Li_4 is antiaromatic as the π-antiaromaticity exceeds the σ-aromaticity by 1 kcal/mol. The energy components also follow very similar trends such as that for C_4H_4 (Figure 1(b), inset). We derive a similar conclusion for the Ga_4Li_4 also; the π-stabilization associated with the distortion is 4 kcal/mol, while σ-destabilization is 2.5 kcal/mol (seen in Figure 1(c)). The distorted π-antiaromatic structure is thus stabilized by an amount of 1.5 kcal/mol, 0.5 kcal/mol more than that for Al_4Li_4. Thus the π-antiaromaticity follows the order $C_4H_4 > Ga_4Li_4 > Al_4Li_4$.

The fact that this simple $\sigma-\pi$ separation gives a very clear picture for the nature of aromaticity/antiaromaticity is evident from Figure 1(d). We retrieve the structure of the Al_4Li_4 unit from the organometallic complex $(Al_4Li_4)Fe(CO)_3$ and consider the Al_4Li_4 dianion system. The interaction of the $Fe(CO)_3$ unit with the Al_4Li_4 converts it into a 6π aromatic system with small BLA. In Figure 1(d) we distort this dianion of Al_4Li_4 and perform a similar analysis. Contrary to the previous cases, in $Al_4Li_4^{2-}$, the stabilization associated with the equalization of the σ-backbone overwhelms the instability due to π-electron localization by 0.5 kcal/mol and forces the system to be aromatic. This is of course true for C_6H_6 where σ-delocalization exceeds the π-localization by 6 kcal/mol.[15,16] In Figure 2 (a), the energy profile is plotted for $C_4H_4^{2+}$ which shows an overwhelming π-delocalization compared to the smaller σ-localization. Similarly, for the all-metal system, Al_4^{2-}, the ground state corresponds to a square

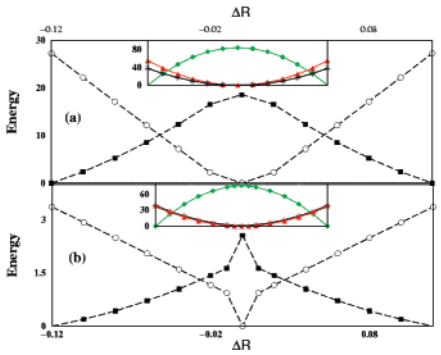


Figure 2. Same variation (including the inset) as in Figure 1 for (a) $C_4H_4^{2+}$ and (b) Al_4^{2-}.

geometry with Al−Al bond length = 2.54 Å. This is readily understood from the plot as the π-destabilization associated with the distortion exceeds the stability in the σ-backbone due to distortion (Figure 2(b)) though again the energy scales for the σ- and π-distortion are comparable. Also, as a general rule, we find that the V_{ne} term favors distortion.

In summary, we have shown that all-metal molecular clusters such as Al_4Li_4 and Ga_4Li_4 are predominately π-antiaromatic although there is a significant contribution from the σ-aromaticity as well due to close proximity in σ/π energy scales compared to the equivalent organic systems. We believe that our analysis provides a tool for assignment of aromaticity/antiaromaticity in all-metal clusters.

Acknowledgment. S.K.P. thanks CSIR and DST, Government of India for research grant.

Supporting Information Available: Structures, Cartesian coordinates, ground-state energies, and complete ref 11. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) . Li, X.; Kuznetsov, A.; Zhang, H.-F.; Boldyrev, A. I.; Wang, L. *Science* **2001**, *291*, 859.

(2) . Li, X.; Zhang, H.-F.; Wang, L.-S.; Kuznetsov, A. E.; Cannon, N. A.; Boldyrev, A. I. *Angew. Chem. Int. Ed.* **2001**, *40*, 1867.

(3) Kuznetsov, A.; Boldyrev, A. I.; Li, X.; Wang, L.-S. *J. Am. Chem. Soc*. **2001**, *123*, 8825.

(4) Kuznetsov, A.; Birch, K.; Boldyrev, A. I.; Li, X.; Zhai, H.; Wang, L. *Science* **2003**, *300*, 622.

(5) Datta, A.; Pati, S. K. *J. Am. Chem. Soc*. **2005**, *127*, 3496.

(6) Chen, Z.; Corminboeuf, C.; Heine, T.; Bohmann, J.; Schleyer, P. V. R. *J. Am. Chem. Soc.* **2003**, *125*, 13930.

(7) Ritter, S. *Chem. Eng. News* **2003**, *81*, 23.

(8) Santos, J. C.; Andres, J.; Aizman, A.; Fuentealba, P. *J. Chem. Theory Comput.* **2005**, *1*, 83.

(9) Shaik, S. S.; Hiberty, P. C. *J. Am. Chem. Soc*. **1985**, *107*, 3089.

(10) Shaik, S. S.; Hiberty, P. C.; Lefour, J.-M.; Ohanessian, G. *J. Am. Chem. Soc*. **1987**, *109*, 363.

(11) Frisch, M. J.; et al. Gaussian 03.

(12) Boldyrev, A. I.; Wang, L.-S. *J. Phys. Chem. A* **2001**, *105*, 10759.

(13) Hilberty, P. C.; Shaik, S. S.; Lefour, J.-M.; Ohanessian, G. *J. Org. Chem*. **1985**, *50*, 4657.

(14) Jug, K.; Hiberty, P. C.; Shaik, S. *Chem. Rev.* **2001**, *101*, 1477.

(15) Shaik, S.; Shurki, A.; Danovich, D.; Hiberty, P. C. *Chem. Rev.* **2001**, *101*, 1501.

(16) Jug, K.; Koster, A. M. *J. Am. Chem. Soc*. **1990**, *112*, 6772.

CT0501351

JCTC Journal of Chemical Theory and Computation

Quantum Chemical Characterization of the Structural and Energetic Properties of HCN−BF_3

James A. Phillips*,† and Christopher J. Cramer‡

Department of Chemistry, University of Wisconsin−Eau Claire, 105 Garfield Avenue, Eau Claire, Wisconsin 54701, and Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455

Received May 3, 2005

Abstract: The structure, dipole moment, binding energy, and vibrational frequencies of HCN−BF_3 are investigated via 12 DFT methods as well as MP2, MC-QCISD, and MCG3 calculations. By comparing the DFT results to both experimental data and results from post-Hartree−Fock molecular orbital methods, we gauge the effectiveness of various density functionals in modeling this fairly weak donor−acceptor system. For structural data, B3PW91, B98, and *m*PWPW91 provide results that compare favorably with experiment. All DFT methods that yield a reasonable structure predict dipole moments that are only slightly larger than the experimental value by 0.1 to 0.2 D. Moreover, to ensure that a comparison of calculated (equilibrium) and experimental (vibrationally averaged) data is indeed valid for this system, the B−N distance potential is calculated using B3PW91, MP2, and MCG3, and the one-dimensional Schrödinger equation for motion along this bond-stretching coordinate is solved to obtain vibrational energy levels, wave functions, and expectation values of the B−N distance and dipole moment. In every instance, average bond lengths differ by only a few thousandths of an angstrom from the corresponding equilibrium values, and dipole moments are unchanged to within hundredths of a debye. For vibrational frequencies, B3PW91 agrees most closely with gas-phase experimental data for BF_3 and also with MP2 calculations of the BF_3-localized modes in the complex; *m*PW1PW91 and B3LYP agree nearly as well. However, despite the effectiveness of DFT for structure, dipole moment, and vibrational frequencies, all DFT methods fail to predict a binding energy that compares favorably to the MCG3//MC-QCISD result of −5.7 kcal/mol.

I. Introduction

Numerous studies throughout the past decade have examined the ability of density functional theory (DFT)[1] to predict the structural and energetic properties of molecular complexes, including van der Waals,[2−7] hydrogen-bonded,[2,3,5,8−11] and donor−acceptor systems.[12−16] In particular, the latter have proven to be quite challenging, and in general, DFT has produced mixed results. Not surprisingly, most shortcomings arise when dispersion forces comprise a significant portion of the total interaction energy as DFT fails to account for it in any formal sense.[1] Most often, hybrid HF/DFT methods have performed better for complexes than have "pure" DFT methods.[4,7,12−16]

Boron−nitrogen complexes are the classic donor acceptor systems, and very recently it was found that pure methods (PBEPBE and PW91PW91) performed quite poorly compared to hybrid methods (MPW1K, *m*PW1PW91, B3PW91, and B3LYP) for amine−borane complexes.[16] Furthermore, it was noted that the very popular B3LYP method performed somewhat poorly compared to MPW1K and *m*PW1PW91.[16]

Our present interest is nitrile−BF_3 systems. Previously, it was noted that B3LYP and MP2 gave very consistent results

* Corresponding author phone: (715)836-5399; fax: (715)836-4979; e-mail: phillija@uwec.edu.
† University of Wisconsin−Eau Claire.
‡ University of Minnesota.

10.1021/ct0501248 CCC: $30.25
Published on Web 06/30/2005

for CH_3CN-BF_3.[17] However, upon proceeding to singly halogenated derivates of this complex (e.g. FCH_2CN-BF_3), B3LYP calculations yielded long B−N bonds (∼2.5 Å), indicative of fairly weak interactions. Given this, and the shortcomings of B3LYP noted in ref 16, we turned to other methods seeking some level of consistency. Unfortunately, we found considerable inconsistency, and moreover, no experimental gas-phase structures were available for these halo−acetonitrile complexes, making validation a difficult task.

In search of an appropriate test system, we turned to $HCN-BF_3$, a similar complex for which an experimental gas-phase structure[18] and dipole moment[19] have been published. The gas-phase B−N distance of 2.473(29) Å is quite similar to those we have calculated for the $X-CH_2CN-BF_3$ species using B3LYP. Furthermore, it is just shorter than the sum of the B and N van der Waals radii, and, as is true for its homologue CH_3CN-BF_3,[17,20,21] the structural properties of $HCN-BF_3$ are very sensitive to chemical environment.[20] For example, the solid-state structure has a B−N distance of 1.638(2) Å and an N−B−F angle of 105.6°.[22] Indeed, adding a single HCN molecule to form $(HCN)_2-BF_3$ causes the B−N distance to contract by 0.17 Å to 2.299(28) Å.[23]

Naturally, these remarkable structural properties have drawn the interest of theoreticians, though prior studies have been mainly concerned with issues secondary to our present interest, such as modeling the dative bond reaction path[24] or examining larger clusters (i.e. $(HCN-BF_3)_n$) in order to evaluate a cooperative mechanism for the large gas−solid structural difference.[25−27] Nevertheless, many equilibrium structures of $HCN-BF_3$ have been obtained in the process. Earlier studies yielded several MP2 results with B−N distances of 2.444 Å[25] and 2.443 Å[28] with the 6-31G** basis set, 2.296 with the D95** basis set,[26] 2.439 with the 6-31++G** basis set,[26] and 2.274 with the d95v++-(2d1f,2p) basis set.[24] Of these, only the 6-31++G** and 6-31G** results agree particularly well with experiment, although the d95v++(2d1f,2p) distance does increase to 2.510 Å after accounting for basis set superposition error.[24] Very recently, the structure was calculated with the MP2, B3LYP, MPW1K, and *m*PWPW91 methods using basis sets up to 6-311+G(d),[27] and the B−N distances obtained were 2.435 Å, 2.483, 1.882, and 2.296 Å, respectively. Clearly, the MP2 and B3LYP results are fairly consistent and agree fairly well with experiment, while the preferred methods for the amine−borane systems[16] do *much* more poorly. It is also worth noting that the MPW1K/6-311G(d) distance is 2.285 Å, indicating a peculiar sensitivity to diffuse functions, similar to that observed previously for CH_3CN-BF_3[17] using MP2 and B3LYP.

Our goal is straightforward; we seek an efficient and accurate DFT method for modeling the structures, binding energies, and vibrational frequencies of nitrile−BF_3 complexes such as $HCN-BF_3$. Below we present results from 12 different DFT methods and the MP2 level with large, triple-ζ basis sets (cc-pVTZ and aug-cc-pVTZ).[1] In addition, we report results from the multicoefficient correlation methods MC-QCISD[29] and MCG3,[30] in which high-level correlation energies are obtained from a scaled, additive approach. In general, such methods have yielded energetic predictions that are quite accurate.[1,29,30] Overall, we find that several hybrid DFT methods perform adequately with respect to the prediction of structure, dipole moment, and vibrational frequencies, but that all DFT methods fail to predict the binding energy accurately compared to the MC-QCISD and MCG3 results.

II. Computational Methods

MP2 and DFT calculations were performed using Gaussian03[31] version b.0.1 for all methods except O3LYP, for which version c.0.3 was used. For optimizations, all geometries were constrained to C_{3v} symmetry, and convergence criteria were set using the "opt=tight" option, which sets the maximum and RMS forces to 1.5×10^{-5} and 1.0×10^{-5} hartrees/bohr, respectively, and the maximum and RMS displacements to 6.0×10^{-5} and 4.0×10^{-5} bohr, respectively. An ultrafine grid was employed for all DFT calculations. The cc-pVTZ and aug-cc-pVTZ basis sets[1] were used almost exclusively, but for one method (B3PW91), the analogous double-ζ and quadruple-ζ basis sets were used to examine convergence behavior. Multi-Coefficient Gaussian-3 (MCG3)[30] energies as well as Multi-Coefficient Quadratic Configuration Interaction (MC-QCISD)[29] geometries and energies were obtained using MULTILEVEL version 4.0.[32] The maximum force criterion for MC-QCISD optimizations in MULTILEVEL was set to 1.0×10^{-4} hartrees/bohr.

To facilitate a direct comparison between computational (equilibrium) and experimental (vibrationally averaged) B−N distances, we calculated B−N potential curves using B3PW91, MP2, and MCG3 and solved the relevant one-dimensional Schrödinger equation to obtain wave functions and energies as well as vibrationally averaged bond lengths and dipole moments as expectation values over the B−N stretching coordinate. B3PW91 and MP2 potentials were calculated by freezing the B−N distance at 0.1 Å intervals from 1.6 to 3.6 Å and at 0.2 Å intervals from 3.6 to 4.0 Å and optimizing the geometry with respect to all remaining internal coordinates. MCG3 energies were then calculated for all MP2 geometries and for comparison a few B3PW91 geometries as well. Vibrational wave functions, energy levels, and average B−N distances were obtained using the FGHEVEN program,[33] which utilizes the Fourier Grid Hamiltonian method[34] to solve a one-dimensional Schrödinger equation for an arbitrary potential. To obtain the potential functions needed for FHGEVEN, we fit each potential to a pair of polynomial functions, fifth order for the points from 1.6 to 2.4 Å and fourth order for points from 2.4 to 4.0 Å. These functions reproduced the calculated energies along the curve to within about 1×10^{-5} hartrees. We considered different options for the reduced mass value in these calculations. Treating the system as a pseudodiatomic molecule, in which the individual atoms of the BF_3 and HCN subunits move coherently with one another throughout a vibrational period of the B−N stretching mode, one might estimate the reduced mass by simply taking the product of the HCN and BF_3 masses and dividing by their sum. Doing so results in a value of 19.25 amu. We compared this to the value used by

Gaussian for the B−N stretching mode in the MP2/aug-cc-pVTZ frequency calculation and found the latter to be much lower (10.16 amu). The difference derives from a coupling of the BF_3 umbrella mode and the B−N stretch. The displacement vectors for the ν_1 mode (predominantly B−N stretch, see below) in the Gaussian output indicate that the F atoms move much less along the B−N bond axis (∼60% relative amplitude) than does the B atom. This is not surprising, since the optimum N−B−F angle does change with B−N distance. Nevertheless it does complicate the use of a simplified one-dimensional model for the B−N stretching motion. In the end, we ran FGHEVEN with both mass values (19.25 and 10.16 amu) and found that there was no significant effect on the vibrationally averaged bond lengths or the qualitative features of the wave functions, despite the significant differences in the energy values. In any event, we note that the results shown below are those obtained with the reduced mass set to 10.16 amu.

III. Results and Discussion

The geometry of the complex is depicted in Figure 1 and calculated values of the B−N distance and N−B−F angle are displayed in Table 1, as are dipole moments and binding energies ($\Delta E = E^{\text{elec}}_{\text{HCN-BF}_3} - (E^{\text{elec}}_{\text{HCN}} + E^{\text{elec}}_{\text{BF}_3})$). The experimentally determined B−N distance[18] and dipole moment[19] are also listed for reference. The difference between the experimental (vibrationally averaged) and theoretical (equilibrium) B−N distances are also included in Table 1 (ΔR), and values for DFT/aug-cc-pVTZ methods that differ from the experimental value by less than 0.1 Å are noted in bold.

Structure and B−N Distance Potential. For comparisons of structural results, we focus initially on the B−N distance, as other structural parameters such as the N−B−F angle track it systematically, and for that matter it dominates the experimental moment of inertia.[18] In fact, the N−B−F angle quoted in the experimental work reflects an average of experiment and theory, and the other parameters were left fixed at the monomer values in the structural analysis.[18]

As a whole, the structural results for $HCN-BF_3$ are much less consistent than those for the amine−borane systems.[16] Nevertheless, results from the various hybrid DFT methods are reasonably consistent and most yielded B−N distances around 2.5 Å. The notable exceptions are MPW1K, which gives a somewhat shorter result of 2.323 Å with the aug-cc-pVTZ basis set, and O3LYP, for which the distances are quite long, especially with the aug-cc-pVTZ basis set. Among the results that compare favorably to experiment, the aug-cc-pVTZ distances are systematically shorter than the cc-pVTZ values by 0.1−0.2 Å. A similar "diffuse function effect" was noted previously in ref 27 for MPW1K and also for CH_3CN-BF_3,[17] though it is much less pronounced in the present case. In contrast, pure DFT methods perform quite poorly, as most of the calculated B−N distances are longer than experiment, especially OLYP. The exception is *m*PWPW91, which agrees quite well; and like most of the hybrid DFT results, the bond length is a bit shorter for the aug-cc-pVTZ basis set. Interestingly, the MP2/aug-cc-pVTZ result is notably shorter than experiment, by about 0.1 Å, and we initially suspected a genuine difference between the equilibrium and vibrationally averaged structures, stemming from a flat B−N distance potential, which had been examined previously at the MP2/ d95v++(2d1f,2p) level.[24] However, the MC-QCISD result for R_{BN} is 2.472 Å, which is almost coincident with experiment.

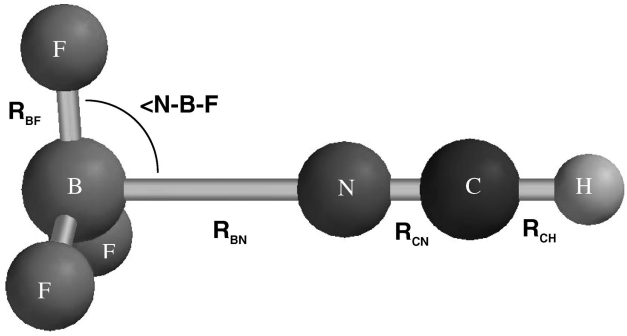


Figure 1. Structural parameters of $HCN-BF_3$.

The apparent inconsistency of the MP2 results relative to experiment and MC-QCISD, together with the very slight difference in MCG3 energies between the MP2 and MC-QCISD minima (0.2 kcal/mol), caused us to undertake a further examination of the intermolecular potential. Figure 2 shows B−N potential curves obtained via B3PW91, MP2, MCG3//B3PW91/aug-cc-pVTZ, and MCG3//MP2/aug-cc-pVTZ calculations. It is worth noting that the MCG3 energies of the B3PW91 (open circles in Figure 2) and MP2 geometries are very similar, within 0.05 kcal/mol in every instance. Indeed, the B3PW91 curve does have a slightly flattened region along the inner wall, around 2.0 Å, which is somewhat reminiscent of the CH_3CN-BF_3 potential, but is much less pronounced in this case. This feature is much less apparent in the MP2 and MCG3 curves.

Vibrational energy levels and wave functions for the B−N stretching mode (from the MCG3//MP2 curve) are shown in Figure 3. For the most part, these functions are quite symmetrical and do not reflect a great deal of anharmonicity. As a result, the average B−N bond distances calculated as expectation values over the vibrational wave functions differ only slightly from the equilibrium values. The results are included in Table 2 (below) with complete structural parameters for B3PW91, MP2, and MC-QCISD. At the B3PW91 level, the average distance is only 0.007 Å longer than the equilibrium distance. At the MP2 level, the average distance is actually 0.005 Å *shorter*. A comparison between the average bond length from the MCG3//MP2 potential curve and the optimized MC-QCISD structure is not entirely direct, but it is worth noting that the MCG3//MC-QCISD point does lie on the MCG3//MP2 curve in Figure 2. In any case, the equilibrium and average B−N distances agree exactly, and they differ from the measured value by only 0.001 Å (though the experimental error bar is actually much larger: 0.029 Å). We also note that the use of ground-state vibrational wave functions to compute the expectation value of R_{BN} is the appropriate choice, rather than a thermal average, since the experimental structure was deduced from an analysis of the microwave spectrum of the ground-state complex obtained in a cold, supersonic expansion.[18]

Table 1: Structural Properties, Dipole Moment, and Binding Energy of HCN−BF_3

method[a]	basis set	R_{BN} (Å)	ΔR (calc-exp)	⟨NBF (deg)	μ (D)	ΔE (kcal/mol)
HF	cc-pVTZ	2.675	0.202	92.2	4.17	
	aug-cc-pVTZ	2.664	0.191	92.3	4.22	−4.9
MP2	cc-pVTZ	2.419	−0.054	93.5	4.44	
	aug-cc-pVTZ	2.361	−0.112	93.9	4.51	−6.3
			Hybrid DFT Methods			
B3LYP	cc-pVTZ	2.552	0.079	92.8	4.14	
	aug-cc-pVTZ	2.535	**0.062**	93.0	4.23	−3.8
MPW1K	cc-pVTZ	2.352	−0.121	94.1	4.62	
	aug-cc-pVTZ	2.323	−0.150	94.4	4.74	−4.7
O3LYP	cc-pVTZ	2.596	0.123	91.4	3.58	
	aug-cc-pVTZ	2.996	0.523	91.4	3.61	−2.2
B3PW91	cc-pVTZ	2.493	0.020	93.2	4.26	
	aug-cc-pVTZ	2.465	**−0.008**	93.5	4.37	−3.2
*m*PW1PW91	cc-pVTZ	2.422	−0.051	93.6	4.41	
	aug-cc-pVTZ	2.390	−0.083	93.9	4.54	−4.1
B98	cc-pVTZ	2.539	0.066	92.9	4.17	
	aug-cc-pVTZ	2.513	**0.040**	93.1	4.28	−4.3
B97−2	cc-pVTZ	2.599	0.126	92.6	4.06	
	aug-cc-pVTZ	2.586	0.113	92.7	4.13	−3.3
			Pure DFT Methods			
BLYP	cc-pVTZ	2.655	0.182	92.4	3.93	
	aug-cc-pVTZ	2.642	0.169	92.5	4.02	−7.4
OLYP	cc-pVTZ	3.132	0.659	91.1	3.39	
	aug-cc-pVTZ	3.203	0.730	91.0	3.43	−1.8
BPW91	cc-pVTZ	2.597	0.124	92.7	4.02	
	aug-cc-pVTZ	2.655	0.182	93.0	4.13	−2.3
*m*PWPW91	cc-pVTZ	2.503	0.030	93.2	4.18	
	aug-cc-pVTZ	2.467	**−0.006**	93.5	4.32	−3.5
HCTH	cc-pVTZ	2.896	0.423	91.5	3.60	
	aug-cc-pVTZ	2.922	0.449	91.5	3.64	−3.4
			Multicoefficient Methods			
MC-QCISD[b]		2.472	−0.001	92.9		−5.6
MCG3[c]//MCQISD						−5.7
MC-QCISD//MP2						−5.5
MCG3//MP2						−5.6
experiment[q]		2.473(29)			4.135(7)	

[a] See ref 1. [b] Reference 29. [c] Reference 30. [d] References 18 and 19.

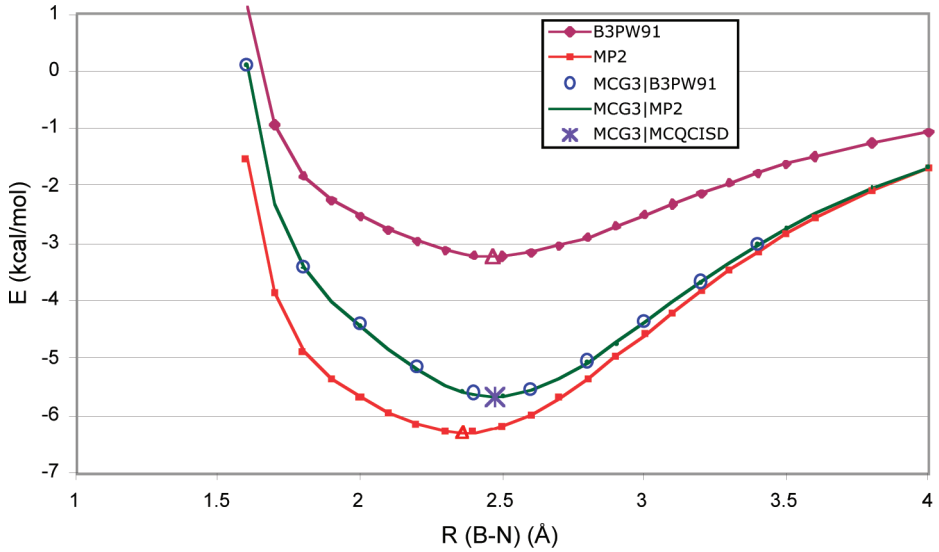


Figure 2. The B−N distance potential of HCN−BF_3. The zero of energy is set to that of the infinitely separated monomers. The aug-cc-pVTZ basis set was used for the B3PW91 and MP2 calculations. Minimum energy points on the MP2 and B3PW91 curves are noted with open triangles.

Thus, it appears that a comparison of equilibrium and vibrationally averaged structures is valid for this system, and we conclude that the best DFT methods for structure prediction seem to be the hybrid methods B3PW91 and B98 as well as the pure method *m*PWPW91. As such, we tested for convergence of the structural results by calculating B3PW91 structures with the cc-pVDZ, aug-cc-pVDZ, cc-pVQZ, and aug-cc-pVQZ basis sets. The B−N distances obtained were 2.312 Å, 2.305 Å, 2.504 Å, and 2.494 Å, respectively. Thus, the structure is nearly converged at the aug-cc-pVTZ level, but the B−N distance may indeed be a bit longer at the infinite basis set limit.

Complete structural results for B3PW91/aug-cc-pVTZ, MP2/aug-cc-pVTZ, and MC-QCISD (which includes the

Table 2: Experimental and Theoretical Values of Structural Parameters for HCN−BF_3

method	R_{BN}(eq) (Å)	$\langle R_{BN}\rangle$ (Å)	R_{CN} (Å)	R_{BF} (Å)	R_{CH} (Å)	<N−B−F (deg)
B3PW91/aug-cc-pVTZ	2.465	2.472	1.143	1.319	1.068	93.5
MP2/aug-cc-pVTZ	2.361	2.356	1.164	1.324	1.107	93.9
MC-QCISD[a]	2.472	2.472	1.155	1.317	1.069	93.0
experiment[b]		2.473(29)	1.153	1.310	1.065	90−93

[a] The equilibrium geometry was determined from an optimization via MC-QCISD. The average bond length $\langle R_{BN}\rangle$ was determined from the B−N distance potential calculated via MCG3 energies at the MP2 geometries. See text for discussion. [b] Note that in the experimental structure analysis, most of the parameters were fixed at the values of free HCN, and BF_3. Also, there was very little sensitivity to the N−B−F angle, so a value of 91.5° was chosen as it was an average of experiment and theory. See ref 18 for discussion.

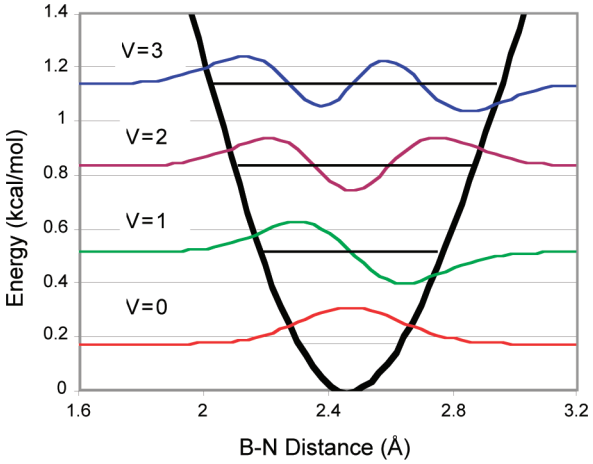


Figure 3. Energy levels and wave functions for the B−N stretching mode in HCN−BF_3.

$\langle R_{BN}\rangle$ value from the analysis of the MCG3//MP2 potential curve) as well as the experimental values[18] are provided in Table 2. It is worth noting that beyond the nonplanar distortion of the BF_3 unit, there are some other very minor structural changes that occur within the two monomer units upon complexation. Specifically, the B−F distance increases, and the C−N distance contracts. We note that the C−N distance in free HCN is calculated to be 1.167 Å by MP2/aug-cc-pVTZ and 1.146 Å by B3PW91/aug-cc-pVTZ, in both cases 0.003 Å longer than in the complex. Similarly, the B−F distances in free BF_3 are 1.317 and 1.314 Å for MP2/aug-cc-pVTZ and B3PW91/aug-cc-pVTZ, respectively, which is in both cases a bit shorter than in the complex. These complex-induced changes are quite small and indicate that the assumptions made in the experimental structural analysis were valid.

Dipole Moment. The dipole moment data in Table 1 demonstrate that the dipole increases monotonically with the B−N distance, as was noted previously in ref 27. In addition, it appears that this trend persists across all methods studied herein. This occurs because the nonplanar distortion of the BF_3 subunit increases with decreasing B−N and also because there is a steady increase in charge transfer as the bond shortens. Among methods that yield reasonable structures, all predict dipoles that are larger than experiment[19] by about 0.1−0.2 D. Furthermore, the MP2 results are much larger than experiment, consistent with the B−N distances being predicted to be too short. To determine whether this effect was structural or a result of method, we also calculated an MP2 dipole moment at the B3PW91 structure. The result, 4.40 D, is still significantly larger than the experimental dipole but only differs from the B3PW91 result by 0.03 D. We also considered the possibility that there might be a genuine difference between the experimental, vibrationally averaged dipole moment and the value that corresponds to the equilibrium bond length. Thus, using the vibrational wave functions determined from the analysis of the MCG3//MP2 potential, together with the dipole moment values for the points along the B3PW91/aug-cc-pVTZ curve, we calculated the expectation value of the dipole for the ground vibrational state. The dependence of the dipole moment on B−N distance is displayed in Figure 4, and since an analytical function was needed for the grid-based routine,[33] we fit the data from 1.7 to 3.6 Å to a fifth-order polynomial. This function reproduced the calculated dipole moment values to within about 0.01 D throughout the range of the fitted data (the average magnitude of the residuals was 0.0075 D). The vibrationally averaged dipole moment obtained in this analysis was 4.38 D, only 0.01 D larger than the equilibrium value of 4.37 D (for B3PW91 at 2.465 Å). The vibrationally averaged value is slightly larger because, about the average B−N distance (2.472 Å), the dipole moment function increases more rapidly toward shorter B−N distances than it decreases toward longer distances. It is also interesting to note that the point at which the dipole moment function

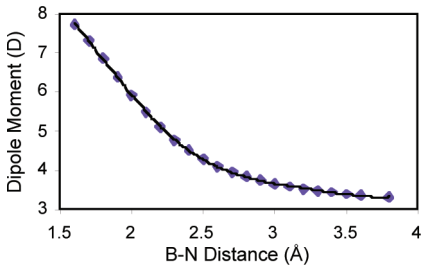


Figure 4. Calculated dipole moment (B3PW91/aug-cc-pVTZ) of HCN−BF_3 versus B−N distance.

Table 3: Observed and Calculated Frequencies[a] of BF_3

	mode[c]				
method[b]	ν_1 (SS)	ν_2 (SD)	ν_3 (AS)	ν_4 (AD)	RMS diff
experiment[d]	888	691	1449	480	
MP2	888	691	1463	475	**7.6**
B3LYP	885	681	1447	472	**6.3**
MPW1K	921	623	1509	460	49.3
O3LYP	877	676	1437	469	12.4
B3PW91	889	683	1457	473	**6.7**
*m*PW1PW91	896	686	1470	476	11.9
B98	890	683	1459	474	**7.0**
B97−2	888	685	1458	475	**6.0**
BLYP	851	658	1384	457	42.7
OLYP	851	658	1384	457	42.7
BPW91	856	660	1401	458	34.3
*m*PWPW91	858	660	1403	458	33.3
HCTH	862	669	1412	464	26.5

[a] In units of cm^{-1}. [b] The aug-cc-pVTZ basis set was used for all methods. [c] Modes 1−4 are the symmetric stretch (SS), the symmetric deformation (SD) (or "umbrella" mode), the asymmetric stretch (AS), and the asymmetric deformation (AD), respectively. [d] Reference 36.

Table 4: Calculated Vibrational Frequencies[a] of HCN−BF_3

mode no.	symmetry[b]	description	MP2[c]	B3LYP[c]	B3PW91[c]	*m*PW1PW91[c]
ν_1	A_1	BN stretch	98	87	77	79
ν_2	A_1	BF_3 "umbrella"	609	622	614	608
ν_3	A_1	BF symmetric stretch	870	873	874	878
ν_4	A_1	CN stretch	2055	2223	2230	2251
ν_5	A_1	CH stretch	3457	3440	3443	3461
ν_6	E	intermolecular bend	68	62	65	69
ν_7	E	BF_3/HCN wobble	201	162	177	194
ν_8	E	BF_3 asymmetric bend	474	471	472	475
ν_9	E	HCN bend	729	770	774	784
ν_{10}	E	BF_3 asymmetric stretch	1432	1424	1431	1438

[a] In units of cm^{-1}. [b] In the C_{3v} point group. [c] The aug-cc-pVTZ basis set was used for all methods.

begins to increase quite rapidly is very near the gas-phase B−N distance. A marked increase in charge separation would certainly impede further contraction of the B−N bond in the absence of a stabilizing medium of some form.

Binding Energy. The binding energy of the complex at the MCG3//MC-QCISD level is −5.7 kcal/mol. By contrast, all DFT methods compare quite poorly to MCG3//MC-QCISD as they predict binding energies that are smaller and indeed no closer to the MCG3 than the HF result. This is not surprising given that the complex is rather weak and DFT often fails in such instances.[1,11] Furthermore, the discrepancy must not stem from basis-set superposition error (BSSE), as the counterpoise correction typically reduces the value of the binding energy. The MP2 binding energy is slightly higher than MCG3, and therefore agreement with MCG3 could possibly be improved by correcting for BSSE. However, a counterpoise correction on the B3LYP/aug-cc-pVTZ structure of CH_3CN-BF_3 raised the energy by only 0.1 kcal/mol.[17]

Vibrational Frequencies. Experimental vibrational frequencies have provided key insight into the effects of bulk condensed phases on the structural properties of nitrile−BF_3 systems,[21] but gas-phase frequencies are needed for comparison, and we have relied on computations for these data in previous studies.[17] In this respect, there are two key criteria by which we judge model performance. First of all, a given model must adequately predict the gas-phase structure of the complex. In addition, since we have found that the BF_3-localized modes are most sensitive to structural changes,[17,21] modeling the BF_3 force field is key as well. Since gas-phase frequencies of HCN−BF_3 have not been measured (aside from the C−H stretch[35]), we turn to the frequencies of the BF_3 subunit. Table 3 displays observed[36] and calculated frequencies of BF_3, together with RMS differences between experiment and theory. We have *not* scaled these frequencies. Typically, computational frequencies are scaled by a factor slightly less than 1.0 to facilitate agreement with experiment[1] (e.g. 0.9806 for B3LYP/6-31G(d)[37]). In this instance, such a scaling would actually decrease agreement with experiment. A similar trend was noted previously in the study of CH_3-CN−BF_3 frequencies,[17] and therein it was (arbitrarily) decided to apply a scale factor only to frequencies above 2000 cm^{-1}, which only affected the CN and C−H stretches (of which the former seemed to be problematic, see below).

It is clear that, in addition to MP2, the hybrid methods B3LYP, B3PW91, B98, and B97-2 all perform quite well (RMS differences less than 10 cm^{-1}), and *m*PW1PW91 is nearly as good. Pure DFT methods perform quite poorly. For the complex, we selected three DFT methods that performed well on free BF_3 and bracketed the experimental structural results: B3LYP, B3PW91, and *m*PWPW91. These frequencies, together with MP2 data, are shown in Table 4. For the most part the results are consistent, especially for the BF_3-localized modes (ν_2, ν_3, ν_8, and ν_{10}). We note also that the differences among methods parallel those of the free BF_3 calculations, rather than the expected shifts based on the slight structural differences.[17] This most likely reflects the very small degree of nonplanar distortion in the BF_3 subunit. A few of the intermolecular modes are a bit higher for MP2, which reflects a tighter intermolecular potential (as is reflected in Figure 2). We also note that there is a significant difference between DFT and MP2 for the C−N stretch. The measured CN stretch of HCN is 2097 cm^{-1},[36] and we calculate 2021 cm^{-1} using MP2/aug-cc-pVTZ and 2204 cm^{-1} using B3PW91. Scaling the latter by a factor of 0.9772 (recommended for B3PW91/6-31G(d)[37]) would change the result to 2153 cm^{-1}, which is only a modest improvement. Thus, while DFT seems to do well for the BF_3 force field, it performs poorly for the HCN moiety, and the MP2 results are at best marginal.

IV. Conclusions

We have shown that certain hybrid DFT methods accurately predict the structure of HCN−BF_3 relative to experiment and demonstrated that calculated equilibrium structures can be compared to experimental, vibrationally averaged structures for this system. We also find that several hybrid DFT methods accurately predict the vibrational frequencies of BF_3 relative to experiment[36] and frequencies of the complex relative to MP2. The possible exception to this success is the CN stretching mode of the complex, for which the MP2 result also seems marginal. Overall, B3PW91 seems to perform the best when considering both structure and vibrational frequencies. We also found that both DFT and MP2 predict dipole moments that are quite reasonable, i.e., within about 0.2 D, of the experimental value. It was also demonstrated that the difference between the equilibrium and vibrationally averaged dipole moment values was negligible. However, despite their success with structure, frequencies, and dipole moment, all DFT methods significantly underestimated the complex binding energy, which was −5.7 kcal/mol at the MCG3//MC-QCISD level.

Acknowledgment. This work was supported by the National Science Foundation (RSEC CHE-0137438 and

CHE-0203346 to C.J.C. and CHE-0407824 to J.A.P.). J.A.P. also acknowledges The Henry Dreyfus Teacher-Scholar Program administered by the Camille and Henry Dreyfus Foundation as well as sabbatical support from the Faculty Sabbatical Leave Program administered by the Office of Research and Sponsored Programs at the University of Wisconsin—Eau Claire and the NSF-RSEC program at the University of Minnesota, Department of Chemistry.

References

(1) For a recent overview of DFT and other computational methods used and discussed herein, see: Cramer, C. J. *Essentials of Computational Chemistry*, 2nd ed.; John Wiley and Sons: Chichester, 2004; and references therein.

(2) Hobza, P.; Sponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315.

(3) Garcia, A.; Cruz, E. M.; Sarasola, C.; Uglade, J. M. *J. Mol. Struct. (THEOCHEM)* **1997**, *397*, 191.

(4) Couronne, O.; Ellinger, Y. *Chem. Phys. Lett.* **1999**, *306*, 71.

(5) Tsuzuki, S.; Lüthi, H. P. *J. Chem. Phys.* **2001**, *114*, 949.

(6) Salazar, M. C.; Paz, J. L.; Hernandez, A. J.; Manzanares, C. I.; Ludena, E. V. *Theor. Chem. Acc.* **2001**, *106*, 218.

(7) Salazar, M. C.; Paz, J. L.; Hernandez, A. J.; Manzanares, C. I.; Ludena, E. V. *Int. J. Quantum Chem.* **2003**, *95*, 177.

(8) Kim, K.; Jordan, K. D. *J. Phys. Chem.* **1994**, *98*, 10089.

(9) Latajka, Z.; Bouteiller, Y. *J. Chem. Phys.* **1994**,*101*, 9793.

(10) Del Bene, J. E.; Person, W. B.; Szczpaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705.

(11) Cybilski, S. M.; Seversen, C. E. *J. Chem. Phys.* **2005**, *122*, 014117.

(12) Ruiz, E.; Salahub, D. R.; Vela, A. *J. Am. Chem. Soc.* **1995**, *117*, 1141.

(13) Ruiz, E.; Salahub, D. R.; Vela, A. *J. Phys. Chem.* **1996**, *100*, 12265.

(14) Latajka, Z.; Berski; S. *J. Mol. Struct. (THEOCHEM)* **1996**, *371*, 11.

(15) Garcia, A.; Cruz, E. M.; Sarasola, C.; Ugalde, J. M. *J. Phys. Chem. A* **1997**, *101*, 3021.

(16) Gilbert, T. M. *J. Phys. Chem. A* **2004**, *108*, 2550.

(17) Giesen, D. J.; Phillips, J. A. *J. Phys Chem. A* **2003**, *107*, 4009.

(18) Reeve, S. W.; Burns, W. A.; Lovas, F. J.; Suenram, R. D.; Leopold, K. R. *J. Phys. Chem.* **1993**, *97*, 10630.

(19) Fiacco, D. L.; Mo, Y.; Hunt, S. W.; Ott, M. E.; Roberts, A.; Leopold, K. R. *J. Phys. Chem. A* **2001**, *105*, 484.

(20) Leopold, K. R.; Canagaratna, M.; Phillips, J. A. *Acc. Chem. Res.* **1997**, *30*, 57, and references therein.

(21) Wells, N. P.; Phillips, J. A. *J. Phys. Chem. A* **2002**, *106*, 1518.

(22) Burns, W. A.; Leopold, K. R. *J. Am. Chem. Soc.* **1993**, *115*, 11622.

(23) Fiacco, D. L.; Leopold, K. R. *J. Phys. Chem. A* **2003**, *107*, 2808.

(24) Hankinson, D. J.; Amlöf, J.; Leopold, K. R. *J. Phys. Chem.* **1996**, *100*, 6904.

(25) Iglesias, E.; Sordo, T. L.; Sordo, J. A. *Chem. Phys. Lett.* **1996**, *248*, 179.

(26) Cabaleiro-Lago, E. M.; Ríos, M. A. *Chem Phys. Lett.* **1998**, *294*, 272.

(27) Venter, G.; Dillen, J. *J. Phys. Chem. A* **2004**, *108*, 8378.

(28) Nxumalo, L. M.; Andrezejak, M.; Ford, T. A. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 377.

(29) Fast, P. L.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 6111.

(30) Tratz, C. M.; Fast, P. L.; Truhlar, D. G. *PhysChemComm* **1999**, 2/14, 1.

(31) 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.; Sarkas, O. F.; 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.

(32) MULTILEVEL-version 4.0/G03, by Rodgers, J. M.; Lynch, B. J.; Fast, P. L.; Zhao, Y.; Pu, J.; Chuang, Y.-Y.; Truhlar, D. G. University of Minnesota, Minneapolis, 2002, based on GAUSSIAN03, ref 31.

(33) http://www.tlchm.bris.ac.uk/dynamics/fghqcpe.for, accessed April, 2005.

(34) (a) Marston, C. C.; Balint-Kurti, G. G. *J. Chem. Phys.* **1989**, *91*, 3571. (b) Balint-Kurti, G. G.; Dixon, R. N.; Marston, C. C. *Intl. Rev. Phys. Chem.* **1992**, *11*, 317.

(35) Kerstel, E. R. Th.; Pate, B. H.; Mentel, T. F; Scoles, G. *J. Chem. Phys.* **1994**, *101*, 2762.

(36) Herzberg, G. *Molecular Spectra and Molecular Structure, Volume II, Infrared and Raman Spectra of Polyatomic Molecules*; Krieger Publishing Co.: Malabar, FL, 1991.

(37) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. http:/comp.chem.umn.edu/database/freq_scale.htm, accessed April, 2005.

CT0501248

JCTC Journal of Chemical Theory and Computation

Quantitative Ranking of Crystal Packing Modes by Systematic Calculations on Potential Energies and Vibrational Amplitudes of Molecular Dimers

A. Gavezzotti*

Dipartimento di Chimica Strutturale e Stereochimica Inorganica, University of Milano, Via Venezian 21, 20133 Milano, Italy

Received April 6, 2005

Abstract: Quantum chemical calculations including electron correlation and calculations with the density sums (Pixel) method have been performed on a variety of molecular dimers representing some frequently observed recognition modes in molecular condensed phases. Notwithstanding some individual fluctuations when different computational methods are used, there is a general agreement for the relative orders of magnitude. The results have been collected in a table that ranks the interaction energies and amplitudes of the energy wells for the recognition between molecular fragments, providing a quantitative guideline to assess the relative importance of hydrogen bonding, aromatic ring stacking, antiparallel arrangements of polar moieties, weak Coulombic C−H···X interactions, and dispersive interactions between nonpolar groups. Since the Pixel method naturally allows for a separation between Coulombic, dispersion, polarization, and repulsion energy contributions, their relative importance can be analyzed with respect to the chemical constitution of the interacting partners. The relevance of these results to the current ideas and methods of "crystal engineering" is discussed.

1. Introduction

Qualitative concepts, related to intuitive permanent polarizations of functional groups and recognizing selected contacts between atom spheres or molecular fragments of supposed relevance, are often used for the analysis, prediction, and control of the architecture of organic crystals. The identification of these "synthons" usually relies on contact distance criteria; a cursory, random search through the latest issues of some major journals found several reports of accurate crystallographic structural results, interpreted in terms of short atom−atom distances.[1] These analyses are seldom, if ever, accompanied by a quantitative calculation of the implied energies. In fact, when one takes into proper account the physical foundations of the intermolecular potential field in organic crystals, the problem reveals its complexity,[2] because packing forces are weak and scarcely selective, so that many different structures of nearly equal energy are possible. A quantitative calculation of the intermolecular potentials becomes, then, indispensable; even that may not be sufficient to resolve all ambiguities, but a scale of orders of magnitude of involved energies would be a great help in setting at least a reproducible background to packing analysis. The nature of the problem is the same when the same molecular recognition concepts are applied to the analysis of the structure of biological macromolecules or of interactions within the biological world in general, such as drug−receptor or enzyme−DNA binding and the like.

The idea that the key interactions between complex molecular systems can be modeled by choosing appropriate molecular dimers with a confrontation of selected interaction units goes back to the times when the first reliable, all-electron quantum chemical calculations became possible.[3] In the past years, electron correlation gradually became tractable for molecular dimers, and the number of calculations of that kind that have appeared in the literature is very large and ever increasing; systems of interest here that have

* Corresponding author e-mail: angelo.gavezzotti@unimi.it.

10.1021/ct050091w CCC: $30.25
Published on Web 06/30/2005

been treated by high-level quantum chemical methods include the water dimer;[4,5] small alcohols;[6,7] fluoromethanes with water, alcohols, and formaldehyde[8] or with ammonia and small amines;[9] water and other hydrides with small hydrocarbons, acetylene, ethylene, methane, ethane,[10] or their fluorinated derivatives;[11] ethylene and benzene π systems against first row hydrides[12] or other hydrocarbons;[13] formic acid hydrogen bonds;[14] benzene[15,16] and its fluorinated dimers;[17,18] pyrimidine−benzoquinone;[19] pyrimidine−water;[20] and the hydrogen bond between small amides,[21] this being a sample and by no means exhaustive list. Because of computational limitations, such studies usually consider rather small systems, but one can also find essays in the calculation of dimerization energies for large molecules such as naphthalenes.[22] In this paper, we show that the results of the recently proposed density sums (Pixel) method[23−25] match very well those of accurate quantum chemical calculations that require a 100-fold larger computing effort. We thus have systematically applied the Pixel method to the calculation of potential energy profiles for the production of a quantitative compendium of the interaction energies for typical molecular recognition modes. Perhaps even more important than the total binding energy is the width of the corresponding energy well, a measure of the expected vibrational amplitudes and, hence, of the reliability of the corresponding approach mode as a true packing binder. In addition, the relative importance of Coulombic-polarization contributions against dispersive-repulsion effects can be analyzed and discussed.[2,25]

2. Computation Methods

Besides standard MP2 and DFT techniques,[26] we use a method based on numerical integration over the whole molecular charge density, the Pixel-SCDS (semiclassical density sums) method.[23−25] In the latter approach, a molecule is a rigid object and no intramolecular energies are calculated. For the evaluation of intermolecular interactions, the valence molecular charge density is calculated[26] from the MP2/6-31G** wave function on a grid with a step of 0.08 Å. These original charge pixels, typically around 10^6 in number, are then condensed so that each condensed pixel contains the sum of an $n \times n \times n$ block from the original charge density. On the basis of previous experience,[25] n, called the condensation level, is taken as equal to 4 for crystal calculations and equal to 3 for dimer calculations, and those pixels whose charges are below the threshold of 10^{-6} electrons are screened out. Each molecule is, thus, eventually represented by atomic nuclear locations and charges and by 5000−20 000 condensed electron charge pixels.

Both the original and the condensed charge pixels are assigned to atomic basins by a new procedure, as follows. Let p be the number of atoms for which the nucleus−pixel distance is smaller than the atomic radius. If $p = 1$, the charge pixel is within one atomic sphere only, and it is assigned to that atom. If $p > 1$, the pixel is assigned to the atom from which the distance is the smallest fraction of the atomic radius. If $p = 0$, the pixel is assigned to the atom whose atomic surface is nearest. Atomic polarizabilities α_{atom} are reassigned in a consistent manner (H, 0.39; aliphatic C, 1.05; aromatic C, 1.35; N, 0.95; O, 0.75; F, 0.50 Å^3), and the local polarizability at a pixel of charge q_i belonging to the basin of an atom with a number of valence electrons equal to Z_{atom} is calculated as before:[24]

$$\alpha_i = (q_i/Z_{atom})\alpha_{atom}$$

Intermolecular Coulombic energies, E_{coul}, are calculated as sums over pixel−pixel, pixel−nucleus, and nucleus−nucleus charges, without any empirical parameterization. By definition, these Coulombic energies include penetration energies. Accuracy here is inversely proportional to the condensation level, but with $n = 3$, the numbers thus obtained are in good agreement with those obtained by accurate integration or by other energy partitioning schemes.[2]

The electric field, ϵ_i, generated by the pixel distribution in all surrounding molecules at any pixel of the molecule under consideration is then calculated, and the total polarization energy, E_{pol}, of that molecule is evaluated as a sum over pixel contributions, $E_{POL,i}$, each of which is calculated by a damped linear polarization formula (damping parameter $\epsilon_0 = 150 \times 10^{10}$ V m^{-1}):

$$E_{POL,i} = -{}^1/_2\alpha_i(\epsilon_i\, d_i)^2 \text{ with } d_i = \exp -[\epsilon_i/(\epsilon_0 - \epsilon_i)] \text{ for } \epsilon < \epsilon_0$$

$$E_{POL,i} = 0 \text{ for } \epsilon > \epsilon_0$$

The Pixel polarization energy is, thus, a many-body, nonpairwise additive energy.

Dispersion energies, E_{disp}, are evaluated as sums of pixel−pixel dispersion contributions, $E_{i,j}$, by a London-type formula, also damped to avoid divergence at short intermolecular distances between overlapping electron densities (damping parameter $D = 3.5$ Å):

$$E_{i,j} = -{}^3/_4 f(R)\, \omega\alpha_i\alpha_j/[(4\pi\epsilon^\circ)^2(R_{ij})^6] \text{ with } f(R) = \exp[-(D/R_{ij} - 1)^2] \text{ for } R_{ij} < D \text{ and } f(R) = 1 \text{ for } R_{ij} > D$$

The "oscillator strength" ω is taken here as the molecular first ionization potential, in turn taken as its Koopmans' theorem equivalent, the energy of the highest occupied molecular orbital. This is an acceptable approximation whenever the HOMO is indeed populated by the electrons involved in the intermolecular contact but might fail when the dispersion effect concerns inner electrons. A refinement of the method, using a variable ω for each intermolecular contact on the basis of a weighted average of atomic ionization potentials is presently being considered.

The repulsion energy, E_{rep}, is modeled as proportional to intermolecular overlap. The total overlap integral between the charge densities of any two molecules is calculated by numerical integration over the original uncondensed densities and is then subdivided into contributions from pairs of atomic species, S_{mn}, using the assignment of pixels to atomic basins. For each $m-n$ pair, the repulsion energy is evaluated as

$$E_{REP,mn} = (K_1 - K_2\Delta\chi_{mn})S_{mn}$$

where $\Delta\chi_{mn}$ is the corresponding difference in Pauling electronegativity (C, 2.5; N, 3.0; O, 3.5; H, 2.1; Cl, 3.0). The values of K_1 and K_2 were optimized at 4800 and 1200,

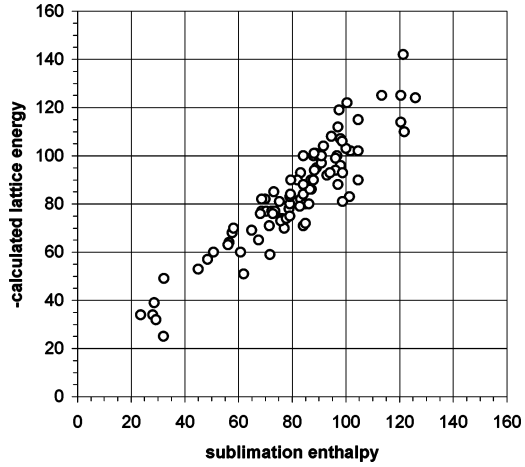


Figure 1. Negative of Pixel-calculated lattice energies versus experimental heats of sublimation for 91 organic crystals (kJ mol^{-1}). The least-squares line is $y = 1.03x$; $R^2 = 0.84$.

respectively (for energies in kJ mol^{-1} with electron densities in electrons Å^{-3}). The rationale behind this approach is that when atoms of different electronegativity meet, a larger reorganization of the electron density occurs and the overlap repulsion must be smaller. The total repulsion energy is then the sum over all $m-n$ pairs.

The total Pixel energy is then the sum of the various contributions:

$$E_{\text{Pixel}} = E_{\text{coul}} + E_{\text{pol}} + E_{\text{disp}} + E_{\text{rep}}$$

The present form of the Pixel method differs from previous applications[23−25] in the new assignment of pixels to atomic basins and in the new form of the repulsion energy. As an overall check of performance, lattice energies calculated for 91 crystal structures retrieved from the Cambridge Structural Database[27] for molecules containing C, H, N, O, and Cl atoms, including most of the common chemical functions with or without hydrogen-bonding capability, were compared with measured enthalpies of sublimation.[28] As seen in Figure 1, the overall agreement between calculated lattice energies and sublimation enthalpies is acceptable. Some discrepancies, some of which are systematic over certain chemical functionalities, do however remain. The difficulties in the comparison between calculated lattice energies and experimental heats of sublimation have been discussed previously;[29−31] they stem from the temperature dependence of the experimental value versus the theoretically "no-temperature" nature of the calculation, in comparison also with the temperature at which the crystal structure determination (from which the starting data for the calculation were taken) has been carried out; from the intrinsic awkwardness of the experimental determinations, which require the evaluation of extremely low vapor pressures, witness the spread of values often reported for the same compound, quite often spanning as much as 5−10 kJ mol^{-1}, and from other less-predictable factors. For a perhaps extreme example, the sublimation enthalpy entry for anthracene in ref 28 has 24 different values in the range 84−105 kJ mol^{-1}. Another measure of the reliability of a computational method can be drawn from a comparison of its performance with the performance of other well-established methods for the same calculation; the results of the present Pixel calculations are slightly better than those obtained on a more restricted set of crystal structures with the previous formulation of the Pixel method and are also slightly better than those obtained by computing lattice energies with a heavily parameterized atom−atom force field[29,30] calibrated using the same thermodynamic database. In addition, it should be recalled that these discrepancies are the result of an accumulation of errors over many molecular dimers in the crystal. The adverse effects of the remaining deficiencies of the Pixel formulation are expected to be less relevant for single dimer calculations.

We have selected 55 molecular dimers that, in our opinion, represent the most common recognition modes in organic crystals as well as the effects of electron-withdrawing substituents in increasing the "acidity" of C−H groups. Pixel total energies were then calculated as a function of intermolecular separation, along the most obvious approach coordinates (interplanar distance for aromatic ring dimers, X···H separation for hydrogen bonds) using rigid, standardized monomer geometries.[32] Equilibrium intermolecular distances and binding energies calculated by the Pixel method were compared with literature ones, when a high quality quantum chemical calculation was available; otherwise, potential-energy curves were recalculated at the MP2 or DFT 6-31G** level. The results are obviously very sensitive to basis set superposition error (BSSE), which depends on the chemical constitution of the interacting moieties at short range, so that its entity may be different even for two different orientations of the same dimer. Although the most frequent procedure is to consider both corrected and uncorrected energies, attaching more physical significance to the latter, sometimes the attitude of the quantum chemical community toward the counterpoise correction oscillates and uncorrected energies are considered for some purposes,[8,9] whereas it has also been suggested that the full counterpoise correction may, in some cases, overestimate the error[12] and that the use of 50% counterpoise correction may be advisable.[33] We have applied the standard counterpoise correction procedure ("counterpoise = 2" option[26]), and we report both uncorrected and corrected interaction energies. The width of the potential-energy well was estimated as corresponding to an energy increase of 5 kJ mol^{-1}, about 1.5RT at 400 K, an average melting temperature for an organic crystal, or about 2RT at room conditions.

3. Results and Discussion

For hydrogen-bonded dimers (see Table 1), the Pixel results are in very good agreement with the results of high-level quantum mechanical calculations as well as with the experimental results available for the water dimer.[34] In addition, the Pixel-calculated dimerization energies are in good agreement with the results of counterpoise-corrected MP2 ab initio calculations. For the $\pi-\pi$ interactions between parallel aromatic rings, the Pixel results are more similar to those obtained by MP2 calculations without counterpoise correction (see Table 2). This may suggest that the Pixel dimerization energies are overestimated, but the comparison with very high quality calculations for the benzene dimer[15]

Table 1. Equilibrium Distances, Binding Energies, and Widths of the Potential-Energy Well for Hydrogen Bonds

dimer	R^o [a]	ΔE[b]	width[c]	literature R^o and ΔE
benzoic acid, cyclic O−H···O=C	1.75	67	1.65−1.95	
formic acid, cyclic O−H···O=C	1.75	67	1.65−1.95	1.68, 68[14]
formic acid, single O−H···O=C	1.75	38	1.55−2.05	1.76, 40[14]
pyridone, cyclic N−H···O=C	2.05	48	1.85−2.35	2.05, 42; 1.90, 68[d]
formamide, cyclic N−H···O=C	2.0	52	1.85−2.2	2.0, 55[21]
formamide, single N−H···O=C	2.05	25	1.8−2.45	
methanol O−H···O−H	1.95	23	1.7−2.3	2.1, 22; 1.95, 28[d]; 1.9, 23[6]
methanol-acetone O−H···O=C	2.0	19	1.7−2.5	
formamide-acetone N−H···O=C	2.0	19	1.75−2.45	
formamide-acetonitrile N−H···N	2.1	22	1.8−2.5	
formic acid-acetone O−H···O=C	1.8	31	1.55−2.1	1.9, 28; 1.9, 36[d]
water	2.0	23	1.85−2.2	1.96, 22.6 exptl[34]

[a] O···H or N···H hydrogen-bonding distance, Å. [b] Binding energy, kJ mol^{-1}. [c] Range of the distance coordinate for an energy increase of 5 kJ mol^{-1} from the minimum. [d] MP2/6-31G** with and without counterpoise correction.

shows that the best estimate of the interaction energy is even larger than the uncorrected MP2 estimate and is close to the Pixel value. A more serious difficulty is encountered for the interaction between the π system of benzene and highly polar OH or NH bonds in water and ammonia; the strength of these interactions is clearly overestimated by Pixel. On the other hand, the C−H···π dimerization energies are well-represented by the Pixel formulation, the results being similar to those of high-level quantum chemical calculations and being between those of counterpoise-corrected and -uncorrected MP2 calculations.

The dimerization energies calculated by the Pixel method for dimers constructed over C−H···O and C−H···N contacts are in good agreement with quantum mechanical results (see Table 3), being always between counterpoise-corrected and -uncorrected MP2 energies, and as such, they are presumably slightly overestimated. Table 4 shows that the repulsion between approaching hydrogen-rich molecular rims is disfavored, as expected. This table also shows that the interaction between parallel aliphatic chains is too attractive in the Pixel representation. This effect is visible also in the calculated lattice energies of saturated hydrocarbon crystals, which systematically exceed the corresponding sublimation enthalpies.

All considered, one can see a consistent agreement between Pixel and quantum mechanics results for a wide variety of chemical environments, although the computational cost of Pixel calculations is a very small fraction of that of quantum chemical calculations.

Tables 1−4 can be used as working tables for the estimation and ranking of packing forces. The largest potential-energy gain is for the acid O−H···O=C hydrogen bond (30−40 kJ mol^{-1}), followed by the stacking of condensed aromatic rings or benzenes with polar substituents (20−30 kJ mol^{-1}) and by the hydrogen bond in amides (15−25 kJ mol^{-1}) or alcohols (20−25 kJ mol^{-1}). Contrary to

Table 2. Equilibrium Distances, Binding Energies, and Widths of the Potential-Energy Well for Recognition Modes Involving Aromatic $\pi-\pi$ or H···π Interactions

dimer	R^o [a]	ΔE[b]	width[c]	quantum mechanical results for R^o and ΔE
		Parallel[d]		
$C_6H_6-C_6F_6$	3.5	23	3.25−4.0	3.7, 15; 3.4, 28[e]
1,3,5-$C_6H_3F_3$	3.6	20	3.3−4.0	3.7, 12; 3.5, 23[e]
pyrimidine-benzoquinone	3.45	22	3.2−3.85	ΔE = 23 at 50% BSSE[19]
chlorobenzene	3.6	18	3.35−4.0	3.9, 7; 3.6, 17[e]
C_6F_6	3.55	17	3.28−4.0	
C_6H_6	3.70	10	3.4−4.4	4.1, 3; 3.8, 10;[e] 3.75, 13[15]
nitrobenzene	3.45	29	3.15−3.8	3.5, 20; 3.3, 41[e]
		Offset Parallel[f]		
benzene	3.6	12	3.3−4.25	3.4, 18[15]
naphthalene	3.6	26	3.35−3.9	≈3.7, ≈25[22]
benzoquinone	3.35	18	3.15−3.85	3.6, 9; 3.2, 23[e]
		H···π Interactions		
benzene π−water[g]	2.65	17	1.4−2.35	3.3, 12[12]
benzene π−ammonia[g]	3.05	10	1.6−2.7	3.4, 8[12]
C−H···π acetylene−acetylene[h]	2.6	8	2.15−3.5	2.7, 5; 2.7, 9[e]
C−H···π benzene−benzene[i]	2.65	11	2.25−3.5	2.4, 15; 2.7, 7;[e] 2.45, 14[15]
C−H···π acetylene−benzene[i]	2.6	12	2.1−3.4	2.9, 8; 2.7, 16;[e] ΔE = 13[13]
CH_3···π ethane−benzene[j]	3.2	7	2.75−4.3	3.45, 2; 3.2, 7;[e] ΔE = 6[13]

[a] Equilibrium distance, Å. [b] Binding energy, kJ mol^{-1}. [c] Range of the distance coordinate for an energy increase of 5 kJ mol^{-1} from the minimum, Å. [d] Two rings on top of one another; R is the distance between ring planes. [e] MP2−6-31G** with and without counterpoise correction. [f] Two parallel rings, offset by half a ring length; R is the distance between ring planes. [g] O−H or N−H bond pointing to the center of the benzene ring; R is the O···ring or N···ring distance. [h] C−H of one acetylene pointing to the midpoint of the triple bond of the other; R is the H···midpoint distance. [i] C−H bond pointing to the center of the benzene ring; R is the H···center distance. [j] Ethane C−C bond perpendicular to the center of the benzene ring; R is the distance between the plane of methyl hydrogens and that of the ring.

Table 3. Equilibrium Distances, Binding Energies, and Widths of the Potential-Energy Well for Recognition Modes Involving C−H···N or C−H···O Contacts

acceptor[a] C−H donor	R°	ΔE[b]	width[c]	quantum mechanical results for R° and ΔE
water acetylene	2.1	14	1.85−2.75	2.2, 12; 2.2, 17;[d] 2.2, 11[10]
water benzene	2.4	7	2.05−3.6	
water CF_3H	2.35	14	2.0−3.0	2.2, 21, no counterpoise correction[11]
water methane	2.6	3		2.7, 4, no counterpoise correction[11]
acetone acetylene	2.2	11	1.85−3.0	
acetone benzene	2.45	6	2.05−4.0	2.5, 5; 2.4, 9[d]
acetone CF_3H	2.4	12	2.0−3.2	
acetone methane	2.6	2.5		
ammonia acetylene	2.15	20	1.8−2.65	2.3, 15; 2.2, 22;[d] 2.3, 14[10]
ammonia benzene	2.45	9	2.15−4.2	
ammonia CF_3H	2.35	18	2.0−3.2	2.3, 24, no counterpoise correction[11]
ammonia methane	2.6	4		2.7, 3, no counterpoise correction[11]
acetonitrile acetylene	2.25	13	1.9−2.9	
acetonitrile benzene	2.5	6.5	2.15−4.2	2.7, 5; 2.6, 8[d]
acetonitrile CF_3H	2.4	14	2.0−3.1	
acetonitrile methane	2.6	2.5		
benzoquinone C═O···H−C[e]	2.35	15	2.1−3.0	
pyridimide-benzoquinone[f]	2.5	15	2.2−3.0	2.7, 11; 2.5,21;[d] $\Delta E = 19$ at 50% BSSE[19]

[a] A C−H bond points at the O or N atom of the acceptor molecule; R is the O···H or N···H distance, Å. [b] Binding energy, kJ mol^{-1}. [c] Range of the distance coordinate for an energy increase of 5 kJ mol^{-1} from the minimum, Å. [d] MP2−6-31G** with and without counterpoise correction. [e] Parallel coplanar rings, O···H cyclic double contact; R is the O···H distance, Å. [f] Parallel coplanar aromatic rings, C−H···O and C−H···N interaction; R is the O···H and N···H distance, Å.

Table 4. Analysis of Contacts between Nonpolar Atoms or Moieties

dimer	R°[a]	ΔE[b]	quantum mechanical results for R° and ΔE
$H_3C-H\cdots H-CCH$	broad	≈1	2.9, 2[13]
$CH_2\cdots H_2C$ ethylene	broad	≈2	
C−H···H−C coplanar benzene	2.5	3	2.4, 1; 2.4, 3[c]
Ar_2	3.9	1.2	
parallel ethane	2.7	5	3.0, 1.5; 2.8, 3[c]
parallel chain butane	2.5	9	2.7, 3; 2.5, 6[c]
parallel chain hexane	2.6	14	
$CH_3-Cl\cdots Cl-CH_3$	no minimum		no minimum[c]
$CH_3-Cl\cdots CH_3Cl$	3.2[d]	5.5	3.3, 3; 3.3, 4[c]

[a] H···H, Ar···Ar, or Cl···Cl distance, Å. Equilibrium distances are only approximate because the minima are always very broad. [b] Binding energy, kJ mol^{-1}. [c] MP2−6-31G** with and without counterpoirse correction. [d] Distance between Cl and the plane of methyl hydrogens, Å.

widespread conceptions, the hydrogen bond is not always the strongest intermolecular interaction. Nonpolar interactions among benzene rings are at a lower level in the hierarchy (10−15 kJ mol^{-1}), the three modes for pure benzene—on-top parallel, offset parallel, and perpendicular—being almost isoenergetic, a result that does not support simple-minded H(+)···π(−) or quadrupole−quadrupole interaction models. In the same energy range, one finds many interactions among C−H···X moieties (X = O, N, or a π system), but only when the donor hydrogen is highly acidic (acetylene, trifluoromethane). Lateral interactions between aliphatic chains are worth about 1.5 kJ mol^{-1} for each carbon atom, so that a hexane chain has about the same cohesive energy as a benzene ring. At a still lower level (<10 kJ mol^{-1}), one finds interactions between benzene C−H groups and O or N atoms, and the experimental results concur, if the total interaction energy between an oxyrane and a trifluoromethane molecule including three C−H···O or C−H···F short contacts is worth about 6.7 kJ mol^{-1}.[35] Interactions between aliphatic C−H groups and all kinds of polar acceptors involve less than 5 kJ mol^{-1}, whereas Cl···Cl interactions are hardly cohesive at all, as already recognized.[36] The use of these last interactions to categorize intermolecular recognition leads to the oxymoron of a zero-energy bond.

The widths of the corresponding energy wells are even more revealing than the binding energies. These results are quite similar in the Pixel calculations and in all sorts of corresponding quantum chemical calculations. The expected vibrational amplitude for the stronger hydrogen bond in average thermal conditions is 0.3−0.5 Å, increasing to 0.5−0.7 Å for the weaker hydrogen bond or for aromatic ring stacking. These data are still compatible with a definite bonding interaction (the corresponding quantity for an average intramolecular single bond in organic compounds is 0.2−0.3 Å). On the contrary, the vibrational amplitudes of even the strongest C−H···O, C−H···N, or C−H···π "bonds", those formed by acetylenic protons, already exceed 1 Å, and those involving benzene or aliphatic hydrogens put these interactions at the limit of spontaneous dissociation. Unless a careful analysis is conducted to ensure that other contributions are even less relevant or are constant over all possible structural variations, the description and prediction (let alone control) of the crystal structures of complex and multifunctional organic compounds as depending on these weaker interactions is not to be encouraged.

4. Orientation Dependence

The results shown in the previous section concern binding curves for molecular dimers as a function of separation between centers of mass. The Pixel method also gives a clear picture of the angular dependence on the interaction, thanks to its partitioning into recognizable Coulombic-polarization and dispersion terms (Table 5). The benzene stacked dimer

Table 5. Angular Dependence of Some Typical Interactions between Aromatic Ring Systems[a]

	E_{coul}	E_{pol}	E_{disp}	E_{rep}	E_{tot}
			Benzene		
$\theta = 0$	3.7	−1.5	−20.6	8.8	−9.6
$\theta = 15$	3.7	−1.5	−20.6	9.1	−9.4
$\theta = 30$	3.7	−1.5	−20.6	9.1	−9.4
			1,3,5-TFB		
$\theta = 0$	−0.6	−0.7	−26.9	11.5	−16.7
$\theta = 20$	−1.4	−0.7	−26.8	12.0	−16.9
$\theta = 40$	−3.0	−0.7	−26.8	11.9	−18.6
$\theta = 60$	−3.8	−0.7	−26.9	11.9	−19.4
			1,4-Dinitrobenzene		
$\theta = 0$	5.0	−1.5	−37.4	18.9	−15.0
$\theta = 180$	−7.9	−1.7	−33.3	13.5	−29.5

[a] θ is the in plane rotation angle at fixed inter-ring distance equal to the equilibrium value in Table 1. $\theta = 0$ corresponds to exact on-top parallel stacking, with eclipsed substituents. Binding energy = $-E_{tot}$.

is quite insensitive to the reciprocal orientation of the two rings. The 1,3,5-trifluorobenzene stacked dimer is, instead, sensitive to orientation, because the Coulombic term is unfavorable when fluorine atoms are on top of fluorine atoms and becomes favorable when fluorine atoms are on top of hydrogen atoms. The effect is obviously magnified in the case of the more strongly polar nitrobenzene molecule; in the dimer, an antiparallel orientation of the nitro substituents is largely favored on a purely Coulombic basis, as expected.

5. Conclusions

Our results reveal that, even for strongly bound dimers, like stacked polar aromatic rings or hydrogen-bonded species, many different interaction modes ensure very similar binding energies, and identifying the leading cohesion energy contributions may be difficult. When dimerization energies fall below 20 kJ mol^{-1}, a large pool of different interaction modes, many of which are bound to appear in nearly all crystal packings of ordinary organic molecules, have approximately equal energies and are highly "stretchable" at ordinary temperatures. In this light, special caution should be exerted in the structural analysis of organic crystals in terms of the structure-defining power of localized interactions between certain groups. In particular, the role of the so-called "weak hydrogen bonds" has been greatly overemphasized, as already suggested in critical examinations.[37]

A necessary (although maybe not sufficient) condition for consistent predictions is a careful evaluation of small differences in cohesion energies arising from interactions between delocalized electron densities. Also important must be an even more careful consideration of cooperative effects in the propagation of long-range order in determining the amount of stretch from the free dimer minimum energy structure (a planned extension of the present work is the analysis of destabilizing options, or "antirecognition" interaction modes, that packing molecules cannot take). Besides, crystallization out of this near continuum of structural possibilities has a good chance of being dictated by kinetics rather than by thermodynamics.

Absolute values in our tables do change by a few kJ mol^{-1} upon different theoretical treatments, but the agreement between the different methods, some of which are of indisputable maturity, gives confidence in the Pixel results and strongly suggests that relative orders of magnitude must not change significantly, so that the above conclusions may be refined but not substantially changed.

Supporting Information Available: Cambridge Structural Database ref codes, heats of sublimation, and calculated lattice energies of the crystal structures in Figure 1 (Table S1) and atomic coordinates for all the minima in the dimer calculations (Table S2) have been deposited. This material is available free of charge via the Internet at http://pubs.acs.org. For reproducibility, the Pixel-SCDS Fortran code and instruction manual, the Gaussian input files, and the Pixel-SCDS input files are available for distribution upon request from the author at angelo.gavezzotti@unimi.it.

References

(1) Munro, O. Q.; Mariah, L. *Acta Crystallogr., Sect. B* **2004**, *B60*, 598−608. Balamurugan, V.; Hundal, M. S.; Mukherjee, R. *Chem.—Eur. J.* **2004**, *10*, 1683−1690. Caronna, T.; Liantonio, R.; Logothetis, T. A.; Petrongolo, P.; Pilati, T.; Resnati, G. *J. Am. Chem. Soc.* **2004**, *126*, 4500−4501. Nguyen, H. L.; Horton, P. N.; Hursthouse, M. B.; Legon, A. C.; Bruce, D. W. *J. Am. Chem. Soc.* **2004**, *126*, 16−17.

(2) Dunitz, J. D.; Gavezzotti, A. *Angew. Chem., Int. Ed.* **2005**, *44*, 1766−1787.

(3) Del Bene, J. *Chem. Phys. Lett.* **1974**, *24*, 203−207.

(4) Halkier, A.; Klopper, W.; Helgaker, T.; Jorgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157−9167.

(5) Feyereisen, M. W.; Feller, D.; Dixon, D. A. *J. Phys. Chem.* **1996**, *100*, 2993−2997.

(6) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872−9882.

(7) Bako, I.; Radnai, T.; Funel, M. C. B. *J. Chem. Phys.* **2004**, *121*, 12472−12480.

(8) Gu, Y.; Kar, T.; Scheiner, S. *J. Am. Chem. Soc.* **1999**, *121*, 9411−9422.

(9) Gu, Y.; Kar, T.; Scheiner, S. *J. Mol. Struct.* **2000**, *552*, 17−31.

(10) Hartmann, M.; Wetmore, S. D.; Radom, L. *J. Phys. Chem. A* **2001**, *105*, 4470−4479.

(11) Wetmore, S. D.; Schofield, R.; Smith, D. M.; Radom, L. *J. Phys. Chem. A* **2001**, *105*, 8718−8726.

(12) Tarakeshwar, P.; Choi H. S.; Kim, K. S. *J. Am. Chem. Soc.* **2001**, *123*, 3323−3331.

(13) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2000**, *122*, 3746−3753.

(14) Brinkmann, N. R.; Tschumper, G. S.; Yan, G.; Schaefer, H. F., III. *J. Phys. Chem. A* **2003**, *107*, 10208−10216.

(15) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887−10893.

(16) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790−18794.

(17) West, A. P.; Mecozzi, S.; Dougherty, D. A. *J. Phys. Org. Chem.* **1997**, *10*, 347−350

(18) Hernandez-Trujillo, J.; Colmenares, F.; Cuevas, G.; Costas, M. *Chem. Phys. Lett.* **1997**, *265*, 503−507.

(19) McCarthy, W.; Plokhotnichenko, A. M.; Radchenko, E. D.; Smets, J.; Smith, D. M. A.; Stepanian, S. G.; Adamowick, L. *J. Phys. Chem. A* **1997**, *101*, 7208−7216.

(20) Melandri, S.; Sanz, M. E.; Caminati, W.; Favero, P. G.; Kisiel, Z. *J. Am. Chem. Soc.* **1998**, *120*, 11504−11509.

(21) Mitchell, J. B. O.; Price, S. L. *J. Phys. Chem. A* **2000**, *104*, 10958−10971.

(22) Gonzalez, C.; Lim, E. C. *J. Phys. Chem. A* **2003**, *107*, 10105−10110.

(23) Gavezzotti, A. *J. Phys. Chem. B* **2002**, *106*, 4145−4154.

(24) Gavezzotti, A. *J. Phys. Chem. B* **2003**, *107*, 2344−2353.

(25) Gavezzotti, A. *CrystEngComm* **2003**, *5*, 429−438.

(26) 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.; Kiene, 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 A.l; Gaussian, Inc.: Pittsburgh, PA, 2003.

(27) Allen, F. H.; Kennard, O. *Chem. Des. Autom. News* **1993**, *8*, 31−37.

(28) Afeefy, H. Y.; Liebman, J. F.; Stein, S. E. Neutral Thermochemical Data.; Chickos, J. S. Heat of Sublimation Data. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Ed.; National Institute of Standards and Technology: Gaithersburg, MD, 2003 (http://webbook.nist.gov).

(29) Filippini, G.; Gavezzotti, A. *Acta Crystallogr., Sect. B* **1993**, *B49*, 868−880.

(30) Gavezzotti, A.; Filippini, G. *J. Phys. Chem.* **1994**, *98*, 4831−4837.

(31) Gavezzotti, A.; Filippini, G. Energetic aspects of crystal packing: experiment and computer simulations. In *Theoretical aspects and computer modeling of the molecular solid state*; Gavezzotti, A., Ed.; Wiley and Sons: Chichester, U. K., 1997.

(32) Callomon, J. H.; Hirota, E.; Iijima, T.; Kuchitsu, K.; Lafferty, W. J. In *Landolt-Boernstein, Numerical Data and Functional Relationships in Science and Technology*, New Series; Springer-Verlag: Berlin, 1987; Vol. 15, Supplement to Vol. 2/7.

(33) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. *Chem. Rev.* **2000**, *100*, 4145−4186.

(34) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104−6115.

(35) Alonso, J. L.; Antolinez, S.; Blanco, S.; Lesarri, A.; Lopez, J. C.; Caminati, W. *J. Am. Chem. Soc.* **2004**, *126*, 3244−3249.

(36) Price, S. L.; Stone, A. J.; Lucas, J.; Rowland, R. S.; Thornley, A. E. *J. Am. Chem. Soc.* **1994**, *116*, 4910−4918.

(37) Cotton, F. A.; Daniels, L. M.; Jordan, G. T., IV; Murillo, C. A. *Chem. Commun.* **1997**, 1673−1674.

CT050091W

JCTC Journal of Chemical Theory and Computation

An Efficient *a Posteriori* Treatment for Dispersion Interaction in Density-Functional-Based Tight Binding

Lyuben Zhechkov,[†] Thomas Heine,*[,†] Serguei Patchkovskii,[‡] Gotthard Seifert,[†] and Helio A. Duarte[§]

Institut für Physikalische Chemie und Elektrochemie, TU Dresden, D-01062 Dresden, Germany, Steacie Institute for Molecular Sciences, NRC, 100 Sussex Dr., Ottawa, Ontario K1A 0R6, Canada, and Departamento de Química-ICEx, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, Brazil

Received March 14, 2005

Abstract: The performance of density functional theory (DFT) (VWN-LDA, PBE-GGA, and B3LYP hybrid functionals), density-functional-based tight binding (DFTB), and ab initio methods [HF, MP2, CCSD, and CCSD(T)] for the treatment of London dispersion is investigated. Although highly correlated ab initio methods are capable of describing this phenomenon, if they are used with rather large basis sets, DFT methods are found to be inadequate for the description of H_2/PAH (polycyclic aromatic hydrocarbon) interactions. As an alternative approach, an a posteriori addition of a van der Waals term to DFTB is proposed. This method provides results for H_2/PAH interactions in close agreement with MP2 and higher-level ab initio methods. Bulk properties of graphite also compare well with the experimental data.

I. Introduction

One of the weakest, qualitatively important interactions between molecules, fragments of molecules, solutes, and solids is the dispersion interaction, named after F. London, who gave the first physical description of this phenomenon. In his famous formula,[1] dispersion is described as the interaction of two neutral, separated particles with a nonoverlapping charge density and without a permanent dipole moment. The dispersion energy depends on the distance between the particles r, the polarizabilities α_1' and α_2', and the ionization energies of the interacting particles I_1 and I_2: Hence, dispersion between nonpolar molecules is always

$$U_{\text{London}} = -\frac{2}{3}\frac{I_1 I_2}{I_1 + I_2}\alpha_1'\alpha_2'\frac{1}{r^6} \qquad (1)$$

attractive and slowly falls off as r^{-6}. It is several orders of magnitude weaker than typical covalent or ionic interactions and still a factor of 10 smaller than hydrogen bridge bonds. Despite this simple physical picture, a straightforward *first principle* quantum-mechanical description of the interaction is not trivial.

Density functional theory (DFT) within the Kohn−Sham formulation and with presently available exchange-correlation functionals does not describe the dispersion interaction correctly.[2] Within the local density approximation (LDA), the inherent overbinding of the method leads to a strong overestimation of the interaction around the van der Waals minimum and does not provide the correct long-range behavior.[2−4] On the other hand, GGA DFT is able to reproduce the order of magnitude of the dispersion interaction close to the van der Waals minimum.[5] At the same time, it also fails to handle the long-range part of the potential and is, hence, not appropriate, for instance, in the simulation of physisorption processes.[4] The true correlation energy functional must include the van der Waals interaction,[6] and future generations of optimized effective potentials[2,7−10] may correct this deficiency of DFT. Indeed, considerable progress has been achieved toward a better description of weak interactions within DFT by the TPSS family of meta-GGAs, which

* Corresponding author e-mail: thomas.heine@chemie.tu-dresden.de.

[†] Institut für Physikalische Chemie und Elektrochemie, TU Dresden.

[‡] Steacie Institute for Molecular Sciences, NRC.

[§] Universidade Federal de Minas Gerais.

10.1021/ct050065y CCC: $30.25
Published on Web 06/30/2005

improve results, in particular, for the short range.[11,12] Direct inclusion of London forces into the nonlocal part of the correlation potential[13] and into pseudopotentials[14] has been suggested recently. An a posteriori correction to the BLYP functional has been proposed by Grimme.[7] His method focuses on the correction of the long-range part of the interaction by employing a C_6 term, while the interaction is switched off by an expression that has been tuned for the PBE and the BLYP functionals. However, at present, only post-Hartree−Fock (HF) ab initio computations with a reasonable description of correlation {for example, second-order Møller−Plesset perturbation theory (MP2), or coupled cluster theory including single, double (CCSD), and (partially) triple excitations [CCSD(T)]} are capable of treating this interaction properly for the short and, in particular, the long distance. Unfortunately, such calculations require nearly converged basis sets and, therefore, a high computational effort.[15,16] On the other hand, the density-functional-based tight-binding (DFTB) methods with and without self-consistent charge (SCC) correction[17−19] are known to completely exclude van der Waals interactions, and a correction by the a posteriori addition of a C_6 term has been proposed for DFTB.[20] For an overview of the DFTB method and its applications, see ref 21.

A working methodology that includes the dispersion interaction is a necessity, as many fundamental processes in chemistry, physics, and biology are strongly affected by this weak long-range interaction, for example, the geometry of molecular crystals, the structure of biological molecules, surface−adsorbate interactions, $\pi-\pi$ stacking interactions, and so forth. In contrast to other long-range interactions, such as the Coulomb interaction of charged particles, the van der Waals interaction between two nonpolar particles is always attractive. Although the van der Waals term falls off as r^{-6} with the distance r, in the region of up to ∼6 Å, its contributions to the interparticle interactions is still not negligible. Although the dispersion interaction influences the long-range interaction energy, its short-range effect on covalent bond parameters and on certain electronic properties can be neglected—van der Waals type interactions are orders of magnitude smaller than covalent bonds.[15,16,22,23] Several methods take advantage of this fact and treat long-range interaction by adding a van der Waals term to energies and gradients.[20,24] This strategy is easy to follow for the attractive, long-range part of the potential energy surface (PES). Indeed, in this part of the PES, the dispersion or the general van der Waals interaction is clearly dominating part of the correlation energy and may be described independently from all other contributions.

At shorter distances, the interaction energy is dominated by electronic overlap effects. The a posteriori term should have no influence on those interatomic distances that are characteristic for covalent bonds. Several techniques have been introduced to avoid problems with the short-range part of an a posteriori-added van der Waals potential: In Rappé's universal force field (UFF),[24] the van der Waals contribution is only considered for interatomic interactions where the two atoms are not adjacent to each other or to a common neighbor. Other implementations use a switching function to turn off the van der Waals potential at short distances or employ lists of "van der Waals-active" interatomic contacts.[20]

The aim of this work is to provide a quantum-chemical method that is extended with a parameterized van der Waals potential in such a way that no contact lists need to be employed and that the van der Waals parameters are available throughout the periodic table. As our main target application is to describe carbon nanostructures, we have chosen the DFTB method as a quantum-mechanical basis, as it clearly excludes van der Waals interaction and is capable of simulating large structures. This aim is achieved by switching between two different analytical descriptions for the van der Waals term for the short-and long-range parts. Our implementation takes advantage of available Lennard−Jones parameters of the UFF force field,[24] which are optimized to reproduce van der Waals distance, well depth, and local curvature and which are available from H to Lw.[24]

In the first part of this article, we compare the performance of DFTB, ab initio, and DFT methods for the dispersion interaction of H_2 with PAHs. Then, we describe augmentation of the DFTB method with a dispersion term and its implementation in the deMon code.[25]

Finally, we evaluate the performance of the new technique for the interaction of H_2 with PAHs and for the structural and elastic properties of graphite.

II. Benchmark Calculations

A. Geometries and Computational Methodology. The interaction of H_2 with some PAHs has been studied previously, and we take all PAH geometries from previous work.[15] The PAH structures have been optimized at the B3LYP/6-31G* level, previously shown to give geometries very close to those of MP2 computations.[15] The basis set superposition errors (BSSE) have been corrected by applying the counterpoise methodology.[26] All DFT- and DFTB-based computations except those with hybrid functionals have been performed using deMon.[25] MP2 calculations and DFT with hybrid functionals have been carried out using Gaussian 03 software,[27] whereas CCSD and CCSD(T) energies were obtained using NWChem.[28]

B. Ab Initio Methods. In Figure 1 and Table 1, a series of ab initio computations on H_2 motion along the normal to benzene plane are compared, including MP2 with several basis sets with and without correcting for BSSE and two coupled cluster approaches (with and without the perturbative triples correction). Strong dependence of the MP2 results with respect to the orbital basis shows the importance of the basis set convergence. Thus, only with large augmented basis sets does MP2 approach a converged result. For the largest basis set studied, aug-cc-pVQZ,[29] the BSSE is smaller than 1 kJ mol^{-1}.[15] On the other hand, MP2 calculations of H_2/benzene[15] and N_2/benzene[22] for various distances and orientations of the physisorbed molecule with respect to benzene showed that the cc-pVTZ basis, without correcting for BSSE, gives very similar results to those of BSSE-corrected MP2/aug-cc-pVQZ interaction energies for a large range of interatomic distances (see Figure 1). Employing the cc-pVTZ basis, the level of correlation has been increased in coupled cluster calculations with single and double

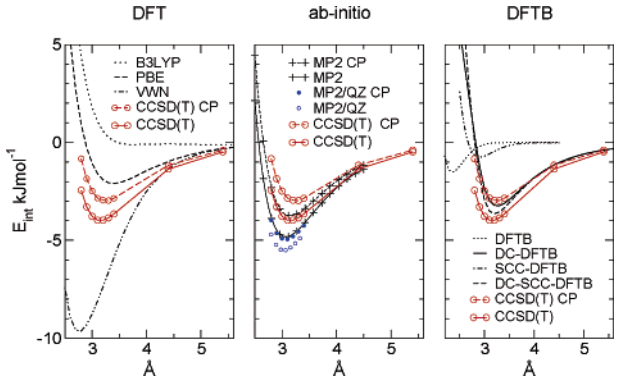


Figure 1. Potential energy surface for H_2 motion along the C_6 axis normal to benzene plane at various levels of ab initio and DF theory, with and without counterpoise (CP) corrections for the BSSE, and dispersion-corrected (DC) and uncorrected DFTB and SCC DFTB. All ab initio and DFT calculations employed the cc-pVTZ basis set, except for QZ, which denotes the aug-cc-pVQZ basis set. Results for other basis sets are given in Table 1.

Table 1. Interaction Energies of H_2 with Benzene[a]

method	R_{int}	E_{int}
VWN/DZVP	2.8(2.8)	−8.77(−8.22)
VWN/TZVP	2.8(2.8)	−9.34(−8.69)
VWN/cc-pVTZ	2.8(2.8)	−9.62(−9.28)
PBE/DZVP	3.4(3.4)	−2.08(−1.62)
PBE/TZVP	3.4(3.5)	−2.12(−1.70)
PBE/cc-pVTZ	3.4(3.4)	−2.10(−1.90)
B3LYP/DZVP	3.7(3.7)	−0.04(+0.25)
B3LYP/TZVP	3.7(3.7)	−0.17(+0.10)
B3LYP/cc-pVTZ	3.7(3.7)	−0.11(+0.04)
HF/cc-pVTZ	4.0(4.0)	−0.55(−0.48)
MP2/6-31G*	3.7(3.4)	−2.56(−0.96)
MP2/6-311G**	3.2(3.4)	−4.16(−2.07)
MP2/pVTZ	3.1(3.4)	−3.78(−2.16)
MP2/cc-pVTZ	3.1(3.2)	−4.82(−3.74)
MP2/aug-cc-pVTZ	3.0(3.1)	−6.64(−4.72)
MP2/aug-cc-pVQZ	3.0(3.1)	−5.51(−4.94)
CCSD/cc-pVTZ	3.2(3.3)	−3.33(−2.53)
CCSD(T)/cc-pVTZ	3.2(3.3)	−3.97(−2.97)
DFTB	2.4	−1.51
DC DFTB	3.3	−3.25
SCC DFTB	2.9	−0.75
DC SCC DFTB	3.2	−3.64

[a] Interaction energies are given in kJ mol^{-1} and equilibrium distances in Å. Numbers in parentheses include counterpoise corrections for BSSE.

excitations (CCSD) and perturbation of triple excitations [CCSD(T)]. For the final estimation of the H_2/benzene interaction, we used the best level of correlation treatment of our computations [CCSD(T)], employing a cc-pVTZ basis, because, for the H_2/benzene system, interaction energies calculated with this basis nearly coincide with the extrapolation of the basis set limit. We found the interaction energy slightly reduced (by ∼0.8 kJ mol^{-1} with respect to MP2/cc-pVTZ), and the overall interaction in Figure 1 shows the same trend as that for MP2. With these results, we estimate the H_2/benzene interaction energy to be 4.2−4.7 kJ mol^{-1} and the intermolecular equilibrium distance to be 3.1−3.3 Å.

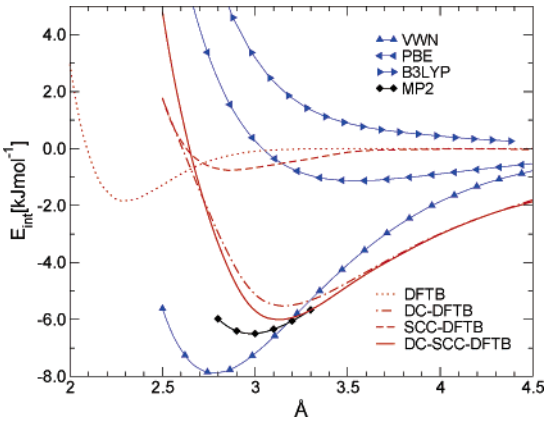


Figure 2. Potential energy surface of H_2 motion normal to the center of coronene. Details and conventions are the same as those in Figure 1.

C. Density Functional Theory. Even though the unknown exact correlation energy functional must include dispersion,[6] all present popular functionals fail to describe this type of interaction for some parts of the potential energy surface.[2] Figures 1 and 2 illustrate the performance of three representative DFT methods with a double-ζ (DZVP[30]) and two different triple-ζ basis sets including polarization functions (TZVP[30] and cc-pVTZ[31]). The three types of functionals are the LDA proposed by Vosko, Wilk, and Nusair (VWN);[32] a recent highly respected GGA (PBE);[33] and the B3LYP hybrid functional.[34,35] In all three functionals, the performance is independent of the employed basis set, and they have only small BSSEs, as can be seen in Table 1. For the H_2/benzene complex (Figure 1), LDA overestimates the attraction by a factor of more than 2 and finds the equilibrium distance at 2.8 Å—approximately 0.3 Å shorter than more elaborate theories. On the other hand, PBE slightly underestimates the interaction and finds the van der Waals minimum at distance 3.4 Å, too long by about 0.3 Å, so this method performs rather well for H_2/benzene. The admixture of the exact HF exchange in the B3LYP functional is clearly not beneficial for treatment of dispersion interactions. When considering coronene, a larger PAH, the failure of popular DFT methods to describe the dispersion interaction is even more evident (see Figure 2). All three DFT variants examined here, including the PBE functional, not only fail to give a reasonable description of the long-range potential but even find the interaction of H_2 with coronene less attractive than with benzene—opposite of the ab initio results and physical expectations based on eq 1. In summary, modern GGAs can give equilibrium distances of van der Waals complexes to good accuracy, a conclusion that is supported by our PBE calculations and by more extensive studies employing the somewhat similar PW91 functional.[5,15] The three functionals compared in this study are unsuitable for the study of large van der Waals complexes, as they show unphysical trends for adsorption energy when comparing PAHs of different size and incorrect long-range behavior.

D. DFTB and its van der Waals Extension. DFTB can be considered as an approximation to DFT.[17−19] It treats short-range atomic potentials and neglects three-center terms in the Hamiltonian. Hamilton matrix elements fall off quickly

and become negligible for interatomic distances typical for the region of the van der Waals minimum. Hence, dispersion is not included in the DFTB electronic Hamiltonian. In both available parameterizations (non-SCC[18] and SCC[19]), DFTB shows only a very small attractive potential at distances shorter than the van der Waals minimum (see Figures 1 and 2). Therefore, the addition of an empirical van der Waals correction does not introduce artificial "double counting" of dispersion in the vicinity of the van der Waals minimum, as it could for DFT computations that already include some dispersion. For practical reasons, we decided to use the parameterization of UFF for the dispersion correction.[24] The dispersion interaction U_{ij} between atoms i and j at distance r is given in Lennard−Jones-type form, which includes two parameters: van der Waals distance (r_{ij}) and well depth (d_{ij}). These UFF parameters are reported in the original paper[24]

$$U_{ij}(r) = d_{ij}\left[-2\left(\frac{r_{ij}}{r}\right)^6 + \left(\frac{r_{ij}}{r}\right)^{12}\right] \tag{2}$$

and are available throughout the periodic table. In UFF, the van der Waals term is set to zero for those interatomic interactions where the atoms are adjacent either to each other or to a common neighbor atom. Interaction cutoffs based on adjacency information postulate an inflexible topology of the system during the whole simulation and would impose unnecessary restrictions to a quantum-mechanical method. Therefore, a different technique to change the asymptotic behavior for short distances in the analytic form of eq 2 is necessary. As can be inferred from Figures 1 and 2, the additional van der Waals term must include the long-range part and the area around the van der Waals distance to provide a reasonable description of the dispersion interaction. However, the empirical r^{-12} repulsive term conflicts with the quantum-mechanical description for particles with overlapping charge densities, and application of eq 2 for short interatomic distances remains useless. Instead, the van der Waals correction at small interatomic distances should converge to a finite limit. We have chosen to use the Lennard−Jones representation of eq 2 for the whole range of r where U_{ij} is attractive. For a shorter range of r, U_{ij} becomes repulsive. This appears, at first glance, to be surprising, but the necessity to have a repulsive, short-range term is obvious when looking at Figures 1 and 2; the Pauli repulsion between two molecules does not yet appear in DFTB and SCC DFTB for distances at the van der Waals minimum, or even at $r_0 = 2^{-1/6}r_{ij}$, the value at which the Lennard−Jones energy is zero ($U_{ij} = 0$). To maintain numerical stability, we continue the description of the additional potential for distances smaller than r_0 by a polynomial.

$$U_{ij}^{(\text{short}-\text{range})}(r) = U_0 - U_1 r^n - U_2 r^{2n} \tag{3}$$

The parameters U_0, U_1, and U_2 are determined such that the interaction energy and its first and second derivatives match those of eq 2 at r_0. In this way, we maintain continuity for energies, forces, and Hessians. Tests for various systems, including the test cases, a series of DNA−carbon-nanotube interactions, water clusters, and cyclodextrine in water—which will be published separately—suggested $n = 5$ as an appropriate choice for the short-range parametrization. For $n = 5$, the parameters U_0, U_1, and U_2 of eq 2 are given by Note that d_{ij} is positive in eqs 4−6. The repulsion of our

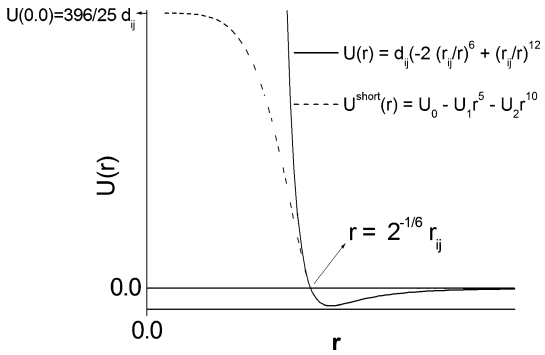


Figure 3. Potential for the dispersion correction to DFTB: The long-range contribution (solid line) is identical for the van der Waals parametrization of UFF (eq 2) up to the point where $U(r) = 0$. The short-range interaction (dashed line) is given by eq 3.

$$U_0 = \frac{396}{25} d_{ij} \tag{4}$$

$$U_1 = 2^{5/6}\frac{672}{25}\frac{d_{ij}}{r_{ij}^5} \tag{5}$$

$$U_2 = -2^{2/3}\frac{552}{25}\frac{d_{ij}}{r_{ij}^{10}} \tag{6}$$

van der Waals correction is limited by the value of $U_0 \approx 16 d_{ij}$ (see Figure 3), a value that is small compared to typical covalent bond energies. Figures 1 and 2 show the performance of dispersion-corrected (DC) and uncorrected DFTB with and without SCC correction compared to our best ab initio reference computations. The agreement of the parameterization and ab initio computations is remarkably good, so that reparameterization of the UFF van der Waals parameters is not necessary, at least for the presently studied systems. The short-range interaction is slightly too repulsive but is dominated in the covalent area by the quantum part of the potential and, hence, is unproblematic. The interaction energy of 3−3.5 kJ mol^{-1} for H_2/benzene is close to the CCSD(T) value, and the equilibrium distance is found to be slightly overestimated by 0.1 Å. More importantly, the long-range part of the potential is described correctly.

The DFTB method is known as a fast quantum-chemical computational tool. It produces good results for various kinds of carbon-based systems.[18,36,37] For such systems, we now propose to use the dispersion-corrected variant, allowing extension to systems with non-negligible dispersion interactions. We have implemented this as an extension in the experimental version of the deMon computer code.[25]

III. Test Applications

A. H_2/PAH and H_2/Graphene Interactions. The DC methodology is applied to a series of H_2/PAH complexes. The MP2 physisorption energies for benzene, naphthalene, anthracene, coronene, and triphenylene have been reported

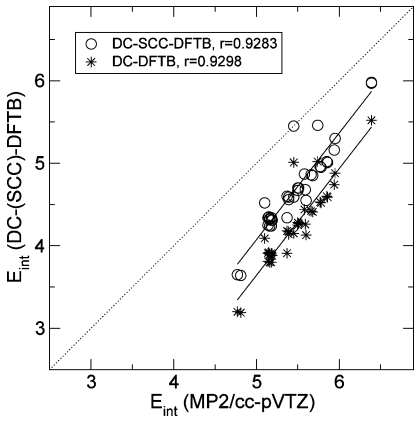


Figure 4. Correlation between MP2 and DC DFTB. SCC DFTB results are given as circles, and DFTB results are given as asterisks. All values are in kJ mol^{-1}. *r* denotes the correlation coefficient.

in an earlier paper[15] and are used to benchmark our new method. One of the key observations of our earlier work was that the interaction energies of such complexes can easily be obtained with single-point calculations for a system where H_2 is coordinated to the ring center, perpendicular to the molecular plane, with an intermolecular distance of ~3.2 Å between the centers of H_2 and the ring. The good performance of the empirical correction is evident when it is compared with the ab initio results in a scatter plot, as shown in Figure 4.

A real advantage of the DC DFTB approach is the possibility to optimize the geometry of such complexes at a low computational cost. Full geometry optimization of the whole set of H_2/PAH complexes shows that all structures correspond to local minima. In general, the DC SCC DFTB bond lengths are found to be slightly longer compared with MP2 computations (≈0.02 Å for C−C bonds, ≈0.05 Å for H−H bonds, and ≈0.1 Å for intermolecular distances). Nevertheless, in the fully optimized structures, the H_2 molecules are slightly tilted, pointing from the ring centers away from the molecule, if they are not physisorbed at the central ring of the PAH. Such trends have been previously reported for H_2/C_6H_5X (X = −F, −OH, −CN, −NH_2) systems using coupled cluster theory.[16] For the fully optimized structures, the DC DFTB schemes give interaction energies slightly lower than those of MP2/cc-pVTZ, which is, at least for the studied H_2/benzene complex, an excellent result, as it matches our highest-level computation at the CCSD(T) level within 0.5 kJ mol^{-1} (see Figure 1).

Since the dispersion interaction of H_2 with PAHs increases with the PAH size, the upper limit for this interaction is expected for the infinite graphene sheet.[15] To test the present methodology, the interaction energy per molecule between a single graphene layer and H_2 molecules, covering it in such way that at least one empty hexagon lies between two H_2 adsorption sites, has been calculated. The calculations have been carried out using periodic boundary conditions applied for super cells containing 196 carbon atoms. The calculated dispersion energy per H_2 is found to be 6.75 kJ mol^{-1} for DFTB and 7.03 kJ mol^{-1} for SCC DFTB. This result is in excellent agreement with our previous estimate of 7.6 kJ mol^{-1}, based on MP2/cc-pVTZ calculations.[15]

B. Bulk Properties of Graphite. The van der Waals interaction plays a crucial role for the bulk properties of graphite. If the graphite interlayer interaction is described properly, the method may be recommended for applications in the popular area of nanostructures, allowing the study of multiwalled nanotubes, nano-onions, nanotube junctions, peapods, DNA stacking, and so forth. To derive the equilibrium lattice parameters of a single graphene layer, we computed the cohesive energy (see Figure 5) applying periodic boundary conditions on a hexagonal graphene sheet containing 140 atoms. Large super cells are required to justify the Γ-point approximation. We found the value of 2.465 Å for the lattice parameter *a*, which is in excellent agreement with the X-ray experiment, which yields a value of 2.462 Å,[38] and corresponds to a C−C distance of 1.421 Å. As a result of the repulsive character of our correction for short-range diatomic interactions, the cohesive energy of −9.01 eV per atom is slightly lower than the reported −9.24 eV per atom for the same method without van der Waals correction.[18] The equilibrium interlayer distance has been calculated in a super cell containing 560 atoms, that is, with 140 atoms per layer and four layers in an ABABAB arrangement, also known as Bernal graphite. The dependence of the interaction energy on the lattice vector *c*, and hence on the interlayer distance, is given in Figure 5. The equilibrium lattice vector *c* is found to be 6.76 Å, which corresponds to the interlayer distance of 3.38 Å. This value is slightly bigger than the experimental value of 3.35 Å (with lattice parameter 6.707 Å).[38]

Unfortunately, no meaningful comparison with theoretical first-principles simulations is possible at present. Reasonable results of DFT−LDA computations, as performed by Telling and Heggie,[39] probably benefit from error compensation: the interaction around the van der Waals minimum is subject to a too-strong overbinding, whereas the long-range attractive interaction is not present. However, two different experimental approaches have been used to determine the interlayer cohesive energy of graphite. Already in 1956, Girifalco and Lad reported a value of 41 meV per atom,[40] obtained by a heat of wetting experiment. In 1998, using nanotube collapse experiments, Benedict et al. determined a similar value of 35 ± 10 meV/atom.[41] The DC DFTB method gives a value of 38.5 meV atom^{-1} for the cohesive energy, which is in excellent agreement with these experimental values.

We determined the lowest-energy stacking geometry by the dislocation glide of the graphene layers, one to each other, in the basal plane along two high-symmetry directions along the lattice vector *a* (denoted as X) and along the vector perpendicular to *a*, parallel to the layers (Y). We find two stacking-fault energies in X and one in Y with displacements that are the same as those reported in a DFT−LDA study[39] and that are in qualitative agreement with experiments.[42,43] The relative stacking-fault energies in the basal plane can be found in Figure 6. The highest barrier is found at 6.7 meV/Å^2 in the X direction and at 2.8 meV/Å^2 in the Y direction. These values are ~10% smaller than those obtained

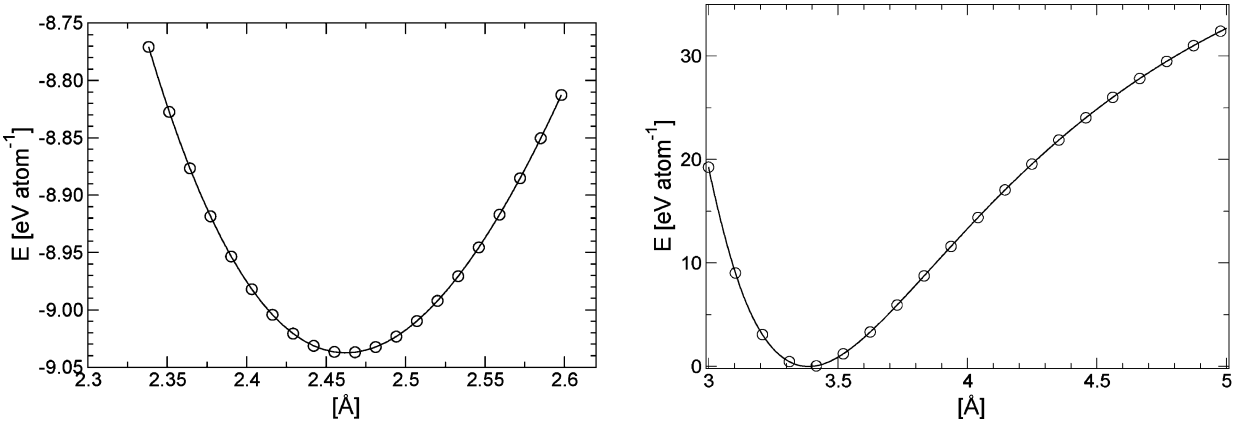


Figure 5. Dependence of the cohesion energy of graphite on the *a* (left) and *c*/2 (right) lattice parameters of the hexagonal unit cell. Equilibrium lattice parameters are $a = 2.465$ Å and $c/2 = 3.38$ Å ($c/2$ coincides with the interlayer distance).

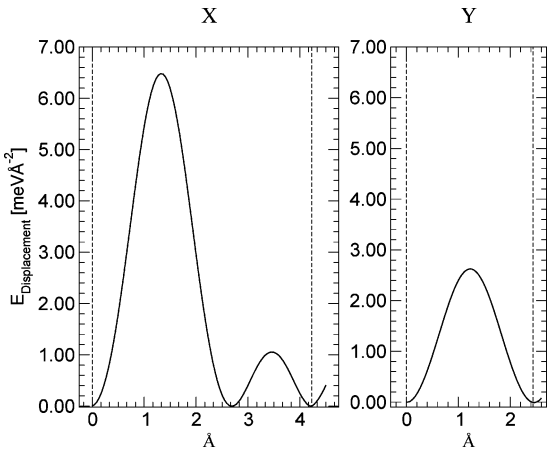


Figure 6. Intrinsic stacking-fault energy (in meV Å^{-2}) for relative layer displacement. Axis direction X denotes movement along the *a* lattice vector, and Y is the movement perpendicular to X parallel to the graphene planes. Unshifted hexagonal graphite is at 0.0 Å. The dashed lines on both sides of the graph indicate the periodicity for X and Y displacements.

by DFT−LDA calculations,[39] which probably reflects the LDA overbinding.

As the final test, we have calculated the isotropic bulk modulus for graphite. This quantity is defined as a second derivative of the energy with respect to the volume. It provides a very sensitive measure of the potential energy curvature close to the van der Waals minimum and, thus, is a good criterion for evaluating the accuracy of DFTB with dispersion correction applied to carbon nanostructures. Our calculated isotropic bulk modulus has been obtained from a volume change at 0 K with a constant c/a ratio (c and a are the lengths of the lattice vectors) and is found to be 330 GPa. This value almost matches the experimental value of 318 ± 11 GPa.[44,45]

IV. Conclusions

The performance of ab initio, DFT, and DFTB methods for treating physisorption of H_2 on PAHs and on graphene was studied. Present LDA, GGA, and hybrid functionals are not adequate for the description of the long-range part of the dispersion interaction, whereas newer GGAs can describe short-range dispersion correctly. LDA overestimates the short-range interaction; PBE-GGA is getting interaction energy and distances qualitatively correct, and B3LYP is even finding the interaction to be slightly repulsive. Only ab initio computations with a reasonable treatment of correlation and large basis sets provide a correct description of this interaction on the basis of a first-principle theory. However, if an a posteriori dispersion potential is added to the DFTB method, physisorption can be described adequately, whereas short-range interactions can still be treated quantum mechanically. Hence, the modified DFTB method allows simulations of processes where the topology of the atoms changes, including the influence of the dispersion interaction. This is particularly important for simulations of nanostructures, especially those including aromatic carbon systems as in fullerenes, nanotubes, and nano-onions.

Acknowledgment. We thank the Deutsche Forschungsgemeinschaft (DFG) and the joint PROBRAL action of DAAD (Germany)−CAPES (Brazil) for financial support. We thank Prof. Dr. Helio F. dos Santos and Nitesh Ranjan for helpful discussions and sharing their unpublished benchmark results. We also thank Dr. Annick Goursot and Dr. Christof Köhler for stimulating discussions.

References

(1) London, F. *Z. Phys.* **1930**, *63*, 245.

(2) Lein, M.; Dobson, J. F.; Gross, E. K. U. *J. Comput. Chem.* **1999**, *20*, 12−22.

(3) Valdes, H.; Sordo, J. A. *J. Comput. Chem.* **2002**, *23*, 444−55.

(4) Jensen, F. *Introduction to Computational Chemistry*; Odense University: Odense, Denmark, 1999.

(5) Tran, F.; Weber, J.; Wesolowski, T. A.; Cheikh, F.; Ellinger, Y.; Pauzat, F. *J. Phys. Chem. B* **2002**, *106*, 8689−8696.

(6) Dobson, J. F.; Das, M. P. *Topics in Condensed Matter Physics*; Nova: New York, 1994.

(7) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

(8) Chakarova, S. D.; Schröder, E. *J. Chem. Phys.* **2005**, *122*, 054102.

(9) Cybulski, S. M.; Seversen, C. E. *J. Chem. Phys.* **2005**, *122*, 014117.

(10) Xu, X.; Goddard, W. A. *J. Chem. Phys.* **2004**, *121*, 4068−4082.

(11) Tao, J.; Perdew, P.; Staroverov, S.; Scuseria, G. *Phys. Rev. Lett.* **2003**, *91*, 416401.

(12) Tao, J.; Perdew, P. *J. Chem. Phys.* **2005**, *122*, 114102.

(13) Dion, M.; Rydberg, H.; Schröder, E.; Landgreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(14) von Lilienfeld, O. A.; Tavernelli, I.; Röthlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.

(15) Heine, T.; Zhechkov, L.; Seifert, G. *Phys. Chem. Chem. Phys.* **2004**, *6*, 980−984.

(16) Hübner, O.; Gröss, A.; Fichtner, M.; Klopper, W. *J. Phys. Chem. A* **2004**, *108*, 3019−3023.

(17) Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185−192.

(18) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1995**, *51*, 12947−12957.

(19) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7260−7268.

(20) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149−5155.

(21) Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Köhler, C.; Amkreutz, M.; Sternberg, M.; Hajnal, Z.; Di Carlo, A.; Suhai, S. *J. Phys. C: Solid State Phys.* **2002**, *14*, 3015−3047.

(22) Zhechkov, L.; Heine, T.; Seifert, G. *Int. J. Quantum Chem.* **2005**, in press.

(23) Schönborn, F.; Schmitt, H.; Zimmermann, H.; Haeberlen, U.; Corminboeuf, C.; Grossmann, G.; Heine, T. *J. Magn. Reson.* **2005**, in press.

(24) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(25) Köster, A. M.; Flores, R.; Geudtner, G.; Goursot, A.; Heine, T.; Patchkovskii, S.; Reveles, J. U.; Vela, A.; Salahub, D. R. *deMon*; NRC: Ottawa, Canada, 2003.

(26) Boys, F. S.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T., Jr.; 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 B.05; Gaussian, Inc.: Pittsburgh, PA, 2003.

(28) Straatsma, T. P.; Aprà, E.; Windus, T. L.; Bylaska, E. J.; de Jong, W.; Hirata, S.; Valiev, M.; Hackler, M. T.; Pollack, L.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, A. A.; Brown, E.; Cisneros, G.; Fann, G. I.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. *NWChem, A Computational Chemistry Package for Parallel Computers*, version 4.6; Pacific Northwest National Laboratory: Richland, WA, 2004.

(29) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6769−6806.

(30) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560−571.

(31) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(32) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.

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

(34) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(35) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(36) Albertazzi, E.; Domene, C.; Fowler, P. W.; Heine, T.; Seifert, G.; Van Alsenoy, C.; Zerbetto, F. *Phys. Chem. Chem. Phys.* **1999**, *1*, 2913−2918.

(37) Porezag, D.; Jungnickel, G.; Frauenheim, T.; Seifert, G.; Ayuela, A.; Pederson, M. R. *Appl. Phys. A* **1997**, *64*, 321−326.

(38) Zhao, Y. X.; Spain, I. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1989**, *40*, 993997.

(39) Telling, R. B.; Heggie, M. I. *Philos. Mag. Lett.* **2003**, *83*, 411−421.

(40) Girifalco, L. A.; Lad, R. A. *J. Chem. Phys.* **1956**, *25*, 693−697.

(41) Benedict, L. X.; Chopra, N. G.; Cohen, M. L.; Zettl, A.; Louie, S. G.; Crespi, V. H. *Chem. Phys. Lett.* **1998**, *286*, 490−496.

(42) Snyder, S. R.; Gerberich, W. W.; White, H. S. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 10823−10831.

(43) Ouseph, P. J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *53*, 9610−9613.

(44) Gauster, W. B.; Fritz, I. J. *J. Appl. Phys.* **1974**, *45*, 3309−3314.

(45) Jansen, H. J. F.; Freeman, A. J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1987**, *35*, 8207−8214.

CT050065Y

JCTC Journal of Chemical Theory and Computation

Evaluation of Functionals O3LYP, KMLYP, and MPW1K in Comparison to B3LYP for Selected Transition-Metal Compounds

Thomas Strassner* and Maria A. Taige

Institute of Organic Chemistry, Technical University Dresden, Mommsenstrasse 13, D-01062 Dresden, Germany

Received December 21, 2004

Abstract: We have investigated the performance of the hybrid density functionals O3LYP, KMLYP, and MPW1K in comparison to B3LYP for predicting geometrical parameters of selected transition-metal compounds. We used different combinations of ECPs and basis sets (Stuttgart/Dresden 1997, LANL2DZ, 6-31G(d), 6-311++G(d,p)) to analyze the effect on selected transition-metal compounds. B3LYP is superior to the other three density functionals for the split valence double-ζ basis set 6-31G(d), contrary to what was reported for organic reactions. But in combination with the split valence triple-ζ basis set 6-311++G(d,p) O3LYP was clearly superior to all the other functionals. LANL2DZ performed worse in every case.

Introduction

During the past decades, density functional theory (DFT)[1,2] has become the most widely used method in computational quantum chemistry. It has been proven in many cases that it is the most suitable method for calculating transition-metal compounds.[3−8] An exchange-correlation matrix replaces the exchange term in standard Hartree−Fock (HF)[9] theory, and the progress in density functional theory is based on the development of better approximations toward the "real" exchange-correlation functional.

The exchange-correlation density functional $f^{XC}(\rho)$ is written as the sum of individual exchange $f^{X}(\rho)$ and correlation $f^{C}(\rho)$ components in nearly all functionals used in the last years. Thereby the approximations are made separately for each of the contributions.

One of the most common and widely used density functionals is the hybrid functional B3LYP,[10] which is a combination of Becke's 1988 nonlocal exchange functional B88,[11] $\rho^{4/3}$ Dirac slater exchange term,[12] about 20% of Hartree−Fock exchange functional,[10] Vosko, Wilk, and Nusair's local correlation functional (VWN),[13] and Lee, Yang, and Parr's nonlocal correlation functional (LYP).[14] Though B3LYP usually gives reliable results, there have been cases where the results from B3LYP density functional calculations showed some discrepancies to experimentally measured values. Therefore the search for improved density functionals is ongoing. Lately Handy and Cohen have developed the new exchange functional OPTX.[15] By replacing B88 through OPTX the density functional O3LYP is obtained. It is claimed to be superior to B3LYP in predicting geometries of organic molecules and heats of reactions as well as activation parameters for organic reactions.[16−20] Contrary to these results, it was reported that O3LYP is improper for the calculation of dispersion-bond homomolecular dimers.[21] But this study did not compare the results against those of the density functional B3LYP. Studying the performance of O3LYP for anticipating bond dissociation energies, geometries, and vibrational frequencies for first-row transition metals, Baker and Pulay[22] found that the values obtained with O3LYP are of equal quality as those achieved with B3LYP, while for predicting atomic excitation energies and ionization potentials they obtained worse results with O3LYP compared to B3LYP.[22] Calculating the activation barriers for hydrocarbon pericyclic reactions, Guner et al.[23] found that the results obtained with O3LYP were comparable to the B3LYP results.

Another new density functional, developed by Musgrave and co-workers,[24] is the hybrid KMLYP. Here the exchange

* Corresponding author phone: +49-351-46338571; fax: +49-351-46339679; e-mail: thomas.strassner@chemie.tu-dresden.de.

10.1021/ct049846+ CCC: $30.25
Published on Web 07/07/2005

functional is a similar combination of the Slater exchange functional[12] and the HF[9] exchange functional, while for the correlation part the functionals of Lee, Yang, and Parr[14] and of Vosko, Wilk, and Nusair are combined.[13] The KMLYP functional has been shown to be a tool of higher precision for calculating enthalpies of reaction and transition-state barriers than the B3LYP method, G2 theories, QCISD(T), and CCSD(T).[24] Studying the geometry as well as the electron affinity of SF_6, Brinkmann et al.[25] found that calculations with the KMLYP functional provided good geometrical parameters for SF_6 and for the corresponding anion. They also observed an improved electron affinity in comparison to standard DFT functionals. Structures and energetics for transition states for H-abstraction reactions are also anticipated well using the KMLYP functional.[26] Predicting the cumulene/poly-yne isomerization energies Pomerantz et al.[27] obtained clearly better results with KMLYP than with B3LYP. KMLYP was also observed to be much better than B3LYP in modeling the aromaticity of 10π-annulenes,[28] while for pericyclic reactions of hydrocarbons, Guner et al.[29] found that KMLYP tends to overestimate activation enthalpies. They could also show that the calculated bond lengths of the transition states are only comparable to the accuracy of those calculated with the B3LYP functional when using high-quality basis sets.

Recently the MPW1K hybrid density functional was developed by Truhlar and co-workers.[30] It uses the modified Perdew-Wang gradient-corrected exchange functional and the correlation functionals of Adamo and Barone.[31] Comparing MPW1K to B3LYP,[10] it is notable that it includes a larger proportion of the HF[9] exchange functional.[30,32,33] It was shown that geometries and vibrational frequencies of stationary points, enthalpies, and barrier heights for organic reactions obtained with the MPW1K method are in good agreement with experimental values or with higher order calculations.[19,29,30,32−48] MPW1K is claimed to be especially superior to other DFT methods for the calculation of large systems.[35] Li et al.[37] investigated several hydrogen atom transfer reactions involved in atmospherical chemistry with hybrid DFT methods. They found that MPW1K provides very accurate stationary point geometries and energetics also in combination with small basis sets. Studying the unimolecular decomposition reaction of C_2H_5O, Zhang et al.[49] could show that MPW1K is also well suited for the calculation of reaction rates. Predicting the structures and B−N bond dissociation energies of amine-boranes with various DFT methods, Gilbert found that using B3LYP poor results were obtained, whereas the MPW1K functional gave good results.[50] It was also shown that using MPW1K the predicted potential energy surface for S_N2 reactions was in good agreement with the one achieved with G2 theory and slightly better than applying the B3LYP functional.[43] Iron et al.[51] demonstrated that the reactivities of organometallic complexes obtained with the MPW1K functional are of comparable accuracy with experimental results, while Kuwata et al.[52] reported the observation that MPW1K/6-31+G(d,p) calculations are less accurate and often disagree with model chemistry predictions investigating the reactivity of syn acetaldehyde oxide. Saeys at al. came to a similar result studying radical reactions of hydrocarbons.[53] Using different DFT functionals including MPW1K, Cohen et al. estimated an inaccurate heat of formation of diazomethane compared to experimental results.[54]

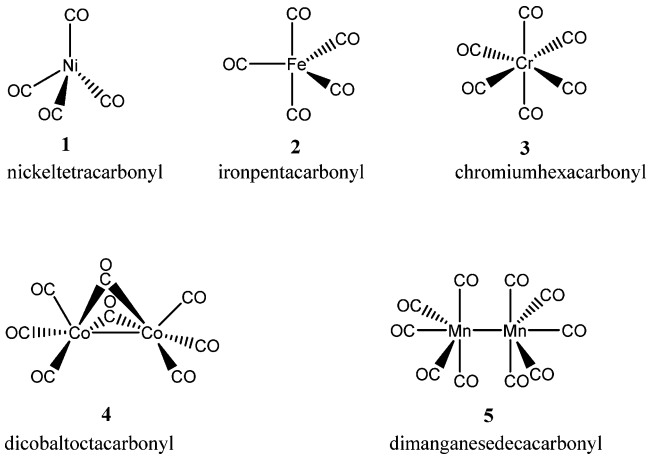
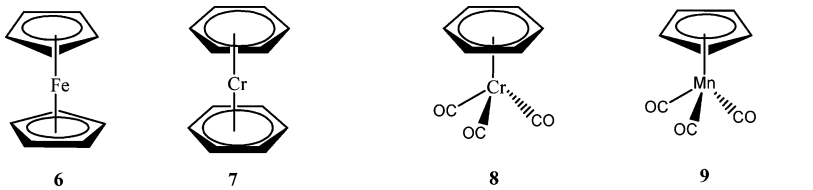


Figure 1. Structures of the selected transition-metal carbonyl compounds.

As described above, many investigations have been published on the suitability of the new hybrid functionals O3LYP, KMLYP, and MPW1K for the calculation of organic molecules and organic reactions. Most of these studies showed that these relatively new functionals are clearly superior to B3LYP, but so far their usefulness for the calculation of geometries and reactivities of transition-metal compounds by these functionals was not clear. Therefore we decided to study the geometries and vibrational frequencies of several transition-metal compounds using the density functionals O3LYP, KMLYP, MPW1K, and B3LYP in comparison to experimental results. Because of the large body of experimental data available, carbonyl complexes are especially well suited for this test,[55] e.g. the calculation of the CO frequencies as an indication for the quality of the method.[56,57]

Computational Details

All calculations were carried out using Gaussian 03.[58] B3LYP, O3LYP, KMLYP, and MPW1K were combined with the split valence double-ζ (DZ) basis set 6-31G(d),[59,60] with the split valence triple-ζ (TZ) basis set 6-311++G(d,p),[61−63] and with LANL2DZ.[64−67] LANL2DZ consists of the D95V[64] basis set for the first- and second-row elements and a double-ζ basis set containing ECP-representations of core electrons for third-row elements and beyond. For ruthenium and osmium the split valence double-ζ (DZ) basis set 6-31G(d)[59,60] and the split valence triple-ζ (TZ) basis set 6-311++G(d,p)[61−63] were used in combination with the Stuttgart/Dresden 1997 ECP.[68,69] To examine the influence of additional diffuse and polarization functions we also calculated selected transition-metal compounds for comparison. Details are given in the Supporting Information together with a complete list of the basis sets used. No symmetry or internal coordinate constraints were applied during optimizations. Vibrational frequency calculations were carried out for all structures to verify them as true minima by the absence of imaginary eigenvalues, and additionally the stability of



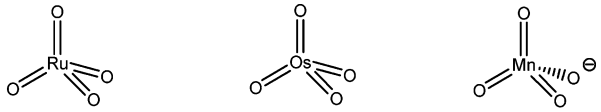
6 ferrocene

7 η^6-dibenzenechromium

8 η^6-benzenechromium-tricarbonyl

9 η^5-cyclopentadienyl-manganesetricarbonyl

Figure 2. Structures of the selected transition-metal sandwich complexes and semisandwich complexes.



10 rutheniumtetraoxide

11 osmiumtetraoxide

12 permanganate

Figure 3. Structures of the selected transition-metal tetraoxides.

the wave function was checked. Different multiplicities were evaluated, and all systems proved to be in the singlet state.[55,56]

Results and Discussion

The suitability of the hybrid density functionals O3LYP, KMLYP, and MPW1K for the prediction of the geometry of transition-metal compounds was investigated. Selected geometrical parameters of the transition-metal carbonyl compounds, transition-metal sandwich complexes, transition-metal semisandwich complexes, and transition-metal tetraoxides were compared to experimental data;[70−82] details are available as Supporting Information. We chose nickel(0)-tetracarbonyl, $Ni(CO)_4$ (**1**), iron(0)pentacarbonyl, $Fe(CO)_5$ (**2**), chromium(0)hexacarbonyl, $Cr(CO)_6$ (**3**), dicobalt(0)-octacarbonyl, $Co_2(CO)_8$ (**4**), and dimanganese(0)decacarbonyl, $Mn_2(CO)_{10}$ (**5**) as examples for transition-metal carbonyl compounds (Figure 1), and ferrocene $Fe(C_5H_5)_2$ (**6**) and η^6-dibenzenechromium, $Cr(C_6H_6)_2$ (**7**) as π-coordinated transition-metal sandwich complexes, while η^6-benzene(0)-chromiumtricarbonyl, $Cr(C_6H_6)(CO)_3$ (**8**) and η^5-cyclopentadienylmanganese(I)tricarbonyl, $Mn(C_5H_5)(CO)_3$ (**9**) represent the transition-metal semisandwich complexes (Figure 2) as well as ruthenium tetraoxide, RuO_4 (**10**), osmium tetraoxide, OsO_4 (**11**), and permanganate, MnO_4^- (**12**) as transition-metal tetraoxides (Figure 3).

We refrained from discussing individual bond lengths and angles. Instead we decided to use a statistical description of the different methods. All details on bond lengths, angles, and frequencies are given in the Supporting Information (Tables S1−S18).

The mean absolute deviations (MAD) and the standard deviations as well as the largest absolute errors for the bond lengths calculated by the split valence double-ζ basis set 6-31G(d) are illustrated in Figures 4−8. A graphical representation of the performance of the four different density functionals together with 6-31G(d) for all calculated compounds except RuO_4 (**10**) and OsO_4 (**11**) is shown in Figure 4. We did not include the differences between computational and experimental values for RuO_4 (**10**) and OsO_4 (**11**) in the overall picture because these two compounds have been calculated using ECPs.

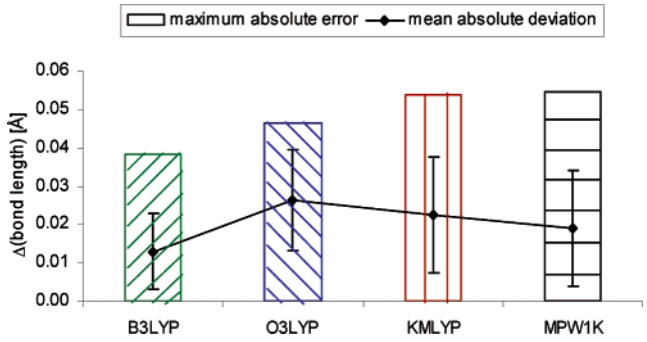


Figure 4. Performance of different hybrid density functionals in combination with 6-31G(d) for the prediction of the bond lengths of compounds **1**−**9** and **12** in comparison to experimentally observed values.

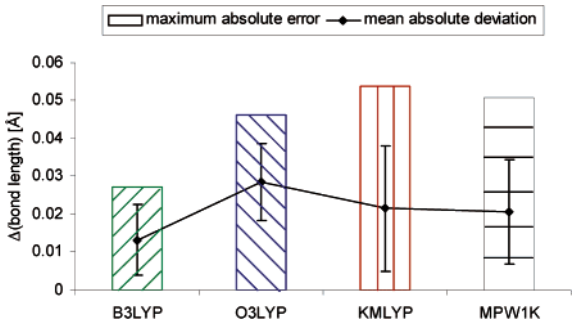


Figure 5. Performance of different functionals in combination with 6-31G(d) for the prediction of the bond lengths of selected first-row transition-metal carbonyl compounds (**1**−**5**) in comparison to experimentally observed values.

Figure 4 shows the overall mean absolute deviation and the maximum absolute error from experimental values for the calculated bond lengths of compounds **1**−**9** and **12** for all four different density functionals in combination with 6-31G(d). B3LYP is clearly superior compared to the other functionals for the calculation of transition-metal compounds together with the basis set 6-31G(d).

The individual results for the different groups of investigated transition-metal compound are summarized below (Figures 5−8), and numerical values are given in the Supporting Information (Table S19).

During this study it was observed that the new hybrid functionals O3LYP, KMLYP, and MPW1K in combination with 6-31G(d) do not perform better than B3LYP in predicting geometries of transition-metal compounds. Instead it could be shown (Figures 5, 7, and 8) that using B3LYP better predictions for almost all bond lengths were obtained. In all cases but for **6** and **7** significantly better results are obtained with B3LYP. For calculating geometries of transi-

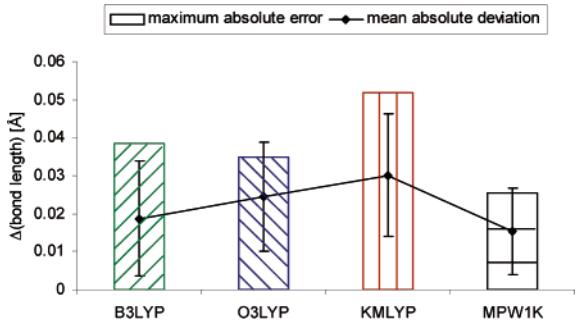


Figure 6. Performance of different functionals in combination with 6-31G(d) for the prediction of the bond lengths of selected first-row transition-metal sandwich complexes (**6**, **7**) vs experimentally observed values.

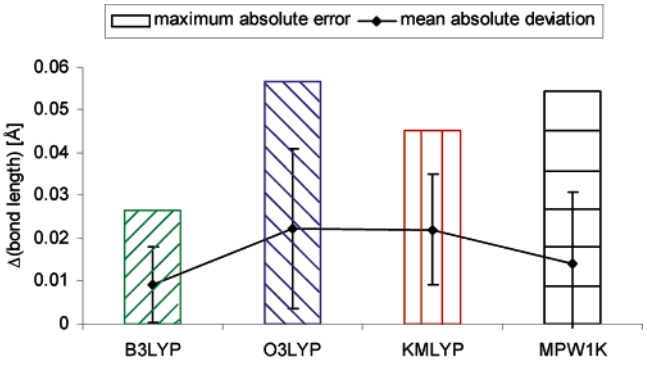


Figure 7. Performance of different functionals in combination with 6-31G(d) for the prediction of the bond lengths of selected first-row transition-metal semisandwich complexes (**8**, **9**) vs experimentally observed values.

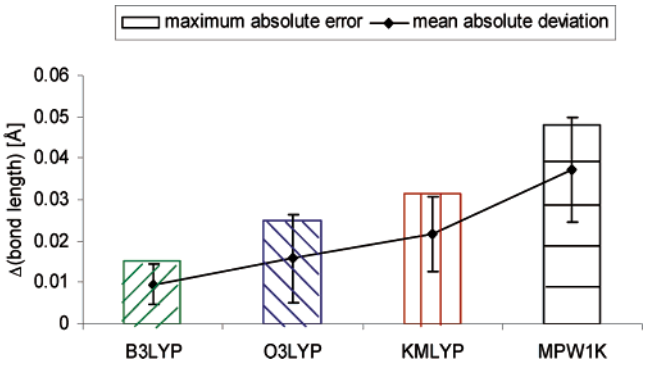


Figure 8. Performance of different functionals in combination with 6-31G(d) for the prediction of the bond lengths of selected transition-metal tetraoxides (**10**−**12**) vs experimentally observed values.

tion-metal sandwich complexes (**6**, **7**) MPW1K performs similar to B3LYP (Figure 6). The maximum absolute errors of the bond lengths of selected first-row transition-metal carbonyl compounds (**1**−**5**) and first-row transition-metal semisandwich complexes (**8**, **9**) using B3LYP are relatively small (in the range of about 0.03 Å, Figures 5 and 7). Applying B3LYP for calculating the bond lengths of transition-metal tetraoxides the maximum error is small, too (≈0.01 Å, Figure 8). For O3LYP or MPW1K the maximum errors of the bond lengths of the compounds studied are in an acceptable range (most 0.06 Å, Figures 5−8). KMLYP

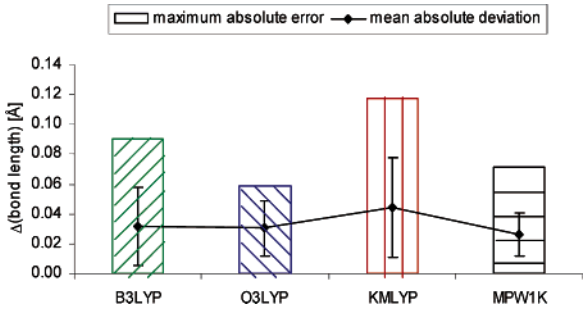


Figure 9. Performance of different functionals in combination with LANL2DZ for the prediction of the bond lengths of compounds **1**−**9** and **12** in comparison to experimentally observed values.

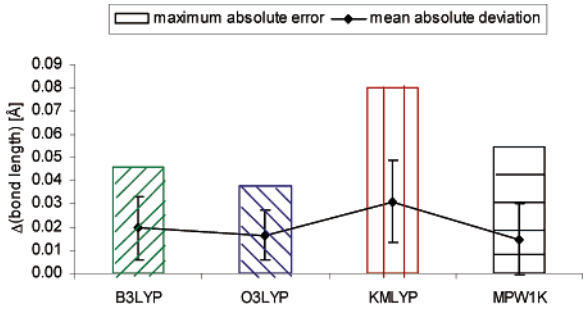


Figure 10. Performance of different functionals in combination with 6-311++G(d,p) for the prediction of the bond lengths of compounds **1**−**9** and **12** in comparison to experimentally observed values.

shows large deviations for the transition-metal carbonyl compounds with up to 0.13 Å (Figure 5).

Basis Set Effects

In addition to 6-31G(d), we also used LANL2DZ and 6-311++G(d,p) as basis sets in combination with the four density functionals. On selected compounds we also checked the influence of additional diffuse and polarization functions (see Supporting Information, Table S20). A statistical description of the LANL2DZ results for compounds **1**−**9** and **12** is given in Figure 9.

In comparison to the results calculated with 6-31G(d) (Figure 8), one notices that a medium-quality basis set like LANL2DZ leads to significant deviations in all cases. They are always larger than the same functional in combination with the split valence double-ζ basis set 6-31G(d) (Figure 4) or the triple-ζ basis set 6-311++G(d,p) (Figure 10).

In combination with 6-311++G(d,p) (Figure 10) the best results were obtained with O3LYP. In agreement with the 6-31G(d) and LANL2DZ results, the largest mean absolute deviation from the experimentally observed values is achieved when using KMLYP.

The results obtained by using the double-ζ basis sets 6-31G(d,p), 6-31+G(d), and 6-31+G(d,p) are similar to those obtained with the double-ζ basis set 6-31G(d), while those achieved with the triple-ζ basis set 6-311+G(3df,2p) are comparable to the results from the calculations with the triple-ζ basis set 6-311++G(d,p). Details of that evaluation are given in the Supporting Information (Table S20).

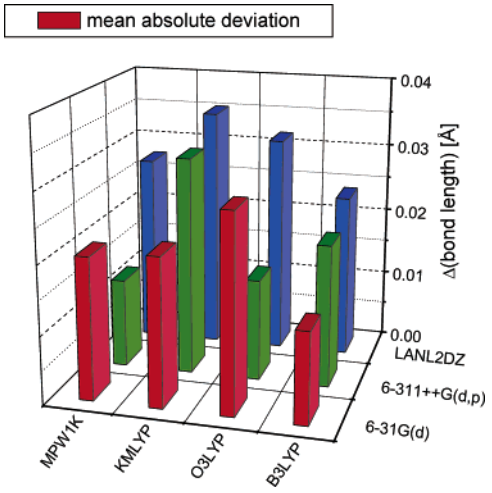


Figure 11. Mean absolute deviation of the calculated bond lengths of selected first-row transition-metal carbonyl compounds (**1**−**5**) from experimentally observed values.

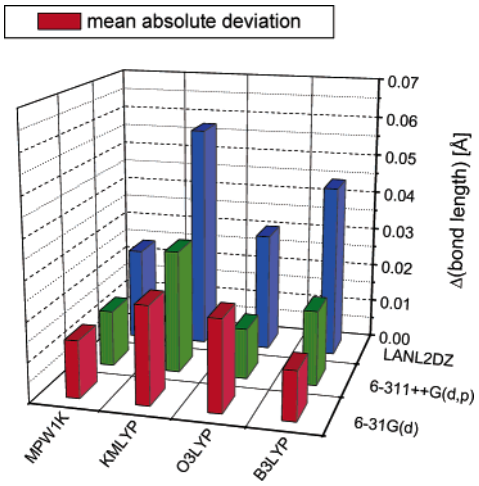


Figure 12. Mean absolute deviation of calculated bond lengths of selected first-row transition-metal sandwich (**6**, **7**) and semisandwich (**8**, **9**) complexes from experimentally observed values.

Taking into account the significant increase in computation time, we recommend to use B3LYP in combination with 6-31G(d) or O3LYP together with 6-311++G(d,p) for calculating transition-metal compounds. Graphical representations of the results are given in Figures 11−13.

Transition-Metal Carbonyl Compounds 1−5. A comparison of the performance of B3LYP, O3LYP, KMLYP, and MPW1K in combination with three different basis sets for the prediction of the bond lengths of selected first-row transition-metal carbonyl compounds (**1**−**5**) is illustrated in Figure 11. The functional/basis set combination B3LYP/6-31G(d) is clearly the best choice for predicting the bond lengths of the transition-metal carbonyl compounds **1**−**5**.

Transition-Metal Sandwich (6, 7) and Semisandwich (8, 9) Complexes. Figure 12 shows the performance of B3LYP, O3LYP, KMLYP, and MPW1K in combination with three different basis sets for the calculation of the bond lengths of the selected transition-metal sandwich (**6**, **7**) and semisandwich (**8**, **9**) complexes. B3LYP/6-31G(d) is the best combination for calculating the geometries of the selected

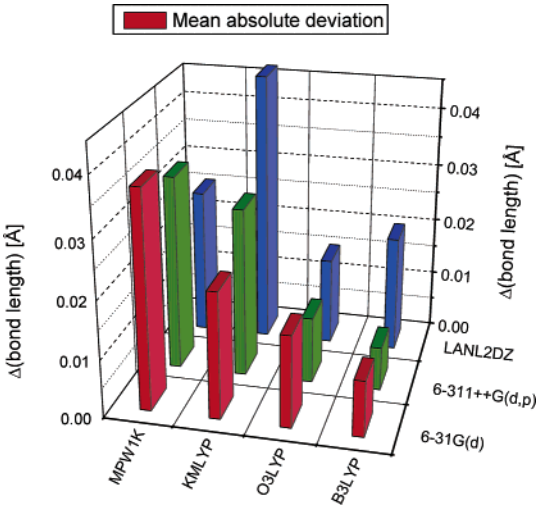


Figure 13. Mean absolute deviation of the calculated bond lengths of selected transition-metal tetraoxides (**10**−**12**) from experimentally observed values.

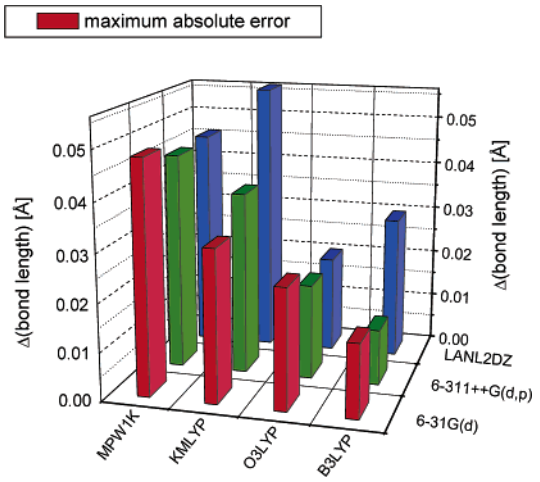


Figure 14. Maximum absolute error of the calculated bond lengths of the investigated transition-metal compounds (**1**−**12**) vs experimentally observed values.

transition-metal sandwich (**6**, **7**) and semisandwich (**8**, **9**) complexes, but together with 6-311++G(d,p), O3LYP and MPW1K are superior to B3LYP. Guner et al.[23,29] already observed a similar effect of higher quality basis sets in combination with the B3LYP functional when they predicted the heats of reactions of hydrocarbon pericyclic reactions with B3LYP, KMLYP, and O3LYP together with double-ζ and triple-ζ basis sets.

Transition-Metal Tetraoxides (10−12). Figure 13 shows the results for the transition-metal tetraoxides (**10**−**12**) for all four functionals in combination with all three basis sets. For compounds **10**−**12** B3LYP/6-31G(d) and B3LYP/6-311++G(d,p) are the favorable combinations.

Maximum Absolute Error. Not only the mean absolute deviation but also the maximum absolute error shows the same trends, given in Figure 14. Applying the split valence double-ζ basis set 6-31G(d), one obtains the smallest maximum absolute error for B3LYP, while in combination with the triple-ζ basis set 6-311++G(d,p) the smallest maximum absolute error is found for O3LYP.

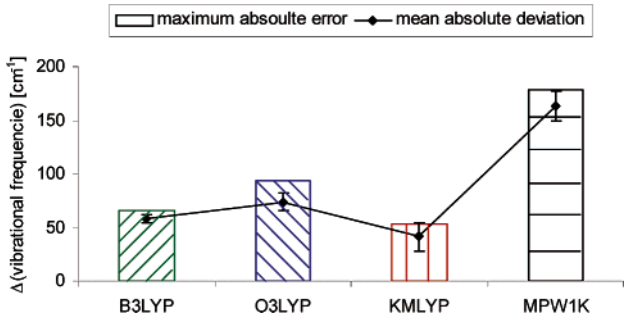


Figure 15. Performance of different functionals with 6-311++G-(d,p) for the prediction of the CO frequencies of selected transition-metal carbonyl compounds (**1**−**5**) vs experimentally observed values.

The maximum negative and positive errors obtained for the different compounds in dependence of functional and basis set are listed in the Supporting Information (Tables S19 and S20). In combination with 6-31G(d) and 6-311++G-(d,p) O3LYP and MPW1K tend to underestimate the bond lengths of the transition-metal compounds, while KMLYP and B3LYP incline to overestimate them.

Vibrational Frequencies. In addition we studied the vibrational frequencies of the selected transition-metal carbonyl compounds (**1**−**5**). Although the accuracy of the calculated frequencies is not perfect, they are still a good indication for the quality of the method used. We compared these calculated frequencies to experimental values from gas-phase data. Selected vibrational frequencies of the transition-metal carbonyl compounds (**1**−**5**) compared to experimental data[83−87] are available as Supporting Information. It is obvious that for the double-ζ calculations larger errors result; therefore, we recommend to use basis sets of triple-ζ quality for these types of calculations. The mean absolute deviations (MAD) and the standard deviations as well as the largest absolute errors of the calculated CO frequencies in comparison to the experimentally obtained values are shown in Figure 15.

For KMLYP/6-31G(d) and KMLYP/6-311++G(d,p) the smallest errors in the CO frequencies were calculated, B3LYP being only the second best method. This result can be explained because the transition-metal atom is not involved in the CO vibration.

Looking at the values in the Supporting Information, Tables S4 and S5, it is obvious that for the triple-ζ basis set the experimentally measured metal−carbon vibrations are reproduced, while the errors in the double-ζ metal−carbon vibrational values are quite large. This is a direct result of the better description of the metal−carbon bonding situation in the higher quality basis set. Only for KMLYP no difference in the quality of the basis sets is observed.

The mean absolute deviations (MAD) and the standard deviations as well as the largest absolute errors of selected metal−carbon vibrational frequencies of $Ni(CO)_4$ (**1**), $Fe(CO)_5$ (**2**), and $Cr(CO)_6$ (**3**) in comparison to experimentally determined values are illustrated in Figure 16. For the transition-metal carbonyl compounds **4** and **5** no experimental data of the metal−carbon vibrational frequencies could be found in the literature.

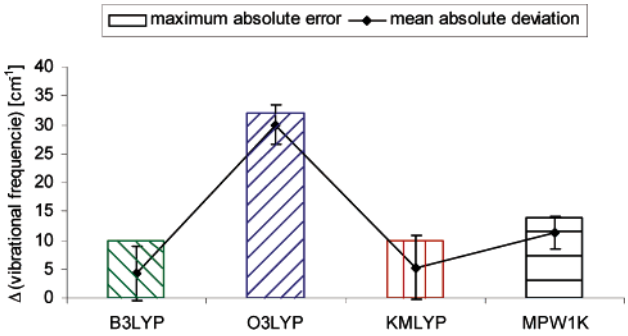


Figure 16. Performance of different functionals in combination with 6-311++G(d,p) for the prediction of selected metal−carbon vibrational frequencies of $Ni(CO)_4$ (**1**), $Fe(CO)_5$ (**2**), and $Cr(CO)_6$ (**3**) vs experimentally observed values.

Figure 16 shows the performance of the four density functionals together with the split valence triple-ζ basis set 6-311++G(d,p) for the calculation of the metal−carbon frequencies of $Ni(CO)_4$ (**1**), $Fe(CO)_5$ (**2**), and $Cr(CO)_6$ (**3**).

Comparing the results obtained with the split valence double-ζ and the triple-ζ basis set, one notices that in combination with 6-311++G(d,p) the errors are smaller for all functionals. B3LYP and KMLYP in combination with 6-311++G(d,p) show similar mean absolute deviations from the experimentally observed values. For B3LYP/6-311++G-(d,p) the mean absolute deviation is slightly smaller. MPW1K and O3LYP show significant mean absolute deviations from the experimentally obtained values, clearly larger than for B3LYP and MPW1K.

The maximum positive and negative errors obtained when predicting the CO and metal−carbon vibrational frequencies of selected transition-metal carbonyl compounds are listed in the Supporting Information (Tables S21 and S22). B3LYP, O3LYP, and MPW1K tend to overestimate the CO vibrational frequencies, while KMLYP tends to underestimate these.

Conclusion

We have investigated the performance of the O3LYP, KMLYP, and MPW1K density functionals in combination with different basis sets for the geometry optimization of several transition-metal compounds as well as for the calculation of the vibrational frequencies of selected transition-metal carbonyl compounds in comparison to the B3LYP density functional. The reliabilities of these methods were evaluated based on their ability to reproduce known geometrical parameters and experimental data.

While O3LYP, KMLYP, and MPW1K have been reported to generally show a better performance than B3LYP for the calculation of organic molecules and organic reactions, this is not true for predicting geometrical parameters of the transition-metal compounds we selected. O3LYP, KMLYP, and MPW1K together with the split valence double-ζ basis set 6-31G(d) gave generally inferior results to B3LYP for the geometry optimization of all transition-metal compounds examined in this study. Only in combination with the split valence triple-ζ basis set 6-311++G(d,p) the results obtained with O3LYP were superior to B3LYP for all investigated

transition-metal compounds. For all studied functionals and transition-metal compounds the mean absolute deviation and the maximum absolute error of the calculated bond lengths is clearly larger when using LANL2DZ as the basis set than for 6-31G(d) or 6-311++G(d,p). Comparing the results obtained with 6-31G(d) and with 6-311++G(d,p) as basis sets, it is important to note that the prediction of the calculated bond lengths of the transition-metal carbonyl compounds (**1**−**5**) in combination with O3LYP and MPW1K as functionals is more precise with 6-311++G(d,p) than with 6-31G(d).

Calculating the CO frequencies of the transition-metal carbonyl compounds (**1**−**5**) together with the split valence double-ζ basis set 6-31G(d) and with the split valence triple-ζ basis set 6-311++G(d,p) the best results were obtained using the KMLYP functional. The prediction of the metal−carbon frequencies of selected transition-metal carbonyl compounds using the B3LYP and KMLYP functionals together with 6-311++G(d,p) we obtained very small deviations from the experimental values, while the results obtained with O3LYP and MPW1K were slightly inferior.

Acknowledgment. We are grateful to the "Fonds der Chemischen Industrie" and to Infineon Technologies AG for support of our research.

Supporting Information Available: Details on calculated and experimentally observed individual bond lengths, angles, and frequencies and maximum absolute errors of the calculated bond lengths in comparison to experimental values as well as a list of the basis sets used. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, 864−871.

(2) Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133−1138.

(3) Ricca, A.; Bauschlicher, C. W. *J. Phys. Chem.* **1994**, *98*, 12899−12903.

(4) Ricca, A.; Bauschlicher, C. W., Jr. *Theor. Chim. Acta* **1995**, *92*, 123−131.

(5) Talarico, G.; Blok, A. N. J.; Woo, T. K.; Cavallo, L. *Organometallics* **2002**, *21*, 4939−4949.

(6) Eriksson, L. A.; Pettersson, L. G. M.; Siegbahn, P. E. M.; Wahlgren, U. *J. Chem. Phys.* **1995**, *102*, 872−878.

(7) Ziegler, T. *Chem. Rev.* **1991**, *91*, 651−667.

(8) Blomberg, M. R. A.; Siegbahn, P. E. M.; Svensson, M. *J. Chem. Phys.* **1996**, *104*, 9546−9554.

(9) Roothaan, C. C. *J. Rev. Mod. Phys.* **1951**, *23*, 69−89.

(10) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(11) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(12) Slater, J. C. *Quantum Theory of Molecules and Solids*, *Vol. 4*: *The Self-Consistent Field for Molecules and* Solids; McGraw-Hill: New York, 1974.

(13) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.

(14) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(15) Handy, N. C.; Cohen, A. *J. Mol. Phys.* **2001**, *99*, 403−412.

(16) Hoe, W.-M.; Cohen, A. J.; Handy, N. C. *Chem. Phys. Lett.* **2001**, *341*, 319−328.

(17) Baker, J.; Pulay, P. *J. Chem. Phys.* **2002**, *117*, 1441−1449.

(18) Stephens, P. J.; Devlin, F.; Chabalowski, M. J.; Frisch, M. J. J. *Phys. Chem.* **1994**, *98*, 11623−11627.

(19) Zhao, Y.; Pu, J.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2004**, *6*, 673−676.

(20) Cohen, A. J.; Handy, N. C. *Mol. Phys.* **2001**, *99*, 607−615.

(21) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334−338.

(22) Baker, J.; Pulay, P. *J. Comput. Chem.* **2003**, *24*, 1184−1191.

(23) Guner, V. A.; Khuong, K. S.; Houk, K. N.; Chuma, A.; Pulay, P. *J. Phys. Chem. A* **2004**, *108*, 2959−2965.

(24) Kang, J. K.; Musgrave, C. B. *J. Chem. Phys.* **2001**, *115*, 11040−11051.

(25) Brinkmann, N. R.; Schaefer III, H. F. *Chem. Phys. Lett.* **2003**, *381*, 123−128.

(26) Song, S.; Golden, D. M.; Hanson, R. K.; Bowman, C. T.; Senosiani, J. P.; Musgrave, C. B.; Friedrichs, G. *Int. J. Chem. Kinet.* **2003**, *35*, 304−309.

(27) Pomerantz, A. E.; Han, J. H.; Musgrave, C. B. *J. Phys. Chem. A* **2004**, *108*, 4030−4035.

(28) Rzepa, H. S.; Sanderson, N. *Phys. Chem. Chem. Phys.* **2004**, *6*, 310−313.

(29) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem. A* **2003**, *107*, 11445−11459.

(30) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811−4815.

(31) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664−675.

(32) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 2936−2941.

(33) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 842−846.

(34) Luo, Q.; Li, Q. S. *J. Phys. Chem. A* **2004**, *108*, 5050−5056.

(35) Coote, M. L. *J. Phys. Chem. A* **2004**, *108*, 3865−3872.

(36) Li, Q. S.; Zhang, Y.; Zhang, S. *J. Phys. Chem. A* **2004**, *108*, 2014−2019.

(37) Li, Q. S.; Xu, X. D.; Zhang, S. *Chem. Phys. Lett.* **2004**, *384*, 20−24.

(38) Bissett, K. M.; Gilbert, T. M. *Organometallics* **2004**, *23*, 850−854.

(39) Claes, L.; Francois, J.-P.; Deleuze, M. S. *J. Comput. Chem.* **2003**, *24*, 2023−2031.

(40) Yang, J.; Ren, Y.; Zhu, H.-j.; Chu, S.-Y. *Int. J. Mass Spectrom.* **2003**, *229*, 199−208.

(41) Claes, L.; Francois, J.-P.; Deleuze, M. S. *J. Am. Chem. Soc.* **2003**, *125*, 7129−7138.

(42) Ren, Y.; Wolk, J. L.; Hoz, S. *Int. J. Mass Spectrom.* **2003**, *225*, 167−176.

(43) Ren, Y.; Wolk, J. L.; Hoz, S. *Int. J. Mass Spectrom.* **2002**, *221*, 59−65.

(44) Claes, L.; Francois, J.-P.; Deleuze, M. S. *J. Am. Chem. Soc.* **2002**, *124*, 7563−7572.

(45) Iron, M. A.; Lo, H. C.; Martin, J. M. L.; Keinan, E. *J. Am. Chem. Soc.* **2002**, *124*, 7041−7054.

(46) Salam, A.; Deleuze, M. S. *J. Chem. Phys.* **2002**, *116*, 1296−1302.

(47) Rybtchinski, B.; Oevers, S.; Montag, M.; Vigalok, A.; Rozenberg, H.; Martin, J. M. L.; Milstein, D. *J. Am.* Chem. Soc. **2001**, *123*, 9064−9077.

(48) Parthiban, S.; De Oliveira, G.; Martin, J. M. L. *J. Phys. Chem. A* **2001**, *105*, 895−904.

(49) Zhang, Y.; Zhang, S.; Li, Q. S. *J. Comput. Chem.* **2004**, *25*, 218−226.

(50) Gilbert, T. M. *J. Phys. Chem. A* **2004**, *108*, 2550−2554.

(51) Iron, M. A.; Martin, J. M. L.; van der Boom, M. E. *J. Am. Chem. Soc.* **2003**, *125*, 11702−11709.

(52) Kuwata, K. T.; Templeton, K. L.; Hasson, A. S. *J. Phys. Chem. A* **2003**, *107*, 11525−11532.

(53) Saeys, M.; Reyniers, M.-F.; Marin, G. B.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. A* **2003**, *107*, 9147−9159.

(54) Cohen, R.; Rybtchinski, B.; Gandelman, M.; Rozenberg, H.; Martin, J. M. L.; Milstein, D. *J. Am. Chem. Soc.* **2003**, *125*, 6532−6546.

(55) Hyla-Kryspin, I.; Grimme, S. *Organometallics* **2004**, *23*, 5581−5592.

(56) Jonas, V.; Thiel, W. *J. Chem. Phys.* **1995**, *102*, 8474−8484.

(57) Ono, Y.; Fujii, Y.; Ishida, T. *Chem. Phys. Lett.* **2004**, *390*, 421−426.

(58) 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 B.05*, Gaussian, Inc., Pittsburgh, PA, 2003.

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

(60) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. *J. Chem. Phys.* **1998**, *109*, 1223−1229.

(61) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650−654.

(62) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033−1036.

(63) Hay, P. J. *J. Chem. Phys.* **1977**, *66*, 4377−4384.

(64) Dunning, T. H.; Hay, P. J. In *Modern Theoretical Chemistry*; Schaefer, H. F., Ed. Plenum: New York, 1976; Vol. 3, p 1.

(65) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270−283.

(66) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284−293.

(67) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299−310.

(68) Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123−141.

(69) Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1991**, *78*, 247−266.

(70) Braga, D.; Grepioni, F.; Orpen, A. G. *Organometallics* **1993**, *12*, 1481−1483.

(71) Jost, A.; Rees, B.; Yelon, W. B. *Acta Crystallogr., Sect. B: Struct. Sci.* **1975**, *31*, 2649−2658.

(72) Sumner, G. G.; Klug, H. P.; Alexander, L. E. *Acta Crystallogr.* **1964**, *17*, 732−742.

(73) Churchill, M. R.; Amoh, K. N.; Wasserman, H. J. *Inorg. Chem.* **1981**, *20*, 1609−1611.

(74) Seiler, P.; Dunitz, J. D. *Acta Crystallogr., Sect. B: Struct. Sci.* **1979**, *35*, 1068−1074.

(75) Keulen, E.; Jellinek, F. *J. Organomet. Chem.* **1966**, *5*, 490−492.

(76) Wang, Y.; Angermund, K.; Goddard, R.; Kruger, C. *J. Am. Chem. Soc.* **1987**, *109*, 587−589.

(77) Cowie, J.; Hamilton, E. J. M.; Laurie, J. C. V.; Welch, A. J. *J. Organomet. Chem.* **1990**, *394*, 1−13.

(78) Whang, D.; Chung, S. K.; Kim, K. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1991**, *47*, 2672−2674.

(79) McGinnety, J. A. *Acta Crystallogr., Sect. B: Struct. Sci.* **1972**, *28*, 2845−2852.

(80) Krivovichev, S. V.; Kir'yanova, E. V.; Filatov, S. K.; Burns, P. C. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **2000**, *56*, 629−630.

(81) Schäfer, L.; Seip, H. M. *Acta Chem. Scand.* **1967**, *21*, 737−744.

(82) Krebs, B.; Hasse, K.-D. *Acta Crystallogr., Sect. B: Struct. Sci.* **1976**, *32*, 1334−1337.

(83) Jones, L. H.; McDowell, R. S.; Goldblatt, M. *Inorg. Chem.* **1969**, *8*, 2349−2363.

(84) Braterman, P. S. *Book Metal Carbonyl Spectra, Vol. Voumen*; Academic: London, 1975.

(85) Sweany, R. L.; Brown, T. L. *Inorg. Chem.* **1977**, *16*, 415−421.

(86) Jones, L. H.; McDowell, R. S.; Goldblatt, M.; Swanson, B. I. *J. Chem. Phys.* **1972**, *57*, 2050−2064.

(87) Parker, D. J. *Spectrochim. Acta* **1983**, *39A*, 463−476.

CT049846+

JCTC Journal of Chemical Theory and Computation

Highly Reduced Polyoxometalates: Ab Initio and DFT Study of $[PMo_8V_4O_{40}(VO)_4]^{5-}$

Xavier López,[†] Coen de Graaf,[‡] Joan Miquel Maestre,[‡] Marc Bénard,[†] Marie-Madeleine Rohmer,[†] Carles Bo,[‡,§] and Josep M. Poblet*[,‡]

Laboratoire de Chimie Quantique, UMR 7551, CNRS and Université Louis Pasteur, 4 rue Blaise Pascal, F-67000 Strasbourg, France, Departament de Química Física i Inorgànica, Universitat Rovira i Virgili, 43005 Tarragona, Spain, and Institut Català d'Investigació Química, Avda, dels Països Catalans s/n, 43007 Tarragona, Spain

Received February 25, 2005

Abstract: DFT and post Hartree−Fock calculations were carried out to characterize the electronic structure of the 10-electron-reduced $[PMo_8V_4O_{40}(VO)_4]^{5-}$ polyoxometalate. This molecule may be viewed as a mixed-metal $PMo_8V_4O_{40}$ Keggin structure capped with four VO units, in which the eight vanadiums form a ring. In mixed V/Mo clusters it is accepted that the first reductions occur at the V^{5+} ions. The BP86 calculations on this modified Keggin anion reveal that the ground state is a septet with the six unpaired electrons delocalized over the eight V centers. The B3LYP calculations and especially the CASSCF technique modify the tendency of the BP86 method, thus reproducing the expected 8/2 distribution. The unpaired electrons residing in the eight vanadiums are antiferromagnetically coupled.

Introduction

The chemistry of medium and large molecular metal oxides, or polyoxometalates (POMs),[1] was mainly restricted for years to the tungsten and molybdenum derivatives, although it has been enriched and broadened in the past decades.[2] New chemical and structural ingredients are incorporated to this field in order to tune the characteristics of such inorganic systems. Among varied strategies driven nowadays with this goal, one of them is incorporating capping units to a parent structure. This method resides in the addition of M_xO_y groups to the surface of a POM cluster. A handful of capped clusters have been synthesized and characterized in the last 10 years, which are mostly restricted to compounds containing Mo, V, and As atoms. A benchmarking bicapped anion was characterized by Hill and co-workers in the middle 1990s,[3] with the formula $[PMo_{12}O_{40}(VO)_2]^{5-}$. The special electronic features of this anion were analyzed by means of the DFT methodology by Maestre et al.,[4] which reproduced and explained the particular electronic nature of such polyanion. Other derivatives with similar chemical compositions containing capping units have been reported, either constructed upon a single-addenda or a substituted parent POM core.[5−14] Novel capped clusters containing V and Ge atoms were reported by Whitfield et al. in 2003.[15] $[Ge_4V_{16}O_{42}(OH)_4]^{8-}$ constitutes the first example of Ge_2O_7 capping units incorporated to vanadate derivatives based on the Keggin anion.

In 1999, Xu et al. reported a novel POM containing for the first time *four* capping groups, namely the $[PMo_8V_4O_{40}(VO)_4]^{5-}$ cluster (**1**),[16] shown in Figure 1A. This species is made up of a $[PMo_8V_4O_{40}]^{7-}$ Keggin-like mixed-metal cluster (Figure 1B) where four VO^{3+} units are added to the 12-metal core. The VO^{3+} entities carry a total charge of +12, which transforms the initial $[PMo_8V_4O_{40}]^{7-}$ anion into the $[PMo_8V_4O_{40}(VO)_4]^{5+}$ cation. The resulting compound was indeed characterized as a $10e^-$-reduced molecule, thus with a net charge of −5. More recently, the same group published a related compound with similar electronic structure[10] and, in 2003, Yuan et al. reported three novel V-capped Keggin derivatives with paramagnetic metal complexes attached to the surface of the cluster.[17] Perhaps, the most important common feature of such compounds is that they are found in highly reduced forms.

* Corresponding author e-mail: josepmaria.poblet@.urv.net.
[†] UMR 7551, CNRS and Université Louis Pasteur.
[‡] Universitat Rovira i Virgili.
[§] Institut Català d'Investigació Química.

10.1021/ct050040z CCC: $30.25
Published on Web 08/06/2005

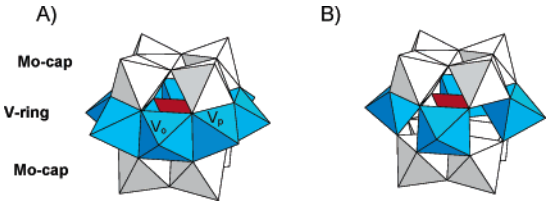


Figure 1. Polyhedral representation of the $[PMo_8V_4O_{40}(VO)_4]^{5-}$ anion (A). Blue polyhedra show the equatorial V_8-ring of alternating octahedral and pyramidal, V_o/V_p, vanadiums, whereas four molybdenums are located in each polar region. The four V_p are the "capping" units of the parent $[PMo_8V_4O_{40}]$ framework (B) and are formally pentacoordinated in the final cluster. The overall symmetry of the cluster is D_{2d}.

Although the structure reported for **1** is formally a tetracapped Keggin cluster, it may be viewed as an eight-membered equatorial ring of vanadium atoms (V_8-ring) together with two polar units made of four MoO_6 units each (Figure 1). Such framework contains all-equivalent Mo atoms, whereas two vanadium types may be distinguished: four pseudo-octahedrally coordinated and four in a square-pyramidal environment, which will be labeled as V_o and V_p from now on, respectively. These two types of V centers are alternated within the V_8-ring. With the analysis of **1** we pursue two objectives: to rationalize reduced mixed-addenda Mo/V derivatives and to progress in the understanding of capped clusters.[3,4] The present work is based on quantum chemical calculations performed upon the title structure and the subsequent analysis and discussion of the data related to the electronic structure.

Results and Discussion

In general, it is difficult to unequivocally characterize the number and nature of the reduced metal centers in an "electron-rich" polyoxometal cluster.[18] The number of electronic configurations that are energy-accessible by a reduced system is, at least, moderately large within a small energy range. This is commonplace in reduced POMs since the metal electrons reside over a large number of chemically similar metal centers. In particular, difficulties come out in the assignment of oxidation states in mixed oxides of vanadium and molybdenum. Electrochemical reductions and EPR spectroscopy have been frequently used to determine the localization/delocalization of the metal electrons in reduced POMs. For example, Hervé and co-workers found after electrochemical reduction of $XMo_2VW_9O_{40}$ and $XMoV_2W_9O_{40}$ anions (X = P, Si) that the V centers are commonly reduced in the first place.[19] In this line of thought, Xu et al., from X-ray diffraction and EPR spectra analysis, claim that the electronic structure for **1** corresponds to the 10 metal electrons delocalized over the whole system. In fact, from valence sum calculations they conclude that all vanadium atoms have an oxidation state of +4. This means that the eight Mo ions share two electrons. The calculations discussed below were performed with the DFT and CASSCF techniques and are aimed at unraveling the electronic structure, e.g., the molecular orbitals sequence and especially the repartition of metal electrons between Mo and V.

DFT Calculations. We carried out a density-functional-theory (DFT) analysis to characterize the electronic structure of **1**. We performed calculations using the exchange-correlation functionals of Becke and Perdew (BP86)[20] and the hybrid B3LYP.[21] The BP86 geometries were optimized, within the constraints of the D_{2d} symmetry, by means of the ADF program[22] with a basis set of triple-ζ + polarization quality. The internal or core electrons (O, 1s; P, 1s−2s; V, 1s−2p; Mo, 1s−4d) were frozen and described by single Slater functions. We applied scalar relativistic corrections to them by means of the zeroth-order regular approximation (ZORA) via the core potentials generated with the DIRAC program.[22] Open-shell configurations were computed within the unrestricted formalism, and some low-spin open-shell configurations are treated within the Broken Symmetry (BS) approach.[23]

The closed-shell configuration, $S = 0$, shows the typical electronic structure common for most POMs, that is, a band of oxo orbitals doubly occupied and a set of metal orbitals well separated in energy. The two lowest metal orbitals, of a_2 and b_1 symmetry, are quasi-degenerate nonbonding combinations of d_{xy}(Mo) orbitals. We found the energy gap between the highest occupied oxo orbital and the first occupied metal orbital to be 1.74 eV (40.1 kcal mol^{-1}), a somewhat smaller value compared to regular Keggin-like molybdates, which feature ∼2 eV (∼46 kcal mol^{-1}) in most cases at this level of computation.[24] At higher energies we find a set of d_{xy}(V)-like orbitals, and the next Mo-like orbital appears more than 1.2 eV (27.7 kcal mol^{-1}) above the first metal orbital (see Figure 2, configuration $S = 0$). The various BP86 calculations performed preferentially give a balanced sharing of the metal electrons, with six for V and four for Mo (6/4), which in principle contradicts the experimental data and also the general assumption that all V^{5+} centers should be reduced before the first Mo center does. The stability of the 6/4 configurations increases, in general, as the number of unpaired electrons grows. Thus, the septet state,[25] with the six unpaired electrons delocalized over the eight V atoms, is the most stable 6/4 configuration and was computed to be 19.1 kcal mol^{-1} below $S = 0$. For this septet state, the Mulliken partition scheme assigns spin populations of 0.90e to V_o and 0.78e to V_p. Several BS low-spin open-shell configurations (not represented in Figure 2) were also found to be more stable than $S = 0$, although invariably with a 6/4 configuration. Among these low-spin open-shell calculations, a common feature was observed, namely, the antiferromagnetic coupling of metal electrons delocalized over the V-ring, as shown in Figure 3.

If we promote all the beta spin electrons of the $S = 0$ configuration and accommodate them in higher orbitals to give a $S = 5$ form, the system increases its energy by 28.1 kcal mol^{-1}, thus lying 47.2 kcal mol^{-1} above the septet state ($S = 3$). This high-spin state is still a 6/4 form as illustrated in Figure 2. The lowest 8/2 state calculated, a triplet ($S = 1$) computed +37.7 kcal mol^{-1} above $S = 0$, confirms that the 6/4 configuration is the most stable with the BP86 functional.

Table 2 contains computed interatomic distances for the title cluster. For comparison, the table also includes the corresponding X-ray values reported in ref 16. The reader

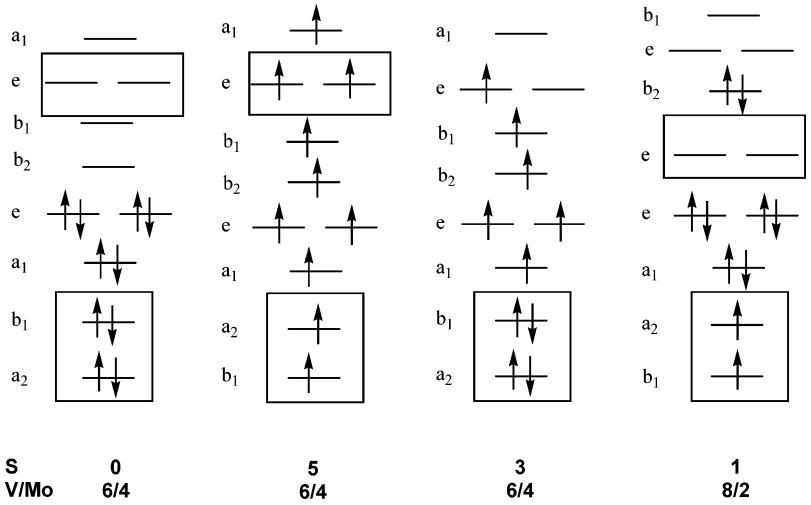


Figure 2. Molecular orbitals representation of the lowest lying metal orbitals of the title compound studied at the BP86 level. The boxes contain orbitals with Mo character. The spin (S) and the V/Mo electron distribution are shown, whereas the relative energies are listed in Table 1. There is no energy scaling in the picture. Only the case $S = 0$ was computed with the restricted DFT formalism.

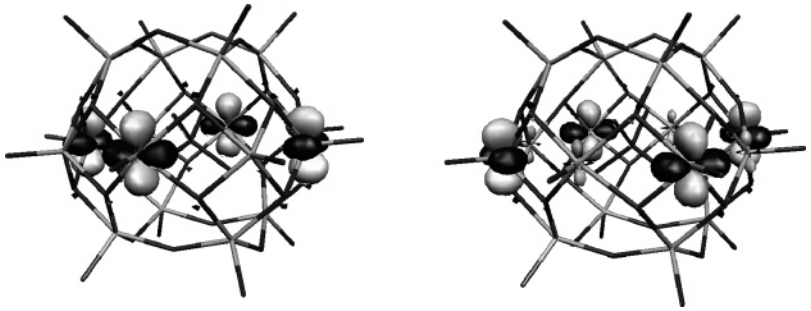


Figure 3. 3D representation of the alpha (left) and beta (right) spatial parts of a symmetry-adapted d_{xy}(V)-like orbital of a_1 symmetry showing the alternate localization in the V_8-ring computed with the unrestricted DFT formalism.

Table 1. Electronic Parameters for $[PMo_8V_4O_{40}(VO)_4]^{5-}$ Computed at the BP86 Level

configuration V/Mo[a]	6/4	6/4	6/4	8/2
spin (S)	0	5	3	1
relative energy[b]	0.00	+28.1	−19.1	+37.7
spin density[c]				
Mo		0.44	0.018	0.27
V_o		0.95	0.90	0.82
V_p		0.89	0.78	−0.78

[a] Number of metal electrons residing on d-V and d-Mo orbitals. [b] kcal mol^{-1}. [c] Spin density per atom.

may observe that the geometry does not change significantly upon variations in the electronic configuration and that the computed distances agree fairly well with the experimental ones. For the type of molecules and computation methods discussed here, the largest deviation between experimental and computed values takes place in the terminal M−O bonds. Hence, for example, the computed terminal Mo−O bond is about 0.06 Å longer than the X-ray.

Taking the reference of the electronic configurations studied with the BP86 functional, we also analyzed the title structure at the B3LYP level. We made use of the Gaussian03 package[26] to obtain energies for several states employing DZ basis sets together with the Los Alamos effective core potentials for the metal atoms. As in the BP86 calculations, the geometries were fully optimized under the constraints of the D_{2d} symmetry group. The values in Table 2 clearly show that the two functionals give similar geometries. On the other hand, the results obtained illustrate that the 6/4 electronic distribution, the lowest in energy with BP86, is not the most stable with the hybrid functional. Using B3LYP, the closed-shell calculation with $S = 0$ also locates six and four electrons over vanadiums and molybdenums, respectively, but this configuration is 125 kcal mol^{-1} less stable than the high-spin 8/2 distribution of the metal electrons ($S = 5$, showed in Figure 4). The $S = 3$ calculation, with formally four paired electrons over Mo atoms and six electrons delocalized over vanadiums, was found 34.5 kcal mol^{-1} above the $S = 5$ state. Finally, the BS triplet state ($S = 1$) strongly competes with $S = 5$ since it was computed only 0.38 kcal mol^{-1} above the all-spin-up form. The atomic spin populations computed for vanadiums with the Mulliken method are alternatively +/−1.14e for V_o and V_p sites. These 8/2 configurations and others with the same electron repartition were computed with the BP86 functional, but they were never found competitive in energy with respect to the 6/4 pattern. Notice from the values in Tables 1 and 3 that the incorporation of the exact exchange functional through a hybrid functional in the calculations favors the localization of the electrons on the vanadium ions. This is an important result since it indicates that the B3LYP functional reproduces the localization of the electrons in mixed V/Mo POMs more

Table 2. Selected Distances (in Å) Computed for the Most Stable States for 6/4 and 8/2 Configurations of the $[PMo_8V_4O_{40}(VO)_4]^{5-}$ Anion

	BP86		B3LYP		X-ray[b]
configuration V/Mo[a]	6/4	8/2	6/4	8/2	
spin (S)	3	1	3	5	
Mo−Mo	3.47−3.74	3.44−3.76	3.48−3.73	3.50−3.76	3.48−3.62
V−V	3.01	3.02	3.02	3.06	2.90−3.03
Mo−V	3.06	3.04	3.09	3.08	3.02
Mo−O(term)	1.73	1.74	1.74	1.73	1.67
V−O(term)	1.62−1.63	1.62−1.63	1.60−1.63	1.61−1.62	1.59−1.66
P−Mo	3.67	3.64	3.67	3.66	3.59−3.64
P−V	3.68−4.11	3.72−4.11	3.62−4.16	3.69−4.20	3.60−4.06

[a] Number of metal electrons residing on d-V and d-Mo orbitals. [b] Reference 16.

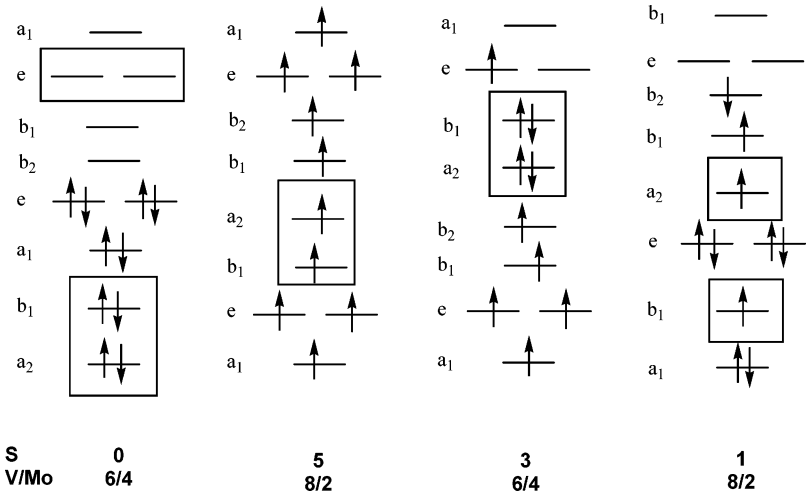


Figure 4. The same scheme as in Figure 2 but using the B3LYP functional. Energy data are listed in Table 2.

Table 3. Electronic Parameters for $[PMo_8V_4O_{40}(VO)_4]^{5-}$ Computed at the B3LYP Level

configuration V/Mo[a]	6/4	6/4	8/2	8/2[d]
spin (S)	0	3	5	1
relative energy[b]	+125	+34.5	0.00	+0.38
spin density[c]				
Mo		0.01	0.32	0.31
V_o		0.74	1.13	1.14
V_p		1.04	1.16	−1.14

[a] Number of metal electrons residing on d-V and d-Mo orbitals. [b] kcal mol^{-1}. [c] Spin density per atom. [d] State computed within the broken symmetry approach.

accurately than pure GGA functionals. For an extensive description of the localization/delocalization problem see for example the papers in ref 27. To evaluate the effect of the basis set upon the results, we tried to obtain some of the configurations given in Figures 3 and 4 with the BP86 functional and the DZ GTO-basis set used in the B3LYP calculations. However, we were unable to converge any 8/2 configuration with the former functional. At this level of computation, the energy difference between the high ($S = 5$) and low spin ($S = 0$) states for the 6/4 distribution (Figure 2) was computed to be 20 kcal mol^{-1}, thus ~8 kcal mol^{-1} smaller than the difference found with the TZP STO-basis set.

The most noteworthy result is that the nature of the orbitals involved in the delocalization of the metal electrons is roughly the same with the two functionals used; however, the relative energy of the electronic configurations varies notably depending on the method applied. Otherwise to the BP86 functional, B3LYP—which includes the exact electron exchange—tends to stabilize the 8/2 configurations with respect to the 6/4 repartition. In addition, the occupied d-orbitals of the Mo atoms appear preferentially at higher energies in comparison to vanadium d-orbitals, except for the closed-shell 6/4 distribution. In this latter configuration, the molecular orbitals sequence and nature are the same, thus indicating that the most remarkable differences between these two functionals appear when open shells are considered and, especially, when the molecular net spin is large.

Ab Initio Post-Hartree−Fock Calculations. To properly represent the distribution of the unpaired metal electrons in $[PMo_8V_8O_{44}]^{5-}$, one may tentatively improve the description by approaching the problem from a different viewpoint. The DFT calculations discussed above indicate that there are many different electronic configurations in a small energy interval. Therefore, it might be more consistent to include all these near-degeneracies in the description of the electronic charge distribution from the start in a well-defined way. Despite the large size of the POM under study, we employ complete active space self-consistent field (CASSCF) calculations as implemented in the MOLCAS 5.4 code.[28−29] This wave function based method has two important advantages above DFT. In the first place, it gives a rigorous description of the spin states, i.e., the N-electron wave function is an eigenfunction of the S^2 operator. Second, it treats all

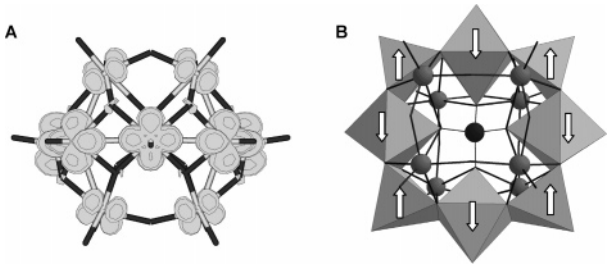


Figure 5. (A) View along the V_8-equatorial plane of the CASSCF alpha spin density for the singlet ground state of $[PMo_8V_4O_{40}(VO)_4]^{5-}$. (B) Perpendicular view of the V_8-ring (polyhedra) showing the antiparallel coupling found for the eight vanadium electrons (grey spheres represent molybdenum centers).

important electronic configurations on an equal footing. This means that all so-called nondynamical electron correlation is accounted for. It is, however, very difficult to include the dynamical electron correlation (the description of the discontinuity in the gradient of the wave function when two electrons are close to each other) in the CASSCF wave function due to the size of the molecule. Nevertheless, the CASSCF approach is usually sufficient for an accurate description of the electron distribution of a given N-electron state. The dynamical electron correlation can have a strong effect on the relative energy of different states but does not change the character of these states.[30−31]

CASSCF calculations were performed on the high-spin coupled ($S = 5$) and the low-spin (singlet) coupled electronic states using the same DZ basis set as in the B3LYP calculations. The active space contains 10 electrons and 12 orbitals. The active orbitals are linear combinations of either V-centered 3d orbitals or combinations of Mo-4d orbitals. In principle, 16 linear combinations can be constructed with these atomic d-orbitals (8 Mo-4d and 8 V-3d). However, the active space has been limited to the 12 molecular orbitals that are lowest in energy, excluding the four out-of-phase linear combinations of Mo-4d orbitals. The optimization of both the CI expansion and the molecular orbital coefficients ensures an unbiased treatment of the different electron distributions over V and Mo. The outcomes of the CASSCF calculations on the high-spin state indicate a 8/2 distribution of the unpaired electrons over the V and Mo centers, respectively. The Mulliken spin populations for the two different V centers are 0.98 e (i.e. there are eight unpaired electrons in the V_8-ring) and 0.22 for Mo, consistent with two electrons delocalized over the eight Mo centers.

The singlet state is slightly lower (1.2 kcal mol^{-1}) in energy with respect to the undecaplet state, although it is expected that this energy difference increases when dynamical electron correlation is accounted for. The character of the unpaired spin distribution in the singlet state cannot be assessed through the Mulliken spin populations since these are strictly zero for a singlet spin eigenfunction. However, the character of the active orbitals and the natural occupation numbers indicate that this state can be characterized by an antiferromagnetic alignment of 10 unpaired electrons with a 8/2 spin distribution over V and Mo. The alpha spin density depicted in Figure 5 clearly shows how each V has an equivalent number of unpaired alpha electrons and that the spin density of the Mo is smaller. Finally, we would like to illustrate the intrinsic multiconfigurational character of the singlet state by reminding that the CI expansion has 80 configuration state functions with $|c_i|^2 > 0.05$ and that the largest CI coefficient is only 0.1737, less than 3% of the total wave function.

Conclusions

The complex problem of the highly reduced polyoxometalates containing V and Mo atoms was faced by means of the ab initio and DFT methodologies. We optimized the geometry of the closed-shell electronic configuration of $[PMo_8V_4O_{40}(VO)_4]^{5-}$ with the DFT/BP86 methodology. At this level of calculation, six out of the 10 metal electrons are delocalized over the V-ring and the other four over the Mo centers (6/4 sharing). Since, in general, it is assumed that V^{5+} ions are reduced before the Mo^{6+} centers, we performed B3LYP and ab initio calculations to check the possible outcomes found in the BP86 description of the electron distribution. Both the B3LYP and CASSCF results eventually change the tendency of the DFT/BP86 method, thus reproducing the expected 8/2 configuration. Although most of the theoretical studies reported on polyoxoanions have been carried out using the BP86 functional the present results suggest that mixed vanadium−molybdenum anions are better described through a hybrid functional. Another appealing characteristic arising from all our calculations is the antiferromagnetic coupling that some metal electrons feature. The eight electrons residing in the equatorial V_8-ring feature a spin-up/spin-down alternate conformation.

Polyoxomolybdates exhibit an extreme variety of complicated structures.[32] Through self-assembly mechanisms mixed-valence giant molecules can be formed. The case herein discussed reveals that the study of the electronic structure of mixed-valence vanadomolybdates is not straightforward and suggests that in the modeling of molybdates constituted by hundreds of atoms the *size* of the anion is not the unique difficulty to overcome.

Acknowledgment. This work was supported by the Spanish MCyT (BQU2002-04110-C02-01) and the CIRIT of the Generalitat of Catalonia (SGR01-00315). Some calculations have been carried out in the Centre Universitaire Régional de Resources Informatiques (CURRI, Université Louis Pasteur). X.L. also thanks the Spanish MEC for a postdoctoral grant (EX2004-0113).

Supporting Information Available: Coordinates for the $[PMo_8V_4O_{40}(VO)_4]^{5-}$ structure computed at the B3LYP level. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Pope, M. T. *Heteropoly and Isopoly Oxometalates*; Springer-Verlag: Berlin, 1983.

(2) Borrás-Almenar, J. J.; Coronado, E.; Müller, A.; Pope, M. T. *Polyoxometalate Molecular Science*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003.

(3) Chen, Q.; Hill, C. L. *Inorg. Chem*. **1996**, *35*, 2403.

(4) Maestre, J. M.; Poblet, J. M.; Bo, C.; Casañ-Pastor, N.; Gómez-Romero, P. *Inorg. Chem*. **1998**, *37*, 3444.

(5) Khan, M. I.; Zubieta, J. *Inorg. Chim. Acta* **1992**, *193*, 17.

(6) Xu, Y.; Xu, J. Q.; Yang, G. Y.; Wang, T. G. Xing, Y.; Lin, Y. H.; Jia, H. Q. *Polyhedron* **1998**, *17*, 2441.

(7) Dolbecq, A.; Cadot, E.; Eisner, D.; Sécheresse, F. *Inorg. Chem*. **1999**, *38*, 4217.

(8) Luan, G. Y.; Li, Y. G.; Wang, E. B.; Han, Z. B. *Inorg. Chem. Commun*. **2001**, *4*, 632.

(9) Müller, A. *Z. Anorg. Allg. Chem*. **1994**, *620*, 599.

(10) Xu, Y.; Zhu, D.-R.; Guo, Z.-J.; Shi, Y.-J.; Zhang, K.-L.; You, X.-Z. *J. Chem. Soc., Dalton Trans*. **2001**, 772.

(11) Huang, G. Q.; Zhang, S. W.; Wei, Y. G.; Shao, M. C. *Polyhedron* **1993**, *12*, 1483.

(12) Khan, M. I.; Chen, Q.; Zubieta, J. *Inorg. Chim. Acta* **1993**, *212*, 199.

(13) Khan, M. I.; Chen, Q.; Zubieta, J. *Inorg. Chem*. **1993**, *32*, 2924.

(14) Müller, A.; Beugholt, C.; Kögerler, P.; Bögge, H.; Bud'ko, S.; Luban, M. *Inorg. Chem*. **2000**, *39*, 5176.

(15) Whitfield, T.; Wang, X.; Jacobson, A. J. *Inorg. Chem*. **2003**, *42*, 3728.

(16) Xu, Y.; Zhu, H.-G.; Cai, H.; You, X.-Z. *Chem. Commun*. **1999**, 787.

(17) Yuan, M.; Li, Y.; Wang, E.; Tian, C.; Wang, L.; Hu, C.; Hu, N.; Jia, H. *Inorg. Chem*. **2003**, *42*, 3670.

(18) Poblet, J. M.; López, X.; Bo, C. *Chem. Soc. Rev*. **2003**, *32*, 297.

(19) Cadot, E.; Fournier, M.; Tézé, A.; Hervé, G. *Inorg. Chem*. **1996**, *35*, 282.

(20) Becke, A. D. *J. Chem. Phys*. **1986**, *84*, 4524; *Phys. Rev. A* **1988**, *38*, 3098. Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822. Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406.

(21) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785.

(22) ADF 2003.01, Department of Theoretical Chemistry, Vrije Universiteit, Amsterdam. Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys*. **1973**, *2*, 41. Versluis, L.; Ziegler, T. *J. Chem. Phys*. **1988**, *88*, 322. Te Velde, G.; Baerends, E. J. *J. Comput. Phys*. **1992**, *99*, 84. Fonseca Guerra, C.; Snijders, J. G.; Te Velde, G.; Baerends, E. J. *Theor. Chem. Acc*. **1998**, *99*, 391.

(23) Noodleman, L. J. *Chem. Phys*. **1981**, *74*, 5737. Noodleman, L.; Davidson, E. R. *Chem. Phys*. **1986**, *109*, 131. Noodleman, L.; Case, D. A. *Adv. Inorg. Chem*. **1992**, *38*, 423. Noodleman, L.; Pen, C. Y.; Case, D. A. Mouesca, J. M. *Coord. Chem. R*ev. **1995**, *144*, 199.

(24) López, X.; Maestre, J. M.; Bo, C.; Poblet, J. M. *J. Am. Chem. Soc*. **2001**, *123*, 9571.

(25) This electronic state could stabilize through a first-order Jahn−Teller effect; however, our experience on similar cases points out that the stabilization upon distortion would be small (Maestre, J. M.; López, X.; Bo, C.; Casañ-Pastor, N.; Poblet, J. M. *J. Am. Chem. Soc.* **2001**, *123*, 3749).

(26) *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.; and Pople, J. A. Gaussian, Inc., Wallingford, CT, 2004.

(27) Pickett, W. E. *Rev. Mod. Phys* **1989**, *61*, 433. Mattheiss, L. F. *Phys. Rev. B* **1994**, *49*, 14050. Cabrero, J.; Calzado, C. J.; Maynau, D.; Caballol, R.; Malrieu, J. P. *J. Phys. Chem. A* **2002**, *106*, 8146. Pacchioni, G.; Frigoli, F.; Ricci, D.; Weil, J. A. *Phys. Rev. B* **2001**, *63*, 054102.

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

(29) Andersson, K.; Barysz, M.; Bernhardsson, A.; Blomberg, M. R. A.; Cooper, D. L.; Fleig, T.; Fülscher, M. P.; de Graaf, C.; Hess, B. A.; Karlström, G.; Lindh, R.; Malmqvist, P.-Å; Neogrády, P.; Olsen, J.; Roos, B. O.; Schimmelpfennig, B.; Schütz, M.; Seijo, L.; Serrano-Andrés, L.; Siegbahn, P. E. M.; Stålring, J.; Thorsteinsson, T.; Veryazov, V.; Widmark, P.-O. MOLCAS version 5.4, Department of Theoretical Chemistry, University of Lund, 2002.

(30) Roos, B. O.; Ryde, U. In *Comprehensive Coordination Chemistry 2;* Lever, B., Ed.; Elsevier: Amsterdam, The Netherlands, 2003; Vol. 1.

(31) de Graaf, C.; Sousa, C.; Broer, R. *Phys. Rev. B* **2004**, *70*, 235104.

(32) Müller, A.; Peters, F.; Pope, M. T.; Gatteschi, D. *Chem. Rev*. **1998**, 239.

CT050040Z

JCTC Journal of Chemical Theory and Computation

A Resolution-Of-The-Identity Implementation of the Local Triatomics-In-Molecules Model for Second-Order Møller−Plesset Perturbation Theory with Application to Alanine Tetrapeptide Conformational Energies

Robert A. DiStasio Jr., Yousung Jung, and Martin Head-Gordon*

Department of Chemistry, University of California, Berkeley, California 94720

Received May 5, 2005

Abstract: In this work, we incorporate the resolution-of-the-identity (RI) approximation into the theoretical framework of the local triatomics-in-molecules (TRIM) second-order Møller−Plesset (MP2) perturbation theory model. The resultant model, RI-TRIM MP2, emerges as a robust fourth-order methodology that extends the regime of practical MP2 calculations. With RI-TRIM MP2, correlation energy corrections can easily be obtained for systems that contain more than 125 heavy atoms with a computational timing cost less than those of the prerequisite self-consistent field procedure and popular density functional theory (DFT) alternatives. In this work, the chemical performance of RI-TRIM MP2 is numerically assessed against untruncated RI-MP2 and DFT (B3LYP) in determining the relative energies of 27 different alanine tetrapeptide conformations at the cc-pVXZ (X = D, T, and Q) levels and the results are T → Q extrapolated to the complete basis set limit. As the quality of the basis set employed increases, we report a significant reduction in the error introduced by the RI-TRIM approximation; at the cc-pVDZ level, the root mean-square (RMS) relative error was found as 0.192 kcal/mol and is decreased to an almost negligible 0.040 kcal/mol at the T → Q extrapolated complete basis set limit. Basis set dependence was investigated by computing the RMS (max) deviations from the extrapolated RI-MP2/cc-pV(TQ)Z data set found as 0.377 (0.944) kcal/mol (MP2/cc-pVTZ) and 0.250 (0.591) kcal/mol (TRIM MP2/cc-pVTZ). These deviations are chemically significant when compared against the conformer energy differences, suggesting that to obtain reliably converged relative conformational energies, computations must be done using the cc-pVTZ and cc-pVQZ basis sets followed by extrapolation to the cc-pV(TQ)Z limit. The findings reported herein also provide the first computational evidence demonstrating that the TRIM model approaches exactness as the one-particle basis approaches completeness.

I. Introduction

Over the past two decades, density functional theory (DFT) has emerged as a robust electronic structure method for treating a myriad of chemical problems, especially those that involve large molecular systems.[1−3] Despite its remarkable success as a moderately accurate method, DFT has several significant weaknesses. Current DFT implementations fail to provide a physically correct description of the dispersion forces[4] responsible for base pair stacking in DNA, dimer formation among rare gases, and various other long-range correlation effects.[5] Furthermore, DFT suffers from the well-documented self-interaction problem[6,7] leading to an overall underestimation of activation energies[8] and a general uncertainty in the vicinity of first-order saddle points on potential energy surfaces. Finally, the numerical quadrature techniques employed in DFT (as well as other methodologies that rely on numerical grids such as the class of pseudospectral

* Corresponding author e-mail: mhg@bastille.cchem.berkeley.edu, tel.: 1-510-642-5957.

10.1021/ct050126s CCC: $30.25
Published on Web 08/06/2005

methods) introduce numerical noise and, therefore, further uncertainty into the characterization of the fine details of potential energy surfaces.

During the same time, several attempts have been made at reducing the computational cost of second-order Møller−Plesset (MP2) perturbation theory[9] calculations. MP2 theory is the simplest and most economical wave-function-based improvement to the mean-field Hartree−Fock (HF) approximation that correctly accounts for the physics of dispersion interactions.[10] Because of simplicity in ansatz, MP2 theory is ideal as a starting point for developing methods and techniques that might be extended to calculations employing higher-level correlation theories, that is, those that involve either the coupled-cluster or configuration interaction formalisms or in "ab initio" approaches to improving the functional forms utilized in DFT. Additionally, equilibrium geometries obtained using MP2 theory are typically more reliable than those obtained using the popular DFT alternatives.[11] However, MP2 theory also has several computational limitations worth mentioning. If MP2 theory is formulated in a basis of canonical molecular orbitals (MOs), that is, the orthonormal set of eigenfunctions that diagonalize the Fock matrix, one finds an unphysical fifth-order computational dependence on the system size (M^5).[12] This relatively high computational cost (DFT formally scales as M^3) coupled with the need for large atomic orbital (AO) basis sets for reliable results further limits MP2 from the regime of large molecular systems.[13] The focus of this work is reducing this unnecessary computational cost and extending the current size limitations to practical MP2 calculations. We also note that the performance of MP2 theory when dealing with radicals and systems with small HOMO/LUMO energy gaps[14] has been somewhat disappointing when compared to its treatment of closed-shell systems.[15] However, most large molecular systems, especially those dealt with in a biochemical context, are closed-shell systems and should be unaffected by this drawback.

Several promising routes to decreasing the cost of MP2 computations have surfaced in the literature and can be divided into four distinct categories. The first approach involves a reduction in the computational prefactor but leaves the underlying fifth-order scaling untouched and includes the resolution-of-the-identity (RI) approximation,[16,17] pseudospectral methods,[18] and others.[19] The second class of methods reformulates MP2 theory in an AO basis and exploits the underlying locality with the Laplace transformation leading to linear scaling among one-dimensional systems using small basis sets.[20,21] The third approach directly attacks the aforementioned scaling by making use of localized MOs and, therefore, introduces fundamental changes by ansatz.[22−25] The fourth class of methods includes hybrid methods that have just recently emerged in the literature. One such method combines prefactor reduction (RI), the Laplace transformation, and other approximations such as separate scaling of the same-spin and opposite-spin components of the MP2 energy,[26,27] on the basis of the findings of Grimme.[28] Other hybrid methods are based on the numerous works of Schütz and co-workers;[29−32] of particular note is the incorporation of the density-fitting approach (synonymous in this context with the RI approximation) with the classic local model of Saebø and Pulay,[22] which has produced remarkably low-scaling methods for use in both single-point energy evaluations[33] and geometry optimizations.[34] In doing so, these novel hybrid MP2 methodologies have demonstrated significantly improved computational performance resulting from the combined reduction in the prefactor as well as a fundamental reduction in the underlying scaling.

The purpose of this paper is to report a new hybrid MP2 method and to thoroughly explore its chemical and computational performance. This new method combines prefactor reduction via incorporation of the RI approximation and a reduction in the underlying scaling by utilizing the local triatomics-in-molecules (TRIM) ansatz. Incorporation of the RI approximation into MP2 theory has demonstrated a marked increase in computational efficiency accompanied by tolerable error (~60 μH/atom).[17] Hence, the RI-MP2 methodology itself has extended the regime of molecular systems that can be investigated at the MP2 level of theory. By combining this powerful technique with a formal reduction in the MP2 correlation space via the TRIM ansatz, the resulting methodology, RI-TRIM MP2, emerges as a computationally efficient fourth-order method with the ability to treat even larger molecular systems. One powerful application of RI-TRIM MP2 would be its use in the design and fundamental parameterization of force fields and other semiempirical methods that can be used to more accurately treat large systems of biochemical and nanotechnological interest.

The remainder of this paper is outlined as follows. In Section II, the theory underlying the TRIM model will be briefly reviewed and the incorporation of the RI approximation into the working TRIM framework will be developed. A generalized algorithm for executing an RI-TRIM MP2 calculation will be presented in Section III, accompanied by a detailed analysis of the associated computational requirements. In Section IV, the numerical performance of the RI-TRIM MP2 model will be assessed against the MP2, TRIM MP2, RI-MP2, and DFT (B3LYP) methodologies in the context of determining the relative energies of 27 different alanine tetrapeptide conformations. Further assessment of this new methodology will follow from extrapolation of the untruncated RI-MP2 and RI-TRIM MP2 results to the complete basis set limit. The paper will then finish with some brief conclusions in Section V.

II. Theory

A. Localization of the Occupied and Virtual Subspaces. MP2 theory begins with a delineation of the occupied and virtual single-particle subspaces. When choosing functions to span these subspaces, a balance is sought between redundancy, overcompleteness, and ease of definition. Additionally, the functions must be atom-centered so that atomic truncations can be made that are smooth and globally defined. In this work, we follow the same procedure as that developed in the original TRIM paper[24] and earlier works.[35,36]

Given an AO basis consisting of N functions (O occupied and V virtual), the localized virtual functions will be defined as the set of AOs projected into the virtual subspace, $|\underline{a}_A\rangle =$

$\hat{Q}|\mu_A\rangle = (\hat{1} - \hat{P})|\mu_A\rangle$, following the original work of Saebø and Pulay.[22] In this representation, there is only a small fractional redundancy for calculations employing high-quality basis sets (where $V \approx N$). Since O does not depend on the basis set for a given molecular system, the analogous definition of the localized occupied functions would be tremendously overcomplete. The least overcomplete atomic representation that still exactly spans the occupied subspace is an orthonormal minimal basis of extracted polarized AOs (EPAOs).[37] Since the EPAOs are linear combinations of AOs, the localized occupied functions will be analogously defined as the set of EPAOs projected into the occupied subspace $|\underline{i}_A\rangle = \hat{P}|\alpha_A\rangle = \hat{P}(\sum_\mu \mathbf{B}_{\mu\alpha}|\mu\rangle)$, where $\mathbf{B}$ is the EPAO coefficient matrix.

B. Review of General MP2 Theory. In this section and the remainder of the paper, indices *a*, *b*, *c*, ... will refer to canonical virtual MOs and *i*, *j*, *k*, ... will refer to canonical occupied MOs. We will begin with the following general expression for the MP2 energy in terms of spin−orbitals

$$E_{\mathrm{MP2}} = \frac{1}{2}\sum_{iajb} t_{ij}^{ab}(ia|jb) \tag{1}$$

where t_{ij}^{ab} are the first-order amplitudes (i.e., $|\Psi^{(1)}\rangle = \sum_{iajb} t_{ij}^{ab}|\Psi_{ij}^{ab}\rangle$) given as the solution to the following set of linear equations:

$$\sum_{kcld} \Delta_{iajb;kcld} t_{kl}^{cd} = (ia||jb) \tag{2}$$

where $(ia||jb)$ are the antisymmeterized two-electron repulsion integrals (ERIs) in Mulliken notation, $(ia||jb) = (ia|jb) - (ib|ja)$, and Δ is the somewhat cumbersome energy-difference matrix. In an arbitrary nonorthogonal basis, Δ is an eighth-rank tensor describing the zeroth-order Hamiltonian and is defined as

$$\Delta_{iajb;i'a'j'b'} = F_{ii'}S_{aa'}S_{jj'}S_{bb'} - S_{ii'}F_{aa'}S_{jj'}S_{bb'} + S_{ii'}S_{aa'}F_{jj'}S_{bb'} - S_{ii'}S_{aa'}S_{jj'}F_{bb'} \tag{3}$$

where $F_{aa'}$ and $S_{aa'}$ are elements of the Fock and overlap matrices in the MO basis, respectively. If one is working in a basis of canonical MOs, Δ takes on the following simpler and more familiar form:

$$\Delta_{iajb;i'a'j'b'} = (\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b)\delta_{ii'}\delta_{aa'}\delta_{jj'}\delta_{bb'} \tag{4}$$

since both the Fock and overlap matrices are diagonal in this representation. Since Δ is diagonal in this representation, the formation of Δ^{-1} is trivial and the first-order amplitudes can easily be obtained via a straightforward evaluation of eq 2.

More generally, the expression for Δ in eq 3 can be simplified using the following matrix-based notation:

$$\Delta = \mathbf{F}_O \otimes \mathbf{S}_V \otimes \mathbf{S}_O \otimes \mathbf{S}_V - \mathbf{S}_O \otimes \mathbf{F}_V \otimes \mathbf{S}_O \otimes \mathbf{S}_V + \mathbf{S}_O \otimes \mathbf{S}_V \otimes \mathbf{F}_O \otimes \mathbf{S}_V - \mathbf{S}_O \otimes \mathbf{S}_V \otimes \mathbf{S}_O \otimes \mathbf{F}_V \tag{5}$$

where $\mathbf{F}_V$ and $\mathbf{F}_O$ ($\mathbf{S}_V$ and $\mathbf{S}_O$) are the MO representations of the Fock (overlap) matrix in the virtual and occupied subspaces, respectively. This expression can be further factorized into a compact direct-product form:

$$\Delta = \tilde{\mathbf{F}}^2 \otimes \tilde{\mathbf{S}}^2 + \tilde{\mathbf{S}}^2 \otimes \tilde{\mathbf{F}}^2 \tag{6}$$

using

$$\tilde{\mathbf{F}}^2 = \mathbf{F}_O\mathbf{S}_V - \mathbf{S}_O\mathbf{F}_V \tag{7}$$

and

$$\tilde{\mathbf{S}}^2 = \mathbf{S}_O\mathbf{S}_V. \tag{8}$$

From eqs 6−8, it is clear that Δ has inherent tensorial structure; its elements can be found as the direct product of two lower-ranked tensors in the mixed canonical occupied and virtual subspaces. Thus, the problem of inverting Δ (and the subsequent evaluation of eq 2 for the amplitudes) can be thought of as being equivalent to the diagonalization of these two lower-rank tensors, namely, $\tilde{\mathbf{F}}^2$ and $\tilde{\mathbf{S}}^2$.

If a local nonorthogonal basis is employed, Δ will not be diagonal and the aforementioned tensorial structure of Δ will be destroyed. Faced with the prohibitively large cost of inverting Δ, the first-order amplitudes must be obtained via an iterative solution of eq 2, as advocated by Saebø and Pulay.[22] However, the iterative solution of eq 2 comes with two fundamental complications. First is the need to store the amplitudes during the iterative procedure. This restriction makes awkward the extension of any local model to higher excitation levels, such as CCSD(T), where there may be an unmanageably large number of local triple amplitudes. Second, the use of redundant nonorthogonal functions followed by local truncations will undoubtedly lead to ill-conditioning in Δ and a myriad of associated convergence issues. Therefore, a viable noniterative solution to eq 2 is sought by retaining the tensorial structure of Δ and developing a methodology that finds a local representation in which it is diagonal.

C. Review of Atom-Centered Truncations. Local approximations to MP2 theory based on atomistic truncations can be characterized in terms of the subset of double substitutions retained in the local ansatz. For reference, the full MP2 correlation space, or the complete set of double substitutions, in an atom-centered basis can be represented in the following symbolic form:

$$\{iajb\}_{\mathrm{FULL}} = \{i\} \otimes \{a\} \otimes \{j\} \otimes \{b\}. \tag{9}$$

Substitutions generally occur from occupied levels centered on two atoms to virtual levels centered on two different atoms. Note that the full set of doubly excited determinants forms a fourth-rank tensor with direct-product structure—a form that enables the indices to be transformed one at a time during the integral transformation from the AO basis to the MO basis, as is done in most conventional MP2 algorithms.

The diatomics-in-molecules (DIM) model retains the following subset of double substitutions:

$$\{iajb\}_{\mathrm{DIM}} = \{(ia)\} \otimes \{(jb)\} \tag{10}$$

where an excitation (ia) occurs from an occupied level to a virtual level on the same atom—and will be referred to as

an atomic single substitution (ASIS). Overall direct-product structure is conserved in the DIM ansatz and is given as the product of the ASIS set with itself.

The TRIM model retains the following subset of double substitutions:

$$\{iajb\}_{\mathrm{TRIM}} = \{(ia)\} \otimes \{j\} \otimes \{b\} \oplus \{i\} \otimes \{a\} \otimes \{(jb)\}. \quad (11)$$

This model is an expansion of the DIM model by inclusion of single-electron transfers (or singly ionic substitutions). Double-electron transfers (or doubly ionic substitutions) are not included in the DIM or TRIM models—as they are often thought of as being the major constituent of basis set superposition error (BSSE).[38,39] Note that direct-product structure in this local ansatz is retained if we consider only half of the substitutions in eq 11. The resultant subset of double substitutions now constitutes the half-TRIM local ansatz.

If one chooses to employ models that retain direct-product structure in the ansatz, namely, the DIM and half-TRIM models described above, the tensorial structure of Δ is conserved and can be exploited during the inversion procedure. The goal now is to develop a formalism that finds a local representation in which Δ is diagonal and also incorporates the powerful RI approximation.

D. Review of the Noniterative Half-TRIM and TRIM Models. Starting with the half-TRIM local correlation ansatz (cf. eq 11)

$$\{iajb\}_{\mathrm{half-TRIM}} = \{(ia)\} \otimes \{j\} \otimes \{b\} \quad (12)$$

the direct-product structure of this model can be exploited using the recanonicalization scheme developed earlier.[24] Indexing the ASIS set by K, the general expression for the energy-difference matrix in eq 3 takes on the following form in the ASIS representation:

$$\Delta^{\mathrm{ASIS}}_{Kjb;K'j'b'} = \mathbf{D}^{\mathrm{ASIS}}_{KK'} S_{jj'} S_{bb'} + \mathbf{S}^{\mathrm{ASIS}}_{KK'} (F_{jj'} S_{bb'} - S_{jj'} F_{bb'}) \quad (13)$$

where $\mathbf{D}^{\mathrm{ASIS}}_{KK'}$ and $\mathbf{S}^{\mathrm{ASIS}}_{KK'}$, the MP2 energy matrix and overlap matrix in the ASIS representation, respectively, are given as

$$\mathbf{D}^{\mathrm{ASIS}}_{KK'} = (\mathbf{FS} - \mathbf{SF})_{KK'} = (\mathbf{B}^{\dagger}\mathbf{SPFPSB})_{ii'}(\mathbf{SQS})_{aa'} - (\mathbf{B}^{\dagger}\mathbf{SPSB})_{ii'}(\mathbf{SQFQS})_{aa'} \quad (14)$$

and

$$\mathbf{S}^{\mathrm{ASIS}}_{KK'} = (\mathbf{B}^{\dagger}\mathbf{SPSB})_{ii'}(\mathbf{SQS})_{aa'} \quad (15)$$

respectively. In eqs 14 and 15, **P**, **Q**, **F**, and **S** are matrix representations in the AO basis and **B** is the matrix defining the transformation from AOs to EPAOs. From here forward, we will use canonical MOs for j and b for computational simplicity. With this simplification, the energy-difference matrix in the ASIS representation in eq 13 takes on the following simpler form:

$$\Delta^{\mathrm{ASIS}}_{Kjb;K'j'b'} = \mathbf{D}^{\mathrm{ASIS}}_{KK'} \delta_{jj'} \delta_{bb'} + \mathbf{S}^{\mathrm{ASIS}}_{KK'} (\epsilon_j \delta_{bb'} - \delta_{jj'} \epsilon_b) \quad (16)$$

since both the Fock and overlap matrices are diagonal in the canonical representation. At this point, we have Δ as a sum of two separable terms, with a side length equal to the total number of ASIS pairs (N_K) given by

$$N_K = \sum_A N_A M_A \quad (17)$$

where N_A and M_A represent the number of local virtuals (atomic basis functions) and local occupieds (minimal basis functions or EPAOs) on atom A, respectively.

To further simplify this expression for Δ, we need to find a local representation in which both of the separable terms in eq 16 are diagonal. To do so, we proceed now in a manner analogous to the derivation of the canonical Hartree−Fock equations.[40] Symmetric orthogonalization (using $\mathbf{S}^{-1/2}_{\bar{K}\bar{K}}$) of the energy-difference matrix in the ASIS representation brings us into the orthogonalized ASIS, or OASIS, representation where Δ takes on the following form:

$$\Delta^{\mathrm{OASIS}}_{\bar{K}jb;\bar{K}'j'b'} = \mathbf{D}^{\mathrm{OASIS}}_{\bar{K}\bar{K}'} \delta_{jj'} \delta_{bb'} + \delta_{\bar{K}\bar{K}'} (\epsilon_j \delta_{bb'} - \delta_{jj'} \epsilon_b). \quad (18)$$

In eq 18, the second separable term (that involving the ASIS overlap matrix) has been brought into diagonal form and $\bar{K}$ has been introduced as the index of substitutions in the OASIS representation. Next, Δ is brought into the desired "canonical" OASIS representation (COASIS) via diagonalization of the MP2 energy matrix in the OASIS representation

$$\mathbf{D}^{\mathrm{OASIS}}_{\bar{K}\bar{K}'} = \sum_{\bar{P}} \mathbf{U}_{\bar{K}\bar{P}} \epsilon_{\bar{P}} \mathbf{U}_{\bar{P}\bar{K}'}. \quad (19)$$

This allows us to write Δ in its final form, in the mixed COASIS and canonical orbital representation

$$\Delta^{\mathrm{COASIS}}_{\bar{P}jb;\bar{P}'j'b'} = (\epsilon_{\bar{P}} + \epsilon_j - \epsilon_b) \delta_{\bar{P}\bar{P}'} \delta_{jj'} \delta_{bb'} \quad (20)$$

where $\bar{P}$ has been introduced as the index of substitutions in the COASIS representation. Now that Δ is diagonal, finding its inverse is trivial, allowing for a noniterative solution to eq 2 for the amplitudes. Knowledge of the amplitudes allows us to write down the following energy expression for the half-TRIM model (cf. eqs 1, 2, and 20):

$$E_{\mathrm{half-TRIM}} = -\frac{1}{2} \sum_{\bar{P}jb} \frac{\bar{V}_{\bar{P}jb} \bar{\bar{V}}_{\bar{P}jb}}{\epsilon_{\bar{P}} + \epsilon_j - \epsilon_b} \quad (21)$$

where $\bar{V}_{\bar{P}jb}$ and $\bar{\bar{V}}_{\bar{P}jb}$ are the appropriately transformed symmeterized and antisymmeterized ERIs in the COASIS representation [i.e., $\bar{V}_{\bar{P}jb} = (\bar{P}|jb)_C$ and $\bar{\bar{V}}_{\bar{P}jb} = (\bar{P}|jb)_C - (\bar{P}|jb)_E$] given as

$$V_{\bar{P}jb} = \sum_{\bar{K}K} \mathbf{U}_{\bar{P}\bar{K}} \mathbf{S}^{-1/2}_{\bar{K}K} V_{Kjb} = \sum_K \mathbf{T}_{\bar{P}K} V_{Kjb} \quad (22)$$

and **T** is a transformation matrix that brings quantities in the ASIS representation into the COASIS representation. (The subscripts C and E, above, refer to Coulomb and exchange integrals, respectively.)

Since the energy expression for the half-TRIM model includes a factor of $^1/_2$ for double-counting energy contributions and the half-TRIM ansatz retains only half of the TRIM subset of double substitutions, the TRIM energy should be double that of the half-TRIM energy. However, this overcounts the DIM contribution to the energy, so we have the

following expression as the working approximation to the final TRIM energy:

$$E_{\mathrm{TRIM}} = 2E_{\mathrm{half-TRIM}} - E_{\mathrm{DIM}}. \tag{23}$$

For completeness, the expression for the DIM energy in the COASIS representation is given as

$$E_{\mathrm{DIM}} = -\frac{1}{2}\sum_{\bar{P}\bar{Q}}\frac{\bar{V}_{\bar{P}\bar{Q}}\bar{\bar{V}}_{\bar{P}\bar{Q}}}{\epsilon_{\bar{P}} + \epsilon_{\bar{Q}}}. \tag{24}$$

At this point, we have developed a formalism that finds a local representation in which Δ is diagonal allowing for a noniterative solution of the MP2 energy; the only task that remains is the formal incorporation of the RI approximation into the theoretical framework of our local MP2 model.

E. Introduction of the RI Approximation. The RI approximation[17] involves the insertion of an approximate resolution-of-the-identity into the Hilbert space of two interacting charge densities, that is,

$$(ia|jb) = (ia|\hat{1}|jb) \approx \sum_{MN}(ia|M)(M|N)^{-1}(N|jb) \tag{25}$$

where the sum is over all members of a fitted auxiliary basis set (typically 2−4 times larger than its complementary AO basis). In other words, the use of the RI approximation involves replacement of four-centered ERIs with two- and three-centered ERIs. Under this approximation, the complete set of four-centered ERIs can now be constructed via a single matrix multiplication step:

$$(ia|jb) = \sum_{N}\mathbf{B}^{N}_{ia}\,\mathbf{B}^{N}_{jb} \tag{26}$$

where

$$\mathbf{B}^{N}_{ia} = \sum_{M}(ia|M)(M|N)^{-1/2} \tag{27}$$

is a modified coefficient matrix that includes the two- and three-centered ERIs in the auxiliary and mixed MO/auxiliary bases, respectively.

To incorporate the RI approximation into the TRIM model, all four-centered ERIs in the localized (DIM) or mixed canonical/localized (half-TRIM) bases must be expressed in terms of **B** matrices. In the expressions for $E_{\mathrm{half-TRIM}}$ (eq 21) and E_{DIM} (eq 24), one can identify four different types of four-centered ERIs, namely, $(\bar{P}|jb)_C$ and $(\bar{P}|jb)_E$ from the half-TRIM model and $(\bar{P}|\bar{Q})_C$ and $(\bar{P}|\bar{Q})_E$ from the DIM model. In the ASIS representation, these ERIs are given in more explicit detail as

$$(K|jb)_C = (\underline{i'a'}|jb), \tag{28}$$

$$(K|jb)_E = (\underline{i'}b|j\underline{a'}), \tag{29}$$

$$(K|L)_C = (\underline{i'a'}|\underline{j''b''}), \tag{30}$$

and

$$(K|L)_E = (\underline{i'b''}|\underline{j''a'}), \tag{31}$$

where underlined orbitals with the same number of primes are localized orbitals centered on the same atom (orbitals without primes and underlines are canonical MOs). To form these integrals within the RI approximation, the prescriptions given in eqs 26 and 27 can be used on eqs 28−31 to determine the corresponding set of **B** matrices that must first be constructed:

$$(K|jb)_C = (\underline{i'a'}|jb) = \sum_{N}\mathbf{B}^{N}_{\underline{i'a'}}\mathbf{B}^{N}_{jb} = \sum_{N}\mathbf{B}^{N}_{K}\mathbf{B}^{N}_{jb}, \tag{32}$$

$$(K|jb)_E = (\underline{i'}b|j\underline{a'}) = \sum_{N}\mathbf{B}^{N}_{\underline{i'}b}\mathbf{B}^{N}_{j\underline{a'}}, \tag{33}$$

$$(K|L)_C = (\underline{i'a'}|\underline{j''b''}) = \sum_{N}\mathbf{B}^{N}_{\underline{i'a'}}\mathbf{B}^{N}_{\underline{j''b''}} = \sum_{N}\mathbf{B}^{N}_{K}\mathbf{B}^{N}_{L}, \tag{34}$$

and

$$(K|L)_E = (\underline{i'b''}|\underline{j''a'}) = \sum_{N}\mathbf{B}^{N}_{\underline{i'b''}}\mathbf{B}^{N}_{\underline{j''a'}}. \tag{35}$$

Using eq 22, the integrals given in eqs 32−35 can be transformed into the COASIS representation, and the energy can then be directly evaluated using eqs 21, 23, and 24. Therefore, the central theme in the evaluation of the RI-TRIM MP2 energy is the computationally efficient formation and subsequent manipulation of the unique set of **B** matrices given by eqs 32−35. Now that we have formally incorporated the RI approximation into the theoretical framework of our local MP2 model, we now describe in some detail the computational algorithm that constructs and manipulates this set of **B** matrices necessary to evaluate E_{TRIM} in the next section.

III. Algorithm

An overview of the RI-TRIM MP2 algorithm is given in Chart 1. In each of the following subsections, the individual functions in Chart 1 will be discussed in greater detail.

A. Function 00: Self-Consistent Field Procedure and EPAO Algorithm. In Function 00, the standard self-consistent field (SCF) procedure is executed, generating the following set of matrices necessary during the post Hartree−Fock computation: the overlap (**S**), Fock (**F**), and density (**P**) matrices in the AO basis and the coefficient (**C**) and orbital energy (ϵ) matrices in the MO basis. Additionally, a separate algorithm that generates the EPAO coefficient (**B**) matrix is called immediately following the generation of a SCF. The computational cost associated with the SCF procedure formally scales cubically with system size, and the associated memory and storage requirements are minimal in the context of our RI-TRIM MP2 algorithm.

B. Functions 01−02: RI Overhead Routines. Function 01 is essentially one of two mandatory overhead routines for computations involving the RI approximation. In Function 01, the two-centered ERIs, $(M|N)$ are assembled using the Coulomb metric in the auxiliary basis. A singular value decomposition is then performed on $(M|N)$ to obtain the desired inverse square-root matrix necessary for constructing the various **B** matrices described earlier. The computational cost of forming $(M|N)^{-1/2}$ is cubic with respect to system size ($\sim X^3$ where X is the number of auxiliary basis functions) and is attributed entirely to the diagonalization of $(M|N)$.

Chart 1. A General Flowchart of the RI-TRIM MP2 Algorithm

Function 00 SCF Procedure; Formation of Overlap (S), MO Coefficient (C), Fock (F), Density (P), Orbital Energy (ε), and EPAO Coefficient (B) Matrices

Function 01 RI- Overhead (Part I): Formation of the $(M\,|\,N)^{-1/2}$ Matrix

Function 02 RI- Overhead (Part II): Formation of the $(ia\,|\,M)$ Matrix

Function 03 TRIM Overhead: Formation of V, O, and T Transformation Matrices and COASIS Energy $(\varepsilon_{\bar{P}})$ Matrix

Function 04 Formation of B Matrices (Part I)

Function 05 Formation of B Matrices (Part II)

Function 06 Construction of the DIM and TRIM Exchange Integrals and the B_K^N Matrix

Function 07 Canonicalization of the DIM Exchange Integrals and the B_K^N Matrix
Construction of the DIM Coulomb Integrals
Evaluation of the DIM Energy

Function 08 Canonicalization of the TRIM Exchange Integrals
Construction of the TRIM Coulomb Integrals
Evaluation of the TRIM Energy

The memory requirements associated with the diagonalization of $(M|N)$, namely, $3X^2$, also sets the minimum overall memory requirements of any algorithm that employs the RI approximation. In our RI-TRIM MP2 algorithm, we have decided on $5X^2$ as the overall memory requirement, which allows for treatment of system sizes on the order of a C_{90} alkane chain at the cc-pVDZ level with 2 GB of available memory. For future use, the $(M|N)^{-1/2}$ matrix is written to disk with quadratic storage requirements (X^2).

In Function 02, the remainder of the overhead associated with the RI approximation is completed as the three-centered ERIs, $(ia|M)$, are constructed in the mixed MO/auxiliary basis. This set of three-centered ERIs is formed in a two-step transformation procedure from the preassembled set of three-centered ERIs in the mixed AO/auxiliary basis $(\mu\nu|M)$. In the first transformation step, $\mu \longmapsto i$ via

$$(i\nu|M) = \sum_{\mu} C_{\mu i}(\mu\nu|M) \tag{36}$$

with an associated quartic computational cost of $\sim O(nb2)X$, where $(nb2)$ is the number of significant AO basis set pairs (determined by the chosen integral threshold value). In the second transformation step, $\nu \longmapsto a$ via

$$(ia|M) = \sum_{\nu} C_{\nu a}(i\nu|M) \tag{37}$$

with an associated quartic computational cost of $\sim NOVX$. The set of $(ia|M)$ is also written to disk with cubic storage requirements (OVX) for use in later functions.

C. Function 03: TRIM Overhead Routine. The use of localized MOs in MP2 theory via utilization of the TRIM ansatz also comes with an associated overhead, all of which is managed in Function 03. In this function, the MP2 energy $(\mathbf{D}_{KK'}^{\mathrm{ASIS}})$ and overlap $(\mathbf{S}_{KK'}^{\mathrm{ASIS}})$ matrices in the ASIS representation are constructed according to the prescriptions given by eqs 14 and 15, respectively. A singular value decomposition is then performed on $\mathbf{S}_{KK'}^{\mathrm{ASIS}}$ to obtain the desired inverse square-root matrix necessary for symmetric orthogonalization of the ASIS representation. The computational cost of forming $\mathbf{S}_{\bar{K}\bar{K}}^{-1/2}$ is cubic with respect to system size $(\sim N_K^3)$, where N_K is the number of ASIS pairs given by eq 17, a quantity that grows linearly with system size. In practice, $N_K \approx X$, so the computational cost of diagonalizing $\mathbf{S}_{KK'}^{\mathrm{ASIS}}$ can directly be compared to that of diagonalizing $(M|N)$ in Function 01. After $\mathbf{D}_{KK'}^{\mathrm{ASIS}}$ is brought into the OASIS representation $(\mathbf{D}_{\bar{K}\bar{K}'}^{\mathrm{OASIS}})$ by both left and right matrix multiplication with $\mathbf{S}_{\bar{K}\bar{K}}^{-1/2}$, it is brought into diagonal form by another diagonalization step (total computational cost of $3N_K^3$ steps). The eigenvector matrix $\mathbf{U}_{\bar{P}\bar{K}}$ which diagonalizes $\mathbf{D}_{\bar{K}\bar{K}'}^{\mathrm{OASIS}}$ is multiplied on the right by $\mathbf{S}_{\bar{K}\bar{K}}^{-1/2}$ to form the desired $\mathbf{T}$ transformation matrix, which brings quantities in the ASIS representation into the COASIS representation, with a cubic computational cost of N_K^3. The corresponding eigenvalue matrix $\epsilon_{\bar{P}}$ provides us with the energies associated with each substitution in the COASIS representation and is used directly in the energy evaluation. Both of these quantities, $\mathbf{T}$ and $\epsilon_{\bar{P}}$, are written to disk for future use with an associated quadratic storage requirement $(\sim N_K^2)$. In addition, two other transformation matrices are necessary for computations involving localized MOs: the $\mathbf{V}$ and $\mathbf{O}$ matrices, which transform canonical MOs to localized virtual and occupied orbitals, respectively. From our definitions of localized virtual and occupied functions given in Part A of Section II, it follows that

$$\mathbf{V} = C_V^{\dagger}\mathbf{S} \tag{38}$$

and

$$\mathbf{O} = C_O^{\dagger}\mathbf{SB} \tag{39}$$

Chart 2. A More Detailed Look at the Algorithm in Function 04. Array and Disk (HD) Ordering Is Given with Index Speeds Decreasing from Left to Right within Square Brackets

Load $(M|N)^{-1/2}$, V, and O matrices from disk

Loop over occupied batches, ob

Loop over $i \in ob$

Load $(ia\|M)\ \forall\ a, M, given\ i$ from disk	$[a, M]$
Make $B_{ia}^{N} = \sum_M (ia\|M)(M\|N)^{-1/2}\ \forall\ a, N, given\ i$	$[a, N]$
Make $\tilde{B}_{ia}^{N} = (B_{ia}^{N})^{\dagger}\ \forall\ a, N, given\ i$	$[N, a]$
Accumulate $B_{ia}^{N}\ \forall\ a, N, i \in ob$	$[a, N, i \in ob]$
Accumulate $\tilde{B}_{ia}^{N}\ \forall\ a, N, i \in ob$	$[N, a, i \in ob]$

End Loop over $i \in ob$

Make $B_{i\underline{a}}^{N} = \sum_a B_{ia}^{N} V_{a\underline{a}}\ \forall\ \underline{a}, N, i \in ob$	$[N, i \in ob, \underline{a}]$
Write $B_{ia}^{N}\ \forall\ a, N, i \in ob$ to disk	HD: $[a, N, i]$

Loop over atoms, A

Extract $B_{i\underline{a}}^{N}\ \forall\ \underline{a} \in A, N, i \in ob$	$[N, i \in ob, \underline{a} \in A]$
Write $B_{i\underline{a}}^{N}\ \forall\ \underline{a} \in A, N, i \in ob$ to disk	HD: $[N, i \in ob, \underline{a} \in A, ob, A]$

End Loop over atoms, A

Loop over virtual batches, vb

Extract $\tilde{B}_{ia}^{N}\ \forall\ a \in vb, N, i \in ob$	$[N, a \in vb, i \in ob]$
Write $\tilde{B}_{ia}^{N}\ \forall\ a \in vb, N, i \in ob$ to disk	HD: $[N, a \in vb, i, vb]$

End Loop over virtual batches, vb

End Loop over occupied batches, ob

where C_V and C_O are the virtual and occupied blocks of the MO coefficient matrix, **S** is the overlap matrix in the AO basis, and **B** is the EPAO coefficient matrix. These quantities are constructed with an associated cubic computational cost ($\sim N^2V$ for **V** and $\sim N^2O + MN^2$ for **O**) and are written to disk with quadratic storage requirements for later use in the construction of the modified RI coefficient matrices.

D. Functions 04−06: Construction of B Matrices and Exchange Integrals. Function 04 is the first of three functions that assembles and stores the various **B** matrices necessary for constructing the four-centered ERIs. The manner in which we proceed uses a fairly involved batching scheme to carefully balance CPU and I/O timings under the working memory constraint of $5X^2$, cubic I/O and disk requirements, and quartic computational cost. Without this batching scheme, our algorithm was susceptible to dominance by I/O time stemming from an unmanageable number of hard drive seeks. Given that the $(ia|M)$ matrix is written to disk with the overall order $[a, M, i]$, that is, a is the fastest index and i is the slowest index, we suggest the algorithm depicted in Chart 2.

The batch sizes involved in the loops over occupied and virtual batches are determined by the overall memory constraint of $5X^2$—a quantity unique to the molecular system being considered. There are two matrix multiplication steps in Function 04: the formation of the canonical $\mathbf{B}_{ia}^{N}$ matrix and the mixed canonical occupied/local virtual $\mathbf{B}_{i\underline{a}}^{N}$ matrix with associated computational costs of $\sim OVX^2$ and $\sim OVNX$, respectively. As shown in eqs 32 and 33, these matrices will be directly used in the construction of the TRIM Coulomb and TRIM exchange integrals. In the next function, the transposed canonical **B** matrix, $\tilde{\mathbf{B}}_{ia}^{N}$, now in the proper order on disk, will be transformed into the mixed local occupied/canonical virtual $\mathbf{B}_{i\underline{a}}^{N}$ matrix via matrix multiplication with **O**. In Function 04, there are associated cubic storage requirements ($2OVX$ and ONX) for writing both canonical **B** matrices and the mixed canonical occupied/local virtual **B** matrix to disk.

In Function 05, another **B** matrix is constructed, namely, the mixed local occupied/canonical virtual $\mathbf{B}_{i\underline{a}}^{N}$ matrix, from the transposed canonical $\tilde{\mathbf{B}}_{ia}^{N}$ matrix formed and written to disk earlier in Function 04 (Chart 3). This transformation step involves a single matrix multiplication with an associated quartic computational cost of $\sim OVMX$. As was done previously in Function 04 with $\mathbf{B}_{ia}^{N}$, the localized index in $\mathbf{B}_{i\underline{a}}^{N}$ is also grouped by atom and written selectively to disk with cubic storage requirements (MVX). This atomistic grouping will allow us to easily construct the four-centered ERIs given in eqs 28−31, where atom-based selection is crucial. From eq 33, it is clear that the $\mathbf{B}_{i\underline{a}}^{N}$ matrix will be directly used in the construction of the TRIM exchange integrals.

Function 06 is the final routine that deals with assembling the unique set of **B** matrices necessary to evaluate E_{TRIM}. In this function (Chart 4), our previous array reordering and selective disk writing is exploited as we construct two

Chart 3. A More Detailed Look at the Algorithm in Function 05

Loop over virtual batches, vb	
Load $\widetilde{B}_{ia}^{N}\ \forall\ a \in vb, N, i$ from disk	$[N, a \in vb, i]$
Loop over $a \in vb$	
Extract $\widetilde{B}_{ia}^{N}\ \forall\ N, i, given\ a$	$[N, i]$
Make $B_{i\underline{a}}^{N} = \sum_i \widetilde{B}_{ia}^{N} O_{i\underline{i}}\ \forall\ \underline{i}, N, given\ a$	$[N, \underline{i}]$
Accumulate $B_{i\underline{a}}^{N}\ \forall\ N, \underline{i}, a \in vb$	$[N, \underline{i}, a \in vb]$
End Loop over $a \in vb$	
Loop over atoms, A	
Extract $B_{i\underline{a}}^{N}\ \forall\ N, \underline{i} \in A, a \in vb$	$[N, \underline{i} \in A, a \in vb]$
Write $B_{i\underline{a}}^{N}\ \forall\ N, \underline{i} \in A, a \in vb$ to disk	HD: $[N, \underline{i} \in A, a, A]$
End Loop over atoms, A	
End Loop over virtual batches, vb	

differently ordered local $\mathbf{B}_{ia}^{N}$ matrices in a fairly complex loop over atoms. These $\mathbf{B}_{i\underline{a}}^{N}$ matrices are not only used to directly assemble the DIM exchange integrals via eq 35 but also contain the $\mathbf{B}_{K}^{N}$ matrix necessary for the construction of the TRIM and DIM Coulomb integrals (cf. eqs 32 and 34). This routine contains four matrix multiplication steps and their associated computational costs: the formation of $\mathbf{B}_{i\underline{a}}^{N}$ from both $\mathbf{B}_{i\underline{a}}^{N}$ ($\sim MNVX$) and $\mathbf{B}_{ia}^{N}$ ($\sim MNOX$) as well as the construction of both the TRIM ($\sim N_K OVX$) and DIM ($\sim N_K^2 X$) exchange integrals. For later retrieval, the $\mathbf{B}_{K}^{N}$ matrix and the set of TRIM exchange integrals, $(K|jb)_E$, are written to disk with associated $N_K X$ and $N_K OV$ storage requirements, respectively. Note that the recently assembled set of DIM exchange integrals is held in memory as it will be used next in Function 07 in the evaluation of E_{DIM}.

E. Functions 07−08: Construction of Coulomb Integrals, Canonicalization, and Energy Evaluation. In Function 07, the transformation matrix, **T**, is read from the disk and used to bring both the DIM exchange integrals and the $\mathbf{B}_{K}^{N}$ matrix from the ASIS representation into the COASIS representation via the prescription given in eq 22. The computational cost associated with bringing the DIM exchange integrals into the COASIS representation is cubic ($2N_K^3$) and is attributed to the left- and right-hand matrix multiplication of $(K|L)_E$ with **T**. We can circumvent multiple *ASIS* $|\rightarrow$ *COASIS* integral transformations by direct canonicalization of the $\mathbf{B}_{K}^{N}$ matrix, which can be directly used in the assembly of the DIM and TRIM Coulomb integrals in the COASIS representation. This transformation (i.e., $\mathbf{B}_{K}^{N}$ $|\rightarrow \mathbf{B}_{\mathrm{P}}^{N}$) has a cubic computational cost of $\sim N_K^2 X$, and $\mathbf{B}_{P}^{N}$ is easily written to disk with a minimal quadratic storage requirement ($N_K X$). The $\mathbf{B}_{P}^{N}$ matrix is also used to construct the DIM Coulomb integrals via eq 34 with cubic computational effort ($\sim N_K^2 X$). With both the DIM Coulomb and exchange integrals in memory, E_{DIM} is evaluated with negligible computational cost using eq 24.

The RI-TRIM MP2 algorithm is then completed in Function 08, as $E_{\mathrm{half-TRIM}}$ is incremented in a loop over occupied batches in which the TRIM exchange integrals are read from disk and brought into the mixed COASIS and canonical representation as the TRIM Coulomb integrals are assembled via matrix multiplication of $\mathbf{B}_{P}^{N}$ and $\mathbf{B}_{ia}^{N}$. The computational costs associated with these steps are both quartic, namely, $\sim N_K^2 OV$ for the canonicalization of the TRIM exchange integrals and $\sim N_K OVX$ for the construction of the TRIM Coulomb integrals.

F. Summary of Computational Requirements. As mentioned earlier, the RI-TRIM MP2 algorithm was written under the following constraints: quadratic memory not to exceed $5X^2$, cubic I/O and storage requirements, and quartic computational cost. Within the eight functions described above, there are nine steps with quartic-associated computational cost and six different quantities are written to disk with cubic storage requirements. The dominant computational steps are attributed to the formation of the canonical $\mathbf{B}_{ia}^{N}$ matrix ($\sim OVX^2$) in Function 04, and the formation of the TRIM exchange and TRIM Coulomb integrals [both ($\sim N_K OVX$)] in Function 06 and Function 08, respectively. Overall savings with respect to RI-MP2, which scales asymptotically with the fifth-power of the system size [the quintic computational cost is $\sim O^2V^2X$ attributed to the construction of the full set of $(ia|jb)$], should be proportional to the number of atoms in the systems under investigation. By introducing the local TRIM ansatz into RI-MP2 theory, we have essentially replaced the aforementioned quintic computational step with six quartic computational steps necessary to build the set of **B** matrices and integrals necessary for the evaluation of the RI-TRIM MP2 energy.

IV. Performance of the Model

One interesting challenge that remains in modern theoretical chemistry is the accurate prediction of the relative energies of a series of polypeptide conformations. More often than not, relative energies among polypeptide conformations are on the order of 1 kcal/mol (or less) and are highly dependent on correlation energy corrections arising from post-HF calculations.[41] In this section, the numerical performance and computational timings of the RI-TRIM MP2 methodology will be assessed against MP2, TRIM MP2, RI-MP2, and DFT (B3LYP) in the prediction of the relative orderings of 27 different alanine tetrapeptide conformations. What one finds

Chart 4. A More Detailed Look at the Algorithm in Function 06

Loop over atoms, A

Load $B_{\underline{i}b}^{N}\ \forall\ N,\underline{i}\in A,b$ from disk $[N,\underline{i}\in A,b]$

Loop over $\underline{i}\in A$

Extract $B_{\underline{i}b}^{N}\ \forall\ N,b,given\ \underline{i}$ $[N,b]$

Make $B_{\underline{i}\underline{b}}^{N}=\sum_{b}B_{\underline{i}b}^{N}V_{b\underline{b}}\ \forall\ N,\underline{b},\underline{i}\in A$ $[N,\underline{b}]$

Accumulate $B_{\underline{i}b}^{N}\ \forall\ N,b,\underline{i}\in A$ $[N,b,\underline{i}\in A]$

Accumulate $B_{\underline{i}\underline{b}}^{N}\ \forall\ N,\underline{b},\underline{i}\in A$ $[N,\underline{b},\underline{i}\in A]$

End Loop over $\underline{i}\in A$

Load $B_{j\underline{a}}^{N}\ \forall\ N,\underline{a}\in A,j$ from disk $[N,j\in ob,\underline{a}\in A,ob]$

Loop over $\underline{a}\in A$

Extract $B_{j\underline{a}}^{N}\ \forall\ N,j,given\ \underline{a}$ $[N,j]$

Make $B_{\underline{j}\underline{a}}^{N}=\sum_{a}B_{j\underline{a}}^{N}O_{j\underline{j}}\ \forall\ N,\underline{j},\underline{a}\in A$ $[N,\underline{j}]$

Accumulate $B_{j\underline{a}}^{N}\ \forall\ N,j,\underline{a}\in A$ $[N,j,\underline{a}\in A]$

Accumulate $B_{\underline{j}\underline{a}}^{N}\ \forall\ N,\underline{j},\underline{a}\in A$ $[N,\underline{j},\underline{a}\in A]$

End Loop over $\underline{a}\in A$

Extract $B_{K}^{N}\ \forall\ N,K\in A$ $[N,P\in A]$

Write $B_{K}^{N}\ \forall\ N,K\in A$ to disk HD: $[N,P]$

Make $(K\mid jb)_{E}=(\underline{i'}b\mid j\underline{a'})=\sum_{N}B_{\underline{i'}b}^{N}B_{j\underline{a'}}^{N}\ \forall\ \underline{i}\in A,\underline{a}\in A,b,j$ $[b,\underline{i}\in A,j,\underline{a}\in A]$

Write $(K\mid jb)_{E}\ \forall\ K\in A,b,j$ to disk in occupied batches HD: $[b,P\in A,j\in ob,A,ob]$

Loop over atoms, B

Extract $B_{\underline{j}\underline{a}}^{N}\ \forall\ N,\underline{j}\in B,\underline{a}\in A$ $[N,\underline{j}\in B,\underline{a}\in A]$

Extract $B_{\underline{i}\underline{b}}^{N}\ \forall\ N,\underline{b}\in B,\underline{i}\in A$ $[N,\underline{b}\in B,\underline{i}\in A]$

Make $(K\mid L)_{E}=(\underline{i'b''}\mid\underline{j''a'})=\sum_{N}B_{\underline{i'b''}}^{N}B_{\underline{j''a'}}^{N}\ \forall\ \underline{i}\,\&\,\underline{a}\in A,\underline{j}\,\&\,\underline{b}\in B$

$[\underline{j}\in B,\underline{a}\in A,\underline{b}\in B,\underline{i}\in A]$

End Loop over atoms, B

End Loop over atoms, A

is that the aforementioned challenge is directly addressed by RI-TRIM MP2, a method that not only accurately provides the relative energies of these conformations but does so with a computational timing cost that is negligible with respect to the iterative procedures necessary in both the SCF and DFT formalisms.

All calculations discussed in this work were performed on a 2 GHz Apple XServe with 2 GB of available memory and a triply striped 7200 rpm hard drive with 100 GB of available disk space. The RI-TRIM MP2 code was implemented into a development version of Q-Chem 2.0[42] and was used for all calculations. The geometries of the alanine tetrapeptide conformations were optimized at the HF/6-31G** level of theory[41] and are provided in the Supporting Information. In all post-HF computations performed, the frozen-core approximation was introduced to reduce computational timings and resources. In all cases, self-consistency was determined by DIIS errors below 10^{-8} au and the numerical threshold utilized in all integral evaluations was set at 10^{-12} au. All calculations were performed utilizing the correlation consistent basis sets (cc-pVXZ where X = D, T, and Q) of Dunning[43] and their respective complementary auxiliary basis sets provided by Weigend et al.[17,44]

A. Chemical Tests. For a series of 27 alanine tetrapeptide conformations, single-point energy calculations were first performed using the MP2, TRIM MP2, RI-MP2, RI-TRIM MP2, and DFT methodologies and the cc-pVDZ and cc-pVTZ basis sets. From this set of calculations, relative conformational energies were computed and are presented in Table 1.

A fairly simple statistical analysis of this set of data (Table 2) yields several quantities worth further discussion. A direct comparison of RI-MP2 with MP2 yields the relative error introduced by the RI approximation. The root-mean-square (RMS) error using both the cc-pVDZ and cc-pVTZ basis sets was computed as 0.003 kcal/mol and is completely negligible with respect to the relative conformational energies given in Table 1. By comparing RI-TRIM MP2 to TRIM

Table 1. Relative Energies (in kcal/mol) for a Series of 27 Different Alanine Tetrapeptide Conformations Using the cc-pVDZ (first column) and cc-pVTZ (second column) Basis Sets (and Their Respective Complementary Auxiliary Basis Sets where Applicable)[a]

conformation	MP2		RI-MP2		TRIM MP2		RI-TRIM MP2		DFT	
1	6.635	5.075	6.637	5.080	6.203	4.722	6.202	4.723	3.175	0.994
2	6.045	4.927	6.047	4.931	5.684	4.614	5.684	4.615	2.904	1.230
3	0.293	0.511	0.294	0.512	0.235	0.449	0.237	0.449	0.000	0.681
4	7.311	6.345	7.313	6.349	6.953	6.025	6.952	6.026	5.126	3.716
5	7.543	5.989	7.543	5.993	7.215	5.717	7.211	5.718	5.441	3.398
6	2.994	2.960	2.994	2.962	2.859	2.839	2.858	2.839	2.681	2.756
7	6.575	6.680	6.573	6.682	6.471	6.570	6.469	6.569	7.445	7.490
8	4.802	4.466	4.798	4.467	4.742	4.461	4.739	4.461	5.852	5.935
9	9.282	8.153	9.288	8.156	9.058	8.040	9.063	8.041	6.756	6.119
10	7.860	7.781	7.859	7.781	7.721	7.726	7.721	7.725	7.716	8.061
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.016	1.626
12	1.158	0.467	1.155	0.466	1.188	0.508	1.186	0.507	2.878	2.788
13	4.706	4.122	4.710	4.125	4.441	3.880	4.442	3.881	2.249	1.210
14	5.983	5.116	5.980	5.118	5.766	4.934	5.761	4.933	4.962	3.716
15	1.708	1.802	1.701	1.800	1.738	1.879	1.733	1.877	4.075	4.641
16	4.701	3.669	4.699	3.670	4.590	3.589	4.588	3.589	4.566	3.695
17	2.801	3.236	2.800	3.236	2.753	3.185	2.753	3.185	2.728	4.007
18	2.182	2.115	2.187	2.118	1.995	1.948	1.998	1.949	0.053	0.000
19	4.630	4.084	4.636	4.087	4.428	3.933	4.432	3.935	2.482	1.985
20	2.785	2.118	2.784	2.120	2.639	2.003	2.637	2.004	2.011	1.188
21	2.724	2.824	2.722	2.827	2.608	2.758	2.605	2.759	2.771	3.041
22	5.535	5.791	5.532	5.791	5.391	5.696	5.389	5.695	5.396	5.893
23	5.972	5.867	5.972	5.868	5.840	5.775	5.839	5.774	5.823	6.006
24	4.012	3.952	4.011	3.953	3.895	3.863	3.894	3.862	4.088	4.148
25	3.578	2.894	3.579	2.897	3.435	2.779	3.434	2.780	3.032	2.431
26	2.046	1.103	2.044	1.105	1.988	1.064	1.984	1.064	2.592	1.779
27	5.621	4.416	5.622	4.419	5.477	4.336	5.476	4.336	5.228	4.079

[a] The number of AO (auxiliary) basis functions employed in the cc-pVDZ and cc-pVTZ calculations was 390 (1428) and 908 (2260), respectively.

Table 2. RMS (Maximum) Errors among Various Different Methodologies (in kcal/mol) Using the cc-pVDZ and cc-pVTZ Basis Sets

	cc-pVDZ	cc-pVTZ
RI-MP2 vs MP2	0.003 (0.007)	0.003 (0.005)
RI-TRIM vs TRIM MP2	0.003 (0.006)	0.001 (0.002)
TRIM MP2 vs MP2	0.192 (0.432)	0.157 (0.353)
RI-TRIM MP2 vs RI-MP2	0.193 (0.435)	0.159 (0.357)
DFT vs RI-MP2	1.574 (3.463)	1.816 (4.086)
DFT vs RI-TRIM MP2	1.420 (3.027)	1.688 (3.729)

MP2, we are now in a position to numerically assess the magnitude of the error introduced by the RI approximation within the TRIM framework. From Table 2, it is clear that RI-TRIM MP2 does so with exceptional consistency with computed RMS errors of 0.003 kcal/mol (cc-pVDZ) and 0.001 kcal/mol (cc-pVTZ). On the basis of the relative magnitude of this error, it is safe to conclude that, within the context of determining the relative energies of these polypeptides, the data obtained from calculations utilizing the RI approximation are indistinguishable from those obtained without using the RI approximation. This allows us the powerful option of pursuing the relative energies of this set of polypeptides at the complete basis set limit by utilizing the higher-quality cc-pVQZ basis set. At the cc-pVQZ level, TRIM MP2 and MP2 calculations on systems of this size are prohibitively expensive and can now be replaced by their computationally efficient RI analogues with no associated loss in accuracy.

Comparing the TRIM MP2 and MP2 data sets allows for quantification of the local error, that is, the error attributed to truncation of the MP2 correlation space introduced by the TRIM model. The RMS local error was computed as 0.192 and 0.157 kcal/mol at the cc-pVDZ and cc-pVTZ levels, respectively. The analogous comparison of RI-TRIM MP2 to RI-MP2 yields a consistent measure of the local error, namely, 0.193 kcal/mol (cc-pVDZ) and 0.159 kcal/mol (cc-pVTZ). Although the magnitude of the local error decreased as the quality of the basis set was increased, the local error is significant enough to introduce relative ordering shifts among close-lying conformers when the cc-pVDZ and cc-pVTZ basis sets are employed. This finding further justifies the need to explore the relative energetics of this system of conformers at the now possible cc-pVQZ level of theory. We would also like to point out that the incorporation of the RI approximation into the theoretical framework of TRIM MP2 does not introduce any additional error; that is, there are no additive or compound errors arising from the simultaneous use of these approximations.

Another interesting quantity of note is the assessment of the popular B3LYP[45,46] functional against the MP2 methods examined in this work. The role of dispersion interactions in the relative energies of polypeptide conformations has been found to be significant in a recent work by Yu et al.[47] In

fact, the electron correlation effects obtained at the MP2 level were found to be critical in determining the binding energies of dipeptide conformations; in treating alanine dipeptide, for example, DFT is unable to predict the formation of the very stable peptide bond.[48] In the case of alanine and glycine dipeptides, it was also shown that the inclusion of higher-order correlation effects [i.e., from CCSD(T)] was not significant with respect to the MP2 results.[48] On the basis of these findings, we feel justified in using the present MP2 results as the benchmark for comparison in this work. We, therefore, conclude that the RMS error computed by comparing DFT to RI-MP2 at the cc-pVDZ level, 1.574 kcal/mol, is completely unacceptable for accurately predicting the relative conformational energies within a biochemical context. In fact, this error increases to 1.816 kcal/mol using the higher-quality cc-pVTZ basis set! This trend is consistently reproduced in the comparison of DFT to RI-TRIM MP2, where the RMS errors were found as 1.420 and 1.688 kcal/mol, at the cc-pVDZ and cc-pVTZ levels, respectively. To put this error analysis into a more global perspective, our findings clearly show that the RMS DFT error is at least 7 times larger than the error introduced by using localized MOs within the MP2 framework.

At the cc-pVQZ level, single-point energy calculations were performed using the RI-MP2, RI-TRIM MP2, and DFT methodologies for the same series of 27 alanine tetrapeptide conformations. From this set of calculations, relative conformational energies were computed and are presented in Table 3. The RMS (max) local error at this level was computed as 0.090 (0.195) kcal/mol; again we observe a substantial decrease in the magnitude of this error with the use of a higher-quality basis set. Although the magnitude of this error is still large enough to occasionally introduce discrepancies in relative orderings (at the cc-pVQZ level, RI-MP2 and RI-TRIM MP2 only disagree once with respect to relative ordering among these 27 different conformations), we emphasize the achievement of error that is less than 0.1 kcal/mol. The RMS (max) DFT errors with respect to RI-MP2 and RI-TRIM MP2 were found as 1.733 (3.212) and 1.674 (3.168) kcal/mol, respectively. Even at the cc-pVQZ level, DFT fails to perform at an acceptable level for accurately treating the relative energetics within this system of polypeptides conformers and will no longer be discussed. These findings provide another instance where DFT fails to converge as rapidly as expected with the use of high-quality correlation consistent basis sets, as was also found when investigating heats of formation for the hydrogenation of N_2[49] and in the determination of atomization energies.[50]

Since all calculations were performed utilizing the correlation consistent basis sets (cc-pVXZ where X = D, T, and Q) of Dunning,[43] we now have the option of utilizing the following prescription:[13]

$$E_{XY} = E_{\mathrm{SCF},Y} + \frac{X^3 E_{\mathrm{CORR},X} - Y^3 E_{\mathrm{CORR},Y}}{X^3 - Y^3} \qquad Y > X \quad (40)$$

to yield the cc-pV(DT)Z and cc-pV(TQ)Z extrapolated data sets (provided in Table 4). At the cc-pV(DT)Z level, the RMS (max) local error introduced by RI-TRIM MP2 has been reduced to 0.146 (0.324) kcal/mol—a modest (~10%) improvement from the RMS error at the cc-pVTZ level, 0.159 kcal/mol. As the complete basis set limit is approached at the cc-pV(TQ)Z level,[13] the RMS (max) error is significantly reduced to 0.040 (0.082) kcal/mol. It should be noted that the magnitude of this error, just slightly below *kT* at room temperature (~0.5 kcal/mol), is not likely to be significant within a biochemical context. Just by using eq 40, we were able to reduce the error introduced by RI-TRIM MP2 by a factor of 2.5 over the error at the cc-pVQZ level. At this level, there are no discrepancies in the RI-TRIM MP2 and RI-MP2 predictions of the relative energy orderings within this series—the first computational evidence that the TRIM model approaches exactness as the complete basis set limit is approached.

At this point, we are also in a position to numerically address the claim that local MP2 methods, which do not include the doubly ionic excitations partly responsible for BSSE, are more accurate than full MP2 when smaller basis sets are employed.[38,39] Taking the RI-MP2/cc-pV(TQ)Z data set as our "best numbers", a direct comparison with MP2/cc-pVDZ and TRIM MP2/cc-pVDZ yields RMS (max) deviations of 1.096 (2.504) and 0.939 (2.072) kcal/mol, respectively. At the cc-pVTZ level, the MP2 and TRIM MP2 methods yield RMS (max) deviations of 0.377 (0.944) and 0.250 (0.591) kcal/mol, respectively. On the basis of these

Table 3. Relative Energies (in kcal/mol) for a Series of 27 Different Alanine Tetrapeptide Conformations Using the cc-pVQZ Basis Set (and the Complementary Auxiliary Basis Set where Applicable)[a]

conformation	RI-MP2	RI-TRIM MP2	DFT
1	4.236	4.042	0.742
2	4.231	4.053	1.056
3	0.534	0.498	1.251
4	5.729	5.545	3.350
5	5.323	5.170	3.263
6	2.879	2.813	3.228
7	6.590	6.522	7.730
8	4.538	4.528	6.333
9	7.859	7.802	6.264
10	7.725	7.707	8.367
11	0.000	0.000	2.184
12	0.340	0.372	3.266
13	3.618	3.474	1.225
14	4.686	4.583	3.862
15	2.080	2.124	5.291
16	3.511	3.462	3.925
17	3.366	3.336	4.638
18	1.811	1.704	0.000
19	3.739	3.648	2.033
20	1.730	1.663	1.270
21	2.802	2.754	3.336
22	5.756	5.708	6.318
23	5.762	5.714	6.335
24	3.899	3.849	4.521
25	2.493	2.430	2.442
26	0.706	0.689	1.806
27	4.014	3.972	4.155

[a] The number of AO (auxiliary) basis functions employed in the cc-pVQZ calculations was 1760 (3850).

Table 4. Relative Energies (in kcal/mol) for a Series of 27 Different Alanine Tetrapeptide Conformations Based on the Extrapolated cc-pV(DT)Z (First Column) and cc-pV(TQ)Z (Second Column) Data Sets

conformation	RI-MP2		RI-TRIM MP2	
1	5.553	4.130	5.229	4.055
2	5.390	4.190	5.095	4.113
3	0.693	0.571	0.628	0.554
4	6.884	5.730	6.577	5.648
5	6.353	5.261	6.101	5.197
6	3.148	2.895	3.031	2.871
7	6.931	6.665	6.816	6.628
8	4.477	4.640	4.492	4.626
9	8.197	7.923	8.129	7.907
10	7.766	7.791	7.745	7.800
11	0.000	0.000	0.000	0.000
12	0.346	0.290	0.391	0.315
13	4.538	3.655	4.304	3.584
14	5.395	4.679	5.224	4.637
15	1.707	2.194	1.803	2.213
16	3.784	3.549	3.715	3.523
17	3.414	3.418	3.360	3.404
18	2.468	1.913	2.308	1.852
19	4.319	3.816	4.188	3.770
20	2.335	1.755	2.231	1.724
21	3.009	2.918	2.963	2.883
22	5.899	5.824	5.823	5.812
23	5.979	5.815	5.901	5.800
24	4.115	3.976	4.036	3.955
25	3.118	2.495	3.012	2.471
26	1.223	0.674	1.189	0.675
27	4.513	4.020	4.456	4.008

deviations, we can report that the local MP2 is only slightly more accurate (∼15%) than full MP2 at the cc-pVDZ level but substantially more accurate (∼34%) at the cc-pVTZ level; any further comment on the relationship between this improvement and BSSE without performing a series of counterpoise-corrected calculations on these conformers would be speculation. For comparison, we present the same error analysis using the LMP2/cc-pVTZ(-f) results of Friesner et al.,[41] which provides a RMS (max) error of 0.865 (1.690) kcal/mol [relative conformational energies at the LMP2/cc-pVTZ(-f) level are provided in Table S4 of the Supporting Information].

We find the basis set dependence of the relative conformational energies at the RI-MP2 and RI-TRIM MP2 levels surprising. Relative conformational energies are usually assumed to be fairly independent of the basis set because of the likelihood of error cancellation. We had performed the cc-pVQZ level calculations and the associated T → Q extrapolation largely for completeness—to demonstrate convergence in our new methodology. However, the RMS (max) deviations from RI-MP2/cc-pV(TQ)Z found as 0.377 (0.944) kcal/mol (MP2/cc-pVTZ) and 0.250 (0.591) kcal/mol (TRIM MP2/cc-pVTZ) are chemically significant when compared against the conformer energy differences. It therefore appears that this level of effort, that is, computations performed systematically at the cc-pVTZ and cc-pVQZ levels followed by extrapolation to cc-pV(TQ)Z, is necessary to obtain reliably converged relative conformational energies.

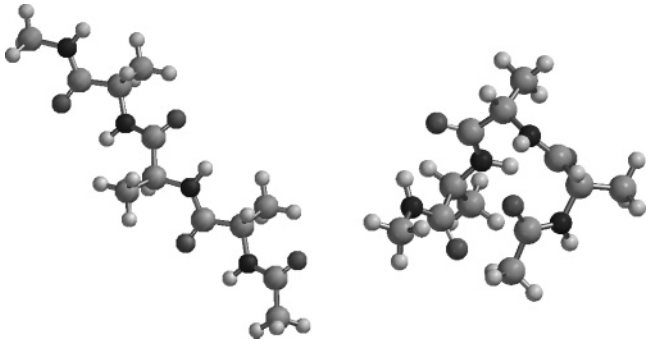


Figure 1. Graphical representations of the linear (1-D, conformation 1) conformation (left) and the globular (3-D, conformation 3) conformation (right) of the alanine tetrapeptide. These figures were created using the Spartan '04 package.[51]

The biophysical implications of our most reliable relative conformational energies are also quite interesting. In particular, among the low energy conformers, conformations 1 and 2 are extended (linear) while conformation 3 retains a compact hairpin geometry (globular). The energy difference between these geometries represents the energy gap between extended and folded backbone conformations—a fundamental quantity that propagates throughout a force field. We find this energy gap to be approximately 3.5 kcal/mol [RI-TRIM MP2/cc-pV(TQ)Z]. This number can be directly contrasted with the small basis canonical MP2 results that are closer to 5.5−6.0 kcal/mol, perhaps associated with intramolecular basis set superposition error. Additionally, previous local MP2 calculations using small basis sets yielded a gap of about 2.5 kcal/mol (Table S4 of the Supporting Information), while B3LYP, even at the cc-pVQZ level, predicts this energy difference as virtually zero. These findings reemphasize that one noteworthy application of the RI-TRIM MP2 methodology would be its use in the fundamental parameterization of force fields that treat systems of biochemical interest.

B. Computational Timings. In this section, we will explore the relationship between computational timings and system size for the following methodologies: SCF, MP2, TRIM MP2, RI-MP2, and RI-TRIM MP2. For clarity, we chose to omit DFT from the plots as the CPU time for DFT is almost always slightly larger (but on the same order of magnitude) than SCF. In doing so, we have considered linear (1-D) and globular (3-D) alanine tetrapeptide (see Figure 1), octapeptide, and hexadecapeptide conformations (the Cartesian coordinates for each of these model systems is provided in the Supporting Information). Using the linear and globular alanine tetrapeptide conformations, the computational scaling of methodology with angular momentum will also be evaluated by a direct comparison of CPU timings and basis set size for the SCF, RI-MP2, and RI-TRIM MP2 methods.

In Figure 2, a plot of log CPU time versus system size is presented for the aforementioned methodologies. This plot clearly shows that the CPU timings associated with the SCF, MP2, and TRIM MP2 methodologies strongly depend on

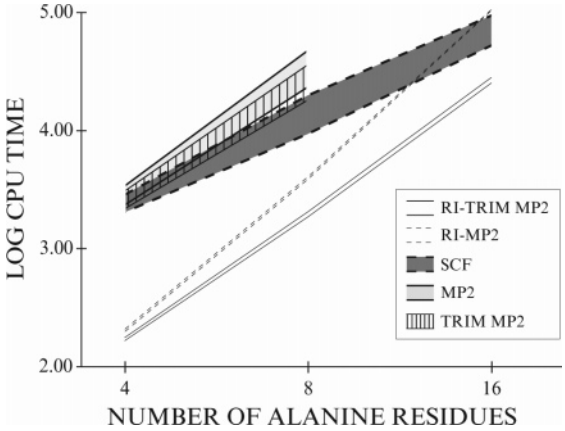


Figure 2. Plot of log CPU time vs system size for the SCF, MP2, TRIM MP2, RI-MP2, and RI-TRIM MP2 methodologies using the cc-pVDZ basis set. Each methodology has two associated lines: the bottom line represents data from linear (1-D) conformations, and the top line represents data from globular (3-D) conformations.

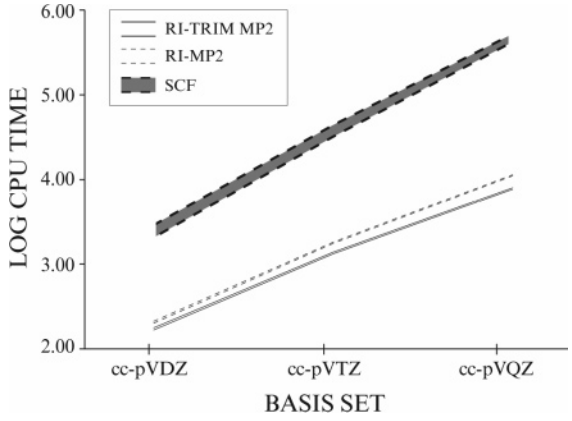


Figure 3. Plot of log CPU time vs basis set size for the SCF, RI-MP2, and RI-TRIM MP2 methodologies on linear (1-D, bottom line) and globular (3-D, top line) alanine tetrapeptide conformations.

system dimensionality, whereas the methods employing the RI approximation do not. From this plot, it is also evident that the iterative SCF procedure is the limiting time factor for calculations involving the RI approximation within a certain regime of system sizes. This trend relies strongly on the fact that we are using relatively tight DIIS convergence and integral screening threshold values; all estimates made in this section are based on these parameters and will need to be adjusted if one is using the default (looser) SCF parameters in most computational software packages. Using linear regression analysis, the crossover point between SCF and RI-MP2 was computed at approximately 75 heavy atoms, just shy of the globular (3-D) alanine hexadecapeptide data point [the linear (1-D) crossover occurs slightly earlier at approximately 64 heavy atoms]. The RI-TRIM MP2 methodology clearly extends the regime of practical MP2 calculations even further, as it is still less expensive than SCF at 125 heavy atoms—a system size corresponding to 25 alanine residues (for the 1-D case, the crossover occurs at approximately 104 heavy atoms, corresponding to a linear chain of 21 alanine residues). Put another way, one has to wait approximately 26 h for the MOs necessary to perform an RI-TRIM MP2 calculation, which takes less than 8 h (alanine hexadecapeptide at the cc-pVDZ level). From these data, it is now evident that some future research efforts should be focused on reducing the computational cost of the SCF procedure.

To explore the relationship between these methodologies and angular momentum in the AO basis set utilized, a plot of log CPU time versus basis set size is presented for SCF, RI-MP2, and RI-TRIM MP2 in Figure 3. This data set clearly shows that methods incorporating the RI approximation are scaling, on average, one power of L (highest angular momentum in the basis set) less than the SCF procedure and are, once again, independent of the system dimensionality. Using linear regression analysis, these scaling factors were found as 2.3 (RI-MP2) and 2.2 (RI-TRIM MP2) after normalization to 3.0 for the SCF methodology. This finding strongly supports the aforementioned point that SCF is the time limiting factor in RI-MP2 and RI-TRIM MP2 calculations. For alanine tetrapeptide at the cc-pVQZ level, a SCF was reached in approximately 6 days; this should be compared to the 2 h necessary to perform an RI-TRIM MP2 calculation. Hence, at the level where correlation energy corrections are most crucial, the RI-TRIM MP2 methodology has emerged with the ability to not only provide them with an acceptable level of accuracy but with very small relative timing costs.

V. Conclusions

In this work, the RI approximation has been incorporated into the theoretical framework of the local TRIM MP2 model to form a novel local fourth-order methodology, RI-TRIM MP2. On the basis of a series of stringent chemical and computational tests, RI-TRIM MP2 has demonstrated the ability to deliver accurate predictions of the relative energies of a series of 27 different alanine tetrapeptide conformations with RMS relative errors found as 0.192 kcal/mol at the cc-pVDZ level and an almost negligible 0.040 kcal/mol at the T → Q extrapolated complete basis set limit. The accurate prediction of the relative energies of this series of 27 alanine tetrapeptide conformations demonstrated a significant dependence on the basis set with RMS (max) deviations from the extrapolated RI-MP2/cc-pV(TQ)Z data set computed as 0.377 (0.944) kcal/mol (MP2/cc-pVTZ) and 0.250 (0.591) kcal/mol (TRIM MP2/cc-pVTZ). These findings strongly suggest that computations must be done using the cc-pVTZ and cc-pVQZ basis sets followed by extrapolation to the cc-pV(TQ)Z limit in order to obtain reliably converged relative conformational energies.

Correlation energy corrections using RI-TRIM MP2 can easily be obtained for systems that contain more than 125 heavy atoms with a computational timing cost less than that of the prerequisite SCF procedure and popular DFT alternatives. The RI-TRIM MP2 methodology reduces the current size limitations on systems that can be treated at the MP2 level and should be used to extend the regime of practical MP2 calculations when RI-MP2 becomes the rate-determining computational step (for systems containing more than 75 heavy atoms). With these findings, some future research

efforts will be directed toward reducing the computational cost of the iterative SCF procedure by utilizing linear scaling algorithms,[52−54] density fitting,[55] or a dual-basis approach.[56] The combination of an appropriately optimized SCF algorithm and the RI-TRIM MP2 gradient would then allow us to perform geometry optimizations on systems containing more than 125 heavy atoms at the MP2 level; this would be highly desirable as small biochemical and nanotechnological systems of interest could then be routinely explored in a number of chemical applications.

Acknowledgment. Funding for this work comes from a subcontract of Q-Chem Inc. as part of an NIH SBIR grant. We gratefully acknowledge useful discussions with Prof. Jay Ponder and Prof. Pengyu Ren, who suggested we re-examine the conformational energies of the alanine tetrapeptides and forwarded the optimized structures to us. We also thank Judy Hicks of Plush Design for her assistance in creating some of the figures in this work and Laura Rutherford for general research assistance. M.H.-G. is a part-owner of Q-Chem Inc.

Supporting Information Available: Absolute energies for the SCF and DFT methods at the cc-pVDZ, cc-pVTZ, and cc-pVQZ levels; absolute correlation energies for the MP2 and TRIM MP2 methods at the cc-pVDZ and cc-pVTZ levels; absolute correlation energies for the RI-MP2 and RI-TRIM MP2 methods at the cc-pVDZ, cc-pVTZ, and cc-pVQZ levels; relative conformational energies at the LMP2/cc-pVTZ(-f) level; and cartesian coordinates for all systems investigated. 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) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974−12980.

(3) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: New York, 2000.

(4) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175−180.

(5) Jung, Y.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2008−2011.

(6) Johnson, B. G.; Gonzales, C. A.; Gill, P. M. W.; Pople, J. A. *Chem. Phys. Lett.* **1994**, *221*, 100−108.

(7) Perdew, J. P.; Zunger, A. *Phys. Rev. B: Condens. Matter* **1981**, *23*, 5048−5079.

(8) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811−4815.

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

(10) Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem.* **1976**, *S10*, 1−19.

(11) For example, see Tables 8−4 of ref 3 and references therein.

(12) Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213−13225.

(13) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639−9646.

(14) Byrd, E. F. C.; Sherrill, C. D.; Head-Gordon, M. *J. Phys. Chem. A* **2001**, *105*, 9736−9747.

(15) Helgaker, T.; Gauss, J.; Jorgensen, P.; Olsen, J. *J. Chem. Phys.* **1997**, *106*, 6430−6440.

(16) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359−363.

(17) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143−152.

(18) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. X. *J. Phys. Chem. A* **1999**, *103*, 1913−1928.

(19) Pulay, P.; Saebø, S.; Wolinski, K. *Chem. Phys. Lett.* **2001**, *344*, 543−552.

(20) Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 3660−3671.

(21) Ayala, P. Y.; Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2001**, *115*, 9698−9707.

(22) Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.

(23) Schütz, M.; Hetzer, G.; Werner, H. J. *J. Chem. Phys.* **1999**, *111*, 5691−5705.

(24) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592−3601.

(25) Subotnik, J. E.; Head-Gordon, M. *J. Chem. Phys.* **2005**, *122*, 034109-1−034109-9.

(26) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793−9802.

(27) Lochan, R. C.; Jung, Y.; Head-Gordon, M. *J. Phys. Chem. A* **2005**, in press.

(28) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095−9102.

(29) Schütz, M.; Hetzer, G.; Werner, H. J. *J. Chem. Phys.* **1999**, *111*, 5691−5705.

(30) Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H. J. *J. Chem. Phys.* **2000**, *113*, 9443−9455.

(31) Schütz, M. *J. Chem. Phys.* **2000**, *113*, 9986−10001.

(32) Schütz, M.; Werner, H. J. *J. Chem. Phys.* **2001**, *114*, 661−681.

(33) Werner, H. J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149−8160.

(34) Schütz, M.; Werner, H. J.; Lindh, R.; Manby, F. R. *J. Chem. Phys.* **2004**, *121*, 737−750.

(35) Maslen, P. E.; Head-Gordon, M. *Chem. Phys. Lett.* **1998**, *283*, 102−108.

(36) Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 7093−7099.

(37) Lee, M. S.; Head-Gordon, M. *Int. J. Quantum Chem.* **2000**, *76*, 169−184.

(38) Saebø, S.; Tong, W.; Pulay, P. *J. Chem. Phys.* **1993**, *98*, 2170−2175.

(39) Schütz, M.; Rauhut, G.; Werner, H. J. *J. Phys. Chem. A* **1998**, *102*, 5997−6003.

(40) For a detailed derivation of the canonical Hartree−Fock equations, see: Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: An Introduction to Advanced Electronic Structure Theory*; Dover: New York, 1989.

(41) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.

(42) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, C. D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Dachsel, H.; Zhang, W. M.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C. P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532−1548.

(43) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(44) Weigend, F.; Kohn, A.; Hattig, C. *J. Chem. Phys.* **2002**, *116*, 3175−3183.

(45) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(46) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter* **1998**, *37*, 785−789.

(47) Yu, C.-H.; Norman, M. A.; Schäfer, L.; Ramek, M.; Peeters, A.; van Alsenoy, C. *J. Mol. Struct.* **2001**, *567*−*568*, 361−374.

(48) Chaudhuri, P.; Canuto, S. *THEOCHEM* **2002**, *577*, 267−279.

(49) Sekusak, S.; Frenking, G. *THEOCHEM* **2001**, *541*, 17−29.

(50) Wang, N. X.; Wilson, A. K. *J. Chem. Phys.* **2004**, *121*, 7632−7646.

(51) *Spartan '04 for Windows*; Wavefunction, Inc.: Irvine, CA.

(52) Challacombe, M. *J. Chem. Phys.* **1999**, *110*, 2332−2342.

(53) Schwegler, E.; Challacombe, M.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *106*, 9708−9717.

(54) Challacombe, M.; Schwegler, E. *J. Chem. Phys.* **1997**, *106*, 5526−5536.

(55) Polly, R.; Werner, H. J.; Manby, F. R.; Knowles, P. J. *Mol. Phys.* **2004**, *102*, 2311−2321.

(56) Liang, W.; Head-Gordon, M. *J. Phys. Chem. A* **2004**, *108*, 3206−3210.

CT050126S

JCTC Journal of Chemical Theory and Computation

The Rotational *g* Tensor as a Benchmark for Density-Functional Theory Calculations of Molecular Magnetic Properties

David J. D. Wilson,*,† Chris E. Mohn, and Trygve Helgaker

Department of Chemistry, University of Oslo, P. O. Box 1033 Blindern, N-0315 Oslo, Norway

Received April 15, 2005

Abstract: The rotational *g* factor for a large number of organic compounds has been investigated with density-functional theory. Rapid convergence toward the basis-set limit is ensured by the use of London atomic orbitals. A statistical analysis of the results has been carried out in comparison with accurate experimental data. It is shown that gradient-corrected and hybrid functionals reproduce experimental results most closely, with the Keal−Tozer KT2 functional being the most accurate.

I. Introduction

Calculations of rotational *g* tensors constitute a particular challenge to ab initio theory, with the errors introduced in the description of the electronic system often being much larger than the experimental standard deviations obtained in highly accurate molecular−beam[1] and microwave Zeeman experiments.[2,3] Still, theoretical studies of the rotational *g* tensor have been presented at various levels of ab initio theory, including Hartree−Fock theory,[4−6] second-order Møller−Plesset (MP2) theory,[7−9] linearized coupled-cluster doubles theory,[9] the second-order polarization propagator approximation,[10−12] multiconfigurational self-consistent field (MCSCF) theory,[13−15] and, for small systems, full configuration-interaction theory.[16,17] However, although the results presented in these wave-function studies are mostly in qualitative agreement with experimental results, they often differ by more than 10 experimental standard deviations from the experimental values, especially at lower levels of theory.

In recent years, density-functional theory (DFT) has emerged as the most popular quantum-mechanical method for studying molecular properties. While the standard local-density approximation (LDA) Slater−Vosko−Wilk−Nusair (SVWN) functional[18] typically gives results of intermediate quality, generalized gradient approximation (GGA) functionals such as the Becke−Lee−Yang−Parr (BLYP) functional[19,20] and hybrid functionals such as the Becke three-parameter Lee−Yang−Parr (B3LYP) functional[21−23] often describe electron correlation effects of molecular properties with an accuracy comparable to that of MP2[24] and coupled-cluster single-and-doubles[25−27] theories. In view of the accurate experimental measurements of *g* tensors, it is therefore of considerable interest to apply DFT to the calculation of this property and, in particular, to benchmark DFT *g* tensors against the available experimental data.

In common with all properties that involve interactions with an external magnetic field, the calculation of rotational *g* factors suffers from the gauge-origin problem and slow basis-set convergence. Despite the many treatments of the gauge-origin problem proposed over the years,[10,28−34] it was, in principle, already solved in 1937 by London, who proposed attaching field-dependent complex phase factors to the atomic orbitals.[35] In London's scheme, each atomic orbital has its own gauge origin: the nucleus to which it is attached. These London orbitals, which are also known as gauge invariant atomic orbitals, are correct to the first order in the magnetic field for a one-electron, one-center problem.[36] Their use in ab initio theory was pioneered by Hameka[37−39] in the 1950s and 1960s and by Ditchfield[40] in the 1970s, but the efficient implementation of London orbitals was not achieved until the 1990s, with the work of Wolinski and co-workers.[41] London orbitals are now widely used in quantum chemistry, with Hartree−Fock,[42] MP2,[43] MCSCF,[44] and

* Corresponding author phone: +47 228 55432; fax: +47 228 55441; e-mail: d.j.wilson@kjemi.uio.no.

† Current affiliation: Chemistry Department, La Trobe University, Australia.

10.1021/ct050101t CCC: $30.25
Published on Web 08/16/2005

Table 1. List of Molecules Considered in This Study[a]

Linear Molecules			
carbonyl sulfide[b]	OCS	hydrogen cyanide	$HC^{15}N$
hydrogen boron sulfide[c]	HBS	fluoro cyanide[d]	$FC^{15}N$
carbonyl selenide[e,f]	OCSe	chloro cyanide[g,d]	$ClC^{15}N$
carbon monoxide	CO	bromo cyanide[d]	$BrC^{15}N$
carbon sulfide	CS	fluoro acetylene[h]	HCCF
carbon selenide[i]	CSe	chloro acetylene[j,h]	HCCCl
methylidene phosphine[k,c]	HCP	bromo acetylene[l,h]	HCCBr
nitrous oxide[m]	N_2O		
Symmetric Top Compounds			
acetonitrile	CH_3CN	fluoromethane	CH_3F
methylisocyanide	CH_3NC[n]	chloromethane	CH_3Cl
ammonia	NH_3	trifluoromethane	CHF_3
Nonring Planar Asymmetric Tops			
2-methyl but-1,3-diene	C_5H_8	difluoromethane	CH_2F_2
acrolein	C_3H_4O	carbonic dufluoride[o]	F_2CO
propene[p]	C_3H_6	formyl fluoride[q]	HFCO
propynal[r]	C_3H_2O	fluoroethylene	C_2H_3F
dimethyl ether	C_2H_6O	1,1-difluoro ethylene	$C_2H_2F_2$
dimethyl sulfide	C_2H_6S	*cis*-difluoro ethylene[r]	$C_2H_2F_2$
acetaldehyde	C_2H_4O	trifluoro ethene[s]	C_2HF_3
formaldehyde[p]	H_2CO	fluoro ethane[o]	C_2H_5F
thioformaldehyde	H_2CS	ozone[s,t]	O_3
formic acid	HCOOH	sulfur dioxide[u]	SO_2
formamide	$HCONH_2$	difluoro oxide	OF_2
glycoaldehyde[r]	$C_2H_4O_2$	hypofluoros acid	HOF
methyl formate	$C_2H_4O_2$	water	H_2O
ketene	$C_2H_2O_2$	hydrogen sulfide	H_2S
Three-Member Ring Compounds			
methylene cyclopropane	C_4H_6	thiirane	C_2H_4S
cyclopropene	C_3H_4	cyclopropenone[v]	C_3H_2O
aziridine	C_2H_5N	methyl cyclopropene	C_4H_6
oxirane	C_2H_4O		
Four-Member Ring Compounds			
cyclobutene	C_4H_6	oxetan	C_4H_6O
oxetane	C_3H_6O	β-pripiolactone	$C_3H_4O_2$
methylene cyclobutane	C_5H_8	diketene	$C_4H_4O_2$

[a] Experimental *g*-tensor data taken from ref 3 except where noted. Experimental geometries taken from ref 68 unless noted. [b] Experimental data from ref 63. Three isotopes considered ($^{16}O^{12}C^{32}S$,-$^{16}O^{12}C^{34}S$, and $^{16}O^{13}C^{32}S$). [c] Geometry from ref 72. [d] Geometry from ref 69. [e] Two isotopes considered ($OC^{76}Se$ and $OC^{80}Se$). [f] Geometry from ref 70. [g] Experimental data from ref 66. [h] Geometry from ref 73. [i] Geometry from ref 71. [j] Two isotopes considered ($^{35}ClCCH$ and $^{37}ClCCH$). [k] Two isotopes considered (HCP and DCP). [l] Two isotopes considered ($^{79}BrCCH$ and $^{81}BrCCH$). [m] Two isotopes considered ($^{15}N_2O$ and $^{14}N_2O$). [n] Geometry from ref 74. [o] Geometry from ref 75. [p] Geometry from ref 14. [q] Geometry from ref 77. [r] Geometry from ref 76. [s] Geometry from ref 78. [t] Experimental data from ref 65. [u] Experimental data from ref 64. [v] Geometry from ref 79.

coupled-cluster[45−47] implementations reported. Particularly impressive is the rapid basis-set convergence of magnetic properties with the use of London orbitals.[5]

To date, very few DFT studies of rotational *g* tensors have been reported,[48] only one of which (from our laboratory) uses London orbitals.[49] Here, we present London DFT calculations of the rotational *g* tensor for the 61 organic molecules listed in Table 1. These molecules differ in the importance and relative magnitudes of static and dynamic electron correlation; hence, they constitute a suitable set for testing the accuracy of various computational approaches in the calculation of *g* tensors. To benchmark the performance of the various density functionals, a statistical analysis of the calculated *g* tensors is carried out, comparing against experimental data taken almost exclusively from the reviews of Flygare and co-workers.[2,3] In all calculations, we use a very recent London orbital implementation of DFT magnetic properties in DALTON.[50]

The paper is organized as follows: In Section 2, we outline the theory of *g*-tensor calculations, with some computational details given in Section 3. In Section 4, the results and statistical analysis from the calculations are presented. Some concluding remarks are given in Section 5.

II. Theory

In an applied external magnetic field, the rotational energy levels of a rotating molecule are shifted relative to those in the absence of such a field. These shifts, which arise from the Zeeman interaction of the external magnetic induction **B** with the magnetic dipole moment generated by the rotating molecule of angular momentum **J** in the center-of-mass coordinate system, are conventionally represented as

$$\Delta E = -\mu_N \mathbf{B}^T \mathbf{g} \mathbf{J}, \tag{1}$$

where the three-by-three matrix **g** is the rotational *g* tensor and μ_N is the nuclear magneton. For computational purposes, it is convenient to express the *g* tensor as a derivative of the molecular electronic energy $\epsilon(\mathbf{B},\mathbf{J})$ with respect to the magnetic field strength and the rotational angular momentum at $\mathbf{B} = \mathbf{0}$ and $\mathbf{J} = \mathbf{0}$:[5]

$$\mathbf{g} = -\frac{\hbar}{\mu_N} \frac{\partial^2 \epsilon(\mathbf{B},\mathbf{J})}{\partial \mathbf{B} \partial \mathbf{J}}\bigg|_{\mathbf{B},\mathbf{J}=\mathbf{0}}. \tag{2}$$

In this manner, the rotational *g* tensor may be obtained in the same way as any other time-independent second-order molecular property, once the expression for $\epsilon(\mathbf{B},\mathbf{J})$ (including the purely nuclear interactions) has been set up for a rotating molecular system in the presence of an external magnetic field. In principle, this is a straightforward process once the expression for the Hamiltonian has been written down.

For a rotating molecule in an external magnetic field, we can write down the Hamiltonian in the center-of-mass (CM) coordinate system as

$$H(\mathbf{B},\mathbf{J}) = H_0 + \frac{1}{2}\mathbf{L}_O\mathbf{B}^T - \mathbf{L}_{CM}^T \mathbf{I}_{nuc}^{-1}\mathbf{J} + \frac{1}{2}\mathbf{B}^T\sum_i(\mathbf{r}_i^T\mathbf{R}_O\mathbf{I}_3 - \mathbf{r}_i\mathbf{R}_O^T)\mathbf{I}_{nuc}^{-1}\mathbf{J} - \frac{1}{2}\mathbf{B}^T\sum_K Z_K(\mathbf{R}_K^T\mathbf{R}_K\mathbf{I}_3 - \mathbf{R}_K\mathbf{R}_K^T)\mathbf{I}_{nuc}^{-1}\mathbf{J} \tag{3}$$

where the first term H_0 is the spin-free nonrelativistic electronic Hamiltonian of a nonrotating diamagnetic molecule in the absence of an external magnetic field. The remaining four terms in $H(\mathbf{B},\mathbf{J})$ are either linear (paramagnetic) or bilinear (diamagnetic) in **B** and **J**. The first paramagnetic term, $^1/_2\mathbf{L}_O\mathbf{B}^T$, where $\mathbf{L}_O$ is the angular momentum operator

of the electrons about the origin $\mathbf{R}_O$ of the external potential $\mathbf{A}(\mathbf{r}) = {}^1/_2\mathbf{B} \times (\mathbf{r} - \mathbf{R}_O)$

$$\mathbf{L}_O = \sum_i (\mathbf{r}_i - \mathbf{R}_O) \times \nabla_i \tag{4}$$

represents the Zeeman interaction of the electrons with the magnetic induction $\mathbf{B}$. In the same manner, the second paramagnetic term $-\mathbf{L}_{CM}^T \mathbf{I}_{nuc}^{-1}\mathbf{J}$ in eq 3, where $\mathbf{I}_{nuc}$ is the moment-of-inertia tensor and $\mathbf{L}_{CM}$ is the electronic angular-momentum operator in the center-of-mass coordinate system

$$\mathbf{L}_{CM} = \sum_i \mathbf{r}_i \times \nabla_i \tag{5}$$

represents the Zeeman interaction of the electrons with the magnetic field $-2\mathbf{I}_{nuc}^{-1}\mathbf{J}$ generated by the rotating molecule. Finally, of the two diamagnetic terms in the Hamiltonian, eq 3, the first involves the electrons and, therefore, enters a one-electron expectation value; the second is a purely nuclear operator whose contribution to the g tensor does not depend on the electronic state. Note that, in $H(\mathbf{B},\mathbf{J})$, we have excluded all terms that do not contribute to the rotational g tensor—for example, the diamagnetic term quadratic in $\mathbf{B}$.

For the calculation of rotational g tensors, the use of London orbitals has been demonstrated not only to ensure gauge-origin independent results but also to accelerate the convergence toward the basis-set limit (compared to conventional field-independent orbitals).[35,36] The London atomic orbitals are here defined as

$$\omega_\mu(\mathbf{B},\mathbf{J}) = \exp[-i(\mathbf{A}_\mu^{\mathbf{B}} + \mathbf{A}_\mu^{\mathbf{J}})\mathbf{r}]\,\chi_\mu \tag{6}$$

with phase factors given by

$$\mathbf{A}_\mu^{\mathbf{B}} = \frac{1}{2}\mathbf{B} \times (\mathbf{R}_\mu - \mathbf{R}_O) \tag{7}$$

$$\mathbf{A}_\mu^{\mathbf{J}} = -\mathbf{I}_{nuc}^{-1}\mathbf{J} \times \mathbf{R}_\mu \tag{8}$$

where $\mathbf{R}_\mu$ is the position of the conventional atomic orbital χ_μ. Note that these orbitals, which are sometimes referred to as rotational London orbitals, constitute a generalization of the standard London orbitals, depending on $\mathbf{A}_\mu^{\mathbf{J}}$ as well as on $\mathbf{A}_\mu^{\mathbf{B}}$.[5] With these orbitals and the Hamiltonian (eq 3), one can derive the expressions for the g tensor using the machinery of time-independent response-theory described elsewhere.[51] The resulting expression for the rotational g tensor contains no reference to the gauge origin of the vector potential representing the external magnetic field but does depend correctly on the center of mass of the molecular system.

We here note the close relationship of the rotational g tensor to the molecular magnetizability, which represents the second derivative of the molecular energy with respect to an external magnetic field at $\mathbf{R} = \mathbf{0}$:

$$\xi = -\mu_0 \left.\frac{\partial^2 \epsilon(\mathbf{B})}{\partial \mathbf{B}^2}\right|_{\mathbf{B}=\mathbf{0}}. \tag{9}$$

In the center-of-mass coordinate system with $\mathbf{R}_{CM}$ as the chosen gauge origin of $\mathbf{L}_O$ in eq 3, the relationship between rotational g tensors and molecular magnetizabilities may be written as[5]

$$\mathbf{g} = -4m_p(\xi^{LAO} - \xi_{CM}^{dia})\mathbf{I}_{nuc}^{-1} + \frac{1}{2\mu_N}\sum_K Z_K(\mathbf{R}_K^T\mathbf{R}_K\mathbf{I}_3 - \mathbf{R}_K\mathbf{R}_K^T)\mathbf{I}_{nuc}^{-1} \tag{10}$$

where m_p is the proton mass, ξ^{LAO} is the magnetizability calculated with London orbitals, and ξ_{CM}^{dia} is the diamagnetic contribution to the magnetizability tensor, calculated using conventional orbitals with the gauge origin at the molecular center of mass. Hence, the rotational g tensor may be programmed and calculated as a simple extension to any code that already calculates the molecular magnetizability.

The extension of a Hartree−Fock magnetizability program to calculate DFT magnetizabilities and rotational g tensors is straightforward. The magnetizability, eq 9, may be calculated using standard response theory.[44] Parametrizing the electronic energy $\epsilon(\mathbf{B},\kappa)$ in terms of the orbital-rotation parameters κ (one parameter κ_{ai} for each pair of occupied and unoccupied orbitals), we then make use of the fact that the optimized electronic energy $\epsilon(\mathbf{B}) = \min_\kappa \epsilon(\mathbf{B},\kappa)$ is variational with respect to rotations between all pairs of occupied and virtual orbitals κ_{ai}, for all values of $\mathbf{B}$:[52−54]

$$\frac{\partial \epsilon(\mathbf{B},\kappa)}{\partial \kappa_{ai}} = \mathbf{0}, \text{ all values of } \mathbf{B} \tag{11}$$

The resulting expression for ξ in eq 9 then follows by differentiation and use of the chain rule:

$$\xi = -\mu_0 \frac{\partial^2 \epsilon(\mathbf{B},\kappa)}{\partial \mathbf{B}^2} - \mu_0 \sum_{ai} \frac{\partial^2 \epsilon(\mathbf{B},\kappa)}{\partial \mathbf{B}\partial \kappa_{ai}} \frac{\partial \kappa_{ai}}{\partial \mathbf{B}} \tag{12}$$

where all derivatives are evaluated at $\mathbf{B} = \mathbf{0}$. The first-order response of the wave function is evaluated from the linear set of equations

$$\sum_{bj} \frac{\partial^2 \epsilon(\mathbf{B},\kappa)}{\partial \kappa_{ai}\partial \kappa_{bj}} \frac{\partial \kappa_{bj}}{\partial \mathbf{B}} = -\frac{\partial^2 \epsilon(\mathbf{B},\kappa)}{\partial \kappa_{ai}\partial \mathbf{B}} \tag{13}$$

obtained by differentiation of the variational conditions, eq 13, with respect to $\mathbf{B}$. In eq 12, the first term is known as the diamagnetic part of the magnetizability and the second term, which contains the response of the wave function to the external magnetic field, is known as the paramagnetic part. The summations are over pairs of virtual and occupied orbitals, with κ_{ai} and κ_{bj} denoting the orbital-rotation parameters as discussed in refs 52−54. Equation 13 is similarly required for the calculation of chemical shieldings and magnetizabilities,[55,56] so a program with these properties implemented only requires the diamagnetic component of the magnetizability (and rotational g tensor) to be coded. The resulting DFT implementation, including the differentiation of London atomic orbitals, follows that for Hartree−Fock theory, which is outlined elsewhere.[44]

III. Computational Details

The above theory for the calculation of molecular magnetizabilities and rotational g tensors has recently been

Table 2. Errors of Calculated *g* Tensors Compared to Experimental Data

	optimized geometry		exptl geometry		optimized geometry		exptl geometry	
method	DZ	TZ	DZ	TZ	DZ	TZ	DZ	TZ
	Mean Relative Deviation, $\bar{\Delta}$				Mean Absolute Deviation, $\bar{\Delta}_{abs}$			
HF	0.005	0.002	−0.005	−0.005	0.029	0.028	0.038	0.036
LDA	−0.020	−0.016	−0.020	−0.014	0.027	0.023	0.028	0.023
BLYP	−0.016	−0.018	−0.014	−0.014	0.017	0.019	0.016	0.015
B3LYP	−0.011	−0.013	−0.009	−0.010	0.014	0.014	0.013	0.012
KT1	−0.006	−0.008	−0.005	−0.005	0.010	0.011	0.010	0.010
KT2	−0.004	−0.006	−0.002	−0.002	0.009	0.010	0.008	0.008
KT3	−0.005	−0.008	−0.003	−0.003	0.010	0.011	0.009	0.009
	Standard Deviation of Errors, $\bar{\Delta}_{std}$				Percentage Mean Absolute Deviation, $\bar{\Delta}_{pabs}$			
HF	0.145	0.141	0.240	0.235	21.1	18.2	21.5	19.1
LDA	0.093	0.075	0.101	0.080	15.3	16.1	14.7	14.1
BLYP	0.056	0.058	0.036	0.032	14.5	18.6	18.8	21.1
B3LYP	0.068	0.069	0.055	0.052	8.9	10.4	9.4	10.9
KT1	0.028	0.028	0.020	0.018	10.7	13.9	13.8	16.1
KT2	0.027	0.026	0.019	0.020	10.2	12.3	9.4	11.2
KT3	0.032	0.031	0.025	0.026	9.4	12.9	10.1	12.1

implemented in the DALTON program for DFT with London orbitals.[50] In this work, we have also taken the opportunity to implement the recently developed Keal−Tozer (KT) functionals KT1,[57] KT2,[58] and KT3,[59] with the implementations tested against CADPAC.[60] The KT functionals were included since they have previously been shown to provide accurate NMR shielding constants.[58]

To benchmark the implemented methods, we have carried out calculations of *g* tensors on a sample of 61 molecules (70 isotopes), representing 155 unique tensor elements. For all molecules, six density functionals were considered: the SVWN[18] functional of LDA theory; the BLYP,[19,20] KT1,[57] KT2,[58] and KT3[59] functionals of GGA theory; and the B3LYP[21−23] hybrid functional. Hartree−Fock results were included for comparison. To assess geometrical effects, the *g* tensors were calculated at both optimized and experimental molecular geometries. Experimental geometries were taken from the literature, with specific references given in Table 1. For a consistent and purely theoretical approach, we have calculated optimized geometries and rotational *g* tensors at the same level of theory and with the same basis set for all molecules. We have not considered zero-point vibrational corrections in this work.

All calculations employ Dunning's correlation-consistent basis sets.[61,62] Since it has previously been noted that diffuse functions are important in calculating magnetizabilities and rotational *g* tensors,[5,14,49] the augmented correlation-consistent double- and triple-ζ basis sets aug-cc-pVDZ and aug-cc-pVTZ were used, with some additional calculations carried out in the corresponding quadruple-ζ aug-cc-pVQZ basis to assess basis-set convergence.

In this work, we have calculated the mean relative deviations, mean absolute deviations, and standard deviations as

$$\bar{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\Delta_i \tag{14}$$

$$\bar{\Delta}_{abs} = \frac{1}{n}\sum_{i=1}^{n}|\Delta_i| \tag{15}$$

$$\Delta_{std} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\Delta_i - \bar{\Delta})^2} \tag{16}$$

where the summations are over the *n* individual errors in the calculated tensor elements of all molecules

$$\Delta_i = g_i^{calc} - g_i^{ref} \tag{17}$$

Since the rotational *g* tensor elements differ widely in magnitude, we have also calculated the percentage mean absolute deviations

$$\bar{\Delta}_{pabs} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{g_i^{calc} - g_i^{ref}}{g_i^{ref}}\right| \tag{18}$$

as an alternative to the mean absolute deviations, eq 15.

IV. Comparison with Experiment

A. Experimental Data. With four exceptions, all experimental gas-phase rotational *g* tensors listed here have been reported by Flygare and co-workers.[2,3] These measurements, which were carried out in the 1960s and early 1970s, still represent the standard compilation of experimental gas-phase rotational *g* tensors. More recent experimental data were included only where they are more accurate than those of Flygare, as defined by smaller error bars and a greater number of significant figures. The four exceptions are OCS, ClCN, O_3, and SO_2, which all arise from the work of Dymanus and co-workers.[63−66]

For OCS, the data of ref 63 is included, since three isotopes are studied whereas Flygare only reported two. The magnitude of the experimental *g* tensors and associated error bars is consistent for the two isotopes reported in both references.

Table 3. Linear Regression Results from a Comparison of Calculated and Experimental Data[a]

	optimized geometry		exptl geometry		optimized geometry		exptl geometry	
method	DZ	TZ	DZ	TZ	DZ	TZ	DZ	TZ
	R^2 Values				Slope of Linear Equation			
HF	0.936	0.941	0.862	0.866	0.97	0.96	0.83	0.84
LDA	0.930	0.997	0.992	0.999	1.10	0.89	0.87	0.88
BLYP	0.940	0.995	0.999	0.999	1.21	0.93	0.95	0.95
B3LYP	0.950	0.990	0.994	0.994	1.30	0.94	0.95	0.95
KT1	0.931	0.998	0.999	0.999	1.28	0.99	0.99	1.00
KT2	0.909	0.998	0.999	0.999	1.24	1.00	1.01	1.02
KT3	0.883	0.997	0.999	0.999	1.18	1.00	1.02	1.03

[a] Linear regression is forced to pass through the origin with the experimental data on the *y* axis.

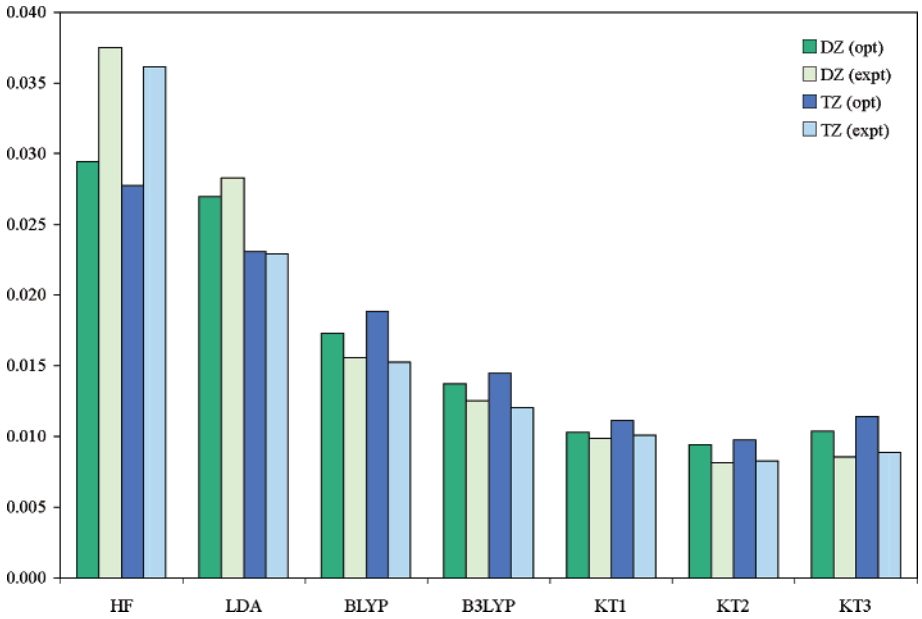


Figure 1. Mean absolute deviation from experimental *g* tensors for Hartree−Fock and DFT methods at both optimized and experimental geometries.

Dynamus' data[65] for O_3 (−2.9877 ± 0.0009, −0.2295 ± 0.0003, −0.0760 ± 0.0003) is used rather than Flygare's (−2.968 ± 0.035, −0.228 ± 0.007, −0.081 ± 0.006) and similarly for SO_2.[64] The ClCN *g* tensor of Reinartz and co-workers[66] (−0.041 21 ± 0.000 13) supersedes the value of Flygare (−0.0384 ± 0.0003).

B. Statistical Analysis. In Table 2, we summarize the calculated rotational *g* tensor elements as deviations from the highly accurate experimental data, and in Table 3, we present data from a linear regression analysis. These results are further illustrated in Figure 1, where $\bar{\Delta}_{abs}$ is plotted for each respective method. The DFT results are clearly superior to the Hartree−Fock results, which is in agreement with that of Wilson and co-workers,[48] who considered a small sample size of seven molecules without the use of London orbitals.

Of the functionals considered, the results in Figure 1 and Table 2 indicate that the KT functionals perform better than the standard SVWN, BLYP, and B3LYP functionals, with the KT2 functional being the best functional of those considered in this work. This observation gives further confirmation of the superior performance of KT functionals for properties involving an external magnetic field.[58] In contrast to the Hartree−Fock calculations, the DFT calculations at the experimental geometry typically produce rotational *g* tensors that are closer to experimental results than those obtained at the optimized geometries.

As in a previous study of the magnetizability of hydrocarbons,[67] we have carried out a linear regression analysis to examine the relationship between the calculated and experimental data. Our results are collected in Table 3, where the superior performance of DFT relative to Hartree−Fock theory is clearly demonstrated, both in terms of the slope of the fitted linear equation and the R^2 values. In particular, the slope of 1.00 ± 0.03 for the KT functionals highlights the ability of these functionals to reproduce experimental results—see Figure 2, where the KT2/aug-cc-pVTZ results (at the experimental geometries) are plotted against the experimental data. Even for the outliers, this functional accurately reproduces experimental results.

From Table 2, we see that basis-set effects are small for rotational *g* tensors calculated with London orbitals. Indeed, it has been shown previously that Hartree−Fock magnetizabilities and *g* tensors converge rapidly with London orbitals, with reasonably accurate results often obtained in the aug-cc-pVDZ basis.[5,44] To test the basis-set convergence, additional aug-cc-pVQZ calculations were performed for the *g* tensors of the linear molecules in Table 1, giving average and maximum differences between the aug-cc-pVTZ and aug-cc-pVQZ basis sets of 0.0002 and 0.0014, respectively.

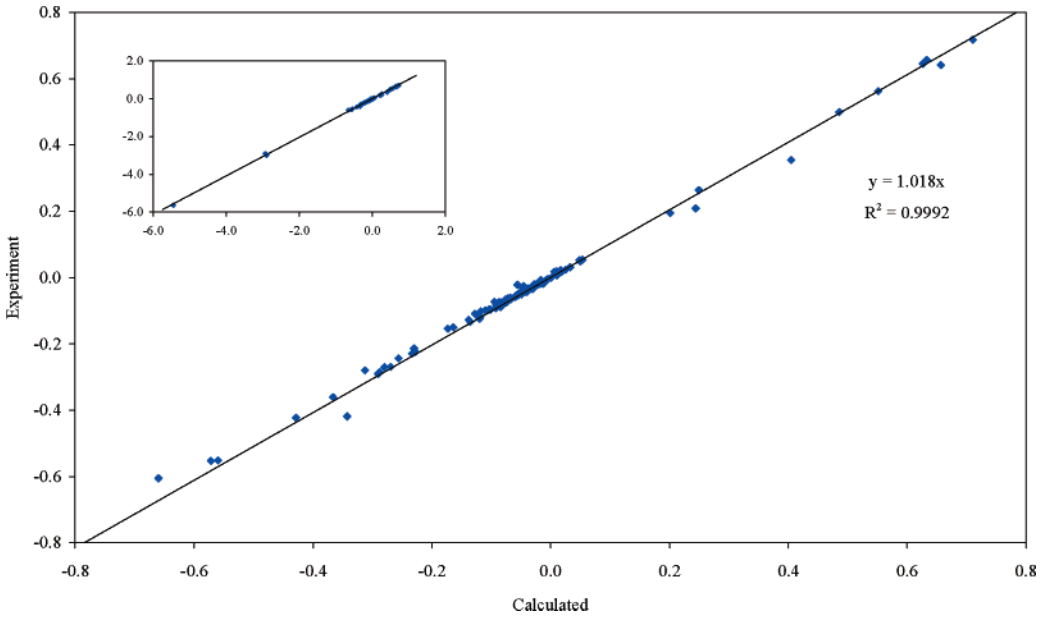


Figure 2. Correlation between KT2/aug-cc-pVTZ calculated and experimental rotational *g* tensors, with all calculations performed at experimental geometries. The inset includes the entire data range of included *g* tensor elements.

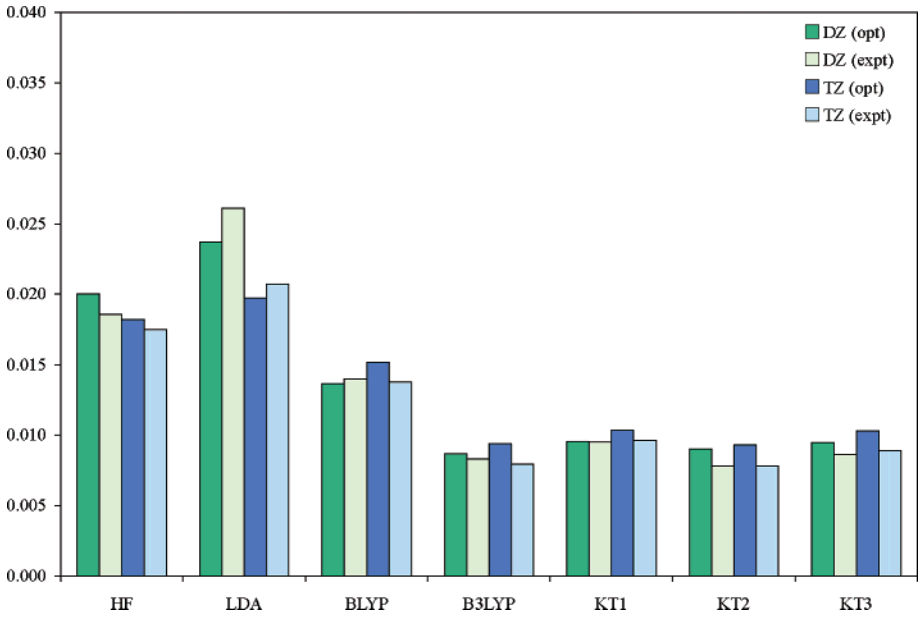


Figure 3. Mean absolute deviation from experimental *g* tensors for Hartree−Fock and DFT methods at both optimized and experimental geometries, excluding ozone.

The average and maximum differences between the aug-cc-pVDZ and aug-cc-pVTZ basis sets are 0.0015 and 0.0044, respectively. This observation is true for all methods (Hartree−Fock as well as DFT) considered here.

C. Ozone. Of the molecules included in our sample, ozone may be considered a special case, representing a system with large static and dynamic correlation corrections for which, in wave function theory, a multiconfigurational treatment is mandatory. Moreover, its rotational *g* tensor is particularly sensitive to changes in the geometry.[49] We note that the magnitude of the largest component of the *g* tensor is very large compared to the average range of values—see Figure 2, where the experimental principal g_{xx} element of O_3 lies well outside the typical range of values. For these reasons, we have listed the ozone *g* tensor elements separately in Table 4.

The poor performance of the Hartree−Fock method for the *g* tensor of ozone is obvious, with a fortuitous cancellation of errors that gives results closer to experimental results at the optimized geometry than at the experimental geometry (as is typical of all Hartree−Fock calculations). The geometry dependence of the rotational *g* tensor for ozone is also noticeable for DFT, with the difference in *g* tensor elements between the optimized and experimental geometries typically several orders of magnitude larger than the experimental standard deviations. Again, we note the superior performance of the KT functionals, even for the difficult ozone system.

To compare with experimental results and to understand the differences in geometry with calculated properties, it is often important to consider zero-point vibrational effects at optimized geometries. In a previous MCSCF study of ozone,

Table 4. Calculated Rotational *g* Tensors of Ozone (aug-cc-pVTZ Basis)

method	g_{xx}	g_{yy}	g_{zz}
	Optimized Geometry		
HF	−4.4843	−0.2578	−0.0681
LDA	−3.5619	−0.2285	−0.0787
BLYP	−3.5878	−0.2398	−0.0811
B3LYP	−3.7853	−0.2460	−0.0767
KT1	−3.1310	−0.2295	−0.0799
KT2	−3.0752	−0.2238	−0.0782
KT3	−3.1809	−0.2233	−0.0773
	Experimental Geometry		
HF	−5.7515	−0.3992	−0.0675
LDA	−3.3687	−0.2473	−0.0808
BLYP	−3.2387	−0.2460	−0.0826
B3LYP	−3.6049	−0.2677	−0.0786
KT1	−2.8966	−0.2335	−0.0812
KT2	−2.9021	−0.2334	−0.0797
KT3	−2.9748	−0.2355	−0.0791
exptl	−2.9877(9)	−0.2295(3)	−0.0760(3)

the zero-point vibrational corrections were found to be −0.0707 (g_{xx}), −0.0018 (g_{yy}), and −0.0014 (g_{zz}).[49] For the principal g_{xx} component, therefore, the inclusion of vibrational corrections worsens the agreement with experiment, for all methods. Moreover, for the g_{zz} component, only the Hartree−Fock result is closer to experimental results with the inclusion of vibrational corrections, whereas for the g_{yy} component, only the KT2 and KT3 functional results are closer to experimental results.

Since ozone is such a difficult special case, we have repeated the statistical analysis with ozone excluded from the sample set. Comparing Figures 1 and 3, we see that the exclusion of ozone markedly reduces the Hartree−Fock deviations from experimental results. Also, with the exclusion of ozone, the B3LYP performance improves markedly relative to that of the KT2 functional. Thus, with the exception of the KT functionals, ozone has a very clear and disproportionate effect on the total calculated statistics.

Table 5. Calculated *g* Tensors at Optimized Geometries, Linear Molecules

method	experiment	HF	LDA	BLYP	B3LYP	KT1	KT2	KT3
CO	−0.2689 ± 0.0001	−0.2736	−0.2864	−0.2832	−0.2811	−0.2723	−0.2698	−0.2744
CS	−0.2702 ± 0.0004	−0.2921	−0.3026	−0.2912	−0.2926	−0.2814	−0.2794	−0.2819
CSe	−0.2431 ± 0.0016	−0.2919	−0.2758	−0.2666	−0.2734	−0.2565	−0.2553	−0.2562
$HC^{15}N$	−0.0904 ± 0.0003	−0.0721	−0.1042	−0.1034	−0.0963	−0.0955	−0.0937	−0.0975
$FC^{15}N$	−0.0504 ± 0.0008	−0.0497	−0.0530	−0.0518	−0.0518	−0.0480	−0.0484	−0.0487
$ClC^{15}N$	−0.04121 ± 0.000 13	−0.0365	−0.0422	−0.0415	−0.0404	−0.0403	−0.0403	−0.0406
$BrC^{15}N$	−0.0325 ± 0.0010	−0.0315	−0.0354	−0.0350	−0.0342	−0.0342	−0.0342	−0.0344
OCS	−0.0288 39 ± 0.000 006	−0.0271	−0.0304	−0.0313	−0.0301	−0.0296	−0.0294	−0.0297
$OC^{34}S$	−0.0282 42 ± 0.000 010	−0.0265	−0.0298	−0.0306	−0.0295	−0.0290	−0.0288	−0.0291
$O^{13}CS$	−0.0287 10 ± 0.000 015	−0.0270	−0.0303	−0.0311	−0.0300	−0.0295	−0.0293	−0.0296
^{15}NNO	−0.076 06 ± 0.000 10	−0.0768	−0.0802	−0.0807	−0.0801	−0.0781	−0.0779	−0.0784
^{14}NNO	−0.078 87 ± 0.000 08	−0.0793	−0.0831	−0.0835	−0.0829	−0.0809	−0.0806	−0.0811
$OC^{80}Se$	−0.019 52 ± 0.000 10	−0.0176	−0.0212	−0.0219	−0.0209	−0.0206	−0.0204	−0.0208
$OC^{76}Se$	−0.019 69 ± 0.000 10	−0.0178	−0.0214	−0.0221	−0.0211	−0.0208	−0.0206	−0.0210
HCP	−0.0430 ± 0.0010	−0.0311	−0.0532	−0.0507	−0.0471	−0.0446	−0.0439	−0.0454
DCP	−0.0353 ± 0.0010	−0.0266	−0.0454	−0.0434	−0.0402	−0.0382	−0.0376	−0.0389
HBS	−0.0414 ± 0.0002	−0.0288	−0.0496	−0.0472	−0.0433	−0.0427	−0.0414	−0.0414
FCCH	−0.0077 ± 0.0002	−0.0050	−0.0098	−0.0106	−0.0093	−0.0077	−0.0078	−0.0084
$^{35}ClCCH$	−0.006 30 ± 0.000 14	−0.0029	−0.0089	−0.0090	−0.0075	−0.0077	−0.0077	−0.0079
$^{37}ClCCH$	−0.006 01 ± 0.000 08	−0.0027	−0.0087	−0.0088	−0.0073	−0.0075	−0.0075	−0.0077
$^{79}BrCCH$	−0.003 95 ± 0.000 32	−0.0017	−0.0066	−0.0070	−0.0057	−0.0061	−0.0061	−0.0063
$^{81}BrCCH$	−0.003 88 ± 0.000 14	−0.0016	−0.0065	−0.0070	−0.0056	−0.0061	−0.0060	−0.0063

Table 6. Calculated *g* Tensors at Optimized Geometries, Symmetric Tops

method	experiment	HF	LDA	BLYP	B3LYP	KT1	KT2	KT3
$^{15}NH_3$	0.563 ± 0.002	0.5906	0.5870	0.5427	0.5626	0.5416	0.5488	0.5386
	0.500 ± 0.002	0.5083	0.5195	0.4780	0.4914	0.4846	0.4879	0.4791
CHF_3	−0.0359 ± 0.0008	−0.0349	−0.0392	−0.0388	−0.0377	−0.0382	−0.0380	−0.0379
	−0.031[a]	−0.0329	−0.0345	−0.0347	−0.0342	−0.0340	−0.0339	−0.0336
$CH_3{}^{14}NC$	−0.0546 ± 0.0015	−0.0548	−0.0643	−0.0631	−0.0612	−0.0611	−0.0607	−0.0614
	0.31[a]	0.2990	0.3275	0.2782	0.2917	0.2862	0.2893	0.2776
$CH_3C^{15}N$	−0.0317 ± 0.0003	−0.0283	−0.0352	−0.0370	−0.0346	−0.0353	−0.0346	−0.0355
	0.31[a]	0.3315	0.3590	0.3056	0.3209	0.3163	0.3198	0.3082
$CH_3C^{14}N$	−0.0338 ± 0.0008	−0.0302	−0.0373	−0.0391	−0.0366	−0.0373	−0.0367	−0.0376
	0.31[a]	0.3315	0.3590	0.3056	0.3209	0.3163	0.3198	0.3082
$CD_3C^{14}N$	−0.0315 ± 0.0008	−0.0290	−0.0349	−0.0364	−0.0343	−0.0348	−0.0343	−0.0351
	0.31[a]	0.1659	0.1797	0.1529	0.1606	0.1583	0.1600	0.1542
CH_3F	−0.062 ± 0.002	−0.0589	−0.0715	−0.0682	−0.0655	−0.0687	−0.0681	−0.0678
	0.265 ± 0.008	0.2654	0.2793	0.2415	0.2555	0.2457	0.2485	0.2374
CH_3Cl	−0.0165 ± 0.003	−0.0095	−0.0202	−0.0193	−0.0166	−0.0196	−0.0192	−0.0189
	0.305[a]	0.2959	0.3095	0.2709	0.2835	0.2744	0.2765	0.2664

[a] Assumed value.

Table 7. Calculated *g* Tensors at Optimized Geometries, Nonring Planar Asymmetric Tops

method	experiment	HF	LDA	BLYP	B3LYP	KT1	KT2	KT3
methyl butadiene	−0.0621 ± 0.0013	−0.0569	−0.0726	−0.0712	−0.0679	−0.0690	−0.0684	−0.0679
C_5H_8	−0.0339 ± 0.0016	−0.0315	−0.0368	−0.0359	−0.0349	−0.0355	−0.0351	−0.0345
	0.0080 ± 0.0016	0.0118	0.0121	0.0078	0.0094	0.0098	0.0101	0.0090
acrolein	−0.5512 ± 0.0019	−0.5019	−0.6409	−0.6171	−0.5917	−0.5874	−0.5820	−0.5842
C_3H_4O	−0.0567 ± 0.0010	−0.0571	−0.0618	−0.0601	−0.0598	−0.0590	−0.0587	−0.0583
	−0.0080 ± 0.0010	−0.0057	−0.0077	−0.0098	−0.0084	−0.0085	−0.0084	−0.0085
propene	−0.0789 ± 0.0006	−0.0660	−0.0981	−0.1063	−0.0953	−0.0960	−0.0949	−0.0983
C_3H_6	−0.0424 ± 0.0004	−0.0401	−0.0468	−0.0470	−0.0452	−0.0466	−0.0459	−0.0456
	0.0107 ± 0.0005	0.0150	0.0120	0.0075	0.0101	0.0088	0.0095	0.0092
propynal	−0.553 ± 0.002	−0.5135	−0.6525	−0.6184	−0.5970	−0.5876	−0.5836	−0.5880
C_3H_2O	−0.040 ± 0.002	−0.0359	−0.0458	−0.0440	−0.0426	−0.0412	−0.0409	−0.0408
	−0.015 ± 0.001	−0.0110	−0.0165	−0.0168	−0.0155	−0.0149	−0.0148	−0.0151
methoxy methane	−0.0214 ± 0.0006	−0.0036	−0.0441	−0.0437	−0.0309	−0.0505	−0.0482	−0.0425
CH_3OCH_3	−0.0093 ± 0.0004	−0.0053	−0.0089	−0.0118	−0.0095	−0.0115	−0.0110	−0.0115
	−0.0210 ± 0.0006	−0.0190	−0.0230	−0.0256	−0.0234	−0.0254	−0.0250	−0.0250
dimethyl sulfane	−0.0193 ± 0.0007	−0.0089	−0.0307	−0.0284	−0.0224	−0.0332	−0.0324	−0.0291
CH_3SCH_3	0.0000 ± 0.0003	0.0052	0.0017	−0.0023	0.0003	−0.0008	−0.0004	−0.0006
	−0.0083 ± 0.0003	−0.0050	−0.0075	−0.0105	−0.0087	−0.0084	−0.0083	−0.0087
acetaldehyde	−0.3609 ± 0.0021	−0.3306	−0.3999	−0.3977	−0.3836	−0.3711	−0.3660	−0.3670
C_2H_4O	−0.0731 ± 0.0003	−0.0718	−0.0783	−0.0782	−0.0770	−0.0762	−0.0752	−0.0755
	−0.0245 ± 0.0006	−0.0222	−0.0251	−0.0278	−0.0259	−0.0249	−0.0243	−0.0247
formaldehyde	−2.9017 ± 0.0008	−2.6654	−3.2558	−3.0868	−3.0179	−2.9677	−2.9331	−2.9174
H_2CO	−0.2243 ± 0.0001	−0.2190	−0.2390	−0.2363	−0.2329	−0.2316	−0.2286	−0.2293
	−0.0994 ± 0.0001	−0.0827	−0.1196	−0.1149	−0.1057	−0.1102	−0.1085	−0.1088
thioformaldehyde	−5.6202 ± 0.0068	−4.7503	−6.1903	−5.8301	−5.6519	−5.5662	−5.5358	−5.4496
H_2CS	−0.1337 ± 0.0004	−0.1335	−0.1452	−0.1412	−0.1403	−0.1371	−0.1354	−0.1352
	−0.0239 ± 0.0004	−0.0108	−0.0292	−0.0294	−0.0241	−0.0286	−0.0274	−0.0265
formic acid	−0.0903 ± 0.0006	−0.0891	−0.0964	−0.0942	−0.0936	−0.0926	−0.0922	−0.0914
HCOOH	−0.2797 ± 0.0060	−0.2511	−0.3002	−0.3003	−0.2913	−0.2840	−0.2811	−0.2877
	−0.0270 ± 0.0006	−0.0267	−0.0271	−0.0290	−0.0282	−0.0280	−0.0278	−0.0281
formamide	−0.0649 ± 0.0004	−0.0627	−0.0685	−0.0686	−0.0675	−0.0671	−0.0665	−0.0662
$HCONH_2$	−0.2843 ± 0.0011	−0.2530	−0.3186	−0.3175	−0.3050	−0.2969	−0.2948	−0.3002
	−0.0117 ± 0.0004	−0.0091	−0.0104	−0.0130	−0.0117	−0.0117	−0.0115	−0.0119
glycoaldehyde	−0.0726 ± 0.0010	−0.0389	−0.0428	−0.0421	−0.0414	−0.0416	−0.0412	−0.0411
$C_2H_4O_2$	−0.1239 ± 0.0013	−0.3136	−0.3990	−0.3754	−0.3644	−0.3621	−0.3589	−0.3575
	−0.0178 ± 0.0010	−0.0198	−0.0231	−0.0234	−0.0223	−0.0228	−0.0227	−0.0225
methyl formate	−0.0391 ± 0.0009	−0.0350	−0.0384	−0.0376	−0.0372	−0.0373	−0.0371	−0.0370
$C_2H_4O_2$	−0.1267 ± 0.0010	−0.2244	−0.2662	−0.2707	−0.2608	−0.2577	−0.2556	−0.2581
	−0.0168 ± 0.0017	−0.0200	−0.0214	−0.0218	−0.0213	−0.0212	−0.0212	−0.0213
ketene	−0.4182 ± 0.0009	−0.3208	−0.2931	−0.4391	−0.4030	−0.3812	−0.3715	−0.4015
$C_2H_2O_2$	−0.0356 ± 0.0013	−0.0349	−0.0363	−0.0383	−0.0372	−0.0359	−0.0354	−0.0361
	−0.0238 ± 0.0006	−0.0243	−0.0241	−0.0284	−0.0269	−0.0262	−0.0256	−0.0266
difluoro methane	−0.0411 ± 0.0004	−0.0407	−0.0441	−0.0427	−0.0421	−0.0422	−0.0421	−0.0419
CH_2F_2	−0.0725 ± 0.0006	−0.0589	−0.0963	−0.0872	−0.0789	−0.0906	−0.0894	−0.0891
	−0.0398 ± 0.0004	−0.0389	−0.0420	−0.0425	−0.0414	−0.0421	−0.0418	−0.0416
carbonic difluoride	−0.0568 ± 0.0006	−0.0538	−0.0618	−0.0604	−0.0594	−0.0586	−0.0585	−0.0582
F_2CO	−0.0747 ± 0.0004	−0.0756	−0.0799	−0.0777	−0.0778	−0.0749	−0.0750	−0.0751
	−0.0328 ± 0.0006	−0.0327	−0.0336	−0.0338	−0.0334	−0.0329	−0.0330	−0.0328
formyl fluoride	−0.4227 ± 0.0007	−0.4108	−0.4719	−0.4550	−0.4495	−0.4417	−0.4389	−0.4422
HFCO	−0.0771 ± 0.0002	−0.0777	−0.0822	−0.0805	−0.0804	−0.0787	−0.0785	−0.0782
	−0.0371 ± 0.0002	−0.0373	−0.0390	−0.0398	−0.0390	−0.0387	−0.0386	−0.0387
fluoro ethene	−0.1533 ± 0.0008	−0.1392	−0.1872	−0.1839	−0.1735	−0.1732	−0.1726	−0.1769
C_2H_3F	−0.0526 ± 0.0001	−0.0526	−0.0575	−0.0568	−0.0560	−0.0554	−0.0551	−0.0551
	−0.0037 ± 0.0001	−0.0011	−0.0034	−0.0066	−0.0047	−0.0055	−0.0051	−0.0055
1,1-difluoro ethene	−0.0421 ± 0.0005	−0.0407	−0.0453	−0.0470	−0.0455	−0.0446	−0.0446	−0.0450
CH_2CF_2	−0.0466 ± 0.0004	−0.0482	−0.0548	−0.0527	−0.0518	−0.0508	−0.0507	−0.0508
	−0.0119 ± 0.0004	−0.0107	−0.0121	−0.0140	−0.0128	−0.0132	−0.0130	−0.0131
cis-difluoro ethene	−0.1015 ± 0.0009	−0.0999	−0.1180	−0.1167	−0.1121	−0.1162	−0.1154	−0.1153
CHFCHF	−0.0296 ± 0.0003	−0.0292	−0.0332	−0.0324	−0.0318	−0.0309	−0.0308	−0.0308
	−0.0158 ± 0.0002	−0.0151	−0.0169	−0.0174	−0.0167	−0.0167	−0.0166	−0.0168
fluoro ethane	0.0185 ± 0.0006	0.0278	0.0119	0.0071	0.0147	0.0056	0.0075	0.0077
C_2H_5F	−0.0124 ± 0.0003	−0.0095	−0.0118	−0.0144	−0.0125	−0.0135	−0.0130	−0.0133
	−0.0197 ± 0.0004	−0.0196	−0.0213	−0.0239	−0.0222	−0.0230	−0.0226	−0.0227
trifluoro ethene	−0.0503 ± 0.0002	−0.0494	−0.0578	−0.0570	−0.0550	−0.0555	−0.0553	−0.0554
C_2HF_2	−0.0321 ± 0.0002	−0.0331	−0.0349	−0.0344	−0.0341	−0.0336	−0.0335	−0.0333
	−0.0170 ± 0.0002	−0.0172	−0.0173	−0.0180	−0.0177	−0.0173	−0.0173	−0.0174
ozone	−2.9877 ± 0.0009	−4.4843	−3.5619	−3.5878	−3.7853	−3.1310	−3.0752	−3.1809

Table 7. Continued

method	experiment	HF	LDA	BLYP	B3LYP	KT1	KT2	KT3
O_3	−0.2295 ± 0.0003	−0.2578	−0.2285	−0.2398	−0.2460	−0.2295	−0.2238	−0.2233
	−0.0760 ± 0.0003	−0.0681	−0.0787	−0.0811	−0.0767	−0.0799	−0.0782	−0.0773
sulfur dioxide	−0.6043 ± 0.0003	−0.6257	−0.6833	−0.7029	−0.6848	−0.6819	−0.6664	−0.6767
SO_2	−0.116 34 ± 0.000 12	−0.1159	−0.1242	−0.1260	−0.1240	−0.1234	−0.1221	−0.1220
	−0.088 65 ± 0.000 10	−0.0889	−0.0839	−0.0830	−0.0846	−0.0816	−0.0821	−0.0818
difluoro oxide	−0.213 ± 0.005	−0.1335	−0.2561	−0.2873	−0.2218	−0.2438	−0.2324	−0.2408
OF_2	−0.058 ± 0.002	−0.0444	−0.0634	−0.0679	−0.0594	−0.0618	−0.0600	−0.0605
	−0.068 ± 0.002	−0.0583	−0.0742	−0.0787	−0.0728	−0.0739	−0.0722	−0.0721
hypofluoros acid	0.642 ± 0.001	0.7050	0.6680	0.6430	0.6641	0.6407	0.6496	0.6450
HOF	−0.119 ± 0.001	−0.0856	−0.1239	−0.1349	−0.1184	−0.1223	−0.1182	−0.1205
	−0.061 ± 0.001	−0.0460	−0.0733	−0.0769	−0.0675	−0.0706	−0.0686	−0.0701
water	0.718 ± 0.007	0.7468	0.7309	0.7001	0.7168	0.6984	0.7051	0.6982
H_2O	0.657 ± 0.001	0.6911	0.6591	0.6271	0.6497	0.6250	0.6356	0.6326
	0.645 ± 0.006	0.6684	0.6534	0.6206	0.6376	0.6212	0.6268	0.6200
hydrogen sulfide	0.355 ± 0.008	0.3893	0.4286	0.3692	0.3844	0.4008	0.4030	0.3907
H_2S	0.195 ± 0.008	0.1767	0.2419	0.1586	0.1756	0.1945	0.1942	0.1747
	0.209 ± 0.008	0.1966	0.2858	0.2151	0.2211	0.2488	0.2467	0.2299

Table 8. Calculated *g* Tensors at Optimized Geometries, Three-Member Rings

method	experiment	HF	LDA	BLYP	B3LYP	KT1	KT2	KT3
methylenecyclopropane	−0.0672 ± 0.0007	−0.0542	−0.0769	−0.0805	−0.0734	−0.0772	−0.0761	−0.0769
C_4H_6	−0.0231 ± 0.0004	−0.0193	−0.0258	−0.0277	−0.0253	−0.0256	−0.0250	−0.0257
	0.0244 ± 0.0004	0.0306	0.0267	0.0227	0.0254	0.0239	0.0245	0.0240
cyclopropene	−0.0897 ± 0.0009	−0.0918	−0.0981	−0.1010	−0.0977	−0.0941	−0.0927	−0.0958
C_3H_4	−0.1492 ± 0.0002	−0.1479	−0.1665	−0.1623	−0.1590	−0.1630	−0.1621	−0.1610
	0.0536 ± 0.0002	0.0629	0.0529	0.0477	0.0527	0.0479	0.0488	0.0473
aziridine	0.0229 ± 0.0009	0.0379	0.0206	0.0157	0.0224	0.0193	0.0204	0.0198
C_2H_5N	−0.0422 ± 0.0008	−0.0254	−0.0503	−0.0505	−0.0433	−0.0460	−0.0453	−0.0453
	0.0539 ± 0.0010	0.0613	0.0550	0.0478	0.0524	0.0525	0.0531	0.0510
oxirane	0.0189 ± 0.0004	0.0329	0.0139	0.0090	0.0164	0.0107	0.0123	0.0113
C_2H_4O	−0.0946 ± 0.0003	−0.0889	−0.1094	−0.1065	−0.1012	−0.1010	−0.1009	−0.1020
	0.0318 ± 0.0006	0.0381	0.0341	0.0277	0.0317	0.0325	0.0328	0.0309
thiirane	−0.0159 ± 0.0021	0.0033	−0.0187	−0.0240	−0.0159	−0.0186	−0.0177	−0.0182
C_2H_4S	−0.0242 ± 0.0003	−0.0149	−0.0302	−0.0308	−0.0262	−0.0300	−0.0294	−0.0291
	0.0487 ± 0.0004	0.0551	0.0501	0.0454	0.0486	0.0493	0.0496	0.0479
cyclopropenone	−0.2900 ± 0.0013	−0.2615	−0.3186	−0.3101	−0.3002	−0.2988	−0.2956	−0.2955
C_3H_2O	−0.0963 ± 0.0004	−0.0938	−0.1058	−0.1035	−0.1014	−0.1005	−0.0999	−0.0997
	−0.0121 ± 0.0004	−0.0062	−0.0128	−0.0154	−0.0124	−0.0165	−0.0161	−0.0157
methyl cyclopropene	−0.0813 ± 0.0070	−0.0816	−0.0927	−0.0968	−0.0734	−0.0951	−0.0940	−0.0954
C_4H_6	−0.0261 ± 0.0040	−0.0276	−0.0296	−0.0320	−0.0253	−0.0312	−0.0307	−0.0309
	0.0166 ± 0.0030	0.0220	0.0188	0.0152	0.0254	0.0155	0.0161	0.0156

D. Comments on Other Specific Molecules. The accuracy and applicability of DFT *g* tensors may further be illustrated by examining some specific cases. In Tables 5−9, we present individual *g* tensors calculated at optimized geometries with the aug-cc-pVTZ basis set.

As discussed above, the two experimental values reported for ClCN are those of Reinartz et al.[66] (−0.041 21 ± 0.000 13) and Flygare (−0.0384 ± 0.0003). Our London DFT results support the more recent value of Reinartz et al., although vibrational corrections are needed for a more definite assessment.

Next, comparing the *g* values of CO, CS, and CSe in Table 5, we note that the GGA and hybrid DFT results are sufficiently accurate to distinguish properly among these analogues, giving the correct ordering $|g_{CSe}| < |g_{CO}| < |g_{CS}|$. By contrast, the Hartree−Fock and LDA models mix CO and CSe, giving instead $|g_{CSe}| < |g_{CS}| < |g_{CO}|$. Again, the best results are provided by the KT2 functional, with all other methods giving *g* values that are too large in magnitude.

The halogenated cyanides HCN, FCN, ClCN, and BrCN may be readily distinguished from their calculated rotational *g* tensors, in agreement with experimental results. By contrast, at the optimized geometries, the results for the halogenated acetylenes FCCH, ClCCH, and BrCCH do not reproduce the experimental trends accurately, except for the usually unreliable Hartree−Fock model (which gets trends right but severely underestimates the *g* values). However, this situation changes when the calculations are carried out at experimental rather than optimized geometries: the Hartree−Fock model then no longer works, but all DFT functionals distinguish correctly between the three species.

In the experimental studies of the CH_3CN isotopes and isomers, the parallel *g* component was given an assumed value of 0.310. Our calculations indicate that this assumption

Table 9. Calculated *g* Tensors at Optimized Geometries, Four-Member Rings

method	experiment	HF	LDA	BLYP	B3LYP	KT1	KT2	KT3
cyclobutene	−0.0516 ± 0.0007	−0.0496	−0.0596	−0.0604	−0.0574	−0.0580	−0.0571	−0.0578
C_4H_6	−0.0663 ± 0.0007	−0.0614	−0.0771	−0.0777	−0.0734	−0.0773	−0.0765	−0.0758
	−0.0219 ± 0.0006	−0.0146	−0.0180	−0.0243	−0.0207	−0.0270	−0.0257	−0.0256
oxetane	−0.0073 ± 0.0005	0.0013	−0.0127	−0.0146	−0.0094	−0.0146	−0.0131	−0.0135
C_3H_6O	−0.0429 ± 0.0007	−0.0357	−0.0497	−0.0501	−0.0455	−0.0491	−0.0481	−0.0482
	−0.0747 ± 0.0005	−0.0683	−0.0806	−0.0799	−0.0768	−0.0847	−0.0832	−0.0813
methylenecyclobutane	−0.0320 ± 0.0008	−0.0243	−0.0364	−0.0420	−0.0368	−0.0381	−0.0372	−0.0385
C_5H_8	−0.0218 ± 0.0010	−0.0196	−0.0241	−0.0258	−0.0239	−0.0241	−0.0235	−0.0240
	−0.0184 ± 0.0011	−0.0132	−0.0151	−0.0193	−0.0172	−0.0185	−0.0177	−0.0180
cyclobutanone	−0.0740 ± 0.0020	−0.0737	−0.0953	−0.0951	−0.0903	−0.0891	−0.0876	−0.0887
C_4H_6O	−0.0325 ± 0.0004	−0.0329	−0.0358	−0.0375	−0.0361	−0.0359	−0.0353	−0.0356
	−0.0279 ± 0.0004	−0.0270	−0.0271	−0.0296	−0.0287	−0.0290	−0.0285	−0.0285
β-propiolactone	−0.0758 ± 0.0005	−0.0708	−0.0862	−0.0893	−0.0845	−0.0853	−0.0841	−0.0850
$C_3H_4O_2$	−0.0356 ± 0.0004	−0.0355	−0.0396	−0.0398	−0.0386	−0.0391	−0.0387	−0.0389
	−0.0319 ± 0.0004	−0.0317	−0.0315	−0.0334	−0.0327	−0.0332	−0.0328	−0.0328
diketene	−0.1091 ± 0.0004	−0.1000	−0.1226	−0.1246	−0.1185	−0.1193	−0.1180	−0.1188
$C_4H_4O_2$	−0.0324 ± 0.0004	−0.0337	−0.0359	−0.0357	−0.0351	−0.0348	−0.0346	−0.0348
	−0.0169 ± 0.0004	−0.0169	−0.0167	−0.0185	−0.0177	−0.0183	−0.0180	−0.0179

is valid for $CH_3{}^{14}NC$, $CH_3C^{14}N$, and $CH_3C^{15}N$ but not for $CD_3C^{14}N$. For all methods, the parallel component for $CD_3C^{14}N$ is $^1/_2$ that for $CH_3C^{14}N$, indicating that a more appropriate value for the parallel component of $CD_3C^{14}N$ is 0.16 ± 0.01. By contrast, our calculations support the assumptions made in the experimental studies of CHF_3 and CH_3Cl, with assumed parallel *g* values of −0.031 and 0.305, respectively.

Apart from ozone, the two statistical outliers are the principal *g* tensor components of H_2CO (−2.9017) and H_2CS (−5.6202). In both cases, the optimized geometries agree with experimental results and, thus, contribute little to the observed differences from experimental *g* tensors. We note the particularly poor performance of the Hartree−Fock and LDA methods for these systems.

For methoxy methane, and to a lesser extent for dimethyl sulfane, all methods poorly reproduce the experimental g_{xx} value. The inclusion of core functions in the augmented correlation-consistent polarized core-valence basis sets aug-cc-pCVDZ and aug-cc-pCVTZ for dimethyl sulfane consistently makes all tensor components more negative. Moreover, the convergence from the aug-cc-pVDZ to aug-cc-pVTZ basis sets similarly makes all tensor components more negative. In light of these results, we suggest that the experimental g_{xx} components of these two molecules may be too positive.

E. Geometry Dependence. It has been shown previously[49] that calculated rotational *g* tensors are sensitive to molecular geometry, necessitating the use of accurate molecular geometries. To illustrate this dependence, the significant deviations between the calculated *g* tensor elements of glycoaldehyde and methyl formate in Table 7 may be attributed to geometry effects, with tensor elements calculated at the experimental geometries typically within 0.005 of experimental results, in contrast to the deviations of greater than 0.2 for the principal *g*-tensor component of glycoaldehyde.

We have considered the geometry dependence of calculated rotational *g* tensors by employing both experimentally determined and theoretically calculated equilibrium molecular geometries; that is, we consider both the intrinsic error of the method and the error associated with an optimized geometry.

From an analysis of the calculated geometries, we noted a clear trend for the methods considered in this work in that going from a double-ζ to a triple-ζ basis, the calculated bond lengths generally increased while bond angles were diminished. From Figures 1 and 3, it would appear that the GGA and hydrid DFT methods considered give more accurate results with a double-ζ basis set than with a triple-ζ basis set for both optimized and experimental geometries. In contrast, for the Hartree−Fock and LDA methods, triple-ζ results are typically closer to experimental results than the double-ζ results. For the calculation of rotational *g* tensors from optimized geometries, all methods, Hartree−Fock and LDA in particular, benefit from a fortuitous cancellation of errors; that is, the intrinsic error of these methods is largely canceled by the error associated with longer optimized bond lengths.

For the GGA and hybrid DFT functionals, the smallest deviation from experimental *g* tensors was calculated at experimental geometries, reinforcing previous conclusions[49] that accurate geometries are necessary for accurate calculations of rotational *g* tensors.

V. Conclusions

We have carried out extensive benchmarking of a number of DFT functionals with respect to the calculation of rotational *g* tensors. The performance and reliability of the DFT approach has been shown to be superior to Hartree−Fock theory in the calculation of this property, with GGA functionals performing better than the LDA functional. The KT functionals show the widest applicability and reliability of the functionals considered in this work, giving good results also for the difficult ozone molecule.

With the use of London orbitals, basis-set convergence is rapid for both Hartree−Fock and DFT methods, with good

agreement with experimental results in the aug-cc-pVDZ basis, and near-basis set limit results obtained in an aug-cc-pVTZ basis.

Acknowledgment. This work has received support from the Norwegian Reseach Council through a Strategic University Program in Quantum Chemistry (Grant 154011/420). D.J.D.W. has been supported by the Norwegian Research Council through a postdoctoral fellowship (Grant 155137/432). We also acknowledge a grant of computer time from the Norwegian Supercomputing Program.

References

(1) Ramsey, N. F. *Molecular Beams*; Clarendon Press: Oxford, 2000.

(2) Flygare, W. H.; Benson, R. C. *Mol. Phys.* **1971**, *20*, 225.

(3) Flygare, W. H. *Chem. Rev.* **1974**, *74*, 653.

(4) Kelly, H. M.; Fowler, P. W. *Chem. Phys. Lett.* **1993**, *206*, 568.

(5) Gauss, J.; Ruud, K.; Helgaker, T. *J. Chem. Phys.* **1996**, *105*, 2804.

(6) Ruud, K.; Helgaker, T. *Chem. Phys. Lett.* **1997**, *264*, 17.

(7) Bishop, D. M.; Cybulski, S. M. *J. Chem. Phys.* **1993**, *98*, 8057.

(8) Cybulski, S. M.; Bishop, D. M. *J. Chem. Phys.* **1994**, *100*, 2019.

(9) Cybulski, S. M.; Bishop, D. M. *J. Chem. Phys.* **1996**, *106*, 4082.

(10) Sauer, S. P. A.; Špirko, V.; Oddershede, J. *Chem. Phys.* **1991**, *153*, 189.

(11) Sauer, S. P. A. *Chem. Phys. Lett.* **1998**, *297*, 475.

(12) Ogilvie, J. F.; Oddershede, J.; Sauer, S. P. A. *Adv. Chem. Phys.* **1998**, *111*, 475.

(13) Åstrand, P. O.; Mikkelsen, K. V.; Ruud, K.; Helgaker, T. *J. Phys. Chem.* **1996**, *100*, 19771.

(14) Ruud, K.; Helgaker, T.; Jørgensen, P. *J. Chem. Phys.* **1997**, *107*, 10599.

(15) Ruud, K.; Vaara, J.; Lounila, J.; Helgaker, T. *Chem. Phys. Lett.* **1998**, *297*, 467.

(16) Ruud, K.; Åstrand, P. O.; Helgaker, T.; Mikkelsen, K. V. *THEOCHEM* **1996**, *388*, 231.

(17) Åstrand, P. O.; Ruud, K.; Mikkelsen, K. V.; Helgaker, T. *Chem. Phys. Lett.* **1997**, *271*, 163.

(18) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(19) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(20) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200.

(21) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(22) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(23) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

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

(25) Paldus, J.; Čížek, J. *Adv. Quantum Chem.* **1975**, *9*, 105.

(26) Bartlett, R. J.; Purvis, G. D. *Int. J. Quantum Chem.* **1978**, *14*, 516.

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

(28) Chan, S. I.; Das, T. P. *J. Chem. Phys.* **1962**, *37*, 1527.

(29) Sadlej, A. J. *Chem. Phys. Lett.* **1975**, *36*, 129.

(30) Lazzaretti, P.; Zanasi, R. *J. Chem. Phys.* **1980**, *72*, 6768.

(31) Kutzelnigg, W. *Isr. J. Chem.* **1980**, *19*, 193.

(32) Schindler, M.; Kutzelnigg, W. *J. Chem. Phys.* **1982**, *76*, 1919.

(33) van Wüllen, C.; Kutzelnigg, W. *Chem. Phys. Lett.* **1993**, *205*, 563.

(34) Hansen, A. E.; Bouman, T. D. *J. Chem. Phys.* **1985**, *82*, 5035.

(35) London, F. *J. Phys. Radium* **1937**, *8*, 397.

(36) Bak, K. L.; Jørgensen, P.; Helgaker, T.; Ruud, K.; Jensen, H. J. A. *J. Chem. Phys.* **1993**, *98*, 8873.

(37) Hameka, H. F. *Mol. Phys.* **1958**, *1*, 203.

(38) Hameka, H. F. *Z. Naturforsch., A: Phys. Sci.* **1959**, *14*, 599.

(39) Hameka, H. F. *Rev. Mod. Phys.* **1962**, *34*, 87.

(40) Ditchfield, R. *J. Chem. Phys.* **1972**, *56*, 5688.

(41) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251.

(42) Häser, M.; Ahlrichs, R.; Baron, H. P.; Weis, P.; Horn, H. *Theor. Chim. Acta* **1992**, *83*, 455.

(43) Gauss, J. *Chem. Phys. Lett.* **1992**, *191*, 614.

(44) Ruud, K.; Helgaker, T.; Bak, K. L.; Jørgensen, P.; Jensen, H. J. A. *J. Chem. Phys.* **1993**, *99*, 3847.

(45) Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **1995**, *103*, 3561.

(46) Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **1996**, *104*, 2574.

(47) Gauss, J. *J. Chem. Phys.* **2002**, *116*, 4773.

(48) Wilson, P. J.; Amos, R. D.; Handy, N. C. *THEOCHEM* **2000**, *506*, 335.

(49) Mohn, C. E.; Wilson, D. J. D.; Lutnæs, O. B.; Helgaker, T.; Ruud, K. *Adv. Quantum Chem.* **2005**, *50*, in press.

(50) *DALTON, a Molecular Electronic Structure Program*, release 2.0; 2005. see http://www.kjemi.uio.no/software/dalton/dalton.html.

(51) Olsen, J.; Jørgensen, P. *J. Chem. Phys.* **1985**, *82*, 3235.

(52) Helgaker, T.; Wilson, P. J.; Amos, R. D.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 2983.

(53) Helgaker, T.; Watson, M.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 9402.

(54) Helgaker, T.; Pecul, M. In *Calculation of NMR and EPR Parameters. Theory and Applications*; Kaupp, M., Bühl, M., Malkin, V. G., Eds.; Wiley-VCH: Weinheim, Germany, 2004.

(55) Lantto, P.; Vaara, J.; Helgaker, T. *J. Chem. Phys.* **2002**, *117*, 5998.

(56) Lutnæs, O. B.; Ruden, T. A.; Helgaker, T. *Magn. Reson. Chem.* **2004**, *42*, S117.

(57) Keal, T. W.; Tozer, D. J. *J. Chem. Phys.* **2003**, *119*, 3015.

(58) Keal, T. W.; Tozer, D. J.; Helgaker, T. *Chem. Phys. Lett.* **2004**, *391*, 374.

(59) Keal, T. W.; Tozer, D. J. *J. Chem. Phys.* **2004**, *121*, 5654.

(60) Amos, R. D.; Alberts, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Laidig, K. E.; Laming, G.; Lee, A. M.; Maslen, P. E.;

Murray, C. W.; Rice, J. E.; Simandiras, E. D.; Stone, A. J.; Su, M.-D.; Tozer, D. J. *CADPAC, the cambridge analytic derivatives package*, issue 6; University of Cambridge: Cambridge, U. K., 2005.

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

(62) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(63) de Leeuw, F. H.; Dymanus, A. *Chem. Phys. Lett.* **1970**, *7*, 288.

(64) Ellenbroek, A. W.; Dymanus, A. *Chem. Phys. Lett.* **1976**, *42*, 303.

(65) Meerts, W. L.; Stolte, S.; Dymanus, A. *Chem. Phys.* **1977**, *19*, 467.

(66) Reinartz, J. M. L. J.; Meerts, W. L.; Dymanus, A. *Chem. Phys.* **1980**, *45*, 387.

(67) Ruud, K.; Skaane, H.; Helgaker, T.; Bak, K. L.; Jørgensen, P. *J. Am. Chem. Soc.* **1994**, *116*, 10135.

(68) NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, release 11. http://srdata.nist.gov/cccbdb (May, 2005).

(69) Maroulis, G.; Pouchan, C. *Chem. Phys.* **1997**, *215*, 67.

(70) LeGuennec, M.; Wlodarczak, G.; Demaison, J.; Büurger, H.; Litz, M.; Willner, H. *J. Mol. Spectrosc.* **1993**, *157*, 419.

(71) O'Neill, D. P.; Gill, P. M. W. *Mol. Phys.* **2005**, *103*, 763.

(72) Puzzarini, C. *Phys. Chem. Chem. Phys.* **2004**, *6*, 344.

(73) Maroulis, G. *J. Comput. Chem.* **2003**, *24*, 443.

(74) Margules, L.; Demaison, J.; Rudolph, H. D. *J. Mol. Struct.* **2001**, *599*, 23.

(75) Nakata, M.; Kohata, K.; Fukuyama, T.; Kuchitzu, K.; Wilkins, C. J. *J. Mol. Spectrosc.* **1980**, *68*, 271.

(76) Harmony, M. D.; Kuczkowski, R. W. L. R. L.; Schwendemann, R. H.; Ramsay, D. A.; Lovas, F. J.; Lafferty, W. J.; Maki, A. G. *J. Phys. Chem. Ref. Data* **1979**, *8*, 619.

(77) Huisman, P. A. G.; Klebe, K. J.; Mijlhoff, F. C.; Renes, G. H. *J. Mol. Struct.* **1979**, *57*, 71.

(78) Mom, V.; Huisman, P. A. G.; Mijlhoff, F. C.; Renes, G. H. *J. Mol. Struct.* **1980**, *62*, 95.

(79) Staley, S. W.; Norden, T. D.; Taylor, W. H.; Harmony, M. D. *J. Am. Chem. Soc.* **1987**, *109*, 7641.

CT050101T

JCTC Journal of Chemical Theory and Computation

Modeling β-Scission Reactions of Peptide Backbone Alkoxy Radicals: Backbone C−C Bond Fission

Geoffrey P. F. Wood,[†,‡] Arvi Rauk,[§] and Leo Radom*[,†,‡]

School of Chemistry, University of Sydney, Sydney, New South Wales 2006, Australia, Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia, and Department of Chemistry, University of Calgary, Calgary, Alberta T2N IN4, Canada

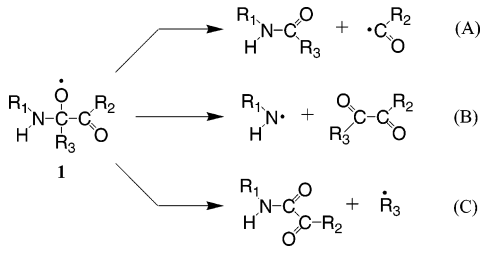
Received May 17, 2005

Abstract: To model the C−C β-scission reactions of backbone peptide alkoxy radicals, enthalpies and barriers for the fragmentation of four substituted alkoxy radicals have been calculated with a variety of ab initio molecular orbital theory and density functional theory procedures. The high-level methods examined include CBS-QB3, variants of the G3 family, and W1. Simpler methods include HF, MP2, QCISD, B3-LYP, BMK, and MPW1K with a range of basis sets. We find that good accuracy can be achieved with the G3(MP2)//B3-LYP and G3X(MP2)-RAD methods. Lower-cost methods producing reasonable results are single-point energy calculations with UB3-LYP/6-311+G(3df,2p), RB3-LYP/6-311+G(3df,2p), UBMK/6-311+G(3df,2p), and RBMK/6-311+G(3df,2p) on geometries optimized with UB3-LYP/6-31G(d) or UBMK/6-31G(d). Heats of formation at 0 K for the alkoxy radicals and their fragmentation products were also calculated. We predict $\Delta_f H_0$ values for the alkoxy radicals of −71.4 ($^{\bullet}OCH_2$-CH=O), −102.5 ($^{\bullet}OCH(CH_3)CH=O$), −176.6 ($^{\bullet}OCH(CH_3)C(NH2)=O$), and −264.6 ($^{\bullet}OC(CH_3)$-(NHCH=O)CH=O) kJ mol^{-1}. For the fragmentation products $NH_2C(^{\bullet})=O$ and CH(=O)NHC-(CH_3)=O, we predict $\Delta_f H_0$ values of −5.9 kJ mol^{-1} and −352.8 kJ mol^{-1}.

1. Introduction

Radical-mediated protein damage has been implicated in a number of diseases such as Alzheimer's disease, atherosclerosis, and diabetes as well as aging.[1−3] Two general observations can be made concerning research dedicated to this subject.[1] First, there are a vast number of possible reactions due to the complexity of proteins and the variety in functionality of peptide residues. Second, the complexity of the species and the possibility of long chain lengths make experimental elucidation of specific processes difficult to achieve.

It has been postulated that alkoxy radicals may form on the peptide backbone via peroxyl radicals through a tetroxide. Alternatively, they may be generated by one-electron reduction of alkyl hydroperoxides or dialkyl peroxides.[1] Once the alkoxy radical (**1**) has formed on a peptide backbone, it may undergo one of three possible fragmentation reactions, namely fission of the C−C bond (A), fission of the C−N bond (B), or loss of the R_3 side-chain (C):



1

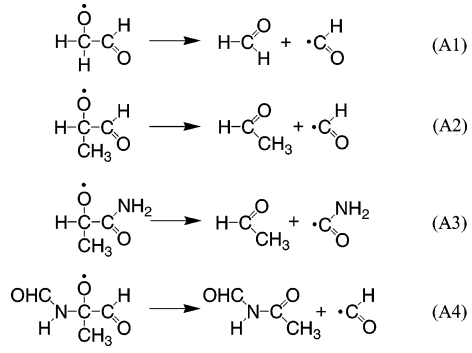
In general, alkoxy cleavage reactions have been found to be endothermic,[4] and the preferred pathway is generally determined by the stability of the radical that is formed.[5] However, this is not always the case, and for one particular class of ring-opening reactions,[6] that has been studied computation-

* Corresponding author e-mail: radom@chem.usyd.edu.au.
† University of Sydney.
‡ Australian National University.
§ University of Calgary.

10.1021/ct050133g CCC: $30.25
Published on Web 08/05/2005

ally,[7] the most stable thermodynamic products are not formed. Other computational studies in this area include investigations of the associated β-scission reactions of a general series of alkoxy radicals.[8,9] Good agreement between theory and experiment was found for reaction enthalpies and heats of formation at the B3-LYP/SVP, CBS-RAD, and G2-(MP2) levels of theory. Dibble[10] has also carried out theoretical studies on cleavage reactions of isoprene-derived alkoxy radicals. Using B3-LYP with small and medium-sized basis sets as well as the CBS-4 and CBS-q model chemistries he found that the cleavage barriers were small.

In a recent study, Huang and Rauk[11] performed an initial investigation of the fragmentation reactions A, B, and C. In that study, the B3-LYP/6-31G(d) method was used to determine free energies of fragmentation for model systems containing an alanine and glycine residue. In the present study, we examine in more detail the pathway leading to C−C bond fission (A) using four models (A1−A4). We use both high-level ab initio procedures and simpler calculations to investigate their reaction enthalpies and barriers.



The first reaction, A1, represents the smallest system that can be used to model reaction A. It is clearly a very simplified model but it allows the assessment of a wide range of theoretical procedures. At the other extreme, reactions A3 and A4 represent more realistic models, but because of their increased size they are more restricted in the range of theoretical procedures that can be applied.

We also make high-level predictions of the heats of formation at 0 K for the alkoxy radicals and for the cleavage products of reactions A1−A4. Heats of formation for some of these species have been previously determined experimentally and computationally, and we use the new calculated values in these cases to further assess the performance of the various theoretical techniques. In other cases, we predict heats of formation for species for which the values are not currently known. For example, reactions A3 and A4 involve substituted alkoxy radicals that have not been studied theoretically before, and for which no thermodynamical data are currently available. On the other hand, the radical product $NH_2C(^{\bullet})=O$, formed in the cleavage reaction A3, has been previously studied theoretically in the context of combustion chemistry[12] and also in the examination of N−H bond energies.[13] However, an accurate heat of formation is lacking.

2. Theoretical Procedures

Standard ab initio molecular orbital theory[14] and density functional theory[15] calculations were carried out with the ACES II,[16] GAUSSIAN 98,[17] GAUSSIAN 03,[18] and MOLPRO 2002.6[19] computer programs.

Unless otherwise stated, calculations have been carried out using an unrestricted reference wave function. Calculations on radicals that have used a restricted-open-shell reference wave function are designated with an "R" prefix. Those that use the partial spin restriction of MOLPRO[20] have a "UR" prefix. We have employed the frozen-core approximation, except in cases where prescribed model chemistries, e.g. G3//B3-LYP,[21] require full calculations to be carried out.

Geometries were initially scanned using the B3-LYP/6-31G(d) level of theory. Geometry optimizations were then carried out at points on the scans that indicated equilibrium structures of reactants and products or transition structures for their interconversion. The optimizations were followed by vibrational frequency analyses to ensure that the selected structures correspond to minima (no imaginary frequencies) or saddle points (one imaginary frequency). Intrinsic reaction coordinate calculations[22,23] were used to ensure that the minima are connected by the appropriate transition structures.

A number of levels of theory were used to optimize geometries. These include the HF, MP2, and QCISD ab initio methods and the B3-LYP, MPW1K, and BMK density functional theory methods, using a range of basis sets. BMK is a hybrid functional that has been recently formulated by Boese and Martin.[24] We found that, for some of the alkoxy radicals, the lowest-energy conformer varies between optimization levels. To resolve these discrepancies, URCCSD-(T) single-point calculations were carried out on the structures optimized at the various levels, with the lowest energy indicating a preferred structure.

A number of levels of theory were also used to determine the energy profiles for reactions A1−A4. These include simpler procedures such as HF, MP2, QCISD, MPW1K, BMK, and B3-LYP with a variety of basis sets. Energies at 0 K were obtained by incorporating zero-point vibrational energies (ZPVEs) using appropriate scale factors.[25] In cases for which a scale factor is not available, these were taken from a closely related method. Specifically, for HF and MP2 in association with the 6-311++G(3df,2p) and 6-31G(2df,p) basis sets, the corresponding scale factors from the 6-311G-(d,p) basis set were used. In the case of B3-LYP in association with the 6-311+G(3df,2p) and cc-pVTZ basis sets, the 6-31G(2df,p) scale factor[26] was used, and in the case of MPWIK/6-31+G(3df,2p), the MPWIK/6-31+G(d,p) scale factor[27] was used. For the QCISD/6-31G(d) energy profiles, we used scaled UB3-LYP/6-31G(d) ZPVEs.

We also examined the performance of simple methods in which the energies and geometries were determined at separate levels of theory. These include UB3-LYP/6-311+G-(3df,2p)//UB3-LYP/6-31G(d), RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), UBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d), RBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d), UBMK/6-311+G(3df,2p)//UBMK/6-31G(d), RBMK/6-311+G(3df,2p)//UBMK/6-31G(d), UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d), and RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d), corrected in each case with either scaled UB3-LYP/6-31G(d) ZPVEs or unscaled UBMK/6-31G(d) ZPVEs where appropriate.

Table 1. Calculated Differences in Energy (kJ mol^{-1}) between Conformers for the Alkoxy Radicals **a2** and **a4**[a]

level of theory	$\mathbf{a2_a} - \mathbf{a2_b}$	$\mathbf{a4_a} - \mathbf{a4_b}$	$\mathbf{a4_a} - \mathbf{a4_c}$
HF/6-31G(d)	−0.7	−0.4	9.6
HF/6-311++G(3df,2p)	0.6	0.8	9.5
MP2/6-31G(d)	−0.9	−1.1	4.9
MP2/6-31+G(d)	−1.0		
MP2/6-31+G(2df,p)	2.4		
B3-LYP/6-31G(d)	−3.0	−2.1	−10.2
B3-LYP/6-311+G(3df,2p)	−3.0	−1.4	−12.6
//HF/6-31G(d)	−0.4[b]	2.1[c]	1.5[c]
//MP2/6-31+G(2df,p)	0.9[b]		
//B3-LYP/6-31G(d)	0.6[b]	−0.7[c]	−5.2[c]
//B3-LYP/6-311+G(3df,2p)	−0.7[b]	−1.3[c]	−8.2[c]

[a] Energy differences are not corrected for zero-point vibrational energy. Geometries optimized at the same level as the energy calculations unless otherwise noted. [b] URCCSD(T)/6-311+G(3df,2p) calculations on the indicated geometry. [c] URCCSD(T)/6-311+G(d,p) calculations on the indicated geometry.

Finally, we used high-level composite methods such as CBS-QB3[28] and U-CBS-QB3[29] from the CBS family of methods, G3(MP2)//B3-LYP,[21] G3X(MP2)-RAD,[30] G3-RAD,[30] and G3//B3-LYP[21] from the G3 group of methods, and also the W1 method of Martin et al.[31] The performance and cost of CBS-QB3 is similar to that of CBS-RAD, which has previously been used to examine the cleavage of alkoxy radicals.[8]

3. Results and Discussion

3.1. Determination of Preferred Conformations. Our initial task was to determine the preferred conformations of the alkoxy radicals, the cleavage products, and the transition structures connecting reactants to products in each of the reactions A1−A4. These were initially screened with geometry optimizations at the B3-LYP/6-31G(d) level of theory. Further optimizations were carried out with B3-LYP and the larger 6-311+G(3df,2p) basis set and with the HF and MP2 methods in conjunction with the small and large basis sets. To examine the sensitivity of the relative energies of alternative possible conformations to the level of geometry optimization, single-point URCCSD(T) calculations were carried out on the various structures. The calculated relative energies are presented in Table 1. Based on the results of this analysis, discussed in more detail below, preferred structures for the alkoxy radicals (Figure 1, **a1**−**a4**), fragmentation products (Figure 2, **fp1**−**fp5**), and fragmentation (addition) transition structures (Figure 3, **ts1**−**ts4**) were identified. Figures 1−3 also include some of the more interesting structural parameters obtained at the B3-LYP/6-311+G(3df,2p) level of theory. Full structural data for equilibrium and transition structures are presented in the Supporting Information in the form of GAUSSIAN archive entries[17,18] for B3-LYP/6-311+G(3df,2p) optimized structures.

We begin our discussion by noting that acetaldehyde (CH_3-CH═O) is well known to have three equivalent conformational minima. Each of these conformers has a C−H bond eclipsing the C═O bond. If one of the methyl hydrogens of acetaldehyde is formally replaced by oxygen to give the

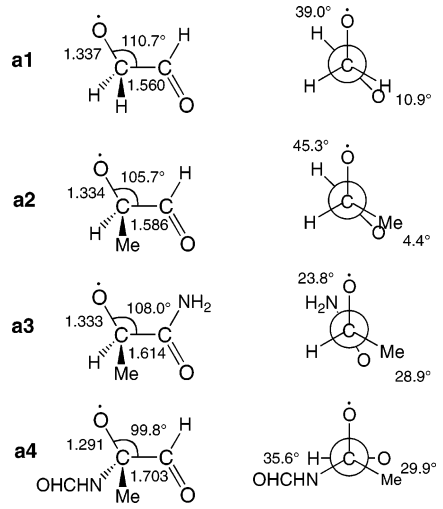


Figure 1. Optimized B3-LYP/6-311+G(3df,2p) structures of preferred conformers of the alkoxy radicals **a1**−**a4**.

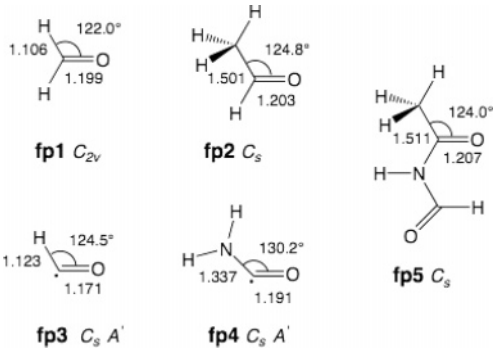


Figure 2. Optimized B3-LYP/6-311+G(3df,2p) structures of preferred conformers of the fragmentation products **fp1**−**fp5** of reactions A1−A4.

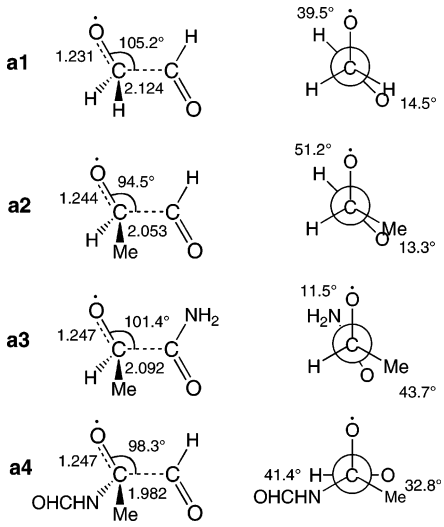


Figure 3. Optimized B3-LYP/6-311+G(3df,2p) transition structures for fragmentation of the C−C bond in reactions A1−A4.

•OCH_2CH═O radical (**a1**), we obtain the rotational potential of Figure 4. It can be seen that with HF/6-31G(d), the rotational potential around the central C−C bond changes from one with three equivalent conformers to one with just two distinct conformers. The preferred conformer has C_s

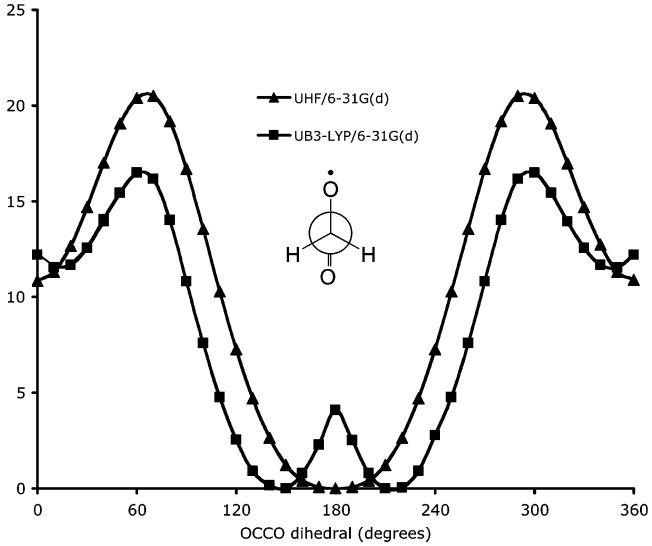


Figure 4. Rotational potential about the central C−C bond of the O═CH−$CH_2O^{\bullet}$ radical (**a1**) with HF/6-31G(d) and B3-LYP/6-31G(d).

symmetry (A″ state) with the oxygen atoms in an anti relationship. The other conformer, which is approximately 12.5 kJ mol^{-1} higher in energy after inclusion of ZPVE, is also C_s symmetric (A″ state) but with the oxygen atoms in a syn relationship. This behavior bears a qualitative similarity to the rotational potential of 2-fluoroacetaldehyde. [32]

At the B3-LYP/6-31G(d) level, there is a distortion from the syn and anti structures. In total, four stable conformers are found, a pair of enantiomeric C_1 structures that have OCCO dihedral angles of 145.6 and 214.4° (C−O• bond approximately anti to the C═O bond), respectively, and another enantiomeric pair that are 10 kJ mol^{-1} higher in energy, with dihedral angels of 13.3 and 346.7° (C−O• bond approximately syn to the C═O bond). Improving the basis set in the B3-LYP calculations to 6-311+G(3df,2p) has a relatively small effect on the rotational potential.

The B3-LYP rotational potential of Figure 4 suggests that alkoxy radical **a1** suffers from second-order Jahn−Teller (SOJT) effects.[33] It has been noted[33c] that density functional theory and coupled-cluster theory can reliably treat SOJT effects, while Hartree−Fock and MP theory do not. In the context of the present study, because of the possibility of SOJT effects, caution will be exercised when considering vibrational frequencies. We find that the energy differences between conformers that are distorted through SOJT effects are quite small. For example, at the B3-LYP/6-31G(d) level, the difference in energy between the constrained C_s (A″ state) conformer with C−O• anti to C═O (which has one imaginary frequency) and the energetically lower C_1 structure (no imaginary frequencies) is just 1.7 kJ mol^{-1}.

The $^{\bullet}OCH(CH_3)CH{=}O$ radical (**a2**) is obtained formally by replacing one of the oxy-carbon hydrogens of **a1** with a methyl group. The possible symmetric structures obtained with HF/6-31G(d) for **a1** cannot be achieved by alkoxy radical **a2**. At this level, three conformers are found for **a2** with OCCO dihedral angles of 5.3° (C−O• bond approximately eclipsing C═O bond), 166.2° (a distortion of the structure with the C−O• bond anti to C═O), and 202.3° (a distortion in the opposite direction of the structure with the C−O• bond anti to C═O), which can be interpreted as a modification of the fluoroacetaldehyde rotational potential.

The conformers with OCCO dihedral angles of 166.2° (**a2**$_a$) and 202.3° (**a2**$_b$) at HF/6-31G(d) were reoptimized at a number of levels of theory. URCCSD(T)/6-311+G-(3df,2p) single points were then carried out on a selection of the optimized geometries.

The results in Table 1 show that the preferred conformer depends on the level of theory, but the energy differences are small. We have taken our best level for geometry optimization throughout this study to be B3-LYP/6-311+G-(3df,2p), and the conformation that gives the lowest coupled-cluster energy at this level was therefore taken to be the preferred structure. For $^{\bullet}OCH(CH_3)CH{=}O$, **a2**$_a$ is thus our predicted structure. At the B3-LYP/6-311+G(3df,2p) level, the OCCO dihedral angle in **a2**$_a$ is 134.2°.

For $^{\bullet}OCH(CH_3)C(NH_2){=}O$ (**a3**), analogies with fluoroacetaldehyde are not apparent. In this case, only one conformation of significance was found (Figure 1), and so no further analysis was carried out.

For the $^{\bullet}OC(CH_3)(NHCH{=}O)CH{=}O$ radical (**a4**), three conformers were examined in detail, namely **a4**$_a$, which has the C═O bond approximately eclipsing the C−CH_3 bond, and **a4**$_b$ and **a4**$_c$, both of which have the C═O bond approximately eclipsing the C−N bond but which differ in the orientation of the NHCH═O group. The detailed geometries for each conformer at the B3-LYP/6-311+G-(3df,2p) level can be found in the Supporting Information. The energetic ordering at various levels of theory of the three conformers are summarized in Table 1. The conformer that gives the lowest coupled-cluster energy for the B3-LYP/6-311+G(3df,2p) geometry is conformer **a4**$_a$, shown as **a4** in Figure 1, and so this was used in the subsequent calculations.

The geometries of fragmentation products were also obtained by initially scanning for the lowest-energy conformations. Unlike the situation for the alkoxy radicals, there was general agreement as to the preferred conformations at the various levels of theory. These are displayed in Figure 2.

Transition structures were found by initially determining single-point energy profiles at the B3-LYP/6-31G(d) level as the cleaving bond was stretched, starting with the preferred conformers of the alkoxy radicals. This procedure identified approximate transition structures, which were refined by full geometry optimizations. Using IRC calculations, we confirmed that the transition structures were connected to the initial alkoxy radicals. However, these transition structures do not necessarily connect directly to the global conformational minima of the products. It is assumed that the energy required for the products to subsequently rearrange to the preferred conformer through torsional motions is small compared with the energy required for the initial bond cleavage.

In summary, assignment of the preferred conformations for the alkoxy radicals **a2**−**a4** is sensitive to the level of theory used for geometry optimization. We have taken the preferred conformation to be the one that gives the lowest CCSD(T) energy with B3-LYP/6-311+G(3df,2p) geometries.

Table 2. Effect of Geometry on Calculated Enthalpy for Reaction A1 and on Cleavage and Addition Barriers (UCCSD(T)6-311+G(3df,2p), kJ mol^{-1})[a]

geometry	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
//UHF/6-31G(d)	20.3	42.4	22.1
//UMP2/6-31G(d)	19.4	42.0	22.6
//UMPW1K/6-31+G(d,p)	20.1	40.5	20.4
//UMPW1K/6-31+G(3df,2p)	21.0	41.2	20.2
//UBMK/6-31G(d)	19.6	40.6	20.9
//UB3-LYP/6-31G(d)	19.7	41.5	21.8
//UB3-LYP/6-31G(2df,p)	18.3	40.4	22.1
//UB3-LYP/6-311+G(3df,2p)	18.3	40.5	22.2
//UB3-LYP/cc-pVTZ	18.6	40.9	22.3
//UQCISD/6-31G(d)	20.1	43.8	23.7

[a] Enthalpies and barriers are not corrected for zero-point vibrational energy.

Table 3. Effect of Geometry on Calculated Enthalpy for Reaction A2 and on Cleavage and Addition Barriers (URCCSD(T)6-311+G(d,p), kJ mol^{-1})[a]

geometry	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
//UHF/6-31G(d)	1.7	32.7	31.0
//UHF/6-311++G(3df,2p)	5.8	34.0	28.2
//UMP2/6-31G(d)	0.6	37.1	36.5
//UMP2/6-31+G(d)	1.0	36.4	35.4
//UMP2/6-31+G(2df,p)	1.2	39.8	38.6
//UBMK/6-31G(d)	−0.2	33.2	33.5
//UB3-LYP/6-31G(d)	−0.2	33.5	33.7
//UB3-LYP/6-31G(2df,p)	−1.7	32.4	34.1
//UB3-LYP/6-311+G(3df,2p)	−2.5	32.4	34.9

[a] Enthalpies and barriers are not corrected for zero-point vibrational energy.

The theoretical determination of the preferred geometries of the transition structures and products is found to be more straightforward than for the alkoxy radical reactants.

3.2. Effect of Level of Geometry Optimization on Calculated Reaction Enthalpies and Barriers. Having selected the preferred conformations of each of the alkoxy radicals, cleavage products (addition reactants), and transition structures, we next sought to determine the sensitivity of the calculated reaction enthalpies and barriers to the level of geometry optimization. To do this, we calculated the enthalpies and barriers for reactions A1−A3 by obtaining CCSD(T) energies for a variety of optimized structures. For reaction A1, calculations were carried out with UCCSD(T)/6-311+G(3df,2p). For reactions A2 and A3, calculations were carried out with URCCSD(T)/6-311+G(d,p) or URCCSD(T)/6-311+G(2df,p). The results of this analysis are summarized in Tables 2−4.

Table 2 shows that the reaction enthalpies and barriers for reaction A1 are not very sensitive to the method used to optimize the structures. In this case, the enthalpies and barriers all agree to within 3.5 kJ mol^{-1}. The enthalpies and barriers for reactions A2 and A3 are somewhat more sensitive to the level of geometry optimization, although for the most part the effects are still not very large. The greater sensitivity may in part be due to the reduced basis set size. It is noteworthy that the basis set effects appear smaller for B3-LYP than for HF or MP2.

Table 4. Effect of Geometry on Enthalpies for Reaction A3 and on Cleavage and Addition Barriers (URCCSD(T)6-311+G(2df,p), kJ mol^{-1})[a]

geometry	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
//UHF/6-31G(d)	−17.7	5.5	23.3
//UHF/6-311++G(3df,2p)	−28.3	3.5	31.8
//UMP2/6-31+G(d)	−25.2	2.7	27.8
//UB3-LYP/6-31G(d)	−20.8	5.4	26.2
//UB3-LYP/6-31G(2df,p)	−19.7	6.3	26.0
//UB3-LYP/6-311+G(3df,2p)	−19.3	6.8	26.1

[a] Enthalpies and barriers are not corrected for zero-point vibrational energy.

Table 5. Calculated Lengths of Cleaving (Forming) Bonds (Å) in Transition Structures for Reactions A1−A4

level of theory	ts1	ts2	ts3	ts4
HF/6-31G(d)	2.078	2.052	2.044	1.975
HF/6-311++G(3df,2p)	2.046	2.020	2.015	
MP2/6-31G(d)	1.961	1.926	1.884	1.801
MP2/6-31+G(d)	1.954	1.917	1.875	
MP2/6-311+G(3df,2p)	1.930			
MPW1K/6-31+G(d,p)	2.174	2.096	2.159	
MPW1K/6-31+G(3df,2p)	2.170			
BMK/6-31G(d)	2.138	2.071	2.165	2.047
B3-LYP/6-31G(d)	2.148	2.068	2.155	2.024
B3-LYP/6-31G(2df,p)	2.103	2.045	2.125	1.998
B3-LYP/6-311+G(3df,2p)	2.124	2.040	2.092	1.981
QCISD/6-31G(d)	2.025			

Table 5 lists the lengths of the fragmenting (forming) bonds in the cleavage (addition) transition structures of reactions A1−A4, as calculated at a variety of theoretical levels. It can be seen that the B3-LYP and MPW1K DFT methods tend to predict longer bonds than the corresponding HF and MP2 bonds, consistent with previous observations.[34] The BMK functional predicts bond lengths similar to those of the other DFT methods, despite its formulation as a functional to give improved kinetic parameters.

In summary, assessment of the effect of geometry on the calculated reaction enthalpies and barriers indicates that the level of theory used for geometry optimization does not have a great effect on these properties. This lends confidence to the use of B3-LYP geometries in calculating improved energies with composite methods.

3.3. Assessment of Reaction Energies and Barriers. Enthalpies and barriers for reactions A1−A4, and barriers for the reverse addition reactions at various levels of theory, are given in Tables 6−9. In this section, we initially describe the results obtained with high-level composite methods. We then examine the results from several computationally more efficient levels that are more amenable to application to larger systems and gauge their performance against that of the high-level methods. All of the high-level composite methods we use prescribe B3-LYP geometries. We found in the previous section that this should not have a large effect on the calculated relative energies.

The highest-level method used in the present study is W1, which was restricted to the calculation of the addition barrier for reaction A1. The highest-level method used for all four

Table 6. Effect of Level of Theory on Calculated Enthalpy of Reaction A1 and on Barriers for Fragmentation and Addition (0 K, kJ mol^{-1})

level of theory	ΔH	$\Delta H^{\ddagger}_{\text{cleavage}}$	$\Delta H^{\ddagger}_{\text{addition}}$
W1			26.6
G3//B3-LYP	2.9	27.5	24.6
G3-RAD	4.2	30.4	26.1
CBS-QB3	3.1	24.4	21.2
U-CBS-QB3	3.2	27.9	24.7
G3X(MP2)-RAD	0.4	29.4	29.0
G3(MP2)//B3-LYP	0.5	26.1	25.7
UB3-LYP/6-311+G(3df,2p)//B3[a]	5.4	23.7	18.2
RB3-LYP/6-311+G(3df,2p)//B3[a]	9.6	23.7	14.1
UBMK/6-311+G(3df,2p)//B3[a]	13.1	35.9	22.7
RBMK/6-311+G(3df,2p)//B3[a]	11.8	37.8	26.0
UBMK/6-311+G(3df,2p)//BM[b]	13.1	36.5	23.3
RBMK/6-311+G(3df,2p)//BM[b]	9.8	36.5	26.5
UMP2/6-311+G(3df,2p)//B3[a]	−23.8	38.9	62.7
RMP2/6-311+G(3df,2p)//B3[a]	−26.9	−0.6	26.3
UQCISD/6-31G(d)	12.5	53.9	41.4
UB3-LYP/cc-pVTZ	10.8	29.0	18.2
UB3-LYP/6-311+G(3df,2p)	9.7	28.3	18.6
UB3-LYP/6-31G(2df,p)	17.0	30.7	13.8
UB3-LYP/6-31G(d)	24.8	36.7	12.0
UMPW1K/6-31+G(3df,2p)	46.2	61.0	14.8
UMPW1K/6-31+G(d,p)	54.9	70.3	15.4
UMP2/6-311+G(3df,2p)	−21.9	46.4	68.3
UMP2/6-31G(2df,p)	−13.7	53.2	66.8
UMP2/6-31G(d)	−5.9	75.1	81.0
UHF/6-311++G(3df,2p)	38.6	96.5	58.0
UHF/6-31G(d)	60.1	107.8	47.7

[a] The notation//B3 is used to designate that the calculations were carried out on UB3-LYP/6-31G(d) optimized structures. [b] The notation//BM is used to designate that the calculations were carried out on UBMK/6-31G(d) optimized structures.

reactions is G3//B3-LYP. We will use this procedure to compare the performance of the other methods. Other high-level methods used for all four reactions are standard CBS-QB3, U-CBS-QB3, G3(MP2)//B3-LYP, and G3X(MP2)-RAD. Mean absolute deviations (MADs) from G3//B3-LYP for all enthalpies and barriers (cleavage and addition) are 2.2 (CBS-QB3), 1.3 (U-CBS-QB3), 1.5 (G3(MP2)//B3-LYP), and 2.8 (G3X(MP2)-RAD) kJ mol^{-1}. The MADs show that the high-level procedures agree well with one another, although it can be argued that a sample size of 12 may not be statistically significant. G3(MP2)//B3-LYP and G3X-(MP2)-RAD are considerably cheaper than the other methods but give comparable performance.

The standard G*n* methods are less reliable when open-shell species have large spin contamination in the unrestricted wave functions. The RAD variants of the G*n* methods[30] were designed specifically to alleviate this situation. In the present study, we used G3-RAD to calculate enthalpies and barriers for reactions A1 and A2 (Tables 6 and 7) and G3X(MP2)-RAD (Tables 6−9) to calculate enthalpies and barriers for reactions A1−A4.

G3-RAD produces enthalpies and barriers that are similar to those from the G3//B3-LYP method. For the enthalpies of reactions A1 and A2, this is not unexpected. The spin contamination of the alkoxy radicals and the product radicals

Table 7. Effect of Level of Theory on Calculated Enthalpy of Reaction A2 and on Barriers for Fragmentation and Addition (0 K, kJ mol^{-1})

level of theory	ΔH	$\Delta H^{\ddagger}_{\text{cleavage}}$	$\Delta H^{\ddagger}_{\text{addition}}$
G3//B3-LYP	−12.3	21.1	33.4
G3-RAD	−12.7	21.6	34.3
CBS-QB3	−10.8	19.2	30.0
U-CBS-QB3	−10.7	22.6	33.3
G3X(MP2)-RAD	−13.6	23.5	37.1
G3(MP2)//B3-LYP	−14.5	19.9	34.4
UB3-LYP/6-311+G(3df,2p)//B3[a]	−14.1	20.4	34.5
RB3-LYP/6-311+G(3df,2p)//B3[a]	−16.2	19.7	35.9
UBMK/6-311+G(3df,2p)//B3[a]	−7.2	28.9	36.1
RBMK/6-311+G(3df,2p)//B3[a]	−8.5	30.2	38.7
UBMK/6-311+G(3df,2p)//BM[b]	−6.9	29.3	36.2
RBMK/6-311+G(3df,2p)//BM[b]	−8.7	30.0	38.7
UMP2/6-311+G(3df,2p)//B3[a]	−42.4	27.5	69.9
RMP2/6-311+G(3df,2p)//B3[a]	−45.7	−13.0	32.7
UB3-LYP/6-311+G(3df,2p)	−13.9	19.3	35.0
UB3-LYP/6-31G(2df,p)	−2.9	23.9	28.7
UB3-LYP/6-31G(d)	4.0	28.7	26.7
UMP2/6-31+G(2df,p)	−37.9	33.8	71.8
UMP2/6-31+G(d)	−53.5	29.6	85.3
UMP2/6-31G(d)	−30.3	51.6	84.1
UHF/6-31G(d)	34.2	97.5	63.4
UHF/6-311++G(3df,2p)	12.2	87.5	75.3

[a] The notation//B3 is used to designate that the calculations were carried out on UB3-LYP/6-31G(d) optimized structures. [b] The notation//BM is used to designate that the calculations were carried out on UBMK/6-31G(d) optimized structures.

Table 8. Effect of Level of Theory on Calculated Enthalpy of Reaction A3 and on Barriers for Fragmentation and Addition (0 K, kJ mol^{-1})

level of theory	ΔH	$\Delta H^{\ddagger}_{\text{cleavage}}$	$\Delta H^{\ddagger}_{\text{addition}}$
G3//B3-LYP	12.7	14.1	1.4
CBS-QB3	15.1	13.9	−1.2
U-CBS-QB3	15.1	16.5	1.4
G3X(MP2)-RAD	13.4	19.1	5.6
G3(MP2)//B3-LYP	10.8	13.5	2.7
UB3-LYP/6-311+G(3df,2p)//B3[a]	5.1	14.4	9.3
RB3-LYP/6-311+G(3df,2p)//B3[a]	3.1	14.1	11.0
UBMK/6-311+G(3df,2p)//B3[a]	16.4	23.8	7.5
RBMK/6-311+G(3df,2p)//B3[a]	14.2	24.4	10.2
UBMK/6-311+G(3df,2p)//BM[b]	15.4	22.7	7.4
RBMK/6-311+G(3df,2p)//BM[b]	9.4	19.6	10.2
UMP2/6-311+G(3df,2p)//B3[a]	−16.0	8.9	24.9
RMP2/6-311+G(3df,2p)//B3[a]	−19.4	−18.6	0.8
UB3-LYP/6-311+G(3df,2p)	2.0	11.6	9.6
UB3-LYP/6-31G(2df,p)	21.8	18.2	−3.6
UB3-LYP/6-31G(d)	26.8	23.2	−3.5
UMP2/6-31+G(d)	−5.7	41.3	47.0
UMP2/6-31G(d)	−2.0	38.6	40.6
UHF/6-311++G(3df,2p)	33.4	89.8	56.4
UHF/6-31G(d)	52.3	96.5	44.2

[a] The notation//B3 is used to designate that the calculations were carried out on UB3-LYP/6-31G(d) optimized structures. [b] The notation//BM is used to designate that the calculations were carried out on UBMK/6-31G(d) optimized structures.

formed in the cleavage reactions is very small. At the UMP2/6-31G(d) level, the $\langle S^2 \rangle$ values are less than 0.77, compared with 0.75 for a pure doublet. The transition structures for

Table 9. Effect of Level of Theory on Calculated Enthalpy of Reaction A4 and on Barriers for Fragmentation and Addition (0 K, kJ mol^{-1})

level of theory	ΔH	$\Delta H^{\ddagger}_{\text{cleavage}}$	$\Delta H^{\ddagger}_{\text{addition}}$
G3//B3-LYP	−49.8	−4.6	45.2
CBS-QB3	−45.8	−4.2	41.6
U-CBS-QB3	−46.3	−2.2	44.1
G3X(MP2)-RAD	−51.9	−2.9	49.1
G3(MP2)//B3-LYP	−52.2	−5.8	46.3
UB3-LYP/6-311+G(3df,2p)//B3[a]	−60.8	−1.1	59.7
RB3-LYP/6-311+G(3df,2p)//B3[a]	−62.9	−2.1	60.8
UBMK/6-311+G(3df,2p)//B3[a]	−54.3	4.6	58.9
RBMK/6-311+G(3df,2p)//B3[a]	−57.7	3.8	61.4
UBMK/6-311+G(3df,2p)//BM[b]	−55.7	1.8	57.6
RBMK/6-311+G(3df,2p)//BM[b]	−60.2	0.0	60.2
UMP2/6-311+G(3df,2p)//B3[a]	−91.5	−18.8	72.7
RMP2/6-311+G(3df,2p)//B3[a]	−98.0	−55.2	42.8
UB3-LYP/6-311+G(3df,2p)	−60.5	0.4	60.8
UB3-LYP/6-31G(2df,p)	−49.1	2.8	51.9
UB3-LYP/6-31G(d)	−44.4	6.0	50.3
UMP2/6-31G(d)	−83.4	9.0	92.5
UHF/6-31G(d)	−43.4	66.7	110.1

[a] The notation//B3 is used to designate that the calculations were carried out on UB3-LYP/6-31G(d) optimized structures. [b] The notation//BM is used to designate that the calculations were carried out on UBMK/6-31G(d) optimized structures.

reactions A1 and A2 have larger $\langle S^2 \rangle$ values (0.89 and 0.88, respectively), and so it is more surprising that the difference between G3-RAD and G3//B3-LYP barriers for these two reactions is small. With G3X(MP2)-RAD, enthalpies and barriers for reactions A1 and A2 as well as for A3 and A4 compare well with G3//B3-LYP. This is characterized by the relatively small MAD of 2.8 kJ mol^{-1}, as quoted above.

Enthalpies and barriers calculated from UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) energies give fairly good agreement with those of the high-level methods. MADs from the G3//B3-LYP results are 5.1 and 5.6 kJ mol^{-1}, and the largest deviations (LDs) are 15.6 and 14.5 kJ mol^{-1} for UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), respectively. The relatively small MADs and LDs make these methods candidates for use with larger systems when G3(MP2)//B3-LYP and G3X(MP2)-RAD become too costly. The very small difference between the two methods is not surprising when considering the $\langle S^2 \rangle$ values for the alkoxy radicals, product radicals, and transition structures. For UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), $\langle S^2 \rangle$ values are always less than 0.76.

We note that the UB3-LYP/6-311+G(3df,2p) enthalpies and barriers obtained from fully optimized UB3-LYP/6-311+G(3df,2p) structures are close to those obtained from UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d). The MAD for the fully optimized UB3-LYP/6-311+G(3df,2p) structures increases slightly to 5.9 kJ mol^{-1} and has an LD of 15.6 kJ mol^{-1}. B3-LYP enthalpies and barriers obtained with smaller basis sets, viz. 6-31G(d) and 6-31G(2df,p), are somewhat poorer. Thus, for B3-LYP/6-31G(d), the MAD and LD are 10.8 and 21.9 kJ mol^{-1}, while for B3-LYP/6-31G(2df,p) they are 6.4 and 14.1 kJ mol^{-1}, respectively.

The MADs of 6.9 and 7.6 kJ mol^{-1} for UBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d), respectively, demonstrate that these two methods, like the analogous B3-LYP calculations, are also potentially cost-effective alternatives to the more sophisticated composite methods. The MADs are improved to 6.5 kJ mol^{-1} for UBMK/6-311+G(3df,2p) and to 6.9 kJ mol^{-1} for RBMK/6-311+G(3df,2p) if UBMK/6-31G(d) geometries are used in place of the UB3LYP/6-31G(d) geometries.

Enthalpies and barriers calculated with HF and MP2 in general are not very satisfactory. The MADs for UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) are 24.1 and 23.7 kJ mol^{-1}, with LDs of 41.7 and 48.2 kJ mol^{-1}, respectively. For UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d), it is interesting that the addition barriers agree reasonably well with G3//B3-LYP with an LD of 14.2 kJ mol^{-1}, while the cleavage barriers are poor with an LD of 38.0 kJ mol^{-1}. On the other hand, the addition barriers are poor with RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) with an LD of 50.7 kJ mol^{-1}, but the cleavage barriers are good with an LD of 2.4 kJ mol^{-1}. For both UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d), the reaction enthalpies are similar and consistently too exothermic.

The differences between the RMP2 and UMP2 results can be rationalized in terms of the general performance of MP2 theory in describing alkoxy radicals and the $\langle S^2 \rangle$ values of the alkoxy radicals and the transition structures. As noted in the section on conformers, MP2 theory predicts higher than expected energies for the alkoxy radicals. However, MP2 theory performs reasonably in the calculation of the fragmentation products. This leads to a greater than expected calculated exothermicity. The UMP2 and RMP2 reaction enthalpies are similar because of low spin contamination in both the alkoxy radicals and their associated products, i.e., the errors in the enthalpies are associated with MP2 rather than spin contamination. Large spin contamination becomes a problem in the transition structures at the UMP2 level. UMP2/6-31G(d) $\langle S^2 \rangle$ values for the transition structures of reactions A1−A4 are 0.89, 0.87, 0.84, and 0.81, respectively. This leads to a fortuitous cancellation of errors when calculating fragmentation barriers with UMP2, but poor fragmentation barriers with RMP2. For the same reasons, addition barriers are poor at the UMP2 level but good with RMP2.

Additional UMP2 and UHF enthalpies and barriers were calculated from optimizations at these levels using a variety of basis sets. These include 6-31G(d), 6-31+G(d), 6-31G(2df,p), 6-31+G(2df,p), 6-31+G(3df,2p), and 6-311++G(3df,2p). As expected from the discussion above, enthalpies and addition barriers are poor with no substantial improvement with basis set size.

The remaining methods that were examined involve full geometry optimizations at the UMPW1K and UQCISD levels of theory with various basis sets for reaction A1. The results are summarized in Table 6. For UMPW1K with small and large basis sets, the deviations from the G3//B3-LYP results are large. UQCISD/6-31G(d) also does not perform well in

Table 10. Effect of Level of Theory on Calculated Heats of Formation (0 K, kJ mol^{-1}) of Fragmentation Products (**fp1**−**fp5**) and of Alkoxy Radicals (**a1**−**a4**)[a]

level of theory	fp1	fp2	fp3	fp4	fp5	a1	a2	a3	a4
B3-LYP/6-311+G(3df,2p)//B3	−107.2	−152.4	31.3	−36.4	−351.8	−81.4	−101.5	−193.9	−259.8
B3-LYP/6-311+G(3df,2p)	−108.4	−154.6	30.3	−37.2	−355.3	−87.9	−109.9	−192.7	−264.6
CBS-QB3	−109.1	−155.7	41.3	−7.6	−354.9	−70.9	−102.1	−178.3	−267.7
U-CBS-QB3	−110.9	−158.1	39.9	−8.9	−359.0	−74.2	−106.1	−182.1	−272.8
G3(MP2)//B3-LYP	−108.9	−156.3	37.4	−5.5	−349.5	−71.9	−101.4	−172.6	−259.9
G3X(MP2)-RAD	−108.3	−155.8	40.1[f]	−3.2	−349.1	−68.6	−100.3	−172.5	−257.1
G3//B3-LYP	−108.7	−157.0	39.1	−7.5	−354.1	−72.6	−102.7	−177.2	−265.3
G3-RAD(6d,10f)	−107.4	−156.2	41.0[f]	−2.8	−350.2	−70.7	−102.6		
W1	−107.3		40.5						
expt	−104.6 ± 0.7[b]	−155.0[b]	41.3 ± 0.8[c]/ 44.6 ± 0.4[d,e]						

[a] See Figures 1 and 2 for B3-LYP/6-311+G(3df,2p) structures of **a1**−**a4** and **fp1**−**fp5**, respectively. [b] From ref . 39. [c] From ref 43. [d] From ref 42. [e] $\Delta_f H_{298}$ back-corrected to 0 K using theoretical temperature correction from B3-LYP/6-31G(d) frequencies (scaled by 0.9989[25]): see ref 41. [f] From ref 41.

comparison with the high-level methods. Contributing factors to this result are most probably the small basis set and the omission of the perturbative triples.

In summary, high-level composite methods are in generally good agreement with one another for the reaction enthalpies and barriers (both cleavage and addition). The best composite methods in terms of accuracy and cost are G3(MP2)//B3-LYP and G3X(MP2)-RAD, which give enthalpies and barriers in good agreement with the more expensive G3//B3-LYP, G3-RAD, CBS-QB3, and U-CBS-QB3 methods. Relatively accurate but low-cost methods suitable for application to larger systems are provided by single-point energy calculations with UB3-LYP/6-311+G(3df,2p), RB3-LYP/6-311+G(3df,2p), UBMK/6-311+G(3df,2p), or RBMK/6-311+G(3df,2p) on UB3-LYP/6-31G(d) geometries. Using UBMK/6-31G(d) geometries in place of UB3-LYP/6-31G(d) geometries has only a small effect. UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) calculations appear to give good results for the cleavage barriers but not for addition barriers. On the other hand RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) calculations appear to give good results for the addition barriers but not for cleavage barriers. Neither of these two methods is suitable for the calculation of the reaction enthalpies.

3.4. Heats of Formation. Heats of formation at 0 K for the fragmentation products (**fp1**−**fp5**) and the four substituted alkoxyl radicals (**a1**−**a4**) of reactions A1−A4 were calculated at various levels of theory. These results are compared with available experimental values in Table 10.

In general, the high-level methods agree well with one another and with available experimental data. B3-LYP gives varied results, with discrepancies of up to 30 kJ mol^{-1} in some cases, and only a few kJ mol^{-1} in others. It has previously been found that B3-LYP/6-311+G(3df,2p) enthalpies of formation for species in the G3/99 test set have an MAD from reliable experimental values of 22.6 kJ mol^{-1}.[35] The B3-LYP results of this section are better than expected on this basis since all but one of the enthalpies of formation have deviations less than this value. It has also previously been found that errors accumulate in the B3-LYP heats of formation as the size of the system increases. This is why the G3/99 test set, with an average of approximately 17 electron pairs per species, has a larger MAD (22.6 kJ mol^{-1}) for heats of formation compared with the G2/97 test set (13.7 kJ mol^{-1}), which has an average of approximately 11 electron pairs per species. We do not find such a correlation for the small number of species in the present study, indicating that other effects are having a more important influence in these cases. For example, the B3-LYP/6-311+G(3df,2p)//B3-LYP/6-31G(d) heat of formation of $^{\bullet}OC(CH_3)(NHCH{=}O)CH{=}O$ (**a4**), which has 22 electron pairs, deviates by 5.5 kJ mol^{-1} from the G3//B3-LYP value. On the other hand, the B3-LYP/6-311+G(3df,2p)//B3-LYP/6-31G(d) heat of formation of $NH_2(C^{\bullet}){=}O$ (**fp4**), which has 8 electron pairs, deviates by 28.9 kJ mol^{-1} from the G3//B3LYP value.

Surprisingly the heats of formation calculated with BMK/6-311+G(3df,2p)//B3-LYP/6-31G(d)[36] differ significantly from the corresponding B3-LYP/6-311+G(3df,2p) values, and this method therefore does not appear to be suitable for calculating heats of formation for the alkoxy radicals.

$CH_2{=}O$ (fp1). Formaldehyde is a well-studied molecule and is primarily included in Table 10 for the sake of completeness. Our high-level methods agree in their predicted heats of formation for formaldehyde to within 1.8 kJ mol^{-1}, with W1, the highest level used in this study, predicting a value of −107.3 kJ mol^{-1}. It is evident that both the current theoretical predictions and other recent high-level theoretical predictions, e.g. the W2 and W3 results of Martin,[37] are consistently more negative than the experimental heat of formation, −104.6 ± 0.7 kJ mol^{-1},[38] by about 3 kJ mol^{-1}. The reason for this apparent discrepancy is not clear.

$CH_3CH{=}O$ (fp2). Acetaldehyde, also has a well-characterized experimental heat of formation (−155.0 kJ mol^{-1}).[39] The high-level theoretical values are all in good agreement with this experimental value. The largest deviation from experiment is 3.1 kJ mol^{-1} at the U-CBS-QB3 level.

$^{\bullet}CH{=}O$ (fp3). The heat of formation of the formyl radical has been theoretically investigated in several high-level studies. Most recently, Feller et al.[40] have used various extrapolation schemes of coupled-cluster energies to obtain a value of 43.5 ± 0.8 kJ mol^{-1}. The associated error has been estimated from the spread of data given by the different extrapolations. Earlier, Henry et al.[41] used a number of high-

level composite methods to calculate $\Delta_f H°_0$(•CHO), obtaining values of 40.1 (G3X(MP2)-RAD) and 41.0 (G3-RAD(6d,10f)) kJ mol^{-1} (Table 10). Two experimental values are available, the more recent value of 44.6 ± 0.4 coming from Becerra et al.[42] and an earlier value of 41.3 ± 0.8 kJ mol^{-1} from Berkowitz et al.[43] The W1, G3-RAD, and G3X(MP2)-RAD estimates all fall within the experimental uncertainty of the Berkowitz et al.[39] result. However, the theoretical value of Feller et al.[40] is in better agreement with the more recent result of Becerra et al.[42]

$NH_2C(•)=O$ (fp4). A heat of formation for the $NH_2C(•)=O$ fragmentation product of reaction A3 has not been reported in the literature. The heats of formation predicted by the high-level methods range between −3.2 kJ mol^{-1} from G3X(MP2)-RAD to −8.9 kJ mol^{-1} from U-CBS-QB3. The previous comparisons with experimental values suggest that U-CBS-QB3 might slightly underestimate heats of formation, and so this value may be too low. The B3-LYP values are more negative by 30 kJ mol^{-1} compared with those from the high-level composite methods. If we average the high-level methods, i.e., all the values in Table 10 except the B3-LYP estimates, we obtain −5.9 kJ mol^{-1} as the 0 K heat of formation for **fp4**.

$CH(=O)NHC(CH_3)=O$ (fp5). The previous observation that U-CBS-QB3 seems to underestimate heats of formation also applies in the case of **fp5**. Compared with the other high-level methods, U-CBS-QB3 gives the most negative heat of formation of −359.0 kJ mol^{-1}. In contrast to **fp4**, the B3-LYP values compare well with the remaining high-level methods. By averaging the results of all the high-level methods in Table 10, we obtain −352.8 kJ mol^{-1} for the 0 K heat of formation for $CH(=O)NHC(CH_3)=O$ (**fp5**).

Alkoxy Radicals (a1−a4). As in the case of the fragmentation products discussed in the preceding paragraphs, the high-level methods are in good agreement with one another. Another continued trend is the underestimation of the heats of formation by U-CBS-QB3. The B3-LYP methods give mixed results. In the case of **a1** and **a3**, the values are somewhat outside the ranges of the high-level methods. However, for alkoxy radicals **a2** and **a4** the B3-LYP heats of formation are in good agreement with the high-level results. If we average the values from the high-level methods in Table 10, we obtain heats of formation at 0 K for the alkoxy radicals of −71.4 (**a1**), −102.5 (**a2**), −176.6 (**a3**), and −264.6 (**a4**) kJ mol^{-1}, respectively.

In summary, heats of formation at 0 K calculated by high-level methods agree well with one another. U-CBS-QB3 seems to underestimate the heats of formation when compared with the other high levels of theory, but the deviation is only small. In recommending heats of formation, we average the high-level results (including U-CBS-QB3). Alkoxy radical **a4** gives the largest range of such values (10.4 kJ mol^{-1}), but for the other species, the range of values is much less (5.8 kJ mol^{-1}). B3-LYP gives varied results, with discrepancies of up to 30 kJ mol^{-1} in some cases, and only a few kJ mol^{-1} in others.

4. Concluding Remarks

The purpose of the present study has been to establish methodologies for investigating the C−C β-scission reactions of substituted alkoxy radicals that might serve as models for peptide backbone alkoxy radicals. We seek methods that provide a suitable compromise between the accuracy in predicting enthalpies and barriers for the β-scission reactions and computational cost.

We find that determination of the lowest-energy conformers of the $•OCH_2CH=O$ (**a1**), $•OCH(CH_3)CH=O$ (**a2**), $•OCH(CH_3)C(NH_2)=O$ (**a3**), and $•OC(CH_3)(NHCH=O)CH=O$ (**a4**) substituted alkoxy radicals is sensitive to the method used to optimize geometries. We have selected the conformer optimized with B3-LYP/6-311+G(3df,2p) that gives the lowest CCSD(T) energy for use in subsequent calculations.

In the calculation of enthalpies and barriers, we find that G3(MP2)//B3-LYP and G3X(MP2)-RAD give similar accuracy to more expensive high-level methods. However, these methods are still computationally too demanding to be used routinely with currently available resources for systems significantly larger than the largest models used in the present study, i.e., reactions A3 and A4. Cost-effective alternative methods are UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) or UBMK/6-311+G(3df,2p) and RBMK/6-311+G(3df,2p) with either UB3-LYP/6-31G(d) or UBMK/6-31G(d) geometries. These give reasonable estimates of enthalpies and barriers that are within about 15 kJ mol^{-1} of the reference G3//B3-LYP values.

Heats of formation for each of the species involved in the four reactions under study were also calculated. The high-level methods perform well for the three species for which experimental values are available for comparison. U-CBS-QB3 seems to underestimate the heats of formation when compared with the other high-level methods. Predicted heats of formation at 0 K for the alkoxy radicals are −71.4 ($•OCH_2CH=O$, **a1**), −102.5 ($•OCH(CH_3)CH=O$, **a2**), −176.6 ($•OCH(CH_3)C(NH2)=O$, **a3**), and −264.6 ($•OC(CH_3)(NHCH=O)CH=O$, **a4**) kJ mol^{-1}. For the fragmentation products, $NH_2C(•)=O$ (**fp4**), and $CH(=O)NHC(CH_3)=O$ (**fp5**), we predict the respective values of −5.3 kJ mol^{-1} and −351.6 kJ mol^{-1}.

Acknowledgment. We gratefully acknowledge generous allocations of computer time from the ANU Supercomputing Facility, the Australian Partnership for Advanced Computing (APAC), and the Australian Centre for Advanced Computing and Communications (AC3) as well as the awards of an Australian Research Council Discovery Grant (to L.R.) and an Australian Postgraduate Award (to G.P.F.W.). We also thank Professor Jan Martin for making the BMK functional available for our use. L.R. is a member of the ARC Centre of Excellence for Free Radical Chemistry and Biotechnology.

Supporting Information Available: B3-LYP/6-311+G(3df,2p) GAUSSIAN archive entries of equilibrium and transition structures (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Davies, M. J.; Dean, R. T. *Radical-Mediated Protein Oxidation: from Chemistry to Medicine*; Oxford University Press: Oxford; New York, 1997; and references therein.

(2) Dean, R. T.; Fu, S.; Stocker, R.; Davies, M. J. *Biochem. J.* **1997**, *324*, 1−18.

(3) Brunelle, P.; Rauk, A. *J. Alzheimer's Dis.* **2002**, *4*, 283−289. (b) Rauk. A. *Can. Chem. News* **2001**, *53*, 20−21. (c) Rauk A.; Armstrong D. A.; Fairlie D. P. *J. Am. Chem. Soc.* **2000**, *122*, 9761−9767.

(4) Gray, P.; Williams, A. *Chem. Rev.* **1959**, *59*, 239−328.

(5) Kochi, J. K. *J. Am. Chem. Soc.* **1962**, *84*, 1193−1197.

(6) Zhang, W.; Dowd, P. *Tetrahedron* **1993**, *49*, 1965−1978.

(7) Wilsey, S.; Dowd, P.; Houk, K. N. *J. Org. Chem.* **1999**, *64*, 8801−8811.

(8) Rauk, A.; Boyd, R. J.; Boyd, S. L.; Henry, D. J.; Radom, L. *Can. J. Chem.* **2003**, *81*, 431−442.

(9) Hippler, H.; Striebel, F.; Viskolcz, B. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2450−2458. (b) Fittschen, C.; Hippler, H.; Viskolcz, B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1677−1683.

(10) Dibble, T. S. *J. Phys. Chem. A* **1999**, *103*, 8559−8565.

(11) Huang, M. L.; Rauk, A. *J. Phys. Org. Chem.* **2004**, *17*, 777−786.

(12) Shapley, W. A.; Bacskay, G. B. *J. Phys. Chem A* **1999**, *103*, 4514−4524.

(13) Song, K.-S.; Cheng, Y.-H.; Fu, Y.; Liu, L.; Li, X.-S.; Guo, Q.-X. *J. Phys. Chem. A* **2002**, *106*, 6651−6658.

(14) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(15) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, 2000.

(16) Stanton, J. F.; Gauss, J.; Watts, J. D.; Nooijen, M.; Oliphant, N.; Perera, S. A.; Szalay, P. G.; Lauderdale, W. J.; Kucharski, S. A.; Gwaltney, S. R.; Beck, S.; Balková A.; Bernholdt, D. E.; Baeck, K. K.; Rozyczko, P.; Sekino, H.; Hober, C.; Bartlett, R. J. ACES II; Quantum Theory Project, University of Florida: Gainsville, 1992.

(17) 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.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. GAUSSIAN 98, Revision A.11.3; Gaussian Inc.: Pittsburgh, PA, 1998.

(18) 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.; and Pople, J. A. Gaussian 03, Revision B.03; Gaussian, Inc.: Pittsburgh, PA, 2003.

(19) Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Knowles, P. J.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schutz, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Werner, H.-J. *MOLPRO 2002.6*; University of Birmingham: Birmingham, U.K., 2002.

(20) Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1993**, *99*, 5219−5227.

(21) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650−7657.

(22) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154−2161.

(23) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523−5527.

(24) Boese, D. A.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405−3416.

(25) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502−16513.

(26) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 108−117.

(27) Lynch, B. J.; Truhlar, D. A. *J. Phys. Chem A* **2001**, *105*, 2936−2941.

(28) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822−2827.

(29) U-CBS-QB3 is a variant of CBS-QB3 that omits the empirical correction for spin contamination in the unrestricted wave function. See, for example: (a) Gómez-Balderas, R.; Coote, M. L.; Henry, D. J.; Radom, L. *J. Phys. Chem. A*, **2004**, *108*, 2874−2883. (b) Wood, G. P. F.; Henry, D. J.; Radom, L. *J. Phys. Chem. A* **2003**, *107*, 7985−7990. (c) Coote, M. L. *J. Phys. Chem. A* **2004**, *108*, 3865−3872.

(30) Henry, D. J.; Sullivan, M. B.; Radom, L. *J. Chem. Phys.* **2003**, *118*, 4849−4860.

(31) Martin, J. M. L.; De Oliveira, G. J. *J. Chem. Phys.* **1999**, *111*, 1843−1856.

(32) Bataev, V. A.; Abramenkov, A. V.; Godunov, I. A. *Russ. Chem. Bull.* **2001**, *50*, 945−951.

(33) (a) Von Neumann, J.; Wigner, E. *Z. Phys.* **1929**, *30*, 467 (b) Stanton, J. F. *J. Chem. Phys.* **2001**, *115*, 10382−10393, and references therein. (c) Russ, N. J.; Crawford, D. T.; Tschumper, G. S. *J. Chem. Phys.* **2004**, *120*, 7298−7306.

(34) See, for example: Coote, M. L.; Wood, G. P. F.; Radom, L. *J. Phys. Chem. A* **2002,** *106*, 12124−12138.

(35) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374−7383.

(36) Heats of formation in (0 K, kJ mol^{-1}) calculated with UBMK/6-311+G(3df,2p)//UBMK/6-31G(d) are as follows: −113.4 (CH_2═O, **fp1**), −114.1 (CH_3CH═O, **fp2**), 25.7 (•CH═O, **fp3**), −42.6 (NH_2C(•)═O, **fp4**), −393.4 (CH(═O)NHC-(CH_3)═O, **fp5**), −100.8 (•OCH_2CH═O, **a1**), −131.5 (•OCH-(CH_3)CH═O, **a2)**, −222.0 (•OCH(CH_3)C(NH2)═O, **a3)**, and −312.0 (•OC(CH_3)(NHCH═O)CH═O, **a4**).

(37) (a) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014−6029. (b) Boese, D. A.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss. J. *J. Chem. Phys.* **2004**, *120*, 4129−4141.

(38) Fletcher, R. A.; Pilcher, G. *Trans. Faraday Soc.* **1970**, *66*, 794−799.

(39) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data* **1988**, *17*, Suppl 1.

(40) Feller, D.; Dixon, D. A.; Francisco, J. S. *J. Phys. Chem. A* **2003**, *107*, 1604−1617.

(41) Henry, D. J.; Parkinson, C. J.; Radom, L. *J. Phys. Chem. A* **2002**, *106*, 7927−7936.

(42) Becerra, R.; Carpenter, I. W.; Walsh, R. *J. Phys. Chem. A* **1997**, *101*, 4185−4190. The reported $\Delta_f H_{298}$ value has been back-corrected to 0 K: see Henry et al.[41]

(43) Berkowitz, J.; Ellison, G. B.; Gutman, D. *J. Phys. Chem.* **1994**, *98*, 2744−2765.

CT050133G

JCTC Journal of Chemical Theory and Computation

Basis Set Requirements for Sulfur Compounds in Density Functional Theory: a Comparison between Correlation-Consistent, Polarized-Consistent, and Pople-Type Basis Sets

Pablo A. Denis*

DEQUIFIM, Facultad de Química, UDELAR, CC 1157, 11800 Montevideo, Uruguay

Received March 20, 2005

Abstract: We have investigated the SX (X = first- or second-row atom), SO_2, and SO_3 molecules employing the correlation-consistent (cc), the recently developed polarization-consistent (pc), and three Pople-type basis sets, in conjunction with the B3LYP functional. The results confirmed that the aug-pc basis sets represent a great contribution in terms of cost−benefits. In the case of the B3LYP functional, when employing the aug-pc-3 and aug-pc-4 basis sets, it is possible to obtain results that are of aug-cc-pV(5+d)Z and aug-cc-pV(6+d)Z quality, respectively, at a much lower cost. The estimations obtained employing smaller members of the family are of nearly double-ζ quality and do not provide reliable results. There is no basis set of quadruple-ζ quality among the polarized-consistent basis sets, although in terms of composition, the aug-pc-3 basis set is a QZ basis set. A precise estimation of the Kohn−Sham complete basis set (CBS) limit with the aug-pc-X basis sets is too difficult for the B3LYP functional because the ∞(aug-pc-4, aug-pc-3, aug-pc-2) extrapolation gives the same results as those of the aug-pc-4 basis set. This is in contrast with the results observed for ab initio methodologies for which the largest basis sets provided the best estimation of the CBS limit. In our opinion, the closest results to the B3LYP/CBS limit are expected to be those obtained with a two-point extrapolation employing the aug-cc-pV(X+d)Z (X = 5, 6) basis sets. The results obtained with this extrapolation are very close to those predicted by the ∞(aug-pc-3, aug-pc-2, aug-pc-1) extrapolation, and that provides a cheaper but more inaccurate alternative to estimate the CBS limit. Minor problems were found for the aug-pc-X basis sets and the B3LYP functional for molecules in which sulfur is bound to a very electronegative element, such as SO, SF, SO_2, and SO_3. For these molecules, the cc basis sets were demonstrated to be more useful. The importance of tight d functions was observed; the total atomization energy of SO_2 employing the cc-pV6Z basis set is 1 kcal/mol lower than that obtained with the aug-cc-pV(6+d)Z basis set.

Introduction

During the past decade, the computational chemistry community experienced significant changes due to the development of reliable density functionals[1]. Several theoretical groups left the traditional wave function methodologies and started to use density functional theory, in particular, the hybrid methods such as B3LYP[1]. A proof of that is that the number of articles published employing this functional, at present time, is higher than 10 000, and that number is seriously being increased every year. However, density functional theory presents a major drawback. We know that the correct functional exists, but we do not know what the correct functional is. For that reason, most of the works that employ the DFT functionals focus on resolving chemical

* Author e-mail: pablod@fq.edu.uy.

10.1021/ct0500702 CCC: $30.25
Published on Web 08/06/2005

Table 1. Distances Obtained for SX, X = First-Row Atom, at the B3LYP Level of Theory Employing Different Basis Sets (Results in Å)

	number of basis functions	SB	SC	SN	SO	SF
6-31+G(d)	42	1.6216	1.5439	1.5114	1.5158	1.6411
6-311+G(3df,2p)	86	1.6101	1.5319	1.4903	1.4884	1.6137
G3Large	108	1.6093	1.5312	1.4889	1.4858	1.6084
aug-cc-pV(D+d)Z	55	1.6218	1.5434	1.5032	1.5080	1.6389
aug-cc-pV(T+d)Z	101	1.6099	1.5320	1.4909	1.4889	1.6106
aug-cc-pV(Q+d)Z	169	1.6086	1.5309	1.4884	1.4861	1.6085
aug-cc-pV(5+d)Z	263	1.6086	1.5307	1.4880	1.4853	1.6072
aug-cc-pV(6+d)Z	387	1.6086	1.5306	1.4876	1.4849	1.6069
aug-pc-0	30		1.6045	1.6113	1.6328	1.7412
aug-pc-1	50		1.5435	1.5107	1.5192	1.6516
aug-pc-2	96		1.5330	1.4930	1.4921	1.6130
aug-pc-3	178		1.5308	1.4880	1.4853	1.6073
aug-pc-4	286		1.5306	1.4877	1.4849	1.6069
∞(aug-pc-4, aug-pc-3)			1.5304	1.4874	1.4845	1.6064
∞(aug-pc-3, aug-pc-2)			1.5292	1.4844	1.4803	1.6031
∞(aug-pc-2, aug-pc-1)			1.5286	1.4855	1.4807	1.5968
∞(aug-pc-4, aug-pc-3, aug-pc-2)			1.5306	1.4877	1.4849	1.6069
∞(aug-pc-3, aug-pc-2, aug-pc-1)			1.5302	1.4860	1.4833	1.6063
∞(aug-pc-2, aug-pc-1, aug-pc-0)			1.5307	1.4890	1.4835	1.5847
experimental		1.6092[a]	1.5439[a]	1.4940[a]	1.48109[a]	1.5962[b]

[a] From ref 28. [b] From ref 29.

problems and not on the assessment of the quality of the method developed. A precise evaluation of the available DFT methods is essential for two reasons. On one hand, it helps in the development of better methodologies. On the other hand, the calibration studies permit the study of large molecular systems with smaller basis sets.

One of the first investigations of the basis set dependency in DFT was performed by Bauschlicher and Partridge in 1995[2]. They employed Pople-type and correlation-consistent basis sets[3] to estimate the total atomization energies (TAEs) of the molecules present in the G2 test set. The results showed that the TAEs determined by employing the 6-311+-(3df,2p) basis set were better than those obtained with the aug-cc-pVTZ basis set. However, for molecules that involved bondings with second-row atoms, the results were not as good as those observed for the first row. Two years later, Martell and co-workers[4] performed another study with the aim of finding the optimal combination of gradient-corrected density functionals and basis sets to study some important reactions in atmospheric chemistry. These reactions included molecules containing only the p-block elements but contained multiple bonds such as S═O. The basis sets employed for that purpose were the 6-31G(d,p), 6-311G(d,p), cc-pVDZ, and cc-pVTZ basis sets. They concluded that the latter basis set provided the best results. Some years later, Raymond and Wheeler[5] studied the compatibility of correlation-consistent basis sets and the B3LYP functional. Their results demonstrated that the minimum basis set to obtain reasonable results was the aug-cc-pVTZ and that calculations with the aug-cc-pVQZ basis set were mandatory to extrapolate to the Kohn−Sham (KS) limit. In 2000, we studied key intermediates in the oxidation of SH_2, namely, SO and HSO[6]. We found that, even at the B3LYP/cc-pV5Z level of theory, the results were not converged (within 2 kcal/mol) for atomization energies and bond distances (within 0.001 Å). Continuing our studies, we published two more articles[7,8] in which we reported B3LYP calculations for more sulfur-containing molecules but increasing the size of basis set up to cc-pV6Z. The molecules studied were HSO_2, SO, SO_2, SO_3,[7] and the SX molecules where X = a first- or second-row atom.[8] As expected, the results confirmed the conclusions reached for HSO. The effect of diffuse functions in the estimation of relative energies employing DFT was investigated by Lynch and co-workers.[9] The authors found that the addition of diffuse functions to a double-ζ basis set was more important than increasing it to a triple-ζ basis set. Very recently, Su and co-workers[10] compared the energetics of large water clusters with basis sets up to aug-cc-pVTZ and Pople-type basis sets, namely, 6-31G** and 6-311++G**. They showed that it is possible to use the X3LYP functional and smaller basis sets while preserving accuracy, in contrast to ab initio methodologies where the basis set requirements are more important. In the organic field, Gregory and Jenks[26] employed the B3LYP functional to study the relative energies of vicinal disulfoxides and other sulfinyl radical dimers. They carried out a comparative investigation employing Pople-type basis sets and correlation-consistent basis sets up to aug-cc-pVQZ. The concluded that the aug-cc-pVQZ basis set gave results that are very close to the KS limit. Very recently, Wang and Wilson[11,22] investigated the behavior of some density functionals and correlation-consistent basis sets for a set of small first-row close-shell molecules. Wang and Wilson showed that there is a smooth convergence toward the Kohn−Sham limit for the hybrid functionals whereas the nonlocal functionals are relatively insensitive to basis set choice.

During the past five years, Jensen et al.[12−15] developed the polarized-consistent basis sets, which are specifically fitted for DFT calculations. The basic concept employed by Jensen et al. to develop the polarized-consistent basis sets is the same as that used by Dunning to construct the correlation-

Table 2. Distances Obtained for SX, X = Second-Row Atom, at the B3LYP Level of Theory Employing Different Basis Sets (Results in Å)

	SAl	SSi	SP	S_2	SCl
6-31+G(d)	2.0560	1.9517	1.9226	1.9279	2.0258
6-311+G(3df,2p)	2.0453	1.9384	1.9019	1.9026	1.9937
G3Large	2.0434	1.9362	1.9004	1.9004	1.9944
aug-cc-pV(D+d)Z	2.0557	1.9485	1.9137	1.9137	2.0076
aug-cc-pV(T+d)Z	2.0455	1.9384	1.9023	1.9023	1.9950
aug-cc-pV(Q+d)Z	2.0421	1.9356	1.9004	1.9003	1.9920
aug-cc-pV(5+d)Z	2.0430	1.9364	1.9008	1.9005	1.9918
aug-cc-pV(6+d)Z	2.0426	1.9362	1.9005	1.9001	1.9913
aug-pc-0		2.0574	2.0368	2.0985	2.2186
aug-pc-1		1.9495	1.9199	1.9267	2.0271
aug-pc-2		1.9386	1.9041	1.9052	1.9981
aug-pc-3		1.9365	1.9008	1.9005	1.9916
aug-pc-4		1.9364	1.9006	1.9001	1.9914
∞(aug-pc-4, aug-pc-3)		1.9363	1.9004	1.8897	1.9912
∞(aug-pc-3, aug-pc-2)		1.9350	1.8984	1.8971	1.9869
∞(aug-pc-2, aug-pc-1)		1.9340	1.8975	1.8962	1.9859
∞(aug-pc-4, aug-pc-3, aug-pc-2)		1.9364	1.9006	1.9001	1.9914
∞(aug-pc-3, aug-pc-2, aug-pc-1)		1.9360	1.8999	1.8992	1.9897
∞(aug-pc-2, aug-pc-1, aug-pc-0)		1.9375	1.9016	1.9019	1.9926
experimental	2.029[a]	1.9293[a]	1.9009,[a] 1.899,[b] 1.897 404[c]	1.8892[a]	1.974 65[d]

[a] From ref 28. [b] From ref 30. [c] From ref 31. [d] From ref 32.

consistent basis sets: functions that contribute similar amounts of the correlation energy are included at the same stage. However, since the convergence of HF and DFT energies is faster than that of correlation energies, the basis set optimized for describing electron correlation may not be the set of optimum exponents for describing DFT energies. To develop the aug-pc-X basis sets, Jensen et al. employed a small set of homonuclear molecules, since it is not possible to employ atoms for analyzing the importance of polarization functions. It is expected that the use of these basis set accelerates the convergence toward the KS limit. Despite the large number of articles published dealing with the basis set requirements in density functional theory, there is no investigation for second-row compounds that compares aug-pc-X and aug-cc-pV(X+d)Z basis sets employing the larger members of the families. As expressed above, the study of the basis set dependence in sulfur compounds is important because it helps in the investigation of the oxidation of sulfur in the atmosphere[4,6−7,27] and organic reactions,[23−26,37] one of our long-term research areas. In the present article, we have studied the convergence toward the KS limit of important sulfur-containing molecules, namely, the SX diatomics (X = a first- or second-row atom), SO_2, and SO_3, employing the B3LYP functional and three different families of basis sets: the augmented and nonaugmented correlation-consistent basis sets, including tight d functions,[16] aug-cc-pV(X+d)Z (X = 2, 3, 4, 5, 6); the Pople-type basis sets[17] 6-31+G(d), 6-311+G(3df), and G3Large; and the augmented polarized-consistent basis sets aug-pc-X recently developed by Jensen.[12−15]

Theoretical Methods

The hybrid DFT method B3LYP[1,18] was employed. The open-shell species have been treated employing the unrestricted formulation UB3LYP. Three different families of basis sets were selected to perform the comparison, (a) the aug-cc-pV(X+d)Z and cc-pV(X+d)Z correlation-consistent basis sets[16] (X = D, T, Q, 5, 6) developed by Dunning et al., (b) the augmented polarized-consistent aug-pc-X basis sets recently published by Jensen et al.,[12−15] and (c) the Pople-type basis sets 6-31+G*, 6-311+G(3df,2p), and G3Large.[17] The G3Large basis set consists of 6s 5p 4d 3f plus one diffuse sp function. Extrapolation to the Kohn−Sham limit was performed employing the two-parameter equation suggested by Halkier et al.,[21] $E_X = E_\infty + AX^{-3}$, and the three-parameter exponential extrapolation suggested by Feller,[38] $E_X = E_\infty + A\,e^{-BX}$. All calculations were carried out with Gaussian 03[19] employing the very tight optimization criteria and the ultrafine grid. The use of the default grid in Gaussian 03 instead of the ultrafine changes the geometries only by 0.0001−0.0002 Å. The basis sets were obtained from the basis set database of the Pacific Northwest National Laboratory[20] except for the aug-pc-X (X = 0, 1, 2, 3, 4) basis sets[12−15].

Results and Discussion

Bond Distances. We present the bond distances obtained for the SX, SO_2, and SO_3 molecules in Tables 1−3. A careful inspection of these tables shows that we have almost reached convergence with respect to the extension of the basis set. The aug-cc-pV(6+d)Z and aug-pc-4 results are essentially the same with only small differences of 0.0001 Å for SN, SP, and SCl and 0.0002 Å for SSi. In those cases, the aug-cc-pV(6+d)Z basis set gives smaller bond distances. Quite remarkable is the fact that the bond distances obtained with the aug-pc-3 basis set are the same as those predicted with the aug-cc-pV(5+d)Z basis set, except for three molecules, SF, SSi, and SCl. For these molecules the results employing

Table 3. Geometrical Parameters Obtained for SO_2 and SO_3 at the B3LYP Level of Theory Employing Different Basis Sets (Results in Å)

	B3LYP			
	SO	SO_2		SO_3
	r_{S-O}	r_{S-O}	O−S−O	r_{S-O}
6-31+G(d)	1.5158	1.4655	118.65	1.4543
6-311+G(3df,2p)	1.4884	1.4367	119.22	1.4253
G3Large	1.4858	1.4346	119.23	1.4238
aug-cc-pV(D+d)Z	1.5080	1.4566	119.08	1.4440
aug-cc-pV(T+d)Z	1.4889	1.4381	119.14	1.4271
aug-cc-pV(Q+d)Z	1.4861	1.4350	119.14	1.4241
aug-cc-pV(5+d)Z	1.4853	1.4341	119.21	1.4234
aug-cc-pV(6+d)Z	1.4849	1.4336	119.23	1.4230
cc-pV(D+d)Z	1.5091	1.4525	119.31	1.4404
cc-pV(T+d)Z	1.4894	1.4370	119.25	1.4262
cc-pV(Q+d)Z	1.4864	1.4348	119.17	1.4240
cc-pV(5+d)Z	1.4854	1.4341	119.20	1.4234
cc-pV(6+d)Z	1.4849	1.4336	119.23	1.4230
aug-pc-0	1.6328	1.6051	113.64	1.6006
aug-pc-1	1.5192	1.4731	118.06	1.4608
aug-pc-2	1.4921	1.4412	118.79	1.4301
aug-pc-3	1.4853	1.4341	119.23	1.4235
aug-pc-4	1.4849	1.4336	119.24	1.4230
∞(aug-pc-4, aug-pc-3)	1.4845	1.4331	119.25	1.4224
∞(aug-pc-3, aug-pc-2)	1.4803	1.4289	119.55	1.4186
∞(aug-pc-2, aug-pc-1)	1.4807	1.4278	119.10	1.4172
∞(aug-pc-4, aug-pc-3, aug-pc-2)	1.4849	1.4336	119.24	1.4230
∞(aug-pc-3, aug-pc-2, aug-pc-1)	1.4833	1.4320	119.92	1.4216
∞(aug-pc-2, aug-pc-1, aug-pc-0)	1.4835	1.4334	118.94	1.4212
experimental	1.481 09[a]	1.4308[a]	119.3[a]	1.4173[b]

[a] From ref 33. [b] From ref 34.

the auc-cc-pV(5+d)Z basis set are shorter by 0.0001, 0.0001, and 0.0002 Å, respectively. This behavior of the aug-pc-3 basis set represents a major advantage of the polarization-consistent basis sets, since the aug-pc-3 set has 89 basis functions for sulfur and the aug-cc-pV(5+d)Z has 136. The results obtained with the aug-pc-2 basis set are of lower quality than those obtained with the aug-cc-pV(T+d)Z basis set, and that is not an advantage but, rather, a disadvantage, since the increment of the cardinal number of the basis set in one changes the results from less than triple-ζ quality to quintuple-ζ quality.

The different behavior of the correlation- and polarization-consistent basis sets is clearly seen in Figure 1 where we plot the r_{S-O} of SO_2 against the cardinal number for both families of basis sets, the correlation- and polarization-consistent basis sets. The slope of the curve fitted with the results obtained with the aug-pc-X family is much higher than that obtained with the correlation-consistent basis sets, although both converge toward the same limit. The higher slope in the curve of the aug-pc-X family makes more difficult the estimation of the Kohn−Sham limit. This is in contrast with the results obtained with the correlation-consistent basis set for which the slope is smaller, and the betterment of the results with respect to the cardinal number of basis sets is less abrupt. However, there is a problem in the convergence of the correlation-consistent basis sets, already noted by several authors.[5−7,11,22] The convergence with the correlation-consistent family is not always well-behaved. This can be illustrated by comparing the results obtained for the bond distances of the SX molecules where X is a first-row atom and those where X is a second-row atom (Tables 1 and 2). The convergence with respect to the basis set is good for the first row, but it is not good for the second row. For the latter, there is a problem with the aug-cc-pV(Q+Z) basis set. The predictions of the latter basis set are too short for SSi, SP, and S_2. Therefore, extrapolation

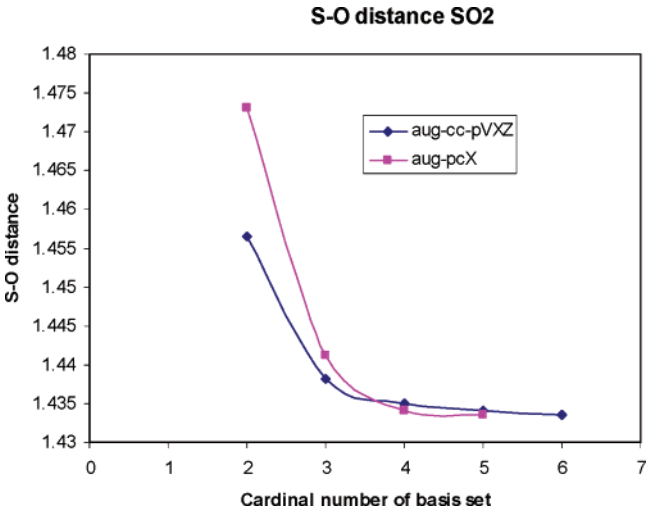


Figure 1. Convergence of the S−O distance for SO_2 with the B3LYP method and the aug-cc-pV(X+d)Z and aug-pc-X basis sets.

Table 4. Estimated TAEs for SX, X = First-Row Atom, at the B3LYP Level of Theory Employing Different Basis Sets (Results in kcal/mol)

	SB	SC	SN	SO	SF
6-31+G(d)	129.54	161.20	106.43	116.74	79.53
6-311+G(3df,2p)	133.08	166.47	115.11	126.41	84.91
G3Large	132.87	166.29	115.17	126.70	84.79
aug-cc-pV(D+d)Z	130.88	164.25	111.02	121.43	82.42
aug-cc-pV(T+d)Z	132.67	166.02	114.62	126.30	85.40
aug-cc-pV(Q+d)Z	133.13	166.53	115.49	126.76	85.94
aug-cc-pV(5+d)Z	132.97	166.50	115.60	127.42	86.00
aug-cc-pV(6+d)Z	133.02	166.56	115.73	127.55	86.34
∞(6, 5)	133.09	166.64	115.91	127.73	86.81
aug-pc-0		132.62	73.03	86.69	65.02
aug-pc-1		162.12	108.22	118.13	80.53
aug-pc-2		165.79	114.00	125.52	85.27
aug-pc-3		166.48	115.51	127.20	86.01
aug-pc-4		166.50	115.63	127.29	86.06
∞(aug-pc-4, aug-pc-3)		166.52	115.76	127.38	86.11
∞(aug-pc-3, aug-pc-2)		166.98	116.61	128.43	86.55
∞(aug-pc-2, aug-pc-1)		167.34	116.43	128.63	87.27
∞(aug-pc-4, aug-pc-3, aug-pc-2)		166.50	115.64	127.30	86.06
∞(aug-pc-3, aug-pc-2, aug-pc-1)		166.64	116.05	127.69	86.15
∞(aug-pc-2, aug-pc-1, aug-pc-0)		166.33	115.14	127.89	87.41
reference value[a]	134.43	171.06	113.39	124.89	84.79

[a] From ref 8. See text for details.

cannot be accomplished employing the aug-cc-pV(Q+d)Z basis set.

We have investigated two different extrapolation schemes, the two-point extrapolation and the three-point extrapolation. It is worth noticing that good results are not expected from the two-point extrapolation, since this formula is expected to be well-behaved only for correlation energies. A three-point extrapolation ∞(aug-pc-4, aug-pc-3, aug-pc-2) does not improve the results obtained with the aug-pc-4 basis set, even in 0.0001 Å. Although we did not perform numerical calculations, it is not expected that the aug-pc4 results are the real complete basis set (CBS) limits because, for some molecules, the aug-cc-pV(6+d)Z basis set provided shorter bond distances. In our opinion, this problem of the aug-pc basis sets is due to the lack of QZ quality results. In contrast, if we employ a two-point extrapolation, ∞(aug-pc-4, aug-pc-3), the bond distances are shortened with respect to the aug-pc4 results. The other two three-point extrapolation schemes investigated, ∞(aug-pc-3, aug-pc-2, aug-pc-1) and ∞(aug-pc-2, aug-pc-1, aug-pc-0), give reasonable results. In particular, the ∞(aug-pc-2, aug-pc-1, aug-pc-0) results are very impressive if we consider, first, that the extrapolated results are very close to those obtained with the aug-pc4 basis set and, second, the basis set results are of less than triple-ζ quality.

Finally, we have performed some comparative calculations with the Pople-type basis sets 6-31+G(d), 6-311+G(3df), and G3Large. The results obtained with these sets are quite good if we consider that they have much fewer basis functions. For example, for the largest basis set, G3large, the results are very close to those obtained with the aug-pc-3 basis set. It is important to remark that the G3large basis set has 66 basis functions for sulfur, only 16 more that the aug-pc-2 basis set and 20 less than the aug-pc-3 basis set. The 6-31+G(d) basis set predicted bond distances that are longer than those obtained with the aug-cc-pV(D+d)Z basis set and very close to those obtained with the aug-pc-1 basis set. Finally, the popular 6-311+(3df,2p) basis set is of nearly aug-cc-pV(T+d)Z quality, and with respect to the polarization-consistent basis sets, it is very close to the aug-pc-2 results.

In Tables 1−3, we have included the experimental bond distances to compare them with the theoretical results obtained in the present article. For 9 of a total of 12 molecules investigated, B3LYP overestimates the bond distances with respect to the experiment. The exceptions are SB, SC, and SN. When employing the aug-cc-pV(6+d)Z basis set, the worst results are obtained for SCl, for which the bond distance predicted is 0.0166 Å larger than the experimental result.

Atomization Energies. The total atomization energies (TAEs) estimated for the molecules investigated in the present article are presented in Tables 4−6. For 8 of the 12 molecules studied in the present article, the TAEs estimated with the aug-pc-4 basis set are within 0.1 kcal/mol of the results obtained with the aug-cc-pV(6+d)Z basis set. This is a tremendous achievement if we consider that a calculation for SN employing the aug-cc-pV(6+d)Z basis set has 387 basis functions and the calculation with the aug-pc-4 basis set has 286 basis functions, 101 less. However, for the most electronegative elements oxygen and fluorine, some differences exist. This is in contrast with the results obtained for bond distances for which both basis set gave almost the same results. Employing the aug-pc-4 basis set, the TAEs obtained for SO, SF, SO_2, and SO_3 are 0.26, 0.28, 0.44, and 0.66 kcal/mol lower, respectively, than those calculated with the largest correlation-consistent basis set. This problem of the aug-pc-4 basis set is also observed when we compare the aug-

Table 5. Estimated TAEs for SX, X = Second-Row Atom, at the B3LYP Level of Theory Employing Different Basis Sets (Results in kcal/mol)

	SAl	SSi	SP	S_2	SCl
6-31+G(d)	91.07	136.38	96.18	94.89	59.63
6-311+G(3df,2p)	94.56	141.93	104.72	103.05	67.02
G3Large	94.58	142.15	104.53	103.08	66.85
aug-cc-pV(D+d)Z	92.56	140.34	101.97	101.27	64.94
aug-cc-pV(T+d)Z	94.35	141.93	104.46	103.24	67.13
aug-cc-pV(Q+d)Z	94.71	142.79	104.84	103.67	67.49
aug-cc-pV(5+d)Z	94.49	142.13	104.81	103.62	67.56
aug-cc-pV(6+d)Z	94.55	142.21	104.95	103.74	67.66
∞(6, 5)	94.63	142.32	105.14	103.90	67.79
aug-pc-0		112.37	67.20	67.30	42.91
aug-pc-1		139.34	99.30	96.97	61.39
aug-pc-2		141.64	103.85	102.46	66.69
aug-pc-3		142.17	104.86	103.68	67.59
aug-pc-4		142.16	104.90	103.71	67.63
∞(aug-pc-4, aug-pc-3)			104.94	103.74	67.67
∞(aug-pc-3, aug-pc-2)		142.56	105.60	103.35	68.25
∞(aug-pc-2, aug-pc-1)		142.61	105.77	104.77	68.92
∞(aug-pc-4, aug-pc-3, aug-pc-2)			104.90	103.71	67.63
∞(aug-pc-3, aug-pc-2, aug-pc-1)		142.34	105.17	104.03	67.77
∞(aug-pc-2, aug-pc-1, aug-pc-0)		141.86	104.72	103.77	68.92
reference value[a]	97.80	146.55	103.87	103.23	68.09

[a] From ref 8. See text for details.

pc-3 and aug-cc-pV(5+d)Z basis sets. These results show that the correlation-consistent basis sets are more robust than the aug-pc-X, perhaps because of the presence of more tight d functions. Indeed, a calculation performed with the cc-pV6Z basis set yields TAE(SO_2) = 252.79 kcal/mol, 1 kcal/mol lower than that obtained with the aug-cc-pV(6+d)Z basis set, showing that tight d functions are also important at the DFT level, as noted by Wang and Wilson for the cc-pV5Z basis set.[36] As observed for the bond distances, the aug-pc-2 basis set gives results that are of a lesser quality than those predicted with the aug-cc-pV(T+d)Z basis set.

In Figure 2, we plot the TAE of SO_3 against the cardinal number of the basis set. The whole picture is similar to that discussed in bond distances. Both families of basis sets converge to the same dissociation limit, but the slope of the aug-pc-X set is larger, showing a much different performance between the members of the series. Basically, we have two zones, one below triple-ζ quality and the other above quintuple-ζ quality. Thus, the quadruple-ζ member is missing. For comparative purposes, we have also extrapolated the TAE to the KS limit for both families of basis sets. It has been observed for the CCSD(T) calculations that the extrapolations employing the largest basis sets, 5Z and 6Z, provide the best estimation of the CBS limit. However, this is not true for the aug-pcX basis sets. The ∞(aug-pc-4, aug-pc-3, aug-pc2) extrapolation does not make an important improvement, and in some cases, for example, SO, SO_2, and SO_3, the results are worse than those obtained with the aug-cc-pV(6+d)Z basis set. The latter statement is also true for a two-point extrapolation employing the aug-pc-4 and aug-pc-3 basis sets. Therefore, we employed another two alternatives to estimate the CBS limit, the use of the correlation-consistent basis sets. The results obtained with the three-point extrapolation ∞(6, 5, Q) are extremely close to those obtained with the aug-cc-pV(6+d)Z basis set, making only a little improvement in the TAE. For example, in the case of SO_2, the TAE is increased by only 0.2 kcal/mol with respect to the aug-cc-pV(6+d)Z results, which is quite unlikely, since the change from aug-cc-pV(5+d)Z to aug-cc-pV(6+d)Z is 0.38 kcal/mol. The ∞(6, 5) extrapolation employing the correlation-consistent basis set appears to be quite reasonable, and it is very close to the results obtained with a three-point extrapolation ∞(aug-pc-3,aug-pc-2,aug-pc1), but because of the lack of numerical DFT calculations, we cannot be 100% sure about our observations; further investigation is required in this field to elucidate how the extrapolation should be performed.

We have investigated other three-point extrapolations, ∞(aug-pc-3, aug-pc-2, aug-pc1) and ∞(aug-pc2, aug-pc-1, aug-pc-0). The latter provides very good results at a small cost and the former, as expressed above, gives results that are very close to those obtained with a two-point extrapolation ∞(6, 5). The Pople-type basis sets give very good results, providing the best cost−results benefits. For the SX molecules, the G3Large basis set is within 1 kcal/mol with respect to the results obtained with the aug-cc-pV(6+d)Z basis set. However, it is important to remark that some unexpected deviations exist. For example, in the case of SF, the TAE estimated with the G3large basis set is 1.55 kcal/mol lower, whereas the smaller 6-311+G(3df,2p) basis set performs better. This result makes the predictions employing the Pople-type basis sets quite unsafe because they do not have uniform behavior.

Finally, it is important to compare the TAEs estimated at the DFT level with the correct results. In a recent article,[8] we investigated the thermochemistry of the SX diatomics at the CCSD(T) level, including corrections for core−valence correlation, scalar relativistic, and spin−orbit effects. The

Table 6. Estimated TAEs for SO, SO_2, and SO_3 at the B3LYP Level of Theory Employing Different Basis Sets (Results in kcal/mol)

	B3LYP		
	SO	SO_2	SO_3
6-31+G(d)	116.74	224.16	290.23
6-311+G(3df,2p)	126.41	250.61	331.78
G3Large	126.70	251.78	333.17
aug-cc-pV(D+d)Z	121.43	236.57	310.99
aug-cc-pV(T+d)Z	126.30	250.35	331.22
aug-cc-pV(Q+d)Z	127.03	252.28	334.18
aug-cc-pV(5+d)Z	127.42	253.41	335.84
aug-cc-pV(6+d)Z	127.55	253.79	336.40
∞(6, 5)	127.73	254.31	337.17
∞(5, 4)	127.89	254.60	337.58
∞(Q, T)	127.09	253.69	336.34
∞(6, 5, Q)	127.61	253.99	336.69
∞(5, Q, T)	127.88	255.04	338.03
cc-pV(D+d)Z	118.57	231.48	307.60
cc-pV(T+d)Z	125.84	250.20	332.77
cc-pV(Q+d)Z	126.85	252.33	334.86
cc-pV(5+d)Z	127.11	252.86	335.19
cc-pV(6+d)Z	127.24	253.23	335.61
aug-pc-0	86.69	136.50	153.87
aug-pc-1	118.13	225.71	293.02
aug-pc-2	125.52	248.12	327.70
aug-pc-3	127.20	253.04	335.24
aug-pc-4	127.29	253.35	335.74
∞(aug-pc-4, aug-pc-3)	127.38	253.68	336.26
∞(aug-pc-3, aug-pc-2)	128.43	256.63	340.74
∞(aug-pc-2, aug-pc-1)	128.63	257.56	342.30
∞(aug-pc-4, aug-pc-3, aug-pc-2)	127.30	253.37	335.78
∞(aug-pc-3, aug-pc-2, aug-pc-1)	127.69	254.47	337.41
∞(aug-pc-2, aug-pc-1, aug-pc-0)	127.89	255.70	339.24
reference value	124.89[a]	258.25[b]	343.72[b]

[a] From ref 8. See text for details. [b] From ref 35.

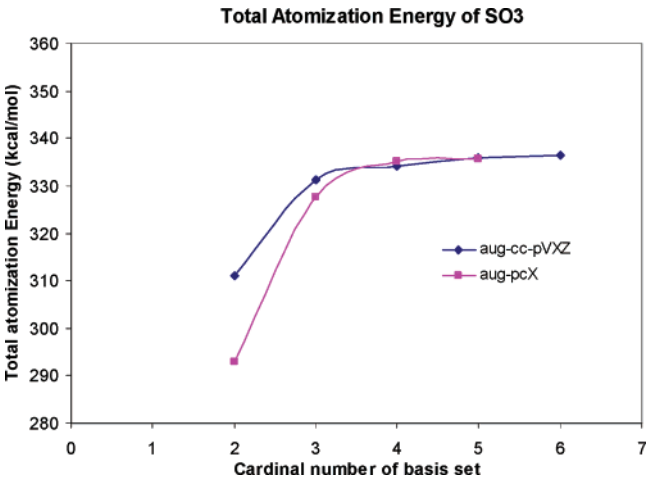


Figure 2. Convergence of the TAE for SO_3 with the B3LYP method and the aug-cc-pV(X+d)Z and aug-pc-X basis sets.

results showed that only the experimental enthalpies of formation of two molecules are known with enough accuracy to be used as a reference, namely, those of SO and SC. Errors as large as 10 kcal/mol have been found for SB, SCl, and SAl. Therefore, we will use, as a reference, the values determined at the CCSD(T) level. In the cases of SO_2 and SO_3, there are available accurate experimental TAEs.[35] These two molecules present the largest deviations. At the B3LYP/aug-cc-pV(6+d)Z level of theory, their TAEs are 4.5 and 7.3 kcal/mol, respectively, smaller than the experimental values. This is in contrast with the results obtained for the TAE of the SX diatomics. At the same level of theory, the mean absolute deviation is 2.2 kcal/mol. Although, it is important to notice that, for three SX diatomics, we have observed deviations similar to that of SO_2, namely, for SC, SAl, and SSi. For those molecules that present larger errors, SO_2, SO_3, SC, SAl, and SSi, the deviation is by defect.

Conclusions

We have investigated the SX (X = first- or second-row atom), SO_2, and SO_3 molecules employing the correlation-consistent (cc), the recently developed polarization-consistent (aug-pc), and three Pople-type basis sets, in conjunction with the B3LYP functional. The results confirmed that the aug-pc basis sets represent a great contribution in terms of cost−benefits. In the case of the B3LYP functional, when the aug-pc-3 and aug-pc-4 basis sets are employed, it is possible to obtain results that are of aug-cc-pV(5+d)Z and aug-cc-pV(6+d)Z quality, respectively, at a much lesser cost. The estimations obtained employing smaller members of the family are of nearly double-ζ quality and do not provide reliable results. There is no basis set of quadruple-ζ quality among the polarized-consistent basis sets, although, in terms of composition, the aug-pc-3 basis set is a QZ basis set. A precise estimation of the Kohn−Sham CBS limit with the aug-pc-X basis sets is too difficult for the B3LYP functional because the ∞(aug-pc-4, aug-pc-3, aug-pc-2) extrapolation gives the same results as the aug-pc-4 basis set. This is in contrast to the result observed for ab initio methodologies for which the largest basis set provided the best estimation of the CBS limit. In our opinion, the closest results to the B3LYP/CBS limit are expected to be those obtained with a two-point extrapolation employing the aug-cc-pV(X+d)Z (X = 5, 6) basis sets. The results obtained with this extrapolation are very close to those predicted by the ∞(aug-pc-3, aug-pc-2, aug-pc-1) extrapolation, and that provides a cheaper but more inaccurate alternative to estimate the CBS limit. Minor problems were found for the aug-pc-X basis sets and the B3LYP functional for molecules in which sulfur is bound to a very electronegative element, such as SO, SF, SO_2, and SO_3. For these molecules, the cc basis sets were demonstrated to be more useful. The importance of tight d functions was observed; the TAE of SO_2 employing the cc-pV6Z basis set is 1 kcal/mol lower than that obtained with the aug-cc-pV(6+d)Z basis set.

Acknowledgment. The author acknowledges PEDECIBA (UNESCO−PNUD) for financial support. The author expresses his gratitude to the reviewers for their helpful criticism, which contributed to improving the article.

References

(1) Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648.

(2) Bauschlicher, C. W.; Partridge, H. *Chem. Phys. Lett.* **1995**, *240*, 553.

(3) Dunning, T. H. *J. Chem. Phys*. **1989**, *90*, 1007.

(4) Martell, J. M.; Goddard, J. D.; Eriksson, L. A. *J. Phys. Chem. A* **1997**, *101*, 1927.

(5) Raymond, K. S.; Wheeler, R. A. *J. Comput. Chem.* **1999**, *20*, 107.

(6) Denis, P. A.; Ventura, O. N. *Int. J. Quantum Chem.* **2000**, 80, 439.

(7) Denis, P. A.; Ventura, O. N. *Chem. Phys. Lett.* **2001**, 344, 221.

(8) Denis, P. A. *J. Phys. Chem. A* **2004**, *108*, 11092.

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

(10) Su, J. T.; Xu, X.; Goddard, W. A. *J. Phys. Chem. A* **2004**, *108*, 10518.

(11) Wang, N. X.; Wilson, A. K. *J. Chem. Phys.* **2004**, *121*, 7632.

(12) Jensen, F. *J. Chem. Phys.* **2001**, *115*, 9113.

(13) Jensen, F. *J. Chem. Phys.* **2002**, *116*, 7372.

(14) Jensen, F. *J. Chem. Phys.* **2002**, 117, 9234.

(15) Jensen, F.; Helgaker, T. *J. Chem. Phys.* **2004**, *121*, 3463.

(16) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys*. **2001**, *114*, 9244.

(17) (a) Hehre, W.; Radom, L.; Schleyer, P. V. R. *Ab initio Molecular Orbital Theory*; Wiley: New York, 1986. (b) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(18) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter* **1988**, B37, 785.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision B.04; Gaussian, Inc.: Wallingford, CT, 2004.

(20) Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 2/25/04, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory, which is part of the Pacific Northwest National Laboratory, P. O. Box 999, Richland, WA 99352, and is funded by the U.S. Department of Energy. The Pacific Northwest National Laboratory is a multiprogram laboratory operated by the Battelle Memorial Institue for the U.S. Department of Energy under contract DE-AC06-76RLO 1830. Contact Karen Schuchardt for further information.

(21) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.

(22) Wang, N. X.; Wilson A. K. *Mol. Phys*. **2005**, *103*, 145.

(23) Gomes, J. R. B.; Ribeiro da Silva, M. A. V. *J. Org. Chem.* **2004**, *108*, 11684.

(24) Gomes, J. R. B, Gomes, P. *Tetrahedron* **2005**, *61*, 2705.

(25) McCulla, R. D.; Jenks, W. S. *J. Org. Chem.* **2003**, *68*, 7871.

(26) Gregory, D. D.; Jenks, W. S. *J. Phys. Chem. A.* **2003**, *107*, 3414.

(27) Denis, P. A. *Chem. Phys. Lett*. **2005**, *402*, 289.

(28) Huber, K. P.; Herzberg, G. *Constants of Diatomic Molecules*; Van Nostrand Reinhold: New York, 1979.

(29) Endo, Y.; Nagai, K.; Yamada, C.; Hirota, E. *J. Mol. Spectrosc.* **1983**, *97*, 212.

(30) Jenouvrier, A.; Pascat, B. *Can. J. Phys*. **1978**, *56*, 1088.

(31) Kawaguchi, K.; Hirota, E.; Ohishi, M.; Suzuki, H.; Takano, S.; Yamamoto, S.; Saito, S. *J. Mol. Spectrosc.* **1988**, *130*, 81.

(32) Yamada, C.; Butler, J. E.; Kawaguchi, K.; Kanamori, H.; Hirota, E. *J. Mol. Spectrosc.* **1986**, *116*, 108.

(33) Morino, Y.; Kikuchi, Y.; Saito, S.; Hirota, R. *J. Mol. Spectrosc.* **1964**, *13*, 95.

(34) Ortigoso, J.; Escribano, R.; Maki, A. G. *J. Mol. Spectrosc.* **1989**, *138*, 602.

(35) Chase, M. W.; Davies, C. A.; Downey, J. R.; Frurip, D. J.; McDonald, R. A.; Syverud, A. N. JANAF Thermochemical Tables Third Edition. *J. Phys. Chem. Ref. Data* **1985**, *14* (Suppl. 1).

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

(37) Denis, P. A.; Ventura, O. N.; Mai, H. T.; Nguyen, M. T. *J. Phys. Chem. A* **2004**, *108*, 5073.

(38) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104.

CT0500702

JCTC Journal of Chemical Theory and Computation

Hydrogen Abstraction Acetylene Addition and Diels−Alder Mechanisms of PAH Formation: A Detailed Study Using First Principles Calculations

V. V. Kislov,[†,‡] N. I. Islamova,[§] A. M. Kolker,[§] S. H. Lin,[‡] and A. M. Mebel*[,†]

Department of Chemistry and Biochemistry, Florida International University, Miami, Florida 33199, Institute of Atomic and Molecular Sciences, Academia Sinica, P.O. Box 23-166, Taipei 10764, Taiwan, and Institute of Solution Chemistry of Russian Academy of Sciences, Akademicheskaya St. 1, Ivanovo 153045, Russia

Received March 1, 2005

Abstract: Extensive ab initio Gaussian-3-type calculations of potential energy surfaces (PES), which are expected to be accurate within 1−2 kcal/mol, combined with statistical theory calculations of reaction rate constants have been applied to study various possible pathways in the hydrogen abstraction acetylene addition (HACA) mechanism of naphthalene and acenaphthalene formation as well as Diels−Alder pathways to acenaphthalene, phenanthrene, and pyrene. The barrier heights; reaction energies; and molecular parameters of the reactants, products, intermediates, and transition states have been generated for all types of reactions involved in the HACA and Diels−Alder mechanisms, including H abstraction from various aromatic intermediates, acetylene addition to radical sites, ring closures leading to the formation of additional aromatic rings, elimination of hydrogen atoms, H disproportionation, C_2H_2 cycloaddition, and H_2 loss. The reactions participating in various HACA sequences (e.g., Frenklach's, alternative Frenklach's, and Bittner and Howard's routes) are demonstrated to have relatively low barriers and high rate constants under combustion conditions. A comparison of the significance of different HACA mechanisms in PAH growth can be made in the future using PES and molecular parameters obtained in the present work. The results show that the Diels−Alder mechanism cannot compete with the HACA pathways even at high combustion temperatures, because of high barriers and consequently low reaction rate constants. The calculated energetic parameters and rate constants have been compared with experimental and theoretical data available in the literature.

1. Introduction

The formation of polycyclic aromatic hydrocarbons (PAHs) has become one of the central areas in combustion research because of the great environmental effect attributed to these compounds. Numerous studies have shown PAHs to be major air pollutants exhibiting mutagenic and tumorogenic activities.[1−3] Having been formed as a result of the incomplete combustion of fossil fuels, PAHs are present in our environment in the form of atmospheric aerosols, soot, and volatile particles, usually together with other airborne toxic chemicals.[4,5] The processes of fullerene and soot formation were shown to involve PAH intermediates at the initial stages of synthesis.[6,7] PAHs and their cations are also found to play an important role in interstellar chemistry.[8,9] According to different estimates, PAH molecules could account for 2−30% of the carbon in the galaxy and may provide nucleation sites for the formation of carbonaceous dust.[8] PAHs are also believed to be responsible for many emission features in

* To whom correspondence should be addressed. E-mail: mebela@fiu.edu.

† Florida International University.

‡ Institute of Atomic and Molecular Sciences, Academia Sinica.

§ Institute of Solution Chemistry of Russian Academy of Sciences.

10.1021/ct0500491 CCC: $30.25
Published on Web 06/21/2005

astronomical infrared spectra.[9] To help researchers and engineers to develop cleaner combustion equipment, which would minimize soot and PAH production, it is imperative to know the mechanisms, rate constants, and product yields for the reactions relevant to PAH synthesis. Experimental and theoretical studies of such mechanisms represent an important source of kinetic and thermodynamic data required for the modeling of complex combustion systems.

Three possible mechanisms relevant to the formation of PAH have been proposed and critically discussed by the combustion community in the past two decades. The first one is known as the hydrogen abstraction acetylene addition (HACA) mechanism. It was introduced by Frenklach and Wang[10] and represents a repetitive sequence of two principal reaction steps, the abstraction of a hydrogen atom from the reacting hydrocarbon by another hydrogen atom followed by the addition (or several additions) of acetylene molecules to the radical site formed in the previous H-abstraction step. The first step activates a singlet hydrocarbon molecule, producing a radical intermediate, which can be involved in further growth by the reaction with acetylene. The production of a higher PAH containing an extra aromatic ring in its core can be accomplished by a ring closure reaction. Two alternative HACA pathways for the sequential additions of two acetylene molecules were proposed by Frenklach et al.[11−13] and by Bittner and Howard.[14] In Frenklach's route, the second acetylene molecule adds to the aromatic ring activated by either a conventional or an internal[15] hydrogen abstraction mechanism, whereas in Bittner−Howard's route, the second acetylene molecule adds to the first one. Several high-level ab initio studies of some important HACA steps were carried out in recent years. The hydrogen abstraction reactions from benzene ($C_6H_6 + H \rightarrow C_6H_5 + H_2$, $C_6H_6 + OH \rightarrow C_6H_5 + H_2O$, and $C_6H_5 + CH_3 \rightarrow C_6H_6 + CH_4$) were studied employing Gaussian-2 and Gaussian-3 (G2 and G3, respectively) model chemistry methods combined with either Rice−Ramsperger−Kassel−Marcus (RRKM) or transition state theory (TST) calculations of rate constants.[16−19] The mechanism of acetylene addition to the phenyl radical was investigated by Tokmakov and Lin,[20] utilizing the G2M scheme to compute the potential energy surface (PES) for the reaction and RRKM theory to calculate reaction rates. However, up to now, there were no systematic and accurate ab initio/statistical theory studies for the complete HACA networks. Bauschlicher and Ricca.[21] computed barrier heights and reaction energies for numerous reaction steps involved in the HACA naphthalene synthesis using a rather modest B3LYP/4-31G level of theory. On the basis of their calculated PES, they concluded that both Frenklach's and Bittner−Howard's routes exhibit low barriers but could not determine which HACA route is the most favorable. In our previous B3LYP/6-31G* study of the HACA phenanthrene synthesis,[22] we computed PESs for all reactions involved in this network and calculated respective rate constants using the canonical transition state theory. Our results also indicated that both HACA routes are probable at combustion temperatures.

An alternative mechanism of PAH formation that does not involve acetylene molecules has been proposed by Miller and Melius on the basis of a series of flame experiments combined with kinetic modeling.[23−25] They introduced several pathways leading to benzene and naphthalene starting from recombination products of two propargyl (C_3H_3)[23] or two cyclopentadienyl (C_5H_5)[24] radicals, respectively. Both propargyl and cyclopentadienyl radicals were shown to be highly abundant in hydrocarbon flames and to play an important role in combustion chemistry.[6] Phenanthrene can be formed in a similar way, through several rearrangements of a recombination product of indenyl and cyclopentadienyl radicals.[25] According to various computational studies of PESs corresponding to these reaction sequences[23,24,26] and recent kinetic studies of propargyl recombination,[27−29] the five-membered ring aromatics (fulvene, cyclopentadiene, and cyclopentadienyl) represent important PAH precursors. In the recent ab initio/QRRK/kinetic modeling study by Richter and Howard,[30] the self-combination of propargyl radicals followed by ring closure and subsequent rearrangements was again shown to be the dominant benzene formation pathway in acetylene and ethylene flames.

The third mechanism of PAH synthesis discussed in the literature can be described as a Diels−Alder-type mechanism, which does not involve radical species during the growth process. It has been proposed by Siegmann and Sattler[31] as the dominant route of PAH growth in methane combustion. In this mechanism, the acetylene molecule acts as a dienophil and effectively closes bay regions in such PAHs as biphenyl, phenanthrene, and so forth. The mechanism (often called a benzogenic Diels−Alder reaction) involves two stages, cycloaddition of C_2H_2 to produce a Diels−Alder adduct, which then undergoes H_2 loss leading to a higher PAH. According to Siegmann and Sattler,[31] this mechanism is favored over the conventional HACA process because it can explain the successive formation of more compact, peri-condensed PAHs observed in methane combustion.

In the present study, we report chemically accurate ab initio G3-type calculations of PESs for possible HACA routes leading from benzene to naphthalene as well as HACA and Diels−Alder routes leading from naphthalene to acenaphthalene. These calculations allow us, for the first time, to consider the HACA network as a whole using most reliable surfaces, which are expected to be accurate within 1−2 kcal/mol. The results of the ab initio calculations, that is, the energies and molecular parameters obtained from the first principles, are utilized in RRKM and TST calculations of reaction rate constants, which are compared with experimental and earlier theoretical data wherever available. Although G3 model chemistry is becoming a standard method for highly accurate predictions of relative energies of intermediates and transition states of chemical reactions, it is conventionally applied to relatively small systems. The applications to larger molecules are scarce so far because of the large computational demands of G3 methods, and systems containing more than 6−7 heavy atoms are most often treated in the literature using much less expensive density functional methods. These methods, unfortunately, are so far unable to provide an accuracy and reliability comparable with those of the G3 family. The accurate energetics are especially important when they are used for the quantitative prediction

of reaction rate constants. Therefore, an additional goal of the present paper is to compare the reaction energies and barrier heights computed by G3 and DFT methods for various types of processes involved in the PAH growth mechanism.

2. Computational Methods

Geometries and vibrational frequencies of all species involved in various reactions were obtained using the hybrid density functional B3LYP[32a−d] method with the 6-311+G(3df,2p) basis set for the HACA naphthalene formation network and the 6-311G** basis set for other reactions, unless mentioned otherwise. Unscaled vibrational frequencies were used to calculate zero-point energy (ZPE) corrections, to characterize the stationary points, and to perform rate constant computations. Although B3LYP as well as other conventional density functionals were shown to produce less accurate results in kinetic calculations than recently developed hybrid DFT methods such as KMLYP,[32e] we used B3LYP instead of those DFT alternatives because the final energies utilized in our kinetics calculations were obtained at a G3-type level, which is considerably more accurate than any available DFT scheme. The differences in calculated frequencies and geometries produced by B3LYP and KMLYP methods differ only slightly, but just the B3LYP method was used in the calibration of several schemes in the G3 family. The optimized geometries of all species (in Cartesian coordinates) involved in the reaction networks, vibrational frequencies, ZPE corrections, and molecular structural parameters (symmetry groups, moments of inertia, and rotational constants) are given in Table S1 of the Supporting Information.

To obtain accurate barrier heights and heats of reactions, we applied the G3 theory[33] for high-level single-point energy calculations. G3 model chemistry has been shown to be less computationally demanding and more accurate than its predecessors, G2 and G2M. We employed the G3(MP2,CC)//B3LYP modification[34] of the original G3 scheme. The final energies at 0 K were obtained using the B3LYP structures, frequencies, and ZPE corrections according to the following formula:

$$E_0[\mathrm{G3(MP2,CC)}] = E[\mathrm{CCSD(T)/6\text{-}31G(d)}] + \Delta E_{\mathrm{MP2}} + \Delta E(\mathrm{SO}) + E(\mathrm{HLC}) + E(\mathrm{ZPE})$$

where ΔE_{MP2} is the basis set correction,

$$\Delta E_{\mathrm{MP2}} = E[\mathrm{MP2/G3MP2large}] - E[\mathrm{MP2/6\text{-}31G(d)}],$$

ΔE(SO) is a spin−orbit correction (significant only for atoms and not included in our calculations), E(HLC) is a high-level correction, and E(ZPE) is the zero-point energy. To further improve the accuracy for reaction species relevant to HACA naphthalene synthesis, we used the 6-311G** and G3large[35] basis sets instead of the 6-31G* and G3MP2large sets,[35] respectively, in the CCSD(T) and MP2 calculations. Here and below, we denote this G3-type scheme applied in our computations as G3 for brevity.

For hydrogen disproportionation steps, which proceed via open-shell singlet transition states, we employed the IRCMax method[36a] where barrier heights were determined as Max{Energy[G3(MP2,CC)]}//IRC{Geom[UMP2/6-31G*]}. The transition state geometries were initially optimized at the UMP2/6-31G* level,[37] then the minimal energy reaction path (MEP) was obtained by intrinsic reaction coordinate (IRC) calculations[38] and energies of the structures along MEP were refined at G3(MP2,CC). We preferred UMP2 geometry optimization over B3LYP in this case because, using the B3LYP method, we were unable to achieve convergence during the saddle-point search on the open-shell singlet PES. It should be noted that the shapes of potential energy curves along MEP appeared to be similar at the UMP2 and G3(MP2,CC) levels, although the exact positions of the transition states and barriers calculated using G3(MP2,CC)//UMP2 and IRCMax differ. Although a more accurate and reliable procedure (VTST-IOC) has been proposed by Chuang et al.[36b] for the reaction path dynamics calculations, we were unable to utilize this scheme for our system because it is significantly more computationally demanding than IRCMax and requires higher-level corrections to energies, frequencies, and moments of inertia, based on stationary-point geometries reoptimized at a higher level than the reaction path calculations. Instead, we utilized the IRCMax scheme considering that it produces more accurate results comparatively with simple MP2 optimization (gradient calculations and frequencies at the G3 level are unavailable so far).

All ab initio calculations were carried out using the Gaussian 98[39] and Molpro 2002[40] program packages. The calculated heats of reactions and barrier heights are summarized in Table 1.

High-pressure thermal rate constants were computed using the RRKM[41−43] theory and conventional TST.[44] The latter was applied for hydrogen abstraction/disproportionation steps, which are bimolecular in both forward and backward directions. Tunneling corrections (Q_{tun}) to the rate constants were calculated using simple Wigner's formula[41] because at combustion temperatures (>1000 K), tunneling does not play a significant role and more sophisticated estimates for tunneling corrections are not necessary.

For unimolecular steps, we carried out conventional RRKM calculations utilizing the energetic and molecular parameters (reaction barriers, moments of inertia, vibrational frequencies) from the ab initio calculations. RRKM rate constants obtained in the present study correspond to the high-pressure limit reaction conditions. The modified Beyer−Swinehart algorithm[45,46] was employed in calculations of the numbers and densities of states. For some intermediates and transition states, which exhibit low-frequency torsional motions, we applied the free internal rotor approximation[41−43] to evaluate their contributions to the partition functions. Such low-frequency modes usually correspond to internal rotations with barriers substantially below $k_{\mathrm{B}}T$, which cannot be properly described by a harmonic oscillator. It should be noted that the free-rotor treatment provides upper estimates for contributions of torsional motions to the partition function. In the free internal rotor calculations, we obtained moments of inertia associated with the free rotor as reduced moments of inertia of two fragments rotating around a single torsional axis.

Table 1. Barrier Heights and Reaction Energies for All Steps Involved in the Studied PAH Formation Mechanisms Computed at the G3(MP2,CC) and B3LYP[a] Levels

		barrier height, kcal/mol		reaction energy, kcal/mol	
raection	TS	G3(MP2,CC)	B3LYP	G3(MP2,CC)	B3LYP
		Hydrogen Abstraction			
A1 + H → A2 + H_2	TS1a	16.5 (17.0)[b]	10.8	8.5 (8.8)[b]	5.2
A1 + OH → A2 + H_2O	TS1b	5.5	0.42	−5.4	−5.5
N1 + OH → N2 + H_2O	TS1a^{AN}	6.3	−1.0[c]	−3.4	−2.4[c]
A1 + CH_3 → A2 + CH_4	TS1c	17.8	15.8	8.8	7.5
A7 + H → A6 + H_2	TS8	17.6	11.5	8.5	4.5
A10 + H → A11 + H_2	TS13	17.9	11.9	10.2	6.3
N1 + H → N2 + H_2	TS1^{AN}	17.5	11.1[c]	9.7	5.6[c]
		Acetylene Addition			
A2 + C_2H_2 → A3	TS2	3.5	4.6	−39.2	−36.9
N2 + C_2H_2 → N3	TS2^{AN}	2.3	4.1[c]	−39.6	−37.0[c]
A3 + C_2H_2 → A4	TS3	4.8	5.2	−39.5	−39.3
A3a + C_2H_2 → A4b	TS7b	5.5	6.8	−32.7	−32.2
A6 + C_2H_2 → A8	TS9	3.9	6.2	−37.6	−33.4
A11 + C_2H_2 → A12	TS14	1.9	3.6	−40.8	−37.9
		Ring Closure (Radical Cyclization)			
A4b → A5	TS5	3.6	4.6	−32.3	−29.3
A8 → A9	TS10	1.8	2.2	−47.6	−49.0
A12 → A13	TS15	5.4	4.8	−53.1	−53.8
N3 → N4	TS3^{AN}	16.6	16.2[c]	−19.9	−16.1[c]
		Hydrogen Atom Loss/Disproportionation			
A3 → A10 + H	TS12	36.6	37.8	30.1	33.3
A5 → N1 + H	TS6	14.8	14.2	2.5	5.2
A9 → N1 + H	TS11	27.8	28.1	21.2	24.8
N4 → AN + H	TS4^{AN}	26.3	27.8[c]	19.2	23.1[c]
A3 + H → A10 + H_2	TS12a	12.0 (6.2)[d]		−74.2[e]	
A5 + H → N1 + H_2	TS6a	2.7 (1.2)[d]		−101.7[e]	
A9 + H → N1 + H_2	TS11a	5.0 (2.3)[d]		−82.5[e]	
N4 + H → AN + H_2	TS4a^{AN}	10.6 (5.3)[d]		−84.5[e]	
		Acetylene Cycloaddition/H_2 Loss (Diels−Alder Mechanism)			
N1 + C_2H_2 → N5	TS5^{AN}	66.3	72.4[c]	47.6	58.0[c]
B1 + C_2H_2 → B2	TS1^{P}	42.2	48.3[c]	3.57	13.4[c]
P + C_2H_2 → P1	TS1^{PYR}	42.0	48.2[c]	4.78	14.6[c]
N5 → AN + H_2	TS6^{AN}	24.2	24.0[c]	−76.8	−82.4[c]
B2 → P + H_2	TS2^{P}	17.9	15.7[c]	−53.9	−60.0[c]
P1 → PYR + H_2	TS2^{PYR}	18.3	16.9[c]	−56.6	−62.5[c]
		Hydrogen Migration/Isomerization			
A3a → A6	TS7	27.6	27.7	1.5	1.2
A3 → A3a	TS7a	2.9	2.2	−1.0	−1.1
A4 → A4a	TS4	4.2	6.4	2.4	3.2
A4a → A4b	TS4a	73.0	37.6	3.3	2.8
		Hydrogen Atom Addition (Radical Recombination)			
A3a + H → A7				−108.1	−107.2
A13 + H → N1				−111.0	−109.3

[a] B3LYP energies correspond to the B3LYP/6-311+G(3df,2p) calculations if not mentioned otherwise. [b] Computed at the G3(CC)//B3LYP and G3(MP2,CC)//B3LYP (in parentheses) levels. [c] Computed at the B3LYP/6-311G** level. [d] Computed at the IRCMax{Energy[G3(MP2,CC)]}// IRC{Geom[UMP2/6-31G*]} and G3(MP2,CC)//UMP2/6-31G* (in parentheses) levels. [e] Computed at the G3(MP2,CC)//MP2/6-31G* level.

All RRKM- and TST-computed rate constants in the form of fitted modified Arrhenius rate expressions $k = AT^n \exp(-E/RT)$ are presented in Table S2 of the Supporting Information. Individual rate constants at temperatures of 1000−3000 K are shown in Table 2.

3. Potential Energy Surfaces and Reaction Rate Constants

Various reaction pathways involved in the HACA and Diels−Alder syntheses of naphthalene and acenaphthalene and also of phenanthrene and pyrene are shown in Figure 1 along with chosen notations for all species and transition states. The A3a → A4b → A5 → N1 and A3 → A4 → A4a → A4b → A5 → N1 branches represent Bittner−Howard's route, whereas the A3 → A10 → A11 → A12 → A13 → N1 sequence represents Frenklach's mechanism of HACA naphthalene synthesis. The A3a → A6(A7) → A8 → A9 → N1 branch is a variation of Frenklach's mechanism, where the second acetylene adds to the aromatic ring of the 1-vinyl-2-phenyl radical (A6) after its prior activation either by conventional (A7 → A6) or internal (A3a → A6) hydrogen abstraction. The N1 → N2 → N3 → N4 → AN and N1 → N5 → AN reaction sequences represent the HACA and Diels−Alder routes, respectively, leading from naphthalene to the acenaphthalene. The other two Diels−Alder routes lead from biphenyl to phenanthrene (B1 → B2 → P) and from phenanthrene to pyrene (P → P1 → PYR).

In subsequent sections, we discuss the results of our ab initio G3-type calculations of PESs for all relevant reaction steps and RRKM/TST computed rate coefficients. We also compare calculated barrier heights and reaction energies with available-in-the-literature experimental activation energies and heats of reactions, respectively. Although phenomenological activation energies normally obtained from the temperature dependence (Arrhenius plots) of measured rate constants often exhibit a complex character, we suppose, on the basis of activated complex theory, that theoretically derived barriers are closely related to them if both correspond

Table 2. Calculated Rate Constants (in Units of s^{-1} and $cm^3\ s^{-1}\ mol^{-1}$ for Unimolecular and Bimolecular Steps, Respectively)

reaction	rate constant	temperature, K				
		1000	1500	2000	2500	3000
Naphthalene Formation Network						
A1 + H → A2 + H_2	k_{TS1}	6.96×10^{10}	1.92×10^{12}	1.18×10^{13}	3.87×10^{13}	9.07×10^{13}
A2 + H_2 → A1 + H	k_{TS-1}	8.1×10^{10}	5.41×10^{11}	1.71×10^{12}	3.86×10^{12}	7.19×10^{12}
A2 + C_2H_2 → A3	k_{TS2}	1.0×10^{12}	3.81×10^{12}	8.86×10^{12}	1.63×10^{13}	2.62×10^{13}
A3 → A2 + C_2H_2	k_{TS-2}	6.37×10^{5}	1.03×10^{9}	4.17×10^{10}	3.83×10^{11}	1.68×10^{12}
A3 → A10 + H	k_{TS12}	6.62×10^{5}	4.93×10^{8}	1.47×10^{10}	1.18×10^{11}	4.92×10^{11}
A10 + H → A3	k_{TS-12}	2.76×10^{12}	1.27×10^{13}	3.09×10^{13}	5.71×10^{13}	9.04×10^{13}
A3 + H → A10 + H_2	k_{TS12a}	2.92×10^{10}	1.96×10^{11}	5.30×10^{11}	9.85×10^{11}	1.51×10^{12}
A10 + H_2 → A3 + H	k_{TS-12a}	4.21×10^{-6}	8.23	1.29×10^{4}	1.14×10^{6}	2.35×10^{7}
A10 + H → A11 + H_2	k_{TS13}	1.24×10^{10}	4.47×10^{11}	3.15×10^{12}	1.12×10^{13}	2.77×10^{13}
A11 + H_2 → A10 + H	k_{TS-13}	8.85×10^{10}	5.76×10^{11}	1.8×10^{12}	4.02×10^{12}	7.45×10^{12}
A11 + C_2H_2 → A12	k_{TS14}	2.27×10^{11}	7.69×10^{11}	1.75×10^{12}	3.25×10^{12}	5.34×10^{12}
A12 → A11 + C_2H_2	k_{TS-14}	3.42×10^{5}	6.48×10^{8}	2.92×10^{10}	2.93×10^{11}	1.38×10^{12}
A12 → A13	k_{TS15}	1.01×10^{11}	2.45×10^{11}	3.83×10^{11}	5.01×10^{11}	6.0×10^{11}
A13 → A12	k_{TS-15}	33.8	1.01×10^{6}	1.86×10^{8}	4.42×10^{9}	3.74×10^{10}
A3 → A3a	k_{TS7a}	2.38×10^{12}	4.39×10^{12}	6.08×10^{12}	7.49×10^{12}	8.67×10^{12}
A3a → A3	k_{TS-7a}	2.07×10^{12}	4.52×10^{12}	6.81×10^{12}	8.82×10^{12}	1.06×10^{13}
A3a → A6	k_{TS7}	4.49×10^{6}	4.66×10^{8}	4.98×10^{9}	2.12×10^{10}	5.66×10^{10}
A6 → A3a	k_{TS-7}	1.06×10^{7}	9.01×10^{8}	8.8×10^{9}	3.58×10^{10}	9.33×10^{10}
A6 + C_2H_2 → A8	k_{TS9}	2.82×10^{10}	1.32×10^{11}	3.54×10^{11}	7.29×10^{11}	1.28×10^{12}
A8 → A6 + C_2H_2	k_{TS-9}	2.95×10^{5}	4.61×10^{8}	1.9×10^{10}	1.81×10^{11}	8.25×10^{11}
A8 → A9	k_{TS10}	2.72×10^{12}	3.91×10^{12}	4.71×10^{12}	5.29×10^{12}	5.72×10^{12}
A9 → A8	k_{TS-10}	4.81×10^{3}	2.84×10^{7}	2.26×10^{9}	3.2×10^{10}	1.9×10^{11}
A9 → N1 + H	k_{TS11}	1.96×10^{7}	2.99×10^{9}	3.98×10^{10}	1.96×10^{11}	5.86×10^{11}
N1 + H → A9	k_{TS-11}	2.29×10^{12}	1.07×10^{13}	2.62×10^{13}	4.86×10^{13}	7.71×10^{13}
A9 + H → N1 + H_2	k_{TS11a}	2.16×10^{11}	6.18×10^{11}	1.16×10^{12}	1.81×10^{12}	2.53×10^{12}
N1 + H_2 → A9 + H	k_{TS-11a}	1.20×10^{-6}	4.46	1.04×10^{4}	1.22×10^{6}	3.14×10^{7}
A3a + C_2H_2 → A4b	k_{TS7b}	5.67×10^{10}	2.85×10^{11}	7.56×10^{11}	1.5×10^{12}	2.53×10^{12}
A4b → A3a + C_2H_2	k_{TS-7b}	1.99×10^{6}	1.98×10^{9}	6.64×10^{10}	5.65×10^{11}	2.41×10^{12}
A4b → A5	k_{TS5}	3.99×10^{10}	9.13×10^{10}	1.45×10^{11}	1.98×10^{11}	2.48×10^{11}
A5 → A4b	k_{TS-5}	5.97×10^{5}	3.22×10^{8}	7.65×10^{9}	5.2×10^{10}	1.88×10^{11}
A5 → N1 + H	k_{TS6}	9.1×10^{9}	1.4×10^{11}	5.87×10^{11}	1.44×10^{12}	2.69×10^{12}
N1 + H → A5	k_{TS-6}	8.94×10^{10}	9.92×10^{11}	3.77×10^{12}	9.07×10^{12}	1.72×10^{13}
A5 + H → N1 + H_2	k_{TS6a}	1.40×10^{12}	3.01×10^{12}	4.82×10^{12}	6.76×10^{12}	8.78×10^{12}
N1 + H_2 → A5 + H	k_{TS-6a}	5.13×10^{-10}	3.68×10^{-2}	3.73×10^{2}	1.01×10^{5}	4.89×10^{6}
Acenaphthalene Formation Network						
N1 + H → N2 + H_2	k_{TS1}^{AN}	2.86×10^{10}	9.45×10^{11}	6.37×10^{12}	2.2×10^{13}	5.35×10^{13}
N2 + H_2 → N1 + H	k_{TS-1}^{AN}	8.30×10^{10}	5.44×10^{11}	1.7×10^{12}	3.81×10^{12}	7.07×10^{12}
N2 + C_2H_2 → N3	k_{TS2}^{AN}	6.97×10^{11}	2.18×10^{12}	4.58×10^{12}	7.94×10^{12}	1.23×10^{13}
N3 → N2 + C_2H_2	k_{TS-2}^{AN}	1.83×10^{6}	2.61×10^{9}	9.84×10^{10}	8.7×10^{11}	3.72×10^{12}
N3 → N4	k_{TS3}^{AN}	3.77×10^{8}	6.34×10^{9}	2.65×10^{10}	6.32×10^{10}	1.14×10^{11}
N4 → N3	k_{TS-3}^{AN}	2.96×10^{5}	1.71×10^{8}	4.25×10^{9}	2.98×10^{10}	1.1×10^{11}
N4 → AN + H	k_{TS4}^{AN}	6.55×10^{7}	7.85×10^{9}	9.24×10^{10}	4.24×10^{11}	1.2×10^{12}
AN + H → N4	k_{TS-4}^{AN}	8.53×10^{11}	4.27×10^{12}	1.09×10^{13}	2.07×10^{13}	3.35×10^{13}
N4 + H → AN + H_2	k_{TS4a}^{AN}	1.09×10^{10}	7.55×10^{10}	2.16×10^{11}	4.27×10^{11}	6.95×10^{11}
AN + H_2 → N4 + H	k_{TS-4a}^{AN}	6.80×10^{-9}	8.41×10^{-2}	3.50×10^{2}	5.75×10^{4}	1.84×10^{6}
N1 + C_2H_2 → N5	k_{TS5}^{AN}	2.29×10^{-4}	34.5	1.68×10^{4}	7.87×10^{5}	1.12×10^{7}
N5 → N1 + C_2H_2	k_{TS-5}^{AN}	4.42×10^{9}	1.26×10^{11}	6.91×10^{11}	1.94×10^{12}	3.89×10^{12}
N5 → AN + H_2	k_{TS6}^{AN}	2.86×10^{7}	1.85×10^{9}	1.59×10^{10}	6.01×10^{10}	1.5×10^{11}
AN + H_2 → N5	k_{TS-6}^{AN}	2.29×10^{-10}	8.08×10^{-3}	59.5	1.41×10^{4}	5.88×10^{5}

to a one-step mechanism, and therefore, they can be directly compared. Most of the reactions considered here are indeed dominated by one-step mechanisms (with few exceptions mentioned in the discussion, such as, for example, the C_6H_5 + OH reaction). The calculated rate constants are further scrutinized by a comparison of their absolute values and temperature dependence with those for experimental rate coefficients.

All networks studied here involve several types of reactions. The HACA mechanism includes hydrogen abstractions (we considered H-abstractions by H, OH, and CH_3 radicals), acetylene additions to the radical site, ring closures, and hydrogen atom loss/disproportionation reactions. Only two reaction steps are involved in the Diels−Alder mechanism, acetylene cycloaddition to a singlet PAH molecule and H_2 elimination from the Diels−Alder adduct. Below, we sequentially discuss these reaction types in more detail.

3.1. Hydrogen Abstraction Steps. A. Ab Initio Calculations. The H atom abstraction from a singlet PAH species can occur through its reactions with H/OH radicals ubiquitous in flames or by the reactions with small hydrocarbon radicals, for example, CH_3, C_2H_5, and so forth. Here, we consider the H-abstraction reactions of benzene (A1) with H, OH, and CH_3 and of naphthalene (N1) with H and OH radicals. For the other H-abstraction steps, we limited our consideration to the reactions with H radicals. The mechanism and kinetics of H abstraction from benzene, C_6H_6 + H, have been studied experimentally employing different techniques[47−49] and theoretically using ab initio G2M(rcc,MP2) calculations and the TST method.[16] Several experimental and theoretical kinetic studies on the C_6H_6 + OH abstraction[50−55] provided a large amount of data for comparison with theoretical results. Recently, this reaction was thoroughly investigated by Tokmakov and Lin[17] using G3 and G2M model chemistry methods in conjunction with the weak collision master equation/RRKM analysis. The kinetics of the C_6H_6 + CH_3 abstraction reaction were studied experimentally by Krech and Price[56] and Zhang et al.[57] Most recently, this reaction

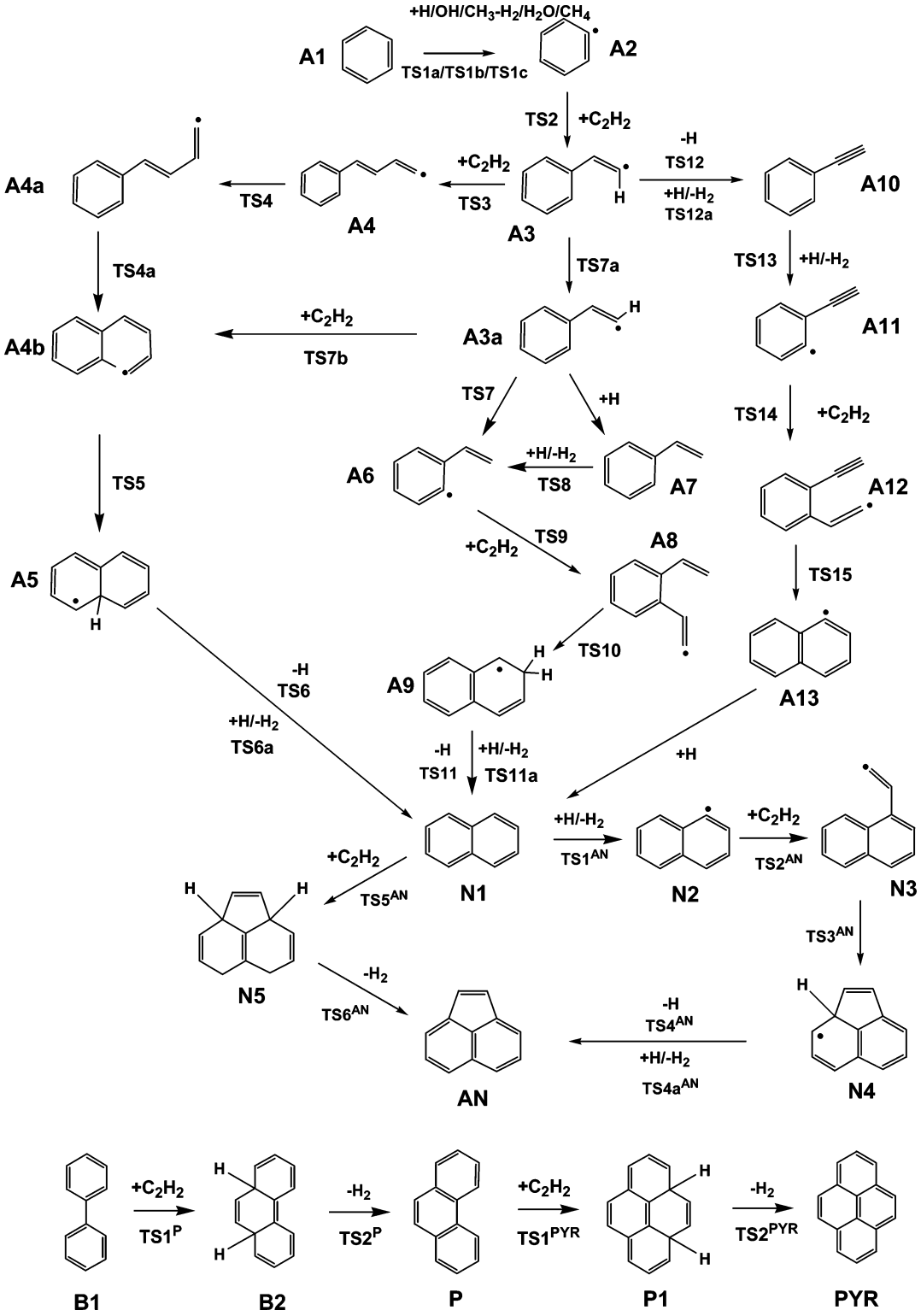


Figure 1. Various reaction pathways involved in the HACA synthesis of naphthalene and acenaphthalene and the Diels−Alder synthesis of acenaphthalene, phenanthrene, and pyrene.

was analyzed both theoretically and experimentally by Tokmakov et al.,[19] who applied two complementary experimental methods, pyrolysis/Fourier transform infrared spectroscopy, and pulsed laser photolysis/mass spectrometry, as well as theoretical G2M(CC,MP2) calculations. For the H-abstraction from naphthalene (N1 → N2), no accurate experimental/theoretical data are available in the literature.

Let us first consider the A1 → A2, A7 → A6, A10 → A11, and N1 → N2 steps, where the abstraction of a H atom from singlet PAH intermediates occurs by the H radical. Because the H-abstraction reactions are endothermic by 8.8−10.2 kcal/mol (see Table 1), all transition states exhibit a late, product-like character. The calculated reaction energies and barrier heights are very similar; the differences between

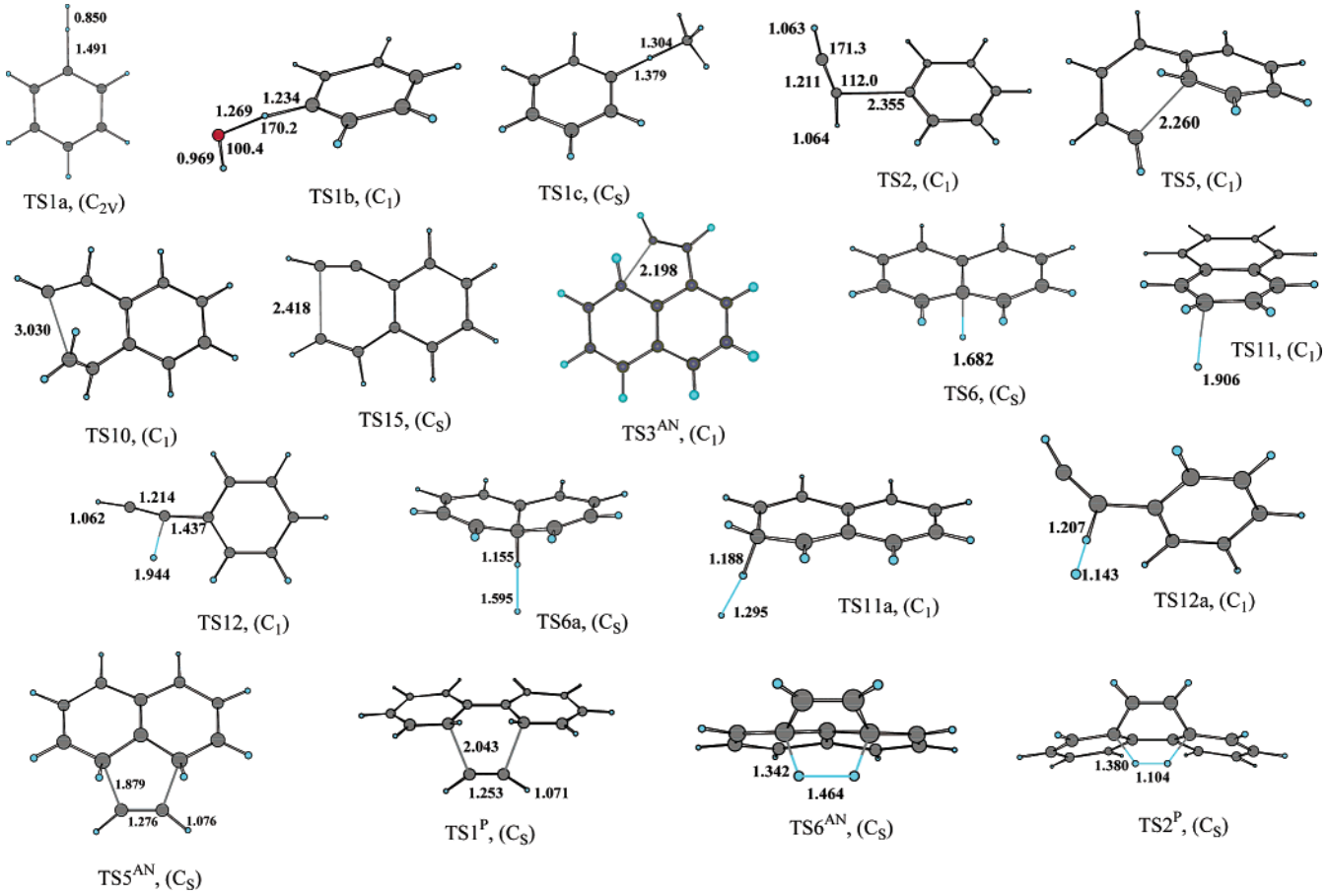


Figure 2. Optimized geometries of various transition states. Selected bond lengths are shown in Å. (For full optimized geometries, see Table S1 in the Supporting Information.)

different reactions do not exceed 1.5 kcal/mol. The highest barrier (17.9 kcal/mol) and heat of reaction (10.2 kcal/mol) are found for the H abstraction from A10 (phenylacetylene), while the lowest barrier (16.5 kcal/mol) is computed for C_6H_6 + H. As seen in Table 1, the B3LYP method significantly underestimates barriers and heats of reaction for the hydrogen-abstraction steps. For example, the B3LYP calculated barrier and heat of the C_6H_6 + H abstraction are found to be 5.7 and 3.3 kcal/mol lower, respectively, than the respective values obtained at the G3 level. The G3 computed barrier height of 16.5 kcal/mol for C_6H_6 + H closely agrees with the experimental activation energy of 16.0 kcal/mol derived from the Arrhenius expression $k(T) = 4.15 \times 10^{-10} \exp[-16\,000/RT]$ cm^3 molecule^{-1} s^{-1} obtained by Kiefer et al.[49] in the high-temperature shock tube pyrolysis of benzene in the temperature range $T = 1900-2200$ K. Similarly close agreement is also found for the heat of reaction, 8.7 ± 0.6 kcal/mol[16] from the experiment and 8.5 kcal/mol at the G3 level. The G2M(rcc,MP2) calculated barrier height and reaction energy for this abstraction step, 19.9 and 11.1 kcal/mol, respectively,[16] are notably higher than the experimental and G3 values. Thus, G3 theory, in this case, provides more accurate energetics than G2M, which tends to overestimate both the barrier and heat of the reaction.

H abstractions by a hydroxyl radical from benzene (A1 + OH) and naphthalene (N1 + OH) exhibit significantly lower barriers (5−6 kcal/mol) than the abstractions by a H radical and are slightly exothermic (by 3−5 kcal/mol). As a result, the structures of corresponding transition states TS1b and $TS1b^{AN}$ exhibit an early character (see Figure 2). As seen in Table S1 of the Supporting Information, TS1b and $TS1b^{AN}$ have low-frequency vibrational modes (lower or about 100 cm^{-1}) corresponding to a torsional motion of the OH group.

Interestingly, the barrier is higher and the exothermicity is lower for the H abstraction by OH from naphthalene than those from benzene. A comparison of B3LYP and G3 barriers and reaction energies for these reactions (Table 1) shows that the B3LYP method significantly underestimates the barrier for the C_6H_6 + OH abstraction, giving an unreasonably low value of 0.4 kcal/mol. For the N1 + OH abstraction, the B3LYP method fails and predicts a negative reaction barrier, which could be caused by the use of a smaller 6-311G** basis set as compared to 6-311+G(3df,-2p) employed for C_6H_6 + OH. The agreement between B3LYP and G3 is good for the heats of reaction, especially, for the C_6H_6 + OH abstraction. While comparing the G3 calculated barrier heights with experimental and theoretical activation energies,[50−55] we have to keep in mind the possibility of the second competitive nonabstraction channel, PAH + OH → PAH−OH, which can contribute to the overall reaction rate and affect the energetic parameters derived from Arrhenius expressions. In some earlier studies,[58,59] the OH addition was recognized as the main reaction channel. However, according to the experimental study by Madronich and Felder,[51] the OH addition channel does not

contribute significantly (<20%) to the overall reaction rate at temperatures relevant to combustion (T > 1000 K). The experimental activation energy of the C_6H_6 + OH abstraction derived in that work is 4.8 kcal/mol, close to our G3-type result of 5.5 kcal/mol and to the G3 value of 5.3 kcal/mol reported by Tokmakov and Lin.[17] Other experimental kinetic measurements for the C_6H_6 + OH reaction[50,52−54] gave different estimates for the activation energy varying from 3.7 to 5.6 kcal/mol. Taking into account the relatively low activation energy and the fact that these experimental works were carried out at different times, using different techniques and temperature ranges, and that they must have different systematic errors, the agreement with the theoretical barrier heights is quite satisfactory. For the reaction energy, our value of −5.4 kcal/mol is very close to the experimental estimate[17] of −5.6 ± 0.8 kcal/mol obtained using thermodynamic parameters from standard reference sources.

Because the H abstraction from benzene by a methyl radical, C_6H_6 + CH_3, was recently studied using the G2M(CC,MP2) method,[19] here we only briefly compare our G3 results with the previous literature data and with the corresponding parameters for the C_6H_6 + H reaction obtained at the same level of theory. The G3 computed reaction barrier of 17.8 kcal/mol and endothermicity of 8.8 kcal/mol for C_6H_6 + CH_3 are close to those for the H abstraction by a H radical, 16.5 and 8.5 kcal/mol, respectively. The transition state TS1c exhibits the same loose, product-like character as that of TS1a for C_6H_6 + H (Figure 2). A specific feature of TS1c is the presence of a very low-frequency mode of 16 cm^{-1} (see Table S1, Supporting Information) corresponding to the internal rotation of the methyl group and two other nearly degenerate modes (86 and 90 cm^{-1}) assigned to bending vibrations of the C−H−C fragment in and out of the symmetry plane. The possibility of the quasi-free methyl rotation makes PESs in the vicinity of TS1c very flat; our B3LYP/6-311+G(3df,2p) optimization gave a C_s-symmetric structure for the transition state, whereas Tokmakov et al.[19] obtained a C_1 structure with a smaller basis set. The difference is, however, insignificant because the energies of the two structures differ only by 0.01 kcal/mol.

The agreement of our G3 computed reaction endothermicity (8.8 kcal/mol) with that of the experiment (8.8 ± 0.9 kcal/mol)[19] is excellent and better than that for the G2M(CC,MP2) value of 10.3 kcal/mol.[19] For the reaction barrier, our value of 17.8 kcal/mol is 1.8 kcal/mol lower than the 19.6 kcal/mol obtained by Tokmakov et al.[19] From the Arrhenius expressions, $k(T) = 6.3 \times 10^{10} \exp[-4680/T]$ cm^3 mol^{-1} s^{-1} by Krech and Price[56] and $k(T) = 2.0 \times 10^{12} \exp[-7580/T]$ cm^3 mol^{-1} s^{-1} by Zhang et al.[57] obtained in early experiments, the activation energies for the C_6H_6 + CH_3 reaction are estimated as 9.7 and 15.8 kcal/mol, both significantly lower than the theoretical predictions. Tokmakov et al.[19] concluded that the kinetic data reported in these works are not reliable and are inconsistent with the higher barrier predicted by accurate ab initio calculations. It is known[16,19] that the G2M method systematically overestimates barrier heights for the C_6H_6 + H and C_6H_6 + CH_3 abstraction reactions because of the unbalanced accuracy of the calculated energies for the reactants and products. To avoid this, the barriers can be corrected for the error in the enthalpy of the reaction. Using such corrections, the G2M values of 19.6 and 19.9[16] kcal/mol for the barrier heights of the C_6H_6 + CH_3 and C_6H_6 + H reactions can be reduced to the values of 18.1 kcal/mol[19] and 17.5 kcal/mol, respectively, which are closer to our G3 results of 17.8 and 16.5 kcal/mol. We conclude again from this comparison that the G3 method could produce highly accurate barriers and enthalpies for H abstraction and, therefore, represents a big improvement for calculations of PESs for these kind of reactions as compared to the G2 theory.

Interestingly, we found a better agreement between G3 and B3LYP calculated barriers and heats of the C_6H_6 + CH_3 reaction than for C_6H_6 + H and similar abstraction reactions by a H radical (N1 + H, A7 + H, and A10 + H). Indeed, for the barrier height for C_6H_6 + CH_3, the difference between G3 and B3LYP/6-311+G(3df,2p) results is only 2 kcal/mol versus 5.7 kcal/mol for the C_6H_6 + H abstraction. For the reaction endothermicity, this difference is even smaller, 1.3 kcal/mol, compared to the 3.3 kcal/mol difference for C_6H_6 + H. Thus, the B3LYP method should perform better for the abstraction reactions by small hydrocarbon radicals than for those by H/OH radicals. We can attribute this difference to the worse ability of B3LYP to correctly predict the H−H bond energies in TS1a and the H_2 molecule.

3.1.B. Rate Constant Calculations. The bimolecular rate constants for the hydrogen-abstraction steps at three temperatures relevant to combustion, 1000, 1500, and 2000 K, and fitted rate expressions are compared in Table 3 with kinetic data available from the literature. For the C_6H_6 + H reaction, the comparison shows a reasonably good agreement (within a factor of 2.25) between our G3/TST computed rate constants, previous G2M/TST results,[16] and the experimental data of Kiefer et al.[49] obtained in the high-temperature shock tube pyrolysis of benzene. The rate expression $4.15 \times 10^{-10} \exp[-16\,000/RT]$ cm^3 $molecule^{-1}$ s^{-1} suggested by Kiefer et al.[49] is the common expression normally used in kinetic modeling studies for the rate constants of H-atom abstraction from different hydrocarbons (see, for example, ref 60). Nicovich and Ravishankara[48] derived a much lower activation energy of 8.1 kcal/mol in their rate expression suggested for 300−1000 K; however, their absolute rate constant at 1000 K agrees with our theoretical data.

Although the calculated barrier heights for the C_6H_6 + H, N1 + H, A7 + H, and A10 + H reactions differ only slightly, the rate constants for the last three reactions are significantly lower than that for C_6H_6 + H. This difference is especially large (nearly 2 orders of magnitude) for A7 + H at low combustion temperatures ($T \sim$ 1000 K). The lower rates for the N1 + H, A7 + H, and A10 + H reactions are due to the lower pre-exponential factors. If we compare rotational symmetry numbers of the reactants, 12 for C_6H_6, 4 for N1, 1 for A7, and 2 for A10, we can see that for benzene, its value is much higher than those for the other species. For the respective transition states, TS1a (2), $TS1^{AN}$ (1), TS8 (1), and TS13 (1), the difference in symmetry numbers is not substantial. The higher value of the rotational symmetry number for benzene significantly affects its partition function and results in a considerable increase of the rate constant

Table 3. Comparison of Available Rate Expressions $k = AT^n \exp(-E/RT)$ and Rate Constants at Several Temperatures Relevant to Combustion for Various HACA and Diels−Alder Steps

	rate expressions			rate constants[a] at T =					
reaction	A[a]	n	E, cal/mol	1000 K	1500 K	2000 K	temperature range, K	method	reference
Hydrogen Abstraction									
A1 + H → A2 + H_2	6.44×10^{-16}	1.86	15 976	1.16×10^{-13}	3.23×10^{-12}	1.99×10^{-11}	300−3000	G3	this work
A1 + H → A2 + H_2	1.00×10^{-15}	1.80	17 107	6.70×10^{-14}	2.16×10^{-12}	1.43×10^{-11}	300−5000	G2M	16
A1 + H → A2 + H_2	4.15×10^{-10}	0	16 000	1.88×10^{-13}	2.45×10^{-12}	8.84×10^{-12}	1900−2200	exp.	49
A1 + H → A2 + H_2	5.00×10^{-12}	0	8100	1.02×10^{-13}	3.72×10^{-13}	7.12×10^{-13}	300−1000	exp.	48
N1 + H → N2 + H_2	4.41×10^{-16}	1.87	17 097	4.74×10^{-14}	1.59×10^{-12}	1.07×10^{-11}	300−3000	G3	this work
A7 + H → A6 + H2	8.68×10^{-19}	2.36	16 917	2.95×10^{-15}	1.17×10^{-13}	8.93×10^{-13}	300−3000	G3[b]	this work
A7 + H → A6 + H2				3.99×10^{-15}	1.30×10^{-13}	8.61×10^{-13}	300−3000	G3	this work
A10 + H → A11 + H_2	2.14×10^{-16}	1.89	17 580	2.06×10^{-14}	7.53×10^{-13}	5.30×10^{-12}	300−3000	G3	this work
A1 + OH → A2 + H_2O	8.62×10^{-21}	3.04	3675	1.93×10^{-12}	1.22×10^{-11}	3.93×10^{-11}	300−3000	G3	this work
A1 + OH → A2 + H_2O	6.70×10^{-22}	3.33	1522	3.15×10^{-12}	1.55×10^{-11}	4.57×10^{-11}	200−2500	G3	17
A1 + OH → A2 + H_2O	3.50×10^{-11}	0	4781	3.51×10^{-12}	7.55×10^{-12}	1.11×10^{-11}	790−1410	exp.	51
A1 + OH → A2 + H_2O	4.67×10^{-18}	2.0	1129	2.71×10^{-12}	7.32×10^{-12}	1.42×10^{-11}	453−1409	exp.[c]	52
A1 + OH → A2 + H_2O	2.70×10^{-16}	1.42	1522	2.32×10^{-12}	5.28×10^{-12}	8.97×10^{-12}		exp.	53
A1 + OH → A2 + H_2O	3.10×10^{-12}	0	5612	2.08×10^{-13}	5.12×10^{-13}	8.04×10^{-13}	213−1150	exp.	50
N1 + OH → N2 + H_2O	1.60×10^{-21}	3.02	4374	2.16×10^{-13}	1.51×10^{-12}	5.12×10^{-12}	300−3000	G3	this work
N1 + OH → N2 + H_2O	1.12×10^{-17}	2.0	2014	4.25×10^{-12}	1.32×10^{-11}	2.76×10^{-11}	≥600	exp.[c]	52
N1 + OH → N2 + H_2O	1.05×10^{-12}	0	1875	4.26×10^{-13}	5.75×10^{-13}	6.69×10^{-13}	≤410	exp.[c]	52
A1 + CH_3 → A2 + CH_4	8.87×10^{-21}	2.89	15 992	1.90×10^{-15}	8.02×10^{-14}	6.72×10^{-13}	300−3000	G3[b]	this work
A1 + CH_3 → A2 + CH_4				2.51×10^{-14}	1.29×10^{-12}	1.25×10^{-11}	300−3000	G3	this work
A1 + CH_3 → A2 + CH_4	2.27×10^{-25}	4.23	14 723	9.32×10^{-16}	5.49×10^{-14}	6.03×10^{-13}	300−2500	G2M[d]	19
A1 + CH_3 → A2 + CH_4	1.05×10^{-13}	0	9728	9.71×10^{-16}	4.62×10^{-15}	1.01×10^{-14}	744−800	exp.	56
A1 + CH_3 → A2 + CH_4	3.32×10^{-12}	0	15 756	1.70×10^{-15}	2.12×10^{-14}	7.50×10^{-14}	650−770	exp.	57
A2 + CH_4 → A1 + CH_3	3.59×10^{-22}	3.22	6948	5.63×10^{-14}	6.47×10^{-13}	2.86×10^{-12}	300−3000	G3[b]	this work
A2 + CH_4 → A1 + CH_3				7.41×10^{-13}	1.04×10^{-11}	5.30×10^{-11}	300−3000	G3	this work
A2 + CH_4 → A1 + CH_3	6.46×10^{-27}	4.57	5498	2.35×10^{-14}	3.62×10^{-13}	2.10×10^{-12}	300−2500	G2M	19
A2 + CH_4 → A1 + CH_3	$10^{12.78}$	0	6201	2.03×10^{-14}	1.60×10^{-13}	4.51×10^{-13}	600−980	exp.	19
Acetylene Addition									
A2 + C_2H_2 → A3	5.46×10^{-18}	2.0484	3162	1.69×10^{-12}	6.40×10^{-12}	1.48×10^{-11}	300−3000	G3[b]	this work
A2 + C_2H_2 → A3				8.52×10^{-13}	3.79×10^{-12}	9.92×10^{-12}	300−3000	G3	this work
A2 + C_2H_2 → A3	2.14×10^{-14}	0.834	4822	6.69×10^{-13}	2.03×10^{-12}	3.80×10^{-12}	>1000	G2M	20
A2 + C_2H_2 → A3	4.10×10^{-18}	1.77	2395	2.65×10^{-13}	7.96×10^{-13}	1.60×10^{-12}	300−2000	exp./BAC-MP4	62
A2 + C_2H_2 → A3	2.63×10^{-11}	0	9832	2.32×10^{-13}	1.12×10^{-12}	2.47×10^{-12}	1000−1330	exp.	63
A2 + C_2H_2 → A3	6.64×10^{-17}	1.56	3800	5.11×10^{-13}	1.77×10^{-12}	3.76×10^{-12}	500−2500	AM1/RRKM	12
N2 + C_2H_2 → N3	2.12×10^{-18}	2.05	1930	1.18×10^{-12}	3.66×10^{-12}	7.67×10^{-12}	300−3000	G3[b]	this work
N2 + C_2H_2 → N3				4.76×10^{-13}	1.80×10^{-12}	4.33×10^{-12}	300−3000	G3	this work
A3 + C_2H_2 → A4	4.35×10^{-18}	2.0	4580	4.85×10^{-13}	2.26×10^{-12}	5.78×10^{-12}	300−3000	G3[b]	this work
A3 + C_2H_2 → A4				4.11×10^{-13}	2.33×10^{-12}	6.89×10^{-12}	300−3000	G3	this work
A3a + C_2H_2 → A4b	1.24×10^{-18}	1.98	5066	9.59×10^{-14}	4.77×10^{-13}	1.26×10^{-12}	300−3000	G3[b]	this work
A3a + C_2H_2 → A4b				6.78×10^{-14}	4.11×10^{-13}	1.25×10^{-12}	300−3000	G3	this work
A6 + C_2H_2 → A8	5.02×10^{-21}	2.55	3181	4.67×10^{-14}	2.19×10^{-13}	5.87×10^{-13}	300−3000	G3	this work
A11 + C_2H_2 → A12	2.23×10^{-20}	2.50	1283	3.76×10^{-13}	1.27×10^{-12}	2.90×10^{-12}	300−3000	G3	this work
Ring Closure (Radical Cyclization)									
A4b → A5	3.61×10^{9}	0.601	3635	4.07×10^{10}	9.27×10^{10}	1.47×10^{11}	300−3000	G3[b]	this work
A4b → A5				5.19×10^{10}	9.74×10^{10}	1.34×10^{11}	300−3000	G3	this work
A8 → A9	5.14×10^{12}	0.056	2127	2.71×10^{12}	3.90×10^{12}	4.70×10^{12}	300−3000	G3	this work
A12 → A13	9.95×10^{11}	0.045	5395	1.01×10^{11}	2.45×10^{11}	3.83×10^{11}	300−3000	G3	this work
N3 → N4	2.88×10^{11}	0.225	17 027	3.76×10^{8}	6.39×10^{9}	2.66×10^{10}	300−3000	G3	this work
cyclization of butylbenzene radical	4.60×10^{9}	0	11 496	1.82×10^{7}	1.15×10^{8}	2.90×10^{8}	100−1100	B3LYP/TST	68
Hydrogen Atom Loss									
A5 → N1 + H	7.16×10^{10}	0.756	15 144	9.09×10^{9}	1.45×10^{11}	6.00×10^{11}	300−3000	G3	this work
A9 → N1 + H	6.09×10^{10}	0.867	29 169	1.96×10^{7}	3.11×10^{9}	4.09×10^{10}	300−3000	G3	this work
A3 → A10 + H	7.18×10^{10}	1.02	38 674	6.66×10^{5}	5.19×10^{8}	1.53×10^{10}	300−3000	G3	this work
A3 → A10 + H	3.80×10^{11}	0.82	38 910	8.13×10^{5}	5.82×10^{8}	1.67×10^{10}	>1000	G2M	20
$C_6H_5CCH_2$ → A10 + H	1.23×10^{13}	0.55	42 580	6.97×10^{5}	8.05×10^{8}	2.87×10^{10}	>1000	G2M	20
N4 → AN + H	1.02×10^{11}	0.862	27 663	6.54×10^{7}	8.14×10^{9}	9.48×10^{10}	300−3000	G3	this work
Hydrogen Disproportionation									
A5 + H → N1 + H_2	3.51×10^{-15}	1.08	2011	2.34×10^{-12}	4.91×10^{-12}	7.92×10^{-12}	300−3000	G3	this work
A9 + H → N1 + H_2	3.40×10^{-16}	1.25	3405	3.56×10^{-13}	1.01×10^{-12}	1.91×10^{-12}	300−3000	G3	this work
A3 + H → A10 + H_2	2.74×10^{-13}	0.490	10 630	4.78×10^{-14}	3.11×10^{-13}	8.55×10^{-13}	300−3000	G3[b]	this work
N4 + H → AN + H_2	2.04×10^{-15}	0.984	9589	1.81×10^{-14}	1.22×10^{-13}	3.54×10^{-13}	300−3000	G3	this work
Acetylene Cycloaddition									
N1 + C_2H_2 → N5	5.18×10^{-22}	2.66	67 609	3.78×10^{-28}	5.70×10^{-23}	2.79×10^{-20}	300−3000	G3	this work
B1 + C_2H_2 → B2	2.81×10^{-23}	2.60	42 193	2.65×10^{-24}	6.60×10^{-21}	4.12×10^{-19}	300−3000	G3	this work
P + C_2H_2 → P1	7.18×10^{-23}	2.58	41 945	6.92×10^{-24}	1.64×10^{-20}	1.0×10^{-18}	300−3000	G3	this work
H_2 Loss									
N5 → AN + H_2	1.33×10^{10}	0.783	24 012	2.82×10^{7}	1.86×10^{9}	1.61×10^{10}	300−3000	G3	this work
B2 → P + H_2	4.73×10^{9}	0.797	17 176	2.95×10^{8}	6.53×10^{9}	3.28×10^{10}	300−3000	G3	this work
P1 → PYR + H_2	4.52×10^{10}	0.533	17 938	3.16×10^{8}	7.29×10^{9}	3.60×10^{10}	300−3000	G3	this work

[a] Units for the A factors and rate constants are s^{-1} and $cm^3\ s^{-1}$ molecule^{-1} for the unimolecular and bimolecular steps, respectively. [b] Computed by treating low-frequency modes as free rotors. Otherwise, all frequency modes were treated as harmonic oscillators during calculation of partition functions. [c] Recommended expressions obtained using combined experimental data and applying the unit-weighted least-squares analysis. [d] Rate expression obtained using the rate expression for the A2 + CH_4 → A1 + CH_3 step, $k(T) = 6.46 \times 10^{-27} T^{4.57} \exp(-5498/RT)$ $cm^3\ s^{-1}$ mol^{-1}, and the expression for the equilibrium constant, $K_{eq}(T) = (2.85 \pm 0.42) \times 10^{-2} T^{0.34} \pm 0.02 \exp(4438.2 \pm 16.1/T)$.[19]

for the $C_6H_6 + H$ reaction. In other words, the higher rate of the $C_6H_6 + H$ abstraction is attributed to its higher reaction path degeneracy. For the same reason, a similar large difference in rate constants is found for the $C_6H_6 + OH$ and N1 + OH abstraction reactions, which also have similar barrier heights of 5.5 and 6.3 kcal/mol, respectively. Thus, the use of the common rate expressions, such as those proposed by Kiefer et al.[49] for $C_6H_6 + H$ and by Madronich and Felder[51] for $C_6H_6 + OH$, is not a good assumption if applied to any H-abstraction process from aromatic hydrocarbons other than C_6H_6. The error in the rate coefficients in this case may exceed an order of magnitude even if the activation energy is expected to be similar to that for the model $C_6H_6 + H/OH$ reactions. In this view, the use of the Reaction Class Transition State Theory/Linear Energy Relationship approach to predict rate constants for a group of similar reactions of related hydrocarbons suggested by Truong and co-workers is a better alternative.[61]

Because the A7 intermediate has one very low-frequency mode of 21 cm^{-1} corresponding to the internal rotation of the vinyl group, we expected a significant effect on the partition function of A7 and, therefore, on the rate constants for the A7 + H abstraction reaction if this mode is treated as a free rotor rather than a harmonic oscillator. However, as seen in Table 3, this is not the case; the rate constants calculated within the two models differ only slightly, indicating that anharmonicity associated with the rotational motion of the vinyl group in A7 does not affect the reaction rate significantly, apparently because of large moments of inertia for rotation around the C−C bond, 1.46×10^{-38} and 3.2×10^{-39} g cm^2 for the phenyl and vinyl fragments, respectively.

The available rate expressions for the H-abstractions by an OH radical are presented in Table 3, along with our fitted expressions for the $C_6H_6 + OH$ and N1 + OH reactions. For $C_6H_6 + OH$, the agreement between our results and the G3/TST computed rate constants of Tokmakov and Lin[17] is found to be within a factor of 1.63 at typical combustion temperatures, although in our treatment, we considered all frequency modes in TS1b, including the low-frequency (100 cm^{-1}) OH-torsional mode, as harmonic oscillators, whereas Tokmakov and Lin[17] applied the hindered rotor treatment for this mode. Also, the agreement with the experimental data of Madronich and Felder,[51] Atkinson,[52] and Baulch et al.[53] is satisfactory; the largest deviation is a factor of 4.38. Significant disagreement (about 1 order of magnitude) is found only with the rate suggested in early experiments.[50] For the N1 + OH abstraction, the agreement of the calculated rate constants with the low-temperature experimental kinetic data of Atkinson[52] is worse, in the range of 1 order of magnitude. According to our G3/TST calculations, the C_6H_6 + OH abstraction reaction exhibits about 1 order of magnitude higher rate constants as compared to those for the abstraction by an H radical, $C_6H_6 + H$, at typical combustion temperatures ($T < 1500$ K), but at higher temperatures, their rates become similar. Such a behavior has an obvious explanation; at lower temperatures, the C_6H_6 + OH abstraction is faster because of a significantly lower barrier height, but as temperature increases, the $C_6H_6 + H$ rate constant grows almost 3 times as fast as that for C_6H_6 + OH because of the higher barrier height. If the concentrations of H and OH radicals are comparable, then C_6H_6 + OH must be the dominant channel for H abstraction at combustion temperatures ($T \sim 1500$ K), whereas at higher temperatures, one should expect a competition between the $C_6H_6 + H$ and $C_6H_6 + OH$ abstractions.

In calculations of rate constants for the $C_6H_6 + CH_3$ reaction, we applied a free internal rotor treatment for the lowest frequency mode (16 cm^{-1}) in TS1c, whereas the other two low-frequency C−H−C bending modes (85−90 cm^{-1}) were treated as harmonic oscillators. Although Tokmakov et al.[19] considered these two modes as hindered rotors, our results coincide with those by Tokmakov et al. within a factor of 2. On the other hand, because the rotation of the CH_3 group in TS1c is almost free, this mode should be treated as a rotor, not a harmonic oscillator, otherwise large errors occur in the rate constants (see Table 3). We were able to find only two sets of low-temperature experimental kinetic data for the $C_6H_6 + CH_3$ abstraction.[56,57] The agreement with these data is close for 1000 K; however, if we extrapolate the experimental rate expressions to high temperatures, the deviations are significant. New experimental measurements are suggested to verify our and Tokmakov et al.'s[19] high-temperature rate constants.

3.2. Acetylene Addition to the Radical Site. A. Ab Initio Calculations. The next step in the HACA PAH growth is acetylene addition (or consecutive additions) to the radical site. In our networks, we considered the addition of an acetylene molecule to phenyl and naphthyl radicals ($C_6H_5 + C_2H_2$ and $N2 + C_2H_2$), the addition of a second acetylene to the aromatic ring of the A6 and A11 radical intermediates ($A6 + C_2H_2$, $A11 + C_2H_2$), which represents Frenklach's variations of the HACA mechanism, and the addition of a second acetylene to the first one ($A3 + C_2H_2$, $A3a + C_2H_2$) taking part in Bittner−Howard's route of HACA synthesis.

All computed acetylene addition steps exhibit low barriers of 2−6 kcal/mol and are highly exothermic, by about 40 kcal/mol (see Table 1). The respective transition states TS2, TS3, TS7b, TS9, TS14, and $TS2^{AN}$ have tight, reactant-like structures (Figure 2). As seen in Table S1 of the Supporting Information, the acetylene addition transition states have several low-frequency modes (<100 cm^{-1}): 25, 64, and 93 cm^{-1} (TS2); 16, 44, 52, and 73 cm^{-1} (TS3); 19 and 52 cm^{-1} (TS7b); 45 and 76 cm^{-1} (TS9); 41, 68, and 89 cm^{-1} (TS14); and 33, 51, and 98 cm^{-1} ($TS2^{AN}$). The lowest mode corresponds to the internal rotation of the HCCH fragment around the forming C−C bond, whereas the other low frequencies are due to bending of the acetylene fragment.

The highest barrier in this reaction class is found for the C_2H_2 addition to the phenylvinyl radical (A3a) and the lowest barrier for the addition to the 2-ethynylphenyl radical (A11). Interestingly, the barrier for the C_2H_2 addition to the 1-naphthyl radical is considerably lower than that for the addition to the phenyl radical. The reaction enthalpies vary slightly, with the lowest exothermicity of 32.7 kcal/mol calculated for $A3a + C_2H_2$. The agreement between the barrier heights and reaction enthalpies computed by the B3LYP and G3 methods (see Table 1) is reasonably good;

as compared to G3, B3LYP overestimates the barriers by about 1−2 kcal/mol and underestimates the reaction exothermicity by 1−3 kcal/mol. Thus, B3LYP performs better in describing the reaction energetics of acetylene addition steps than for the hydrogen abstraction reactions discussed in the previous section.

Recent G2M calculations of Tokmakov and Lin[20] gave the barrier height for the $C_6H_5 + C_2H_2 \rightarrow C_6H_5CH{=}CH$ reaction as 3.5 kcal/mol, the same value as our G3 result. For the reaction enthalpy, the agreement is also close; their G2M value is only 1.2 kcal/mol lower (−40.4 kcal) than our prediction (−39.2 kcal/mol). Utilizing experimental heats of formation available for the phenyl radical and acetylene and isodesmic enthalpy of formation of phenylvinyl radical (A3a), the experimental estimation for the reaction enthalpy at 0 K is -39.6 ± 2.8 kcal/mol,[20] close to our value. Available experimental kinetic data[62−64] provide different values for the activation energy of acetylene addition to a phenyl radical. For example, the activation energy of 3.1 kcal/mol from the Arrhenius expression $2.2 \times 10^{11} \exp(-1560/T)$ cm^3 mol^{-1} s^{-1} ($T = 297-523$ K, $P = 20$ Torr) obtained by Yu et al.[62] in a low-temperature cavity ring-down experiment is in good agreement with our calculated barrier and with the results of Tokmakov and Lin.[20] The other two experimental reports[63,64] give significantly higher values for the barrier height, 9.8 kcal/mol from the very-low-pressure pyrolysis/mass spectroscopy study of Fahr et al.[63] and 8.0 kcal/mol from the shock tube/UV absorption study of Heckmann et al.[64] We tend to believe that, because the results of two accurate theoretical calculations (the present G3 study and the G2M study of Tokmakov and Lin[20]) are reliable, only the experimental data of Yu et al.[62] provide the most reasonable estimate for the barrier height of the $C_6H_5 + C_2H_2$ reaction, although their measurements were performed only at low temperatures and within a narrow temperature range.

3.2.B. Rate Constant Calculations. The calculation of rate constants corresponding to the acetylene addition steps requires taking into account the anharmonic effects due to low-frequency internal rotations of the C_2H_2 fragment in the corresponding transition states. Several transition states, TS2, TS3, TS7b, TS2^{AN}, and intermediate A4b, have very low-frequency modes (~20 cm^{-1}), which can be treated either as free or hindered rotors. As seen in Table 3, the effect of the free-rotor treatment (instead of the use of a harmonic oscillator) is not as dramatic as in the case of the $C_6H_6 + CH_3$ reaction, which can be possibly attributed to significantly larger moments of inertia of the rotating C_2H_2 fragment. The free-rotor treatment increases the rate constants by a factor of 1.5−2 as compared to the harmonic oscillator model, and the effect increases with the temperature decrease. The maximal effect is found for the acetylene addition to the naphthyl radical (N2).

For the $C_6H_5 + C_2H_2$ reaction, our computed rate constants overestimate the recent G2M/RRKM results of Tokmakov and Lin[20] by a factor of 2.5−4. However, to better reproduce their experimental kinetic data, Tokmakov and Lin adjusted the G2M computed barrier height of 3.7 kcal/mol to a slightly higher value of 4.1 kcal/mol, whereas in our calculation, we used a lower value of 3.5 kcal/mol computed at the G3 level. The other sets of available experimental and theoretical data provide even lower values of rate constants in the temperature range of 1000−2000 K. For instance, 6−9 times lower values of rate constants were obtained in the experimental study of Yu et al.[62] combined with theoretical BAC-MP4/RRKM calculations and 5−7 times lower values were measured in the low-pressure pyrolysis/mass spectroscopy study of Fahr et al.[63] It is worth noting that the rate expression by Fahr et al.[63] is commonly used for the rate constants of acetylene addition reactions.

According to our calculations, acetylene addition to a phenyl radical has higher rates as compared to the other acetylene addition reactions. Acetylene addition to the A3 isomer of a phenylvinyl radical is significantly (by a factor of 4.5−5) faster than the addition to the A3a isomer. The acetylene addition to a naphthyl radical (N2) has rate constants 1.5−2 times lower than those for the addition to a phenyl radical. The A6 + C_2H_2 reaction exhibits the largest difference for the rate constants (25−36 times lower values) as compared to the $C_6H_5 + C_2H_2$ reaction. Thus, the individual rates for the acetylene addition steps may significantly vary depending on the structure of the reacting radical, and the use of a common rate expression for all acetylene additions to PAH intermediates may result in significant errors and dramatically affect the kinetic modeling results.

Two acetylene addition reactions considered here, A3a + C_2H_2 (to the vinyl chain of a phenylvinyl radical) and A11 + C_2H_2 (to the aromatic ring of a 2-ethynylphenyl radical) are of a special interest in the HACA synthesis, as the former belongs to Bittner−Howard's route (the second acetylene molecule adds to the first one) and the latter is a part of Frenklach's route (the second acetylene adds to the aromatic ring). Acetylene addition to the A3 isomer of a phenylvinyl radical is not relevant because the subsequent rearrangements to the A4b intermediate are energetically and kinetically unfavorable. According to our calculations, at combustion temperatures, the A11 + C_2H_2 addition (Frenklach's route) exhibits 2.3−3.9 times higher rates as compared to the A3a + C_2H_2 addition (Bittner−Howard route).

3.3. Ring Closure Steps. Cyclization steps in the naphthalene formation network are the following: A4b → A5, A8 → A9, and A12 → A13 (see Figure 1). The N3 → N4 ring closure in the acenaphthalene network leads to the formation of a five-membered ring—a substructure of cyclopentafused PAH. All the cyclization steps can be defined as intramolecular additions either to the aromatic C−C bond (A4b → A5 and N3 → N4) or to the double (A8 → A9) or triple (A12 → A13) bonds. Additionally, the A4b → A5 and N3 → N4 steps represent ring closure between a hydrocarbon radical chain and the aromatic ring, whereas the A8 → A9 and A12 → A13 cyclizations involve reactions between two hydrocarbon chains (vinyl and ethynyl) attached to the aromatic ring.

The A4b → A5, A8 → A9, and A12 → A13 cyclization steps exhibit low barriers (1.8−5.4 kcal/mol) and are highly exothermic. The five-membered ring formation in the acenaphthalene core via N3 → N4 is found to have a much higher barrier of 16.6 kcal/mol and a lower exothermicity

as compared to those of the other cyclization steps. This result can be explained if we analyze the structure of the corresponding transition state $TS3^{AN}$ (see Figure 2); the naphthyl fragment in this transition state is nonplanar and differs significantly from that in the reactant, N3 radical. Thus, the ring closure N3 → N4 requires a significant deformation of the reactant molecule in the transition state, apparently because of the steric repulsion. The lowest barrier of 1.8 kcal/mol is found for the A8 → A9 step, which is an example of intramolecular addition to the vinyl double bond. The addition to the ethynyl triple bond (A12 → A13) has a 3 times higher barrier height, 5.4 kcal/mol. The transition state TS10, which corresponds to the A8 → A9 step, is found to have the lowest imaginary frequency of $90i$ cm^{-1}. The analysis of its structure, vibrational frequencies, and the results of IRC calculations show that the ring closure A8 → A9 occurs simultaneously with the rotation of the vinyl fragment around the C−C bond. In accord with the high exothermicity of the ring closure reactions, all respective transition states exhibit an early, reactant-like character.

The agreement between G3- and B3LYP-computed barrier heights and reaction exothermicities for the A4b → A5, A8 → A9, and A12 → A13 ring closure steps is found to be within 0.5−1.0 and within 3 kcal/mol for barriers and reaction enthalpies, respectively. Therefore, B3LYP can perform well in predicting the energetics for these kind of reactions. This is not surprising as a similarly impressive agreement between the high-level CBS−RAD method and B3LYP was also demonstrated[65] for the ring opening of the cyclopropylcarbinyl radical. Up to now, the mechanism of radical cyclization reactions was scarcely studied by high-level ab initio methods. We were able to find only a few DFT studies of analogous reaction steps relevant to the PAH synthesis. Bauschlicher and Ricca[21] computed barrier heights and reaction enthalpies of the A12 → A13 and A4b → A5 cyclization steps at the B3LYP/4-31G level. For A12 → A13, their barrier height of 4.5 kcal/mol and reaction enthalpy of −56.5 kcal/mol are similar to our B3LYP predictions (4.8 and −53.8 kcal/mol, respectively) obtained with the significantly larger basis set. For the A4b → A5 ring closure, the agreement with our B3LYP calculations is notably worse; their computed barrier and reaction enthalpy are 7.3 and −23.6 kcal/mol, respectively, at the B3LYP/4-31G level versus our values of 4.6 and −29.3 kcal/mol, respectively, at the B3LYP/6-311+G(3df,2p) level.

B3LYP/6-311G** calculations were also applied earlier to investigate the mechanism of cyclization reactions for the alkyl chain radicals bonded to various PAHs including benzene, naphthalene, anthracene, and so forth, containing 1−4 aromatic rings.[66−68] The calculated barriers were found to be significantly higher as compared to those of our predictions for the cyclization involving vinyl-type radical side chains and varied from 8.0 to 12.0 kcal/mol depending on the PAH. For instance, for the $C_6H_5-C_4H_8$ (1-phenylbutyl) radical, the barrier for the ring closure was calculated to be 11.8 kcal/mol,[68] whereas our predictions for the cyclization barriers of the A4b radical are 3.6 and 4.6 kcal/mol at the G3 and B3LYP levels, respectively. Such a difference in the calculated barrier heights may originate from the different reactivities of the C_4H_8 (alkyl, single C−C bond) and C_4H_4 (vinyl-type, double C═C bond) radical chains involved in cyclization.

According to the recent G2M study of Tokmakov and Lin,[20] the formation of a four-membered ring of the benzenecyclobutadiene core in the radical cyclization of phenylvinyl radical (A3) and 1-vinyl-2-phenyl radical (A6) was found to be considerably less energetically favorable than cyclizations leading to five- or six-membered rings in naphthalene or acenaphthalene. Indeed, the ring closure of the phenylvinyl radical giving the bicyclo[4.2.0]octa-2,4,7-trien-1-yl radical intermediate requires overcoming a barrier of 33.4 kcal/mol, and a similar barrier of 32.1 kcal/mol was found for the cyclization of the 2-vinylphenyl radical producing bicyclo[4.2.0]octa-1,3,5-trien-7-yl. Significantly lower barriers of 10.1 and 12.5 kcal/mol at the CBS−RAD and G2M(RCC) levels, respectively, were obtained by Smith et al.[65] for the formation of a three-membered ring in the cyclopropylcarbinyl radical. However, their values are also much higher than the barriers for the formation of the second aromatic ring in naphthalene calculated here.

As seen in Table 3, rate constants for the ring closure steps at combustion-relevant temperatures significantly differ from each other and increase in the row: k(N3 → N4) < k(A4b → A5) < k(A12 → A13) < k(A8 → A9). The highest rate is found for the A8 → A9 reaction, in line with the lowest barrier height for this step. Also, the lowest rate for the N3 → N4 ring closure can be attributed to the highest reaction barrier of 16.6 kcal/mol. However, as the temperature increases, the differences between various cyclization rate constants significantly decrease. Available kinetic data for the radical cyclization of the primary butylbenzene radical[68] are compared with our results in Table 3. It can be seen that cyclization involving a single bond of the alkyl radical chain attached to the aromatic ring exhibits significantly lower reaction rates as compared to cyclizations involving double (A4b → A5) or triple (A12 → A13) bonds. This confirms the advantage of the HACA scheme of PAH synthesis with respect to the other mechanisms involving PAH with side alkyl chains.

3.4. Hydrogen Loss/Disproportionation. The reactions of radical cyclization lead to the formation of an additional ring in the forming PAH core. Depending on the hydrocarbon side chains involved in the ring closure, the resulting PAH molecule may contain "extra" hydrogen atoms, like in A5, A9, and N4 intermediates. Elimination of these hydrogen atoms from the radical intermediates together with reactions of radical recombination represent the final steps in the HACA synthesis of larger, stable PAH species. Another important role of hydrogen elimination is to form singlet PAH intermediates with side chains (e.g., phenylacetylene A10), which are involved in further PAH synthesis or represent abundant byproducts in flame combustion. It is commonly assumed that H elimination from radical intermediates at combustion conditions can be realized by the direct hydrogen loss mechanism, which normally has a high barrier. According to our G3 calculations and results of other ab initio studies of similar reactions,[20] the barriers for the H loss from radical intermediates are in the range of several

tens of kilocalories per molecule (see Table 1) and these reactions are highly endothermic. In our previous study[22] of the HACA phenanthrene synthesis, we proposed an alternative hydrogen disproportionation mechanism, which involves free H radicals in the elimination of extra hydrogen atoms (the other abundant small radicals, like OH, CH_3, etc., may also play a similar role). We showed that the disproportionation mechanism is energetically more favorable than the H-atom loss mechanism; the computed barriers for the former are much lower than those for the latter.

Accurate calculations of PESs for hydrogen disproportionation reactions represent a challenging problem because, during this process, the wave function character changes from open-shell to closed-shell singlet. By applying the UB3LYP method, we were often unable to reach convergence during geometry optimization of the H-disproportionation open-shell singlet transition states. To solve this problem, in our previous study, we employed the multireference CASSCF method with the 6-31G* basis set for calculation of reaction barriers; however, because of the large size of the molecules, the calculations were limited to a small (4,6) active space. To make more reliable predictions of the reaction barriers, more sophisticated methods must be applied. In the present study, we utilized the IRCMax{Energy[G3(MP2,CC)]}//IRC-{Geom[UMP2/6-31G*]} method instead of B3LYP or HF. Using UMP2 for geometry optimization, we were able to locate open-shell singlet transition states TS6a, TS11a, TS12a, and TS4a^{AN}. Then, IRC calculations were performed to confirm that the optimized transition state structures actually correspond to the H-disproportionation reaction path. The single-point energies of the transition states and several structures along the minimal energy reaction path obtained by IRC calculations were refined at the G3 level, with the UMP2 frequencies, geometries, and ZPE corrections. Thus, we expect that the accuracy of barrier heights calculated using the IRCMax approach should be close to that anticipated from G3 calculations.

As seen in Table 1, for the A5 → N1, A9 → N1, A3 → A10, and N4 → AN reaction steps, the calculated barriers for H disproportionation are respectively 12.1, 22.8, 24.6, and 15.7 kcal/mol lower as compared to those for H loss. For the reaction enthalpies, the difference is even more dramatic; the disproportionation mechanism is strongly exothermic, by 75−125 kcal/mol, whereas the H loss exhibits an endothermicity of several tens of kilocalories per molecule, except for the A5 → N1 + H reaction, which is only 2.5 kcal/mol endothermic. The H-disproportionation transition states TS6a, TS11a, TS12a, and TS4a^{AN} are more reactant-like, whereas the direct H-loss transition states TS6, TS11, TS12, and TS4^{AN} exhibit more product-like structures. For TS12a, we found a very low-frequency mode of 12 cm^{-1} corresponding to high-amplitude torsional motion of the C_2H_3 fragment around the $C_{Phenyl}-C_{Vinyl}$ bond. During rate constant calculation, this low frequency was treated as a free internal rotor.

Several theoretical predictions of the reaction barriers and enthalpies for the direct H-atom loss reactions can be found in the literature.[20−22,69] Tokmakov and Lin[20] applied G2M theory to compute the PES for the $C_6H_5 + C_2H_2$ reaction including two H-loss steps, from 1-phenylvinyl ($C_6H_5C=CH_2$) and phenylvinyl (A3) radicals. Both reactions have the same product, phenylacetylene (A10). For the former, their computed barrier and reaction enthalpy are 41 and 38.6 kcal/mol, respectively. The H loss from the A3 radical exhibits a barrier of 38 kcal/mol and a reaction enthalpy of 31.5 kcal/mol at the G2M level, which are slightly higher than our G3 values of 36.6 and 30.1 kcal/mol, respectively. The PES for the hydrogen loss from the allyl (C_3H_5) radical to form of a singlet allene was studied at the CCSD(T)/cc-pVTZ//MP4/6-31G* level,[69] and the reaction barrier and enthalpy were evaluated as 63.3 and 55.4 kcal/mol, respectively. The H loss from a cyclic intermediate, the cyclopropyl radical, to form cyclopropene was found to have ∼10 kcal/mol lower values for the barrier (52.1 kcal/mol) and reaction enthalpy (48.8 kcal/mol). These energies are significantly higher than the barrier and reaction enthalpy for the H loss from A3, indicating that the direct H loss is more favorable energetically from the cyclic PAH intermediates than from chain hydrocarbons.

Earlier,[70] we studied the $C_6H_5 + H$ disproportionation reaction at the G2M(cc,MP2) level and found the barriers of 7.6, 10.9, and 17.1 kcal/mol corresponding to the formation of *o*-C_6H_4, *m*-C_6H_4, and *p*-C_6H_4, respectively, whereas the computed heats of reaction were −29.1, −14.8, and −1.9 kcal/mol, respectively. Thus, the barriers for the disproportionation steps considered here are similar to that for $C_6H_5 + H \rightarrow o\text{-}C_6H_4 + H_2$, but their exothermicities are much higher because the products formed are more stable than benzyne. On the other hand, H elimination from C_6H_5 requires much higher barriers, especially to produce the *m*- and *p*-C_6H_4 isomers. Rate constant calculations[70] demonstrated that the $C_6H_5 + H \rightarrow o\text{-}C_6H_4 + H_2$ disproportionation channel can compete with the $C_6H_5 + H \rightarrow C_6H_6$ recombination only at high temperatures ($T > 2000$ K).

In our previous study of the HACA phenathrene synthesis, we applied lower-level B3LYP/6-31G* and CASSCF(4,6)/6-31G*//UHF/6-31G* calculations and considered both direct H loss and H disproportionation from various PAH radical intermediates. For the H-loss reactions similar to A5 → N1 + H, we obtained barriers of 16.6−19.0 kcal/mol and reaction enthalpies were in the range of 7.5−10.9 kcal/mol. For H disproportionation, barriers and reaction exothermicities were 0.9−5.8 and 92.5−95.9 kcal/mol, respectively. For the reaction steps similar to A9 → N1 + H, the barriers were 27.5−32.3 and 0.7−1.5 kcal/mol for the H loss and H disproportionation, respectively, and reaction enthalpies varied within 24.2−29.8 and −73.6 to −79.2 kcal/mol for the H-loss and disproportionation mechanisms, respectively. The reactions leading to acetylenic intermediates, like A3 → A10 + H, were shown to have barriers within 40.6−40.2 (H loss) and 4.7−9.5 kcal/mol (H disproportionation) and heats of reaction were 35.8−36.0 (H loss) and −67.4 to −67.7 kcal/mol (H disproportionation). These data demonstrate a close agreement between our earlier results for the H-elimination reactions from PAH radical intermediates (both for the H-loss and H-disproportionation mechanisms) obtained using lower-level ab initio methods and the present, supposedly more accurate, G3 values.

The calculated rate constants for the H-loss and H-disproportionation steps involved in the naphthalene and acenaphthalene formation networks are compared in Table 3. For A3 → A10 + H, the agreement of our G3/RRKM calculated rate constants with those reported by Tokmakov and Lin[20] is very close. Rate constants for the H-loss steps increase systematically with decreasing barrier heights; the lowest rate is found for the A3 → A10 + H reaction with the barrier of 36.6 kcal/mol, and the highest rate is calculated for A5 → N1 + H with the lowest barrier of 14.8 kcal/mol. As the hydrogen disproportionation steps have much lower barriers, their rate constants are much higher than those for the H-loss channels. We can conclude this from the comparison of the first-order rate constants for the H-loss steps with the second-order rate constants for the H-disproportionation steps converted to the $cm^3\ mol^{-1}\ s^{-1}$ units (see Table 2). For instance, at 1500 K, the rate constants are 1.4×10^{11} (3.0×10^{12}), 3.0×10^9 (6.2×10^{11}), 4.9×10^8 (2.0×10^{11}), and $7.9 \times 10^9\ s^{-1}$ ($7.6 \times 10^{10}\ cm^3\ mol^{-1}\ s^{-1}$) for the A5 → N1, A9 → N1, A3 → A10, and N4 → AN steps, respectively, where the H-disproportionation rate constants are given in parentheses. Another issue concerns equilibrium constants for the H-loss steps. As seen in Table 2, the rates for the backward reactions (hydrogen atom addition) are significantly faster as compared to those for the forward reactions, which are responsible for the H-atom removal from the PAH intermediate. The computed rate constants for the forward (reverse) reactions at 1500 K are 1.4×10^{11} (9.9×10^{11}), 3.0×10^9 (1.1×10^{13}), 4.9×10^8 (1.3×10^{13}), and $7.9 \times 10^9\ s^{-1}$ ($4.3 \times 10^{12}\ cm^3\ mol^{-1}\ s^{-1}$) for A5 → N1, A9 → N1, A3 → A10, and N4 → AN, respectively. Thus, the equilibrium of H-loss reactions is considerably shifted to the reactant species suppressing the production of singlet PAH molecules unless the H radicals are consumed in some other reactions. On the contrary, equilibrium constants for hydrogen disproportionation reactions are high (see Table 2), and therefore, the H-disproportionation steps can be considered as irreversible. However, the formation of singlet PAH species by the H-disproportionation mechanism can compete with their consumption in the H-addition reactions (reverse reactions for the H-loss steps), leading back to the reactant species. Indeed, if we compare the rate constants of the forward reactions for the H-disproportionation steps, A5 + H → N1 + H_2, A9 + H → N1 + H_2, A3 + H → A10 + H_2, and N4 + H → AN + H_2, with those for the H-atom additions, N1 + H → A5, N1 + H → A9, A10 + H → A3, and AN + H → N4, we find that the latter usually have 1 or 2 orders of magnitude higher rates (except for the A5 → N1 channel) than those of the former.

3.5. Acetylene Cycloaddition (Diels−Alder) Mechanism. Another possible mechanism of PAH formation in combustion flames, which does not involve radical species in any elementary step, is the so-called benzogenic Diels−Alder mechanism,[2] recently proposed to be the dominant route of PAH growth in methane combustion.[31] Together with the conventional HACA synthesis, the Diels−Alder mechanism was postulated as an important process in PAH growth.[31] To prove that the Diels−Alder mechanism is a viable alternative to HACA, accurate calculations of the reaction energetics and rate constants are required. Here, we calculated PESs and reaction rate constants for three Diels−Alder-type acetylene cycloadditions leading to the formation of acenaphthalene (AN), phenanthrene (P), and pyrene (PYR) starting from naphthalene (N1), biphenyl (B1), and phenanthrene (P), respectively (Figure 1). Previously, we studied the B1 → B2 → P reaction sequence at the B3LYP/6-31G* level,[22] and now, we have refined those results at the more accurate G3 level.

The acetylene cycloaddition pathways involve two consecutive elementary steps: [2+4] or [2+3] cycloaddition leading to a Diels−Alder adduct (N5, B2, and P1 intermediates) followed by elimination of H_2. The first step has a high barrier and is endothermic, but the second one is strongly exothermic and exhibits at least 2 times lower barriers. The [2+3] cycloaddition of an acetylene molecule to naphthalene shows a more than 20 kcal/mol higher barrier and a ∼40 kcal/mol higher endothermicity than the [2+4] cycloadditions (B1 + C_2H_2 and P + C_2H_2). The subsequent H_2 elimination from N5 has a 7−8 kcal/mol higher barrier but a considerably higher exothermicity than those for H_2 eliminations from B2 and P1. The trends in the computed rate constants for both acetylene cycloaddition and H_2-loss reactions are parallel to the respective barrier heights. For instance, the [2+3] cycloaddition reaction of acetylene to naphthalene is found to have a lower rate as compared to those for the reactions of C_2H_2 with biphenyl and phenanthrene. The same is true for the subsequent H_2 eliminations; the N5 → AN + H_2 reaction is slower than B2 → P + H_2 and P1 → PYR + H_2. Thus, the [2+3] cycloaddition to naphthalene followed by H_2 elimination is less energetically and kinetically favorable than the [2+4] cycloadditions and H_2 eliminations involving biphenyl and phenanthrene.

We found considerable discrepancies in B3LYP and G3 computed barriers and reaction enthalpies (especially for the latter) for the Diels−Alder steps. B3LYP overestimates barrier heights and reaction endothermicities for acetylene cycloaddition reactions by about 6 and 10 kcal/mol, respectively. For the H_2-elimination steps, the agreement in computed barriers is satisfactory (within 1−2 kcal/mol), but for reaction exothermicities, the B3LYP values overestimate G3 results by 6−8 kcal/mol. Surprisingly few theoretical studies have been carried out on Diels−Alder-type reactions relevant to the formation of PAH or other hydrocarbons abundant in combustion flames. A slightly modified version of the G2M method (G2MS)[71] has been applied by Froese at al. to calculate PESs for the analogous [2+4] cycloaddition of acetylene to butadiene and cyclopentadiene.[72] For these reactions, they found relatively high but almost 2 times lower barriers, 25.8 and 22.0 kcal/mol, respectively, as compared to those of the acetylene cycloaddition steps considered here. For the reaction enthalpies, the difference is even more dramatic; the reactions of acetylene with butadiene and cyclopentadiene are strongly exothermic (by −52.2 and −26.5 kcal/mol, respectively), in contrast to endothermic N1 + C_2H_2 → N5, B1 + C_2H_2 → B2, and P + C_2H_2 → P1 reactions. This can be explained by the higher stability of the reaction products (1,4-cyclohexadiene in the case of the

C_2H_2 + butadiene reaction) as compared to the lower stability of distorted structures of the N5, B2, and P1 Diels−Alder adducts with respect to the aromatic naphthalene, biphenyl, and phenanthrene molecules. Nevertheless, transition states $TS5^{AN}$, $TS1^{P}$, and $TS1^{PYR}$ have structures similar to those for the C_2H_2 + butadiene/cyclopentadiene reactions.

Owing to much higher barriers, all computed Diels−Alder acetylene cycloaddition steps, N1 + C_2H_2 → N5, B1 + C_2H_2 → B2, and P + C_2H_2 →P1, have rate constants several orders of magnitude (12−16, 8−13, and 7−9 orders of magnitude at 1000, 1500, and 2000 K, respectively) lower than the respective values for C_2H_2 addition to radical intermediates A2, A3, N2, A6, and A11. The difference is still large even at higher combustion temperatures of 2500−3000 K. For example, the rate constants for the N2 + C_2H_2 → N3 acetylene addition and the N1 + C_2H_2 → N5 Diels−Alder [2+3] cycloaddition reactions are respectively 1.4×10^{-11} and 1.9×10^{-17} cm^3 s^{-1} $molecule^{-1}$ at 3000 K. This strongly indicates that the Diels−Alder PAH synthesis cannot compete with the conventional HACA mechanism not only at low but also at high combustion temperatures.

4. Conclusions

Ab initio G3-type calculations of PESs and RRKM and TST calculations of reaction rate constants for the HACA mechanism of naphthalene and acenaphthalene synthesis and Diels−Alder reaction pathways leading to acenaphthalene, phenanthrene, and pyrene have been performed. The calculations allowed us, for the first time, to consider the HACA network as a whole, using the most reliable potential energy surfaces, which are expected to be accurate within 1−2 kcal/mol. The barrier heights, reaction energies, and molecular parameters of the reactants, products, intermediates, and transition states have been generated for all types of reactions involved in the HACA mechanism, including H abstraction from various aromatic intermediates, acetylene addition to radical sites, ring closures leading to the formation of additional aromatic rings, elimination of hydrogen atoms, and H disproportionation. The calculated energetic parameters and rate constants are compared with available literature data, both experimentally and theoretically obtained earlier using less accurate computational methods. A good agreement with experimental data has been found wherever the experimental data are available, and in general, the present G3 results are demonstrated to be more accurate than previous theoretical values calculated either by density functional methods or by G2 theory and its modifications.

The results show that the Diels−Alder mechanism cannot compete with the HACA pathways even at high combustion temperatures, because of high barriers and consequently low reaction rate constants. Alternatively, the reactions involved in various HACA sequences are demonstrated to have relatively low barriers and high rate constants under combustion conditions. A comparison of the significance of different HACA mechanisms (e.g., Frenklach's, alternative Frenklach's, and Bittner and Howard's routes) in PAH growth can be made in the future using PESs and molecular parameters obtained in the present work. To achieve this ultimate goal, pressure-dependent rate constants need to be calculated for each reaction step using RRKM theory, and then, kinetic modeling can be performed to estimate relative probabilities of different routes within the PAH formation networks. At real combustion conditions, many other important reaction steps, such as oxidation, radical recombination, and so forth, must be taken into account to predict the system evolution and product yields.

Our results also demonstrate that the use of common rate expressions based on kinetic data for the first aromatic ring (for example, for the C_6H_6 + H and C_6H_6 + OH reactions or acetylene addition to C_6H_5) in kinetic modeling of flame combustion is not a good approximation if applied to any H-abstraction or acetylene addition reactions involving aromatic hydrocarbons other than C_6H_6. The error in the rate coefficients, in this case, may exceed 1 order of magnitude even if the activation energy is expected to be similar to that for the model C_6H_6 + H/OH or C_6H_5 + C_2H_2 reactions. The individual reaction rates vary significantly depending on the structure of the reacting aromatic molecule or radical, and the use of common rate expressions may result in significant errors and dramatically affect the kinetic modeling results. The effect of the free-rotor treatment replacing the harmonic oscillator is significant only in the cases when the moments of inertia of rotating fragments are small, as for the transition state of the C_6H_6 + CH_3 reaction.

Acknowledgment. This work is funded by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Sciences of U.S. Department of Energy (Grant No. DE-FG02-04ER15570). We also thank Academia Sinica, National Science Council of Taiwan, R.O.C., for partial financial support.

Supporting Information Available: Vibrational frequencies, ZPEs, moments of inertia, rotational constants, and optimized Cartesian coordinates of all species involved in the studied PAH formation networks (Table S1); fitted rate expressions in the modified Arrhenius form $k = AT^n \exp(-E/RT)$ for the forward and reverse rate constants for all studied reactions (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Grimmer, G. *Environmental Carcinogens: Polycyclic Aromatic Hydrocarbons: Chemistry, Occurrence, Biochemistry, Carcinogenity*; CRC Press: Boca Raton, FL, 1983.

(2) Ramdahl, T.; Bjorseth, J. *Handbook of Polycyclic Aromatic Hydrocarbons*, 2nd ed.; Marcel Dekker: New York, 1985; p 1.

(3) Perera, F. P. *Science* **1997**, *278*, 1068.

(4) Venkataraman, C.; Friedlander, S. K. *Environ. Sci. Technol.* **1994**, *28*, 563.

(5) Allen, J. O.; Dookeran, K. M.; Smith, K. A.; Sarofim, A. F.; Taghizadeh, K.; Lafleur, A. L. *Environ. Sci. Technol.* **1996**, *30*, 1023.

(6) Richter, H.; Howard, J. B. *Prog. Energy Combust. Sci.* **2000**, *26*, 565.

(7) Frenklach, M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2028.

(8) Cook, D. J.; Schlemmer, S.; Balucani, N.; Wagner, D. R.; Steiner, B.; Saykally, R. J. *Nature* **1996**, *380*, 227.

(9) Snow, T. P.; Le Page, V.; Keheyan, Y.; Bierbaum, V. M. *Nature* **1998**, *391*, 259.

(10) Frenklach, M.; Wang, H. *Proc. Combust. Inst.* **1991**, *23*, 1559.

(11) Frenklach, M.; Clary, D. W.; Gardiner, W. C.; Stein, S. E. *Proc. Int. Symp. Combust.* **1984**, *20*, 887.

(12) Wang, H.; Frenklach, M. *J. Phys. Chem.* **1994**, *98*, 11465.

(13) Appel, J.; Bockhorn, H.; Frenklach, M. *Combust. Flame* **2000**, *121*, 122.

(14) Bittner, J. D.; Howard, J. B. *Proc. Int. Symp. Combust.* **1981**, *18*, 1105.

(15) (a) Frenklach, M.; Moriarty, N. W.; Brown, N. J. *Proc. Combust. Inst.* **1998**, *27*, 1655. (b) Moriarty, N. W.; Brown, N. J.; Frenklach, M. *J. Phys. Chem. A* **1999**, *103*, 7127.

(16) Mebel, A. M.; Lin, M. C.; Yu, T.; Morokuma, K. *J. Phys. Chem. A* **1997**, *101*, 3189.

(17) Tokmakov, I. V.; Lin, M. C. *J. Phys. Chem. A* **2002**, *106*, 11309.

(18) Chen, C.-C.; Bozzelli, J. W. *J. Phys. Chem. A* **2004**, *108*, 4632.

(19) Tokmakov, I. V.; Park, J.; Gheyas, S.; Lin, M. C. *J. Phys. Chem. A* **1999**, *103*, 3636.

(20) Tokmakov, I. V.; Lin, M. C. *J. Am. Chem. Soc.* **2003**, *125*, 11397.

(21) Bauschlicher, C. W.; Ricca, A. *Chem. Phys. Lett.* **2000**, *326*, 283.

(22) Kislov, V. V.; Mebel, A. M.; Lin, S. H. *J. Phys. Chem. A* **2002**, *106*, 6171.

(23) Miller, J. A.; Melius, C. F. *Combust. Flame* **1992**, *91*, 21.

(24) Melius, C. F.; Colvin, M. E.; Marinov, N. M.; Pitz, W. J.; Senkan, S. M. *Proc. Int. Symp. Combust.* **1996**, *26*, 685.

(25) Marinov, N. M.; Pitz, W. J.; Westbrook, C. K.; Vincitore, A. M.; Castaldi, M. J.; Senkan, S. M.; Melius, C. F. *Combust. Flame* **1998**, *114*, 192.

(26) Madden, L. K.; Mebel, A. M.; Lin, M. C.; Melius, C. F. *J. Phys. Org. Chem.* **1996**, *9*, 801.

(27) Miller, J. A.; Klippenstein, S. J. *J. Phys. Chem. A* **2003**, *107*, 7783.

(28) Howe, P.-T.; Fahr, A. *J. Phys. Chem. A* **2003**, *107*, 9603.

(29) Shafir, E. V.; Slagle, I. R.; Knyazev, V. D. *J. Phys. Chem. A* **2003**, *107*, 8893.

(30) Richter, H.; Howard, J. B. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2038.

(31) Siegmann, K.; Sattler, S. *J. Chem. Phys.* **2000**, *112*, 698.

(32) (a) Becke, A. D. *J. Chem. Phys.* **1992**, *96*, 2155. (b) Becke, A. D. *J. Chem. Phys.* **1992**, *97*, 9173. (c) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. (d) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *B37*, 785. (e) Kang, J. K.; Musgrave, C. B. *J. Chem. Phys.* **2001**, *115*, 11040.

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

(34) (a) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650. (b) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Baboul, A. G.; Pople, J. A. *Chem. Phys. Lett.* **1999**, *314*, 101.

(35) The G3large and G3MP2large basis sets can be downloaded from http://chemistry/anl.gov/compmat/g3theory.htm.

(36) (a) Malick, D. K.; Petersson, G. A.; Montgomery, J. A., Jr. *J. Chem. Phys.* **1998**, *108*, 5704. (b) Chuang, Y. Y.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 1140.

(37) Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(38) (a) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys*. **1989**, *90*, 2154. (b) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem*. **1990**, *94*, 5523.

(39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, R. E.; 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.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C. M.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.11; Gaussian, Inc.: Pittsburgh, PA, 2001.

(40) Werner, H.-J.; Knowles, P. J. *MOLPRO*, version 2002.1. Amos, R. D., Bernhardsson, A., Berning, A., Celani, P., Cooper, D. L., Deegan, M. J. O., Dobbyn, A. J., Eckert, F., Hampel, C., Hetzer, G., Knowles, P. J., Korona, T., Lindh, R., Lloyd, A. W., McNicholas, S. J., Manby, F. R., Meyer, W., Mura, M. E., Nicklass, A., Palmieri, P., Pitzer, R., Rauhut, G., Schutz, M., Schumann, U., Stoll, H., Stone, A. J., Tarroni, R., Thorsteinsson, T., Werner, H.-J., Eds.

(41) Steinfield, J.; Francisco, J.; Hase, W. *Chemical Kinetics and Dynamics*; Prentice Hall: Englewood Cliffs, NJ, 1989.

(42) Eyring, H.; Lin, S. H.; Lin, S. M. *Basic Chemical Kinetics*; Wiley: NewYork, 1980.

(43) Robinson, P. J.; Holbrook, K. A. *Unimolecular Reactions*; Wiley: NewYork, 1972.

(44) Glasstone, S.; Laidler, K. J.; Eyring, H. *The Theory of RateProcesses*; McGraw-Hill: New York, 1941.

(45) Stein, S. E.; Rabinovitch, B. S. *J. Phys. Chem.* **1973**, *58*, 2438.

(46) Beyer, T.; Swinehart, D. F. *Commun. ACM* **1973**, *16*, 379.

(47) Rao, V. S.; Skinner, G. B. *J. Phys. Chem.* **1984**, *88*, 5990.

(48) Nicovich, J. M.; Ravishankara, A. R. *J. Phys. Chem.* **1984**, *88*, 2534.

(49) Kiefer, J. H.; Mizerka, L. J.; Patel, M. R.; Wei, H. C. *J. Phys. Chem*. **1985**, *89*, 2013.

(50) Tully, F. P.; Ravishankara, A. R.; Thompson, R. L.; Nicovich, J. M.; Shah, R. C.; Kreutter, N. M.; Wine, P. H. *J. Phys. Chem*. **1981**, *85*, 2262.

(51) (a) Madronich, S.; Felder, W. *J. Phys. Chem.* **1985**, *89*, 3556. (b) Felder, W.; Madronich, S. *Combust. Sci. Technol.* **1986**, *50*, 135.

(52) Atkinson, R. *Chem. Rev.* **1986**, *86*, 69.

(53) Baulch, D. L.; Cobos, C. J.; Cox, R. A.; Esser, C.; Frank, P.; Just, T.; Kerr, J. A.; Pilling, M. J.; Troe, J.; Walker, R. W.; Warnatz, J. *J. Phys. Chem. Ref. Data* **1992**, *21*, 411.

(54) Lorenz, K.; Zellner, R. *Ber. Bunsen-Ges. Phys. Chem.* **1983**, *87*, 629.

(55) Wallington, T. J.; Neuman, D. M.; Kurylo, M. J. *Int. J. Chem. Kinet.* **1987**, *19*, 725.

(56) Krech, M.; Price, J. W. *Can. J. Chem.* **1967**, *45*, 157.

(57) Zhang, H.-X.; Ahonkhai, S. I.; Back. M. H. *Can. J. Chem.* **1989**, *67*, 1541.

(58) Bittner, J. D.; Palmer, H. B.; Howard, J. B. In *Soot Formation in Combustion Systems and Its Toxic Properties*; Lahaye, J., Prado, G., Eds.; Plenum: New York, 1983; p 95.

(59) Hsu, D. S. Y.; Lin, C. Y.; Lin, M. C. *Proc. Int. Symp. Combust.* **1984**, *20*, 623.

(60) (a) Richter, H.; Grieco, W. J.; Howard, J. B. *Combust. Flame* **1999**, *119*, 1. (b) Wang, H.; Frenklach, M. *Combust. Flame* **1997**, *110*, 173.

(61) Violi, A.; Truong, T. N.; Sarofim, A. F. *J. Phys. Chem. A* **2004**, *108*, 4846.

(62) Yu, T.; Lin, M. C.; Melius, C. F. *Int. J. Chem. Kinet.* **1994**, *26*, 1095.

(63) Fahr, A.; Mallard, W. G.; Stein, S. E. *Proc. Int. Symp. Combust.* **1986**, *21*, 825. (b) Fahr, A.; Stein, S. E. *Proc. Int. Symp. Combust.* **1988**, *22*, 1023.

(64) Heckmann, E.; Hippler, H.; Troe, J. *Proc. Int. Symp. Combust.* **1996**, *26*, 543.

(65) Smith, D. M.; Nicolaides, A.; Golding, B. T.; Radom, L. *J. Am. Chem. Soc.* **1998**, *120*, 10223.

(66) Speybroeck, V.; Reyniers, M.-F.; Marin, G. B.; Waroquier, M. *Chem. Phys. Chem.* **2002**, *3*, 863.

(67) Speybroeck, V.; Hemelsoet, K.; Waroquier, M.; Marin, G. B. *Int. J. Quantum Chem.* **2004**, *96*, 568.

(68) Speybroeck, V.; Borremans, Y.; Neck, D.; Waroquier, M.; Wauters, S.; Saeys, M.; Marin, G. B. *J. Phys. Chem. A* **2001**, *105*, 7713.

(69) Deyerl, H.-J.; Fischer, I.; Chen, P. *J. Chem. Phys.* **1999**, *110*, 1450.

(70) Mebel, A. M.; Lin, M. C.; Chakraborty, D.; Park, J.; Lin, S. H.; Lee, Y. T. *J. Chem. Phys.* **2001**, *114*, 8421.

(71) Froese, R. D. J.; Humbel, S.; Svensson, M.; Morokuma, K. *J. Phys. Chem. A* **1997**, *101*, 227.

(72) Froese, R. D. J.; Coxon, J. M.; West, S. C.; Morokuma, K. *J. Org. Chem.* **1997**, *62*, 6991.

CT0500491

JCTC Journal of Chemical Theory and Computation

Double-Metal-Ion/Single-Metal-Ion Mechanisms of the Cleavage Reaction of Ribozymes: First-Principles Molecular Dynamics Simulations of a Fully Hydrated Model System

Mauro Boero,*[,†] Masaru Tateno,[‡,||] Kiyoyuki Terakura,[§,||] and Atsushi Oshiyama[†]

Institute of Physics, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan, Center for Biological Resources and Informatics, Tokyo Institute of Technology, Nagatsuta 4259, Midori-ku, Yokohama 226-8501, Japan, Division of Frontier Research, Creative Research Initiative "Sousei", Hokkaido University, Kita 21, Nishi 10, Kita-ku, Sapporo 001-0021, Japan, and Research Institute for Computational Sciences, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

Received March 14, 2005

Abstract: The role of metal cations (Mg^{2+}) in the cleavage reaction of fully hydrated RNA enzymes is investigated via Car−Parrinello calculations. We find that the action of two metal catalysts is the most efficient way to promote, on one hand, the proton abstraction from $O^{2'}-H$ that triggers the nucleophilic attack and, on the other hand, the weakening and subsequent cleavage of the $P-O^{5'}$ bond. The elimination of one of the two metal cations is shown to lead to an increase in the activation energy. Furthermore, we also find that an OH^- included in the coordination shell of the Mg^{2+} close to $O^{2'}$ promotes the initial proton abstraction and prevents its transfer to the ribozyme in both single- and double-metal-ion pathways, consistently with the experiment. This suggests that in real ribozyme systems, the double-metal-ion reaction mechanism in the presence of an OH^- anion is favored with respect to single-metal-ion mechanisms.

1. Introduction

The great importance of catalytic RNA molecules (ribozymes) stems from their ability to cleave other target RNA molecules with a high specificity and, hence, to inhibit gene expression.[1−18] The hydrolysis of the phosphodiester proceeds by cleaving the $P-O^{5'}$ bonds of a RNA target molecule at a specific site,[1,2] a reaction known as transesterification, giving as a product 2′,3′-cyclic phosphate and a departing 5′-hydroxyl group.[8−14] The reaction path of the transesterification has been discussed by several authors over the years.[1−20] For the ongoing discussion, we shall refer to the generally accepted scheme, reported in Figure 1, shared by all the self-cleaving ribozymes;[1,14] namely, the reactant (1) overcomes an activation barrier passing across the trigonal bipyramidal (TBP) transition state, carrying a pentacoordinated P, (2) before sliding down to the final product (3).[1,2] Two metal cations, indicated in Figure 1 as M_1 and M_2, are expected to act as catalysts, since a large variety of RNA enzymes are known to be metalloenzymes.[19−30] The most common metal used in experiments is Mg^{2+}. However, several divalent metal ions have been used either in vivo or in vitro,[21−27] and the transesterification rate has been shown to depend also on the nature of the metal catalyst adopted.[22−28]

A lot of effort, from both experimentalists[1−28] and theoreticians,[29−43] has been devoted to unraveling the intimate

* To whom correspondence should be addressed. Tel.: +81-29-8535921, e-mail: boero@comas.frsc.tsukuba.ac.jp.
† University of Tsukuba.
‡ Tokyo Institute of Technology.
§ Hokkaido University.
|| National Institute of Advanced Industrial Science and Technology.

10.1021/ct050066q CCC: $30.25
Published on Web 07/07/2005

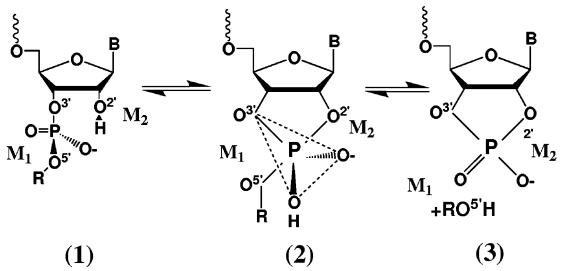


Figure 1. Schematic cleavage reaction path for a general transesterification reaction promoted by two divalent metal ions M_1 and M_2.

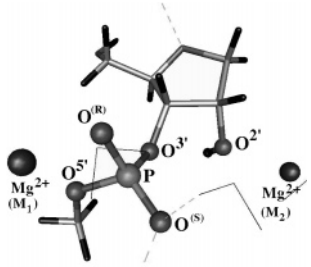


Figure 2. Average configuration of the RNA system used in the simulations. All the most relevant atoms are labeled according to the standard notations and displayed as balls; the metal ions are also indicated as M_1 and M_2 in parentheses for the sake of clarity. Thick sticks show the rest of the structure, with the H atoms in black and the C atoms in gray. Thinner v-shaped sticks are some of the surrounding water molecules, and dashed lines indicate the H bonds.

details of the transesterification reaction. Yet, because of the intrinsic complexity of the mechanism, several questions are still unanswered and many details of the reaction pathway still escape direct experimental probes. Nevertheless, past experimental and theoretical studies suggest that the important elementary processes on which one must focus can be summarized in (i) the role of divalent metal cations as a Lewis acid[27] and the number of ions involved in the reaction, (ii) the deprotonation of $O^{2'}-H$,[43−48] (iii) the nucleophilic attack of $O^{2'}$ to P, (iv) $P-O^{5'}$ bond cleavage, and (v) pseudorotations of the phosphodiester.

The aim of the present work is to analyze the effects of the actors, such as metal cations, water, and OH^-, on the above elementary processes. We can remark, however, that, as the present study does not deal with the system in acidic conditions, pseudorotation is not expected to be an important issue.[1] We focus on two fundamental issues: the nature of the interactions of the RNA moiety with the divalent metal cations and the proton abstraction from the $O^{2'}-H$ group, along with its subsequent transfer either to the ribozyme departing group or to the solvent. We give microscopic insight into these two issues via first-principles molecular dynamics for both a single-metal-ion and a double-metal-ion mechanism, computing and comparing activation barriers and reaction pathways for the various cases, evidencing how two metal catalysts are both important in lowering the overall energy barrier and in promoting the reaction. Simulations in which either M_1 or M_2 is removed show that an energetically more demanding reaction path arises: on one hand, the absence of M_2 makes the proton abstraction from the $O^{2'}-H$ group, and the related nucleophilic attack, more problematic; on the other hand, if M_1 is absent, a larger activation barrier exists for the cleavage of the $P-O^{5'}$ bond, indicating that the partial electron withdrawal from the $P-O^{5'}$ bond operated by M_1 plays a special role in the enhancement of the reaction. It is shown that the unoccupied 3s orbitals of Mg^{2+} withdraw electrons from the chemical bonds of the nearby RNA system, thus weakening them and favoring the above-mentioned two reactions, that is, proton abstraction from $O^{2'}-H$ and cleavage of the $P-O^{5'}$ bond. Since an OH^- is often postulated to be involved in the catalytic reactions of metalloenzymes, we analyze the effects induced by the presence of an OH^- hydroxyl anion in the solvation shell of M_2, both in the single-ion and in the double-ion reaction mechanisms. We find that, besides its noticeable effect in triggering the initial proton abstraction, it is an essential ingredient to explain, for example, the absence of proton transfer in the reaction.[44]

2. Computational Methods

First-principles molecular dynamics simulations were performed within the Car−Parrinello[49,50] (CP) scheme in the framework of Kohn−Sham (KS) density functional theory (DFT),[51] including gradient corrections on the exchange and correlation functional after Hamprecht, Cohen, Tozer, and Handy (HCTH).[52] This particular choice of the exchange correlation has already been assessed and turns out to be comparable to a B3LYP hybrid functional[52,53] in reproducing geometries, energetics, and hydrogen-bonding properties.[31,40,53] Whenever open-shell calculations are required, spin polarization within the same HCTH framework was taken into account, and valence-core interactions are described by Troullier−Martins[54] norm-conserving pseudopotentials. In the case of Mg, nonlinear core corrections[55] were included with cutoff radii of 1.104 and 1.343 Å for the s and p angular moments, respectively, and a core density cutoff radius of 0.762 Å, and valence electrons are represented in a plane wave basis set with an energy cutoff of 70 Ry. This cutoff value has already been shown to be necessary to get a basis set convergence sufficiently accurate to describe the fast oscillations of the O wave functions in the vicinity of the core region and to get structural and energetic parameters comparable to all-electron approaches within the same functional.[40−42,53] The simulated systems consist of (*N*, *V*, *T*) ensembles containing one RNA anion moiety including one ribose unit, as shown in Figure 2, amounting to 26 atoms, plus 60 water molecules and two metal cations or 61 H_2O molecules plus a single Mg^{2+} for the double- and single-metal-ion cases, respectively. The pristine pure water system was per-equilibrated alone at standard density (1.0 g/cm^3) and temperature (300 K) in a cubic supercell of side $L =$ 13.146 Å with periodic boundary conditions for about 6 ps. Then, a space suitable to allocate the solute ribozyme model was cleaved as described in ref 31. The temperature was set to 300 K and controlled via a Nosé−Hoover thermostat chain.[56−58] An integration step of 4.0 au (0.0967 fs) with a fictitious electron mass of 600 au ensured good control of the conserved quantities.

The reaction path was sampled via the metadynamics approach,[59,60] by adding to the CP Lagrangean L^{CP} the harmonic degrees of freedom of the collective variables s_α plus the history-dependent Gaussian potential $V(s_\alpha,t)$

$$L = L^{CP} + \sum_\alpha 1/2 M_\alpha \dot{s}_\alpha{}^2 - \sum_\alpha 1/2 k_\alpha [s_\alpha(q) - s_\alpha]^2 - V(s_\alpha,t) \quad (1)$$

where the argument q of $s_\alpha(q)$ can be any function of an arbitrary set of atomic coordinates, electronic wave functions, and so forth. In our cases, we use collective variables that account for the formation of the bond between P and $O^{2'}$, the cleavage of the $P-O^{5'}$ bond, and any other variable involved in the reaction that is expected to be slow with respect to the ps time scale of ordinary first-principles molecular dynamics. Specific details about the set of s_α adopted will be given for each case to support the discussion. In each simulation, a new Gaussian contribution is added to the potential $V(s_\alpha,t)$ every $\Delta t = 0.012$ ps, amounting to 150 molecular-dynamics-generated configurations, as explained in ref 60. By monitoring the running averages

$$\langle \phi^k \rangle_t = \frac{1}{t}\int_0^{t \le \Delta t} \phi^k(t')\, dt' \quad (2)$$

where the integrands $\phi^\kappa(t')$ ($k = 1, 2, 3$) are $E^{KS} = \phi^1$ (the KS total energy), $K_e = \phi^2$ (the electronic fictitious kinetic energy), and $T_{ion} = \phi^3$ (the ionic temperature), we checked that these quantities are sufficiently well-equilibrated during the short sampling time Δt and preserve the adiabaticity of the CP dynamics. This allows for an accurate exploration of the free energy surface (FES) without requiring a too heavy computational workload. The free energy landscape $F(s_\alpha)$ results as

$$F(s_\alpha) = -\lim_{t\to\infty} V(s_\alpha,t) + \text{const.} \quad (3)$$

where the limit has to be intended in the sense given in ref 60; that is, additional Gaussians are added until the selected portion of the phase space, spanned by the collective variables, is saturated. The history-dependent potential has the explicit form

$$V(s_\alpha,t) = \int_0^t dt' |\dot{\vec{s}}(t')| \delta\left\{\frac{\dot{\vec{s}}(t')}{|\dot{\vec{s}}(t')|}[\vec{s} - \vec{s}(t')]\right\} A(t') \exp\left\{-\frac{[\vec{s} - \vec{s}(t')]^2}{2(\Delta s)^2}\right\} \quad (4)$$

where $\vec{s} = (s_1, ..., s_\alpha, ...)$ and the Gaussian amplitude $A(t')$, having the dimensions of an energy, is sampled in the interval (0.02, 0.4) kcal/mol. Computing free energies is, anyhow, always a very delicate issue. In this work, in an attempt at estimating free energy differences, we simply inspect the overall free energy landscape given by eq 3 and seek for the minimum energy barrier on $V(s_\alpha)$ in an interval $\Omega = [s_1{}^a, s_1{}^b] \times [s_2{}^a, s_2{}^b] \times ...$ around the transition value of each collective variable, separating the reactants from the products, that is

$$V(s_a{}^{TS}) = \min_{\{s_\alpha \in \Omega\}} [V(s_\alpha)] \quad (5)$$

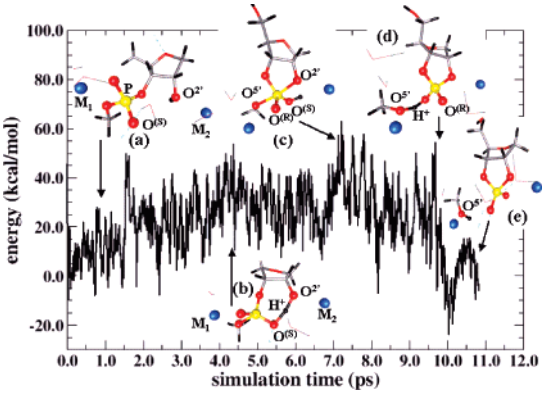


Figure 3. Potential energy during the metadynamics simulation at 300 K. The main steps are (a) the initial configuration, (b) the proton transfer from $O^{2'}-H$, (c) the TBP transition structure, (d) the final proton transfer to the departing group, and (e) the final state. The arrows indicate approximately the location of the different configurations along the metadynamics trajectory, and the initial configuration is assumed as a reference point on the energy axis. The most relevant atoms are labeled in the snapshots according to Figure 2, and the colors adopted are black for H atoms, gray for C, red for O, yellow for P, and blue for Mg^{2+}.

and assuming, in each case, as a reference level, the initial stable local minimum $V(s_\alpha{}^0)$ at $s_\alpha(t = 0) = s_\alpha{}^0$, that is, before any Gaussian penalty potential is added. The latter corresponds to the initial free dynamics of the system equilibrated on the reactants side. We then use $\Delta F = V(s_\alpha{}^{TS}) - V(s_\alpha{}^0)$. Of course, this estimation is biased by how well-resolved the sampled free energy hypersurface is. The dispersions reported in the following paragraphs refer, instead, to the dynamical averages, at a given $s_\alpha(t)$ over the aforementioned 150 CP steps performed before updating $s_\alpha(t)$, that is, $\delta(\Delta F)^2 = \langle V^2(s_\alpha)\rangle - [V(s_\alpha)]^2$. As for the total energy, we simply time-averaged over the trajectory intervals at fixed $V(s_\alpha)$, that is, in between a Gaussian contribution and the subsequent one, along the evolution of the system as displayed in Figures 3 and 7. Further details will be given in the following paragraphs whenever needed to support the discussion.

In some cases, the activation barrier was also checked via thermodynamic integration within the Blue Moon ensemble method,[61] assuming as a reaction coordinate the distance $P-O^{2'}$. In doing so, equilibration times of ~2.0 ps for each sampled constraint ensured meaningful statistics and a good convergence of the running averages of both the constraint force and the KS total energy. Total and free energies are affected by an average error bar typical of first-principles approaches and amounting to 2−3 kcal/mol. Mulliken charges are computed by a standard projection of the KS states onto atomic orbitals.[62]

3. Results and Discussion

Before starting simulations using phase space sampling techniques (either metadynamics or the Blue Moon ensemble method) to analyze the reaction path leading to the transesterification,[1,2] we pre-equilibrated, for about 8 ps via standard unconstrained CP[49,50] molecular dynamics at room temper-

Table 1. Main Geometrical Parameters of the Initial Configurations and TBP Structures as Obtained during the Dynamical Simulations Described in the Text[a]

initial	no metal	one Mg^{2+} close to $O^{5'}$	one Mg^{2+} close to $O^{2'}$	two Mg^{2+}	one Mg^{2+} with OH^-	two Mg^{2+} with OH^-
$P-O^{2'}$	4.134 ± 0.141	3.901 ± 0.135	3.924 ± 0.128	4.202 ± 0.262	3.904 ± 0.095	3.970 ± 0.114
$P-O^{5'}$	1.621 ± 0.050	1.646 ± 0.059	1.630 ± 0.055	1.649 ± 0.064	1.637 ± 0.066	1.662 ± 0.072
$M_1-O^{5'}$		2.156 ± 0.159		2.279 ± 0.107		2.510 ± 0.185
$M_1-O^{(R)}$		2.072 ± 0.122		2.157 ± 0.119		2.052 ± 0.139
$M_2-O^{2'}$			2.188 ± 0.068	2.277 ± 0.106	2.051 ± 0.152	2.598 ± 0.389
$O^{2'}PO^{5'}$	$97.6 \pm 10.1°$	$105.4 \pm 4.9°$	$112.0° \pm 8.6°$	$99.1° \pm 6.7°$	$115.1° \pm 7.9°$	$94.6° \pm 7.2°$
TBP	no metal	one Mg^{2+} close to $O^{5'}$	one Mg^{2+} close to $O^{2'}$	two Mg^{2+}	one Mg^{2+} with OH^-	two Mg^{2+} with OH^-
$P-O^{2'}$	1.689 ± 0.051	1.760 ± 0.063	1.783 ± 0.099	1.809 ± 0.107	1.809 ± 0.098	1.812 ± 0.097
$P-O^{5'}$	1.667 ± 0.067	1.781 ± 0.119	1.805 ± 0.138	1.820 ± 0.137	1.795 ± 0.097	1.821 ± 0.105
$M_1-O^{5'}$		2.170 ± 0.204	3.052 ± 0.489[a]	2.190 ± 0.149		2.212 ± 0.142
$M_1-O^{(R)}$		2.104 ± 0.258		2.089 ± 0.228		2.131 ± 0.232
$M_2-O^{2'}$			3.504 ± 0.590	2.401 ± 0.233	2.018 ± 0.160	3.235 ± 0.365
$O^{2'}PO^{5'}$	$178.1 \pm 6.7°$	$177.0 \pm 7.8°$	172.3 ± 13.9	$179.4 \pm 5.9°$	$178.4° \pm 5.1°$	$178.9° \pm 5.3°$

[a] The value is the $M_2-O^{5'}$ distance since a large distortion, shown in Figure 6, brings M_2 close to $O^{5'}$.

ature $T = 300$ K,[56−58] the RNA model structure in a neutral solution with two Mg^{2+} metal cations in the vicinity of $O^{2'}$ and $O^{5'}$. As expected, the average equilibrium structure (Figure 2) shows that the two metal ions occupy stable positions close to the phosphate; more precisely, the Mg^{2+} corresponding to M_1 includes in its solvation shell the $O^{5'}$ atom, whereas $O^{2'}$ participates with the solvation shell of M_2. The main geometrical parameters of the initial equilibrated structures used in our simulations are reported in Table 1, along with the TBP states, and hereafter, we shall refer to them whenever necessary to support the discussion.

3.1. Double-Metal-Ion Mechanism in the Absence of OH^- Hydroxyl Anion. The reaction path was sampled assuming, as a starting point, this equilibrated configuration and using the metadynamics[59,60] approach. In this case, as inferred from the accepted reaction mechanism and the known experimental evidence,[1,2] the collective variables were chosen to be the coordination number between P and $O^{2'}$ [$s_1(t) = N_{coord}(P-O^{2'})$], the coordination number between P and $O^{5'}$ [$s_2(t) = N_{coord}(P-O^{5'})$], and the coordination number between $O^{2'}$ and its initially bound H atom [$s_3(t) = N_{coord}(O^{2'}-H)$]. Here and in the following, we adopt the definition of N_{coord} introduced in refs 60 and 63, that is, the B coordination around A

$$N_{coord}(A-B) = \frac{1}{N_A}\sum_{I=1}^{N_A}\sum_{J=1}^{N_B}\frac{1-(r_{IJ}/d_{AB})^6}{1-(r_{IJ}/d_{AB})^{12}} \qquad (6)$$

where N_A and N_B are the number of atoms of species A and B, respectively, r_{IJ} is the interatomic distance, and d_{AB} is the typical equilibrium distance or chemical bond between A and B. The fictitious effective masses in the Lagrangean of eq 1 were set to $M_1 = M_2 = 30$ au and $M_3 = 50$ au, and the harmonic coupling constants were $k_1 = k_2 = 0.6$ au and $k_3 = 0.8$ au. The result of the simulation is summarized in Figures 3 and 4. Namely, in a neutral environment, the starting structure (Figure 3a) undergoes, first, a deprotonation of the $-O^{2'}-H$ (Figure 3b) with the transfer of the proton from $O^{2'}$ to the $O^{(S)}$ oxygen, and not to *pro-R* oxygen $O^{(R)}$ (Figure 3c). During the dynamics, the system equilibrates with the *pro-R* oxygen being coordinated to the metal ion M_1 and sharing the solvation shell of this Mg^{2+} with $O^{5'}$, as inferred from experiments[43,44]. $O^{2'}$, in turn, comes close to $O^{(S)}$, and both of them coordinate to M_2. An intramolecular hydrogen bond (H bond) $-O^{2'}-H\cdots O^{(S)}$ is then formed, leading to the proton transfer. This is shown in Figure 4 by the variable s_3—the coordination number of $O^{2'}$—that, during the simulation, jumps between 0 and 1 from 3.4 to 6.2 ps[64] as long as the proton is shared between $O^{2'}$ and $O^{(S)}$. The average configuration is shown in snapshot b in Figure 3. Eventually, when the new $P-O^{2'}$ bond is formed, at about $t = 6.2$ ps, s_1 stabilizes to values around 1 and s_3 decreases to zero. This moment of the simulation represents the nucleophilic attack, leading to the formation of a monoanionic/monoprotic TBP intermediate[2] shown in snapshot c of Figure 3, and the energy barrier for the proton transfer amounts to $\Delta E = 8.1 \pm 1.9$ kcal/mol and $\Delta F = 6.9 \pm 1.6$ kcal/mol for the total and free energies, respectively. In this case, a minimum local barrier was found at $s_3 = 0.48$, whereas $s_1 = 0.55$ (i.e., $P-O^{2'}$ bond not yet formed) and $s_2 = 0.80$ (i.e.,

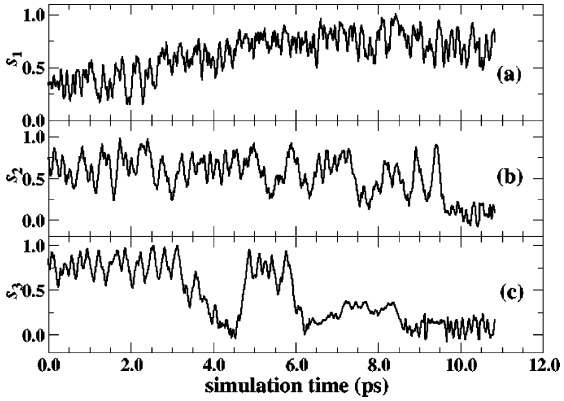


Figure 4. Collective variables as a function of the metadynamics simulation time; s_1, s_2, and s_3 are the coordination numbers of (a) P with $O^{2'}$, (b) P with $O^{5'}$, and (c) $O^{2'}$ with its initially bound H. Further details are given in the text.

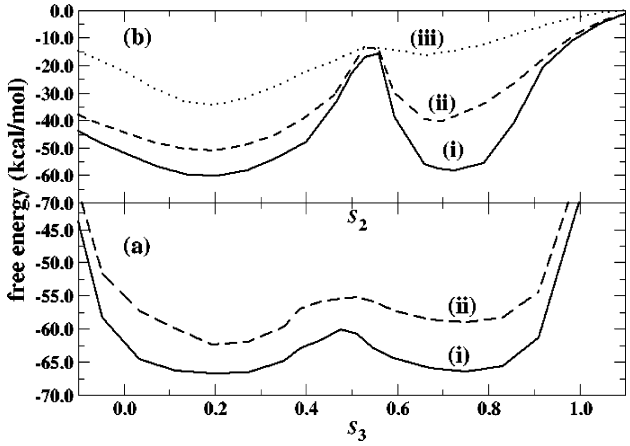


Figure 5. Reconstructed free energy landscape according to eq 4. In panel a, $V(s_1,s_2,s_3)$ is shown as a function of s_3 at (s_1, s_2) = (0.55, 0.80), solid curve (i), and at (s_1, s_2) = (0.65, 0.70), dashed curve (ii), to evidence the $O^{2'}$ deprotonation activation energy. In panel b, $V(s_1,s_2,s_3)$ is shown as a function of s_2 at fixed (s_1, s_3) = (0.80, 0.10), (0.75, 0.15), and (0.70, 0.20) in i (solid curve), ii (dashed curve), and iii (dotted curve), respectively, relevant to the $P-O^{5'}$ bond cleavage.

$P-O^{5'}$ bond not yet broken). This is represented by curve i in panel a of Figure 5 and corresponds roughly to the average configuration (b) of Figure 3. In the same panel, a nearby slice is also reported (dashed curve), representing the free energy profile along the same s_3 collective variable for (s_1, s_2) = (0.65, 0.70) to show how the free energy landscape evolves by moving on the $V(s_\alpha)$ hypersurface in proximity of this transition state.

The system then continues its exploration of the FES until $O^{5'}$ and $O^{2'}$ reach the two axial positions[1] (Table 1), and the $P-O^{5'}$ bond is cleaved before a pseudorotation can occur. Once this bond is cleaved, the departing $R-O^{5'}$ (R, in the present model, being CH_3) group migrates away together with the solvation shell of M_1, to which $O^{5'}$ belongs, and $O^{(R)}$ is no longer coordinated to M_1 (Figure 3e). Since we controlled the coordination number $N_{coord}(O^{2'}-H)$, we see, of course, a drop in its value to zero when the deprotonation occurs (Figure 4).

After the cleavage of the $P-O^{5'}$ bond, the strengthened $P-O^{(S)}$ bond leads to the destabilization of the $O^{(S)}-H$ bond and the proton is released and donated to the $R-O^{5'}$ group that reverts to $R-O^{5'}-H$ (Figure 3d and e). No transfer of the proton to the *pro-R* oxygen was found, since it is hindered by the relatively large distance separating $O^{(R)}$ from H^+. On the other hand, the direct transfer of a proton to the departing group, mediated in our case by $O^{(S)}$, is not totally unexpected and may indeed occur, for example, in the presence of NH_4^+.[43]

By repeating the simulation twice, we noticed that another viable possibility exists. Namely, after the $P-O^{5'}$ bond breaking, the H^+ transferred to the *pro-S* oxygen is spontaneously donated to a nearby H-bound H_2O molecule that, in turn, becomes an OH_3^+, and eventually, the proton is dispersed in the solvent. The departing group forms an H-bonded $R-O^{5'}\cdots H_2O$ complex with another nearby water molecule and eventually gives rise, spontaneously, to the reaction $R-O^{5'} + H_2O \rightarrow R-O^{5'}-H + OH^-$.[65]

Table 2. Computed Activation Barriers (kcal/mol)[a]

	no metal	one Mg^{2+}	two Mg^{2+}	one Mg^{2+} with OH^-	two Mg^{2+} with OH^-
ΔE	60.1 ± 3.1	55.2 ± 2.8[b] 57.3 ± 2.7[c]	46.5 ± 3.0	51.6 ± 2.9[c]	43.8 ± 2.5
ΔF	58.5 ± 2.4	54.0 ± 2.5[b] 55.5 ± 2.5[c]	44.7 ± 2.3	49.2 ± 2.7[c]	41.9 ± 2.2

[a] The error bars reported here are computed as mean square dispersions of the free and total energies during the dynamical sampling of the stationary points. [b] One Mg^{2+} close to $O^{5'}$. [c] One Mg^{2+} close to $O^{2'}$.

The lowest activation barrier for the $P-O^{5'}$ cleavage reaction was identified on the free energy hypersurface at (s_1, s_2, s_3) = (0.80, 0.55, 0.10), as shown in Figure 5b, and this amounts to $\Delta E = 46.5 \pm 3.0$ kcal/mol and $\Delta F = 44.7 \pm 2.3$ kcal/mol for the total and free energies, respectively (Table 2), either in the case in which the released H^+ coming from $O^{(S)}-H$ is transferred to the departing $R-O^{5'}$ group or in the case in which the proton is dispersed in the solvent. In both cases, the collective variables are the same and the final proton transfer is found to be a barrierless process. We remark that the small entropic contribution, in this case, can be ascribed to several causes, namely, the uncertainty affecting the energy barrier estimates due to both the definition adopted and the finite resolution in the free energy reconstruction, the error bars affecting both the total and the free energy, and the fact that, at the transition state, the newly formed $P-O^{2'}$ bond, leading to a pentacoordinated P atom, along with a stabilization of the solvation shells of the two metal ions, makes the system slightly more rigid, thus reducing the entropic contribution that, indeed, is larger by about a factor of two for the initial or final local minima.

A simulation on this same system was performed within the Blue Moon ensemble theory,[61] assuming as a reaction coordinate the distance $P-O^{2'}$. Although the reaction path turned out similar to that of the metadynamics case, as far as the cleavage and proton transfer are concerned, the estimated activation energies turned out to be higher by ~2.0 kcal/mol with respect to the metadynamics estimations. As discussed in ref 63, this seems to be due to the insufficiency of a single reaction coordinate in controlling all the *slow* degrees of freedom involved in the reaction. These activation energies are lower (by about 10 kcal/mol) than analogous barriers computed on this same system in the absence of any metal catalyst (Table 2).[31] Indeed, when metal catalysts are not around, the proton abstraction from $-O^{2'}-H$ occurs simultaneously with the nucleophilic attack and results in a bond switching of $O^{2'}$ from H to P. This is accompanied by several breakings and formations of H bonds with the water molecules of the solvent that contribute to the rising activation barrier. On the other hand, as an inspection of the electronic structure and Mulliken population analysis has evidenced (Table 3), a positive metal ion at the M_2 site withdraws electrons from the $O^{2'}$ oxygen, belonging to its solvation shell, thus weakening the $O^{2'}-H$ bond and favoring the proton release. In an analogous way, the Mg^{2+} located in the vicinity of $O^{5'}$ weakens the $P-O^{5'}$ bond by displacing the electron distribution toward its positive center of charge, thus favoring the cleavage reaction.

Table 3. Mulliken Atomic Charges on the Most Important Atoms in the Initial Configurations and in the Transition States in the Presence of Two Mg^{2+} Metal Cations[a]

atom	initial state	TBP
P	+1.921	+1.786
$H^{2'}$	+0.381	+0.257
$O^{2'}$	−0.389	−0.330
$O^{3'}$	−0.472	−0.355
$O^{5'}$	−0.381	−0.294
M_1	+1.567	+1.442
M_2	+1.558	+1.470

[a] The atom labeling is the same as that in Figure 1.

The electronic structures of the initial (average) configuration and of the TBP transition state were analyzed in terms of Mulliken population,[62] reported in Table 3 for the most relevant atoms involved. As it can be noticed, on one hand, in the initial stage, when the proton abstraction from $O^{2'}-H$ is expected to occur, $O^{2'}$ has a Mulliken charge lower than that of $O^{3'}$ and comparable to that of $O^{5'}$. Indeed, $O^{2'}$ and $O^{5'}$ are part of the solvation shell of M_2 and M_1, respectively. On the other hand, in the TBP configuration, whose main geometrical parameters are listed in Table 1, the Mulliken charges on top of $O^{3'}$ and $O^{5'}$ decrease if compared to the initial state, as a result of the $P-O^{2'}$ bond formation. In particular, the small Mulliken charge of $O^{5'}$ may imply an easier cleavage of $P-O^{5'}$ with respect to a pseudorotation. Let us note also that the presence of the solvent, leading to the H-bonds' formation, alters significantly the Mulliken charges of the O atoms with respect to the vacuum case reported in Table 2 of ref 31. We can infer that spatially extended empty s orbitals, contributed by divalent cations, become partially occupied by withdrawing electrons from those O atoms of the phosphate that belong to the solvation shell of the metal cation. In fact, by analyzing the electronic states of the starting average configuration (Figure 3a) and of the intermediate TBP structure (Figure 3c), we noticed that the spread of the LUMO + 1 orbital,[66] corresponding to the 3s unoccupied states of the metal cations M_1, overlaps $O^{5'}$. As a word of warning, we must acknowledge that,as a result of the energy gap underestimation inherent in these DFT approaches, the location of LUMO + 1 on the energy axis might be lower than in reality and closer to the valence band, thus leading to an overestimation of its interaction with the occupied states. Hence, the picture given above should be intended more as a qualitative trend rather than a precise quantitative estimation.

3.2. Single-Metal-Ion Mechanism in the Absence of OH^- Hydroxyl Anion. A question that remains still unanswered is how many metal ions we really need to enhance the reaction efficiently.[1,30] In an attempt at addressing this issue, we performed two other auxiliary metadynamics simulations by replacing, alternatively, M_1 and M_2 with a water molecule. By separately eliminating the first and the second Mg^{2+}, it is possible to analyze, in detail, the effect of the two metal cations. In our first simulation, we eliminated M_2 from the initial structure of Figure 2 by replacing it with a H_2O molecule and, keeping all the simulation parameters unaltered, we performed a metadynamics run identical to that of the previous case. Contrary to the case in which two metal ions are present, here, the proton abstraction from $-O^{2'}-H$ did not occur until the system reached the TBP structure (Table 1) and occurred in a way similar to the case with an absence of metal ions.[31] The formation of the TBP transition state occurs for a $P-O^{2'}$ bond length of 1.715 Å, shorter than the corresponding TBP formed in the presence of M_2 ($P-O^{2'} = 1.741$ Å), and the release of the proton occurs simultaneously. In fact, only when $O^{2'}$ approaches P, forming a new chemical bond, does the oxygen $O^{2'}$ become over-coordinated (3-fold) and the proton is released. But this is in contrast with the accepted mechanism[1,2] that describes the nucleophilic attack as a consequence of the deprotonation[45,46] of the $O^{2'}-H$ group, which, in this scheme, should precede and not follow the formation of the $P-O^{2'}$ bond. As a consequence, the energy barrier, estimated as described in the previous paragraphs, increases to $\Delta E = 55.2 \pm 2.8$ kcal/mol and $\Delta F = 54.0 \pm 2.5$ kcal/mol. We remark, in passing, that this mechanism seems to be inconsistent with the experiments.[44] The Mg^{2+} located at the M_1 site occupies a dynamically stable catalytic site and does not seem to participate in this first stage of the reaction. Instead, it helps in weakening the $P-O^{5'}$ bond in a way very similar to what was described in the former paragraph and promotes the bond cleavage. Since this bond cleavage is expected to be the rate-limiting step of the reaction, the metal M_1 contributes to the lowering, by about 5 kcal/mol, of the overall energy barrier compared to the case of no metal catalyst. The remaining difference is mostly due to geometrical distortions (see Table 1) and H bonds breaking and forming with the H_2O molecules filling the space formerly occupied by M_2.

On the other hand, when the Mg^{2+} at the M_1 site was eliminated, the deprotonation of $-O^{2'}-H$ was still promoted by the metal cation M_2. However, the proton release occurs close to the transition state and the breaking of the $P-O^{5'}$ bond turns out to be more problematic. In fact, the TBP undergoes a *spontaneous* distortion (Figure 6a and Table 1) that brings both $O^{5'}$ and $O^{2'}$ in the solvation shell of M_2. In these conditions, the proton of $-O^{2'}-H$ is directly transferred to $O^{5'}$ (Figure 6b) and the only available Mg^{2+} also participates in the breaking of the $P-O^{5'}$ bond (Figure 6c). The overall activation energies read $\Delta E = 57.3 \pm 2.7$ kcal/mol and $\Delta F = 55.5 \pm 2.5$ kcal/mol (Table 2); hence, they are not too different with respect to the previous single-metal case. This clearly indicates that the system has a natural tendency to search for available metal cations in order to make the reaction proceed along the smoothest possible path. However, we must remark that the distortion that our model system undergoes in these conditions might well be due to the relatively high mobility of the small $-O^{5'}-CH_3$ group. In a real ribozyme system, such a distortion could be hindered by the more bulky $-O^{5'}-R$, and this suggests that a double-metal-ion mechanism might be favored with respect to a single-metal one. Work is now underway to clarify this point using the whole (hydrated) ribozyme experimental structure.

3.3. The Role of an OH^- Hydroxyl Anion: Double-Metal-Ion Mechanism. In the general reaction of metalloenzymes,[47,48] an OH^- is often postulated to be present in

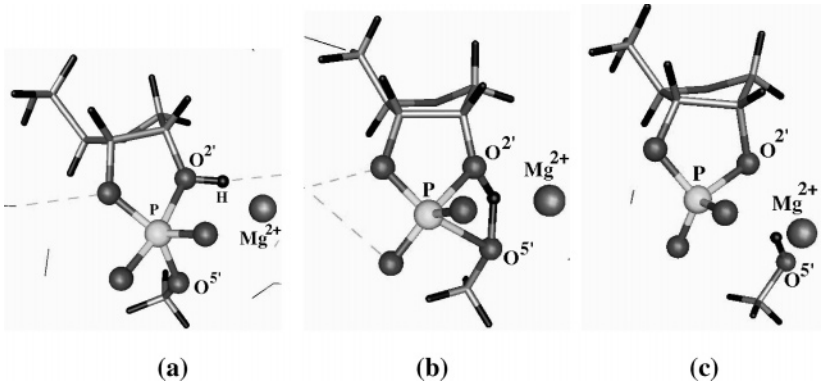


Figure 6. Snapshots of the main steps of the reaction in the presence of a single Mg^{2+} close to $O^{2'}$. The principal atoms are shown as balls and labeled. Color code and graphical representations are the same as those of the preceding figures.

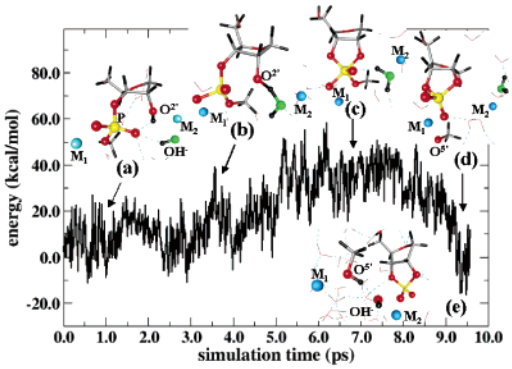


Figure 7. Potential energy for the case of the OH^- hydroxyl anion. The main steps are (a) the initial configuration, (b) the proton abstraction from $O^{2'}-H$, (c) the TBP transition structure, (d) the final $P-O^{5'}$ bond cleavage, and (e) the final product in which a water molecule undergoes a spontaneous dissociation, donating a proton to the departing $O^{5'}$ group. The arrows indicate approximately the location of the various configurations along the metadynamics trajectory, and the initial configuration is assumed as a reference point on the energy axis. Main atoms are labeled and shown as balls; the colors are the same as those of Figure 3, apart from the O belonging to the original OH^- shown in green for the sake of clarity.

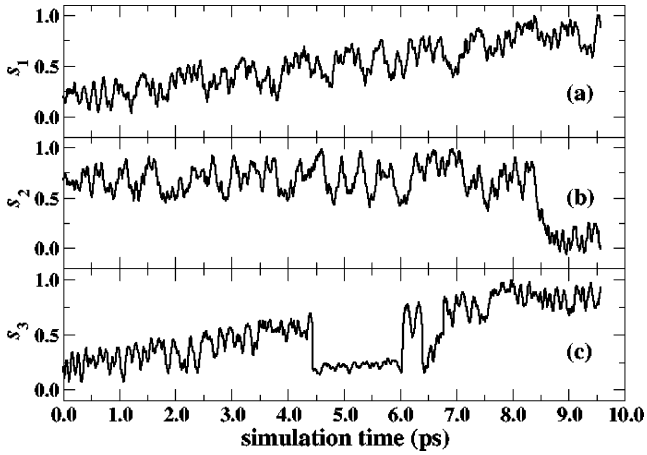


Figure 8. Evolution of the collective variables during the simulation for the case of an OH^- in the coordination shell of M_2. s_1, s_2, and s_3 are the coordination numbers of (a) P with $O^{2'}$, (b) P with $O^{5'}$, and (c) $O^{2'}$ with the nearby OH^- hydroxyl anion. Details are discussed in the text.

the vicinity of the catalytic site M_2. Furthermore, the presence of a hydroxyl anion has been shown to be essential in smoothing the Coulomb repulsion between the two Mg^{2+} ion catalysts when they are too close to each other.[67] In an attempt at understanding its detailed role, we repeated the metadynamics simulation with the two Mg^{2+} metal cations, substituting one of the water molecules in the solvation shell of M_2 with an OH^- (Figure 7, snapshot a). This particular choice is inferred by the known experimental data.[1,2,14,30,44,48] In this case, we assumed as collective variables s_1 and s_2 the coordination number between P and $O^{2'}$ [$s_1(t) = N_{coord}(P-O^{2'})$] and the coordination number between P and $O^{5'}$ [$s_2(t) = N_{coord}(P-O^{5'})$], as in the former simulation. For s_3, instead, we used the coordination number of the H atom of $-O^{2'}-H$ with the O atom of the hydroxyl anion [$s_3(t) = N_{coord}(H^{2'}-OH^-)$], since the detached proton is expected to form a water molecule with this OH^-. The fictitious effective masses were set to $M_1 = M_2 = 30$ au and $M_3 = 50$ au, and the harmonic coupling constants were $k_1 = k_2 = 0.6$ au and $k_3 = 0.8$ au for the three variables, respectively. A total simulation (meta)time of about 9.7 ps was required to explore the FES. The main steps of the reaction and the evolution of the collective variables are reported in Figures 7 and 8, respectively. In summary, between 4.4 and 6.5 ps, the proton of $O^{2'}-H$ jumps between the $O^{2'}$ oxygen and the OH^- group, being substantially a shared proton (Figure 7b), with M_2 playing its "usual" role in the $O^{2'}-H$ bond weakening. Eventually, at 6.7 ps (simulation time), the proton is stabilized on the OH^- anion that becomes a H_2O molecule, and the activation barrier for this process amounts to $\Delta E = 6.7 \pm 2.0$ kcal/mol and $\Delta F = 5.4 \pm 1.6$ kcal/mol for the total and free energies, respectively. This clearly indicates that, in the presence of a hydroxyl anion, this proton is released before the transition state and cannot be transferred to the ribozyme, in agreement with the experimental evidence,[44] and it is only after its release[2,45−48] that the nucleophilic attack of $O^{2'}$ to P can occur, leading to the formation of the TBP structure of Figure 7c. Furthermore, a noticeable decrease of the activation barrier characterizes the proton abstraction from the $-O^{2'}-H$ group if a metal ion and a hydroxyl anion are simultaneously present to cooperate. Once the OH^- is converted into a water molecule via the reaction $-O^{2'}-H + OH^- \rightarrow -O^{2'} + H_2O$, the role of M_2 is

terminated and the metal ion is free to migrate away from the ribozyme moiety keeping inside its solvation shell the newly formed H_2O molecule. Finally, when the evolution of $s_2(t)$ reaches a simulation (meta)time of 8.3 ps (Figure 7), the $P-O^{5'}$ bond undergoes a cleavage (Figure 7d), mediated by M_1, in a way similar to that of the previous cases. As soon as the bond is cleaved, a spontaneous water dissociation, $H_2O \rightarrow H^+ + OH^-$, occurs close to the departing group (Figure 7e), resulting in the protonation reaction $H_2O + R-O^{5'} \rightarrow R-O^{5'}-H + OH^-$. The OH^- anion is left behind, thus restoring the initial conditions, while the departing group, whose oxygen $O^{5'}$ is coordinated to M_1, moves away with the solvation shell of the metal ion. The overall activation barriers estimated are $\Delta E = 43.8 \pm 2.5$ kcal/mol and $\Delta F = 41.9 \pm 2.2$ kcal/mol for the total and free energies, respectively, and it must be remarked that they turn out to be the lowest among all the processes considered in the present work (Table 1). The lowering of the activation barrier can be ascribed to several factors. On one hand, the proton released by the $O^{2'}-H$ group is now significantly stabilized by its reaction with OH^- to form a H_2O molecule. On the other hand, the fact that this proton is not transferred to $O^{(S)}$, contrary to the case of the double-metal-ion reaction path discussed above, induces an intrinsic weakening of the $P-O^{5'}$ bond. This seems to be an important "preconditioning" factor that helps the subsequent bond cleavage.

3.4. The Role of an OH^- Hydroxyl Anion: Single-Metal-Ion Mechanism. Starting from an initial configuration pre-equilibrated for about 5 ps, reported in Figure 9a, a metadynamics simulation was performed in which the Mg^{2+} metal ion corresponding to the M_1 site was eliminated. The system obtained in this way is characterized by a local phosphate−metal−OH^- configuration that closely resembles the system studied in ref 30.

For this simulation, we adopted the same set of collective variables $\{s_1, s_2, s_3\}$ used in the former calculation with the same M_α and k_α ($\alpha = 1, 2, 3$) parameters, for the sake of consistency. The reaction path, in this case, is schematically summarized by the snapshots shown in Figure 9 from a to d. The presence of an OH^- in the solvation shell of the Mg^{2+} in the vicinity of $O^{2'}$ is again the triggering factor that promotes the proton abstraction (Figure 9b) and allows the subsequent nucleophilic attack of the phosphate. The TBP structure of panel c in Figure 9 represents the transition state and corresponds to the local maximum of the free energy surface along the metadynamics reaction path. In this simulation, at variance with the case of the double metal ion, we observed that the Mg^{2+} is still in the vicinity of the ribozyme system also at the end of the reaction (Figure 9 d) and keeps in its solvation shell $O^{2'}$.

The activation barriers were found to be $\Delta E = 51.6 \pm 2.9$ kcal/mol and $\Delta F = 49.2 \pm 2.7$ kcal/mol, thus, slightly lower (by about 5 kcal/mol) than in the case of the single-metal-ion reaction in the absence of OH^- (see Table 2). As in the former cases, the empty 3s orbital of Mg^{2+} overlaps the $O^{2'}$ atom, belonging to the solvation shell of the metal; this results in a weakening of the $O^{2'}-H$ bond and in an easier abstraction of the proton. Calculations on a similar model system have been reported in a recent work.[30]

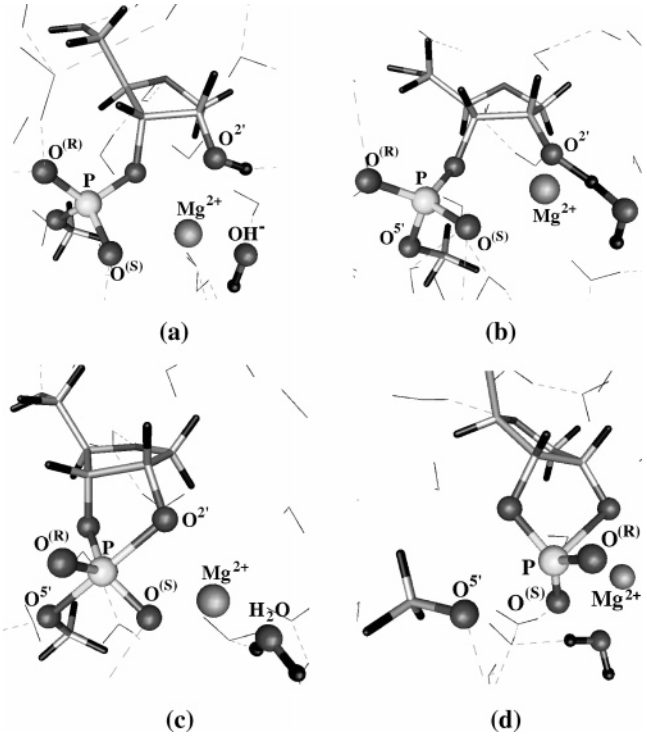


Figure 9. Initial configuration (a), proton abstraction from $O^{2'}$ (b), transition state (c), and final product (d) for the ribozyme system in the presence of a single Mg^{2+} metal ion and one OH^-. Surrounding H_2O molecules are shown as thin v-shaped sticks and H bonds as dashed light blue lines, and the most relevant atoms have been labeled for the sake of clarity.

However, a direct comparison is difficult since, there, static Gaussian calculations were performed, and apart from four H_2O molecules belonging to the solvation shell of the Mg^{2+}, the solvent is not explicitly accounted for in ref 30, but its effects are included by a conductor-like model. Thus, dynamical breaking and formation of H bonds with surrounding water molecules during the reaction are somehow missing. Nevertheless, the energy barriers estimated turn out to be 29.8 and 31.1 kcal/mol before empirical corrections for entropy and solvation effects. Similar results were recently obtained[68] by using a hybrid quantum mechanics/molecular mechanics approach. There, the authors adopt a semiempirical MNDO quantum approach with a CHARMM force field for the ribozyme model and a standard TIP3P model for the solvent water. However, no counterions are included, and this, along with the level of the quantum approach adopted, does not allow for figuring out a definitive picture. In both cases, these calculations seem to be closer to our double-metal-ion case and to the known experimental outcome.[1] As a matter of fact, the experiments are very difficult as well, and their results depend strongly on the environment, the type of ribozyme (hammerhead, hairpin, etc.), the metal cations' concentrations, and the (unknown) number of water molecules participating in the reaction.[1,2] Indeed, it is unclear if, in the ribozyme *pocket* in which the catalytic reaction occurs, the system is fully or only partially solvated. What can be inferred is that the cleavage activation energy can range roughly from 25 to 35 kcal/mol.

Since, as mentioned, the rate-limiting step of the reaction is the cleavage of the $P-O^{5'}$ bond, the computed activation

barriers are of, course higher, than the ones resulting from the double-metal-ion reaction. However, they are significantly lower than the analogous barriers computed for the single-metal-ion pathways in the absence of OH^-. This seems to support the notion that hydroxyl anions are a fundamental ingredient in the reaction mechanism and that at least one divalent metal cation is essential in catalyzing the reaction and enhancing the reaction rate. We observed that, along the whole reaction path, the metal ion does not coordinate directly to the *pro-R* $O^{(R)}$ atom. This is due to the fact that $O^{(R)}$ is, at each step, rather far from Mg^{2+} ($\sim$4.20 Å in the initial configuration, $\sim$3.01 Å at the TBP, and $\sim$3.72 Å in the final state), and its approach to the metal cation would imply a very large distortion of the ribozyme model system. This result seems to be confirmed by the experimental evidence[69,70] and could provide support to the observed absence of thio effects in the hammerhead-ribozyme-mediated cleavage of a phosphorodithionate linkage.

4. Concluding Remarks

We have investigated, via first-principles dynamical simulations, the effects of metal cations in the cleavage reaction of RNA metalloenzymes. The simulations show that two metal catalysts play active roles in enhancing the reaction, and they play separately important roles in two different phases of the transesterification; thus, we can conclude that the major role of the metal ion at the M_1 site is to weaken the $P-O^{5'}$ bond. Conversely, the major contribution of the Mg^{2+} located at the M_2 site is to promote the release of H^+ from the $O^{2'}-H$ group. The absence of one of the two metals has been shown to lead to a distortion of the ribozyme structural model—that is not expected to occur in a more realistic, larger system—and to an increase in the activation barrier, thus lowering the reaction rate. This latter fact is due to the effects induced by divalent metal cations in withdrawing electronic charge from the chemical bonds that undergo a cleaving process. Thus, our analysis indicates rather clearly that a two-metal-ion mechanism is preferred with respect to a single-metal-ion one, leading to a remarkable enhancement of the reaction. Finally, we could shed some light on the proton-abstraction/proton-transfer mechanism from $O^{2'}-H$, underscoring the essential role played by a hydroxyl anion in the reaction, offering an interpretation of the experimental outcome. Finally, since the deprotonation of the $-O^{2'}-H$ group, although not being the rate-limiting step of the reaction, is an important aspect in the reaction pathway, the study presented here can provide a clue in understanding the role of metal ions in the proton-release process of several ribozymes and ribonucleases.[1,2,19−21]

Acknowledgment. We are grateful to Michele Parrinello, Alessandro Laio, Marcella Iannuzzi, Francesco L. Gervasio, Katsumasa Kamiya, and Takashi Ikeda for fruitful discussions. Calculations were performed on the computer facilities of Tsukuba University, ISSP—Tokyo University, and at the Computer Center for Agriculture, Foresty, and Fisheries Research (AFFRC) of the Japan Ministry of Agriculture, Foresty, and Fisheries. We acknowledge support from the Special Nanoscience Project—Tsukuba University and the ACT−JST program.

References

(1) Zhou, D.; Taira, K. *Chem. Rev.* **1998**, *98*, 991−1026.

(2) Perreault, D. M.; Anslyn, E. V. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 432−450.

(3) Hutchins, C. J.; Rathjen, P. D.; Forster, A. C.; Symons, R. H. *Nucleic Acids Res.* **1986**, *14*, 3627−3640.

(4) Uhlenbeck, O. C. *Nature* **1987**, *328*, 596−600.

(5) Haseloff, J.; Gerlach, W. L. *Nature* **1988**, *334*, 585−591.

(6) Koizumi, M.; Iwai, S.; Otsuka, E. *Nucleic Acids Res.* **1989**, *17*, 7059−7071.

(7) Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* **1994**, *372*, 68−74.

(8) Murray, J. B.; Terwey, D. P.; Maloney, L.; Karpeisky, A.; Usman, N.; Biegelman, L.; Scott, W. G. *Cell* **1998**, *92*, 665−673.

(9) Buzayan, J. M.; Gerlach, W. L.; Bruening, G. *Nature* **1986**, *323*, 349−353.

(10) Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4262−4266.

(11) Carola, C.; Eckstein, F. *Curr. Opin. Chem. Biol.* **1999**, *3*, 274−283.

(12) Bramlage, B.; Luzi, E.; Eckstein, F. *Trends Biotechnol.* **1998**, *16*, 434−438.

(13) Hermann, T.; Auffinger, P.; Westhof, E. *Eur. Biophys. J.* **1998**, *27*, 153−165.

(14) Fedor, M. J. *J. Mol. Biol.* **2000**, *297*, 269−291.

(15) Joyce, G. F. *Science* **2000**, *289*, 401−402.

(16) Schlutes, E. A.; Bartel, D. P. *Science* **2000**, *289*, 448−452.

(17) Yoshinari, K.; Taira, K. *Nucleic Acids Res.* **2000**, *28*, 1730−1742.

(18) Taira, K.; Nishikawa, S. In *Gene regulation: biology of antisense RNA and DNA*. Erickson, R. P., Izant, J. G., Eds.; Raven Press: New York, 1992; pp 35−54.

(19) Takagi, Y.; Warashina, M.; Stec, W. J.; Yoshinari, K.; Taira, K. *Nucleic Acids Res.* **2001**, *29*, 1815−1834.

(20) Oshima, K.; Kawasaki, H.; Soda, Y.; Tani, K.; Asano, S.; Taira, K. *Cancer Res.* **2003**, *63*, 6809−6814.

(21) Pyle, A. M. *Science* **1993**, *261*, 709−714.

(22) Steitz, T. A.; Steitz, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6498−6502.

(23) Uchimaru, T.; Uebayashi, M.; Tanabe, K.; Taira, K. *FASEB J.* **1993**, *7*, 137−142.

(24) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065−2069.

(25) Kuimelis, R. G.; McLaughlin, L. W. *Chem. Rev.* **1998**, *98*, 1027−1044.

(26) Zhou, D. M.; Taira, K. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 14343−14348.

(27) Zhou, J.; Zhou, D.; Takagi, Y.; Kasai, Y.; Inoue, A.; Baba, T.; Taira, K. *Nucleic Acids Res.* **2002**, *30*, 2374−2382.

(28) Li, Y.; Breaker, R. R. *J. Am. Chem. Soc.* **1999**, *121*, 5364−5372.

(29) Noodleman, L.; Lovell, T.; Han, W. G.; Li, J.; Himo, F. *Chem. Rev*. **2004**, *104*, 459−508.

(30) Torres, R. A.; Himo, F.; Bruice, T. C.; Noodleman, L.; Lovell, T. *J. Am. Chem. Soc*. **2003**, *125*, 9861−9867.

(31) Boero, M.; Terakura, K.; Tateno, M. *J. Am. Chem. Soc*. **2002**, *124*, 8949−8957.

(32) Zhang, B.; Cech, T. R. *Chem. Biol*. **1998**, *5*, 539−553.

(33) Kuwabara, T.; Warashima, M.; Taira, K. *Trends Biotechnol*. **2000**, *18*, 462−468.

(34) Uchimaru, T.; Tanabe, K.; Nishikawa, S.; Taira, K. *J. Am. Chem. Soc*. **1991**, *113*, 4351−4353.

(35) Dejaegere, A.; Lim, C.; Karplus, M. *J. Am. Chem. Soc*. **1991**, *113*, 4353−4355.

(36) Lim, C.; Karplus, M. *J. Am. Chem. Soc*. **1990**, *112*, 5872−5873.

(37) Storer, J. W.; Uchimaru, T.; Tanabe, K.; Uebayashi, M.; Nishikawa, S.; Taira, K. *J. Am. Chem. Soc*. **1991**, *113*, 5216−5219.

(38) Yliniemela, A.; Uchimaru, T.; Tanabe, K.; Taira, K. *J. Am. Chem. Soc*. **1993**, *115*, 3032−3033.

(39) Uchimaru, T.; Stec, W. J.; Tsuzuki, S.; Hirose, T.; Tanabe, K.; Taira, K. *Chem. Phys. Lett*. **1996**, *263*, 691−696.

(40) Doltsinis, N. L.; Sprik, M. *Phys. Chem. Chem. Phys*. **2003**, *5*, 2612−2618.

(41) Akola, J.; Jones, R. O. *J. Phys. Chem B* **2003**, *107*, 11774−11783.

(42) Alber, F.; Folkers, G.; Carloni, P. *THEOCHEM* **1999**, *489*, 237−245.

(43) Takagi, Y.; Taira, K. *J. Am. Chem. Soc*. **2002**, *124*, 3850−3852.

(44) Sawata, S.; Komiyama, M.; Taira, K. *J. Am. Chem. Soc*. **1995**, *117*, 2357−2358.

(45) Weinstein, L. B.; Jones, B. C.; Cosstick, R.; Cech, T. R. *Nature* **1997**, *388*, 805−808.

(46) Oivanen, M.; Kuusela, S.; Lönnberg, H. *Chem. Rev.* **1998**, *98*, 961−990.

(47) Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc*. **2000**, *122*, 166−167.

(48) Dahm, S. C.; Derrick, W. B.; Uhlenbeck, O. C. *Biochemistry* **1993**, *32*, 13040−13045.

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

(50) CPMD code by J. Hutter et al. at MPI für Festkörperforschung and IBM Zurich Research Laboratory (1990−2004).

(51) Kohn, W.; Sham, L. J. *Phys. Rev. A: At., Mol., Opt. Phys.* **1965**, *140*, A1133−A1138.

(52) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys*. **1998**, *109*, 6264−6271.

(53) Boese, A. D.; Doltsinis, N. L.; Handy, N. C.; Sprik M. *J. Chem. Phys*. **2000**, *112*, 1670−1678.

(54) Troullier, N.; Martins, J. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993−2006.

(55) Louie, S. G.; Froyen, F.; Cohen, M. L. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1982**, *26*, 1738−1742.

(56) Nosé, S. *Mol. Phys*. **1984**, *52*, 255−268.

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

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

(59) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.

(60) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.

(61) Sprik, M.; Ciccotti, G. *J. Chem. Phys*. **1998**, *109*, 7737−7744.

(62) Segall, M. D.; Shah, R.; Pickard, C. J.; Payne, M. C. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 16317−16320.

(63) Boero, M.; Ikeshoji, T.; Teralura, K.; Liew, C. C.; Parrinello, M. *J. Am. Chem. Soc*. **2004**, *126*, 6280−6286.

(64) Here and in every metadynamics simulation, the time has to be intended as the progress in the FES exploration and not as a "real" time, as explained in refs 59 and 60. Hence, it gives the energetically ordered sequence of the various steps of the reaction path along with the energy difference with respect to the initial potential well of the reactants.

(65) Because of the relatively large simulation cell, despite the periodic boundary conditions, on the time scale of the simulation, this OH^- does not recombine with the OH_3^+.

(66) In each case, the LUMO is a diffuse orbital, precursor state of a solvated electron. See, e.g., Boero, M.; Parrinello, M.; Terakura, K.; Ikeshoji, T.; Liew, C. C. *Phys. Rev. Lett*. **2003**, *90*, 226403.

(67) Hermann, T.; Auffinger, P.; Scott, W. G.; Westhof, E. *Nucleic Acids Res*. **1997**, *25*, 3421−3427.

(68) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 7504−7513.

(69) Warashima, M.; Takagi, Y.; Stec, W. J.; Taira, K. *Curr. Opin. Biotechnol*. **2000**, *11*, 354−362.

(70) Murray, J. B.; Scott, W. G. *J. Mol. Biol.* **2000**, *296*, 33−41.

CT050066Q

JCTC Journal of Chemical Theory and Computation

Potential Energy Surface Profile of the Oxygen Reduction Reaction on a Pt Cluster: Adsorption and Decomposition of OOH and H_2O_2

Yixuan Wang* and Perla B. Balbuena*

Department of Chemical Engineering, Texas A&M University, 3122 TAMU, College Station, Texas 77843

Received March 26, 2005

Abstract: Because of their essential roles on determining pathways of the oxygen reduction reaction (ORR), the adsorption behavior and decomposition of the radical OOH and hydrogen peroxide on Pt clusters (Pt_n, $n = 3$, 6, and 10) are extensively investigated using density functional theory. Two types of adsorption of the radical OOH on Pt clusters are found. One-end adsorbed hydrogen peroxide H_2O_2, arising from reduction of adsorbed OOH, is also located on Pt_3, with an adsorption energy of −0.63 eV, suggesting that the ORR may proceed via a series pathway generating H_2O_2 as an intermediate. However, since OOH readily decomposes on Pt_3 into atop adsorbed atomic oxygen and hydroxyl with an activation energy of only ∼0.25 eV, the OOR may take place preferentially via a direct pathway without H_2O_2 produced. A potential energy surface profile for the ORR is proposed, and the adsorption properties of other involved oxide species are characterized.

1. Introduction

The oxygen reduction reaction (ORR), taking place at the cathode of polymer-electrolyte-membrane fuel cells (PEMFC), has been attracting much attention[1−7] in the recent decades because of the desire to elucidate its slow kinetics on electrocatalyst surfaces, which is one of the bottlenecks for achieving improved efficiencies in the fuel cell operation.[8] On the other hand, understanding the ORR mechanisms on a Pt(111) surface, currently known as one of the most active electrocatalysts of PEMFCs, undoubtedly has also significant implications to designing alternative cathode catalysts, aiming to decrease the required amount of the expensive noble metal and to improve the ORR reaction kinetics.

The *overall* ORR on Pt surfaces is a multielectron complex reaction that includes a number of elementary reactions. Yeager proposed two pathways for the *overall* OER in acid medium:[2] (a) a *direct* 4-electron pathway where O_2 is reduced directly to water without involvement of hydrogen peroxide (H_2O_2), $O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$, and (b) a *series* pathway in which O_2 is reduced to H_2O_2, $O_2 + 2H^+ + 2e^- \rightarrow H_2O_2$, followed by its further reduction, $H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O$. It was thereafter suggested that the O_2 reduction on a Pt surface proceeds by a *parallel* pathway, the *direct* and *series* mechanisms occurring simultaneously, with the *direct* as the dominant one.[3] Recent studies from Marković et al.[9] somewhat favored and suggested that a *series* pathway via an *adsorbed* H_2O_2 intermediate may apply to Pt and Pt-based bimetallic catalysts.

Quantum mechanics provides an alternative way to explore the ORR, which is a good complement to state-of-the-art experimental techniques. Using a single Pt atom[5] and two Pt atoms (Pt_2)[6] to coordinate the species most likely to be found on the potential energy surface of the ORR, Anderson and co-workers extensively analyzed the mechanism of the ORR at a molecular level. Adsorbed OOH was proposed as an intermediate for the first electron-transfer step, and its formation has a lower activation energy than the dissociation of the adsorbed O_2 (0.6 eV vs 0.7 eV), implying that O_2 may not dissociate on the Pt surface before the first electron transfer. They also found that OOH dissociation on Pt has a rather small activation barrier of ∼0.06 eV. Our Car-Parrinello molecular dynamics (CPMD) simulations of the ORR on a Pt(111) surface in the presence of hydrated protons

* Corresponding author e-mail: yixuan.wang@chemail.tamu.edu (Y.W.) and balbuena@tamu.edu (P.B.B.).

10.1021/ct0500794 CCC: $30.25
Published on Web 08/06/2005

at 350K basically support their findings.[10] Although dissociative adsorption of O_2 may take place, the adsorbed O_2 would combine with the transferred proton to form OOH_{ads} rather than decompose. Also, OOH_{ads} readily decomposes and is unable to yield H_2O_2. These findings seem to be inconsistent with those from experiments indicating that a certain amount (0−20%, depending on the applied potential) of hydrogen peroxide is formed as either a product or an intermediate of the ORR on Pt(111) surfaces.[11,12] A slight attempt is made here to review the previous work in detail, but it is necessary to mention that there is still a lack of agreement with respect to the nature of the overall ORR mechanism, which motivates us to carry out the present investigation.

A variety of intermediates can be present in the process of the ORR, such as atomic O, radicals OH and OOH, and H_2O_2. The adsorption properties of these intermediates on the Pt electrode are therefore very important to obtain insights into the ORR mechanism. Balbuena et al. investigated O, OH, and H_2O adsorption on Pt and Pt-based bimetallic clusters.[13] Using planewave-based periodic density functional theory, Panchenko et al.[14] very recently performed systematic calculations for adsorption behavior of ORR intermediates on low-index Pt surfaces. In the present study, we first focus on adsorption and decomposition of OOH and H_2O_2 on Pt clusters. The properties of the transition states for the decomposition of H_2O_2 on Pt clusters are reported for the first time to the best of our knowledge. Then, coadsorptions of the relevant oxides are systematically investigated, e.g., O/OH, O/H_2O, OH/OH, OH/H_2O, and H_2O/H_2O. On the basis of this information, a potential energy surface profile of the ORR is established, which allows discussing a couple of very important issues about the reaction mechanisms of the ORR. Individual adsorptions on Pt clusters of some relevant species have been previously investigated by DFT, including O, OH, O_2, and H_2O on Pt_2, Pt_3, and bigger clusters,[6,13−17] for the sake of self-consistency, particularly with respect to necessary corrections, like zero point energy (ZPE) and basis set superposition error (BSSE) corrections; we also did our own calculations for adsorptions of O, OH, O_2, and H_2O on Pt_3.

2. Computational Details

Geometries were fully optimized with the hybrid B3LYP density functional[18,19] as implemented in Gaussian03 RevC.02[20] together with the LANL2DZ pseudopotential and the corresponding double-ζ basis set for Pt[21] and with the 6-311++G** basis set for oxygen and hydrogen. The S^2 expectation values for the involved open-shell species with B3LYP are nearly identical to the exact value, indicating that B3LYP is proper to treat the present open shell systems. The relevant transition states were located with conventional linear synchronous transit QST2 or quadratic synchronous transit QST3 methods. To characterize the stationary points and make the zero point energy (ZPE) corrections, a frequency analysis was done for all stationary points at this level. All minima and transition states have the proper number of imaginary frequencies. To explicitly establish the relevant species, the intrinsic reaction coordinate (IRC) pathway was also run for all the transition states presented here. Since the default grid for the integral did not give sufficient numerical accuracy for some calculations, causing convergence difficulties, "ultrafine" grid was used in all of the present DFT calculations. If not noted otherwise, adsorption energies in the text usually refer to values with unscaled ZPE correction. BSSE corrections were estimated for some complexes.[22] The results show that BSSE is not significant at the level of B3LYP/6-311++G**. For example, BSSE corrections for adsorption energies of O, H_2O, and H_2O_2 on Pt_3 are 0.06, 0.08, and 0.08 eV, respectively. Refined energies were attempted with a sophisticated electron correlation method, CCSD(T), for all of the B3LYP-optimized structures, yet they failed for a few open shell complexes probably due to too high spin contamination.

3. Results and Discussion

3.1. Adsorptions of O, OH, O_2, and H_2O on Pt_3. Consistent with a sophisticated ab initio method, relativistic CI calculation including spin−orbit effects,[23] the present optimization of Pt_3 finds that the triplet state $^3B_1$ (C_{2v} symmetry with Pt−Pt bond lengths of 2.499 and 2.583 Å) is the ground state yet is almost degenerate with the singlet state $^1A_1$ (D_{3h} symmetry with Pt−Pt bond lengths of 2.521 Å) by a small energy difference of 0.011 eV. A very recent B3LYP calculation with Pt−Pt bond lengths fixed at 2.775 Å also predicted that the triplet state is the ground state of Pt_3.[17] Despite small differences arising from the employed methods and calculation details, adsorption energies and sites of the oxide species on Pt_3 are in reasonable agreement between the present and previous results, which are summarized in Table 1. Adsorption energies are defined as the difference between the energy of the complex (e.g., Pt_3−O) and the sum of the energies of the adsorbate (O) and adsorbent (Pt_3). O is most stably adsorbed on a bridge site with a singlet ground state and an adsorption energy of −3.10 eV, closely followed by a top adsorption with a triplet ground state and an adsorption energy of −3.06 eV. The OH radical is preferentially adsorbed atop on Pt_3 with an adsorption energy of −2.62 eV. Previous theoretical studies[14,24] show that OH strongly binds on top as well as on the bridge site of Pt(111) with an adsorption energy between −2.14 and −2.52 eV at low coverage. The adsorption of H_2O on Pt_3 (adsorption energy ∼0.6 eV) with a closed electronic shell is much weaker than those of O and OH; however, it is still one time stronger than regular hydrogen bonding.

3.2. •OOH Adsorption and Dissociation on Pt_n (n = 3, 4, 6, and 10). Two types of •OOH adsorption on Pt_3 have been located. As shown in Figure 1, the structure **S1** is most likely to be considered as a 1-fold adsorption with oxygen connecting to Pt, and structure **S2** is a 2-fold adsorption with Pt−O shorter than Pt−O(H) (2.017 vs 2.418 Å). Adsorption energies of OOH to Pt_3 in **S1** and **S2** are −1.64 and −1.26 eV, respectively. **S2** corresponds to the bridge adsorption conformer of •OOH on Pt(111) reported by Panchenko et al.,[14] where the Pt−O bond distances are 2.03 and 2.64 Å with an adsorption energy of −1.07 eV. Mulliken population analysis indicates that the charge is transferred from the Pt_3 cluster to •OOH by 0.38 and 0.20e in **S1** and **S2**, respectively. Specifically, the charge is mostly transferred from *d* orbitals of the Pt atoms to *p* orbitals of oxygen. The strong adsorption

Table 1. Adsorption and Coadsorption Properties of Oxides on Pt_3: Adsorption Energies (AE/eV), Distances between Pt and O unless Those Noted for O−O, and Vibrational Frequencies of Pt−O Stretch unless Those Specified for O−O Stretch

system	multiplicity	adsorption site[a]	−AE,[b] present	−AE, references	distances/Å	vibrational frequency/ cm^{-1}
O	1	top	2.90/2.98	2.812[f]	1.753	861
	3	top	3.06/3.12	3.455,[f]3.22[c]	1.787	763
	1	bridge	3.10/3.17	3.269,[f]3.24[c]	1.918, 1.918	619
	3	bridge	2.81/2.87		1.960, 1.959	566
OH	2	top	2.62/2.74	2.92[c]	1.943	584
	4	bridge	1.93/2.07	2.10[c]	2.203, 2.203	398
				2.23 (atop),[g] 2.14(bridge)[g]		
O_2	1	bridge	0.02	0.58;[d]	2.08, 1.316(O−O)	541, 1052(O−O)
	3	bridge	0.56/0.60	1.08;[d]	2.102, 2.102, 1.318(O−O)	538, 1050(O−O)
				0.4−0.5[e]		
				0.63(bridge)[g]		
H_2O	1	top	0.64	1.02;[c]	2.25	293
OOH	2	top	1.57/1.64	1.07[g]	1.954, 1.420(O−O)	504 , 896(O−O)
		bridge	1.18/1.26		2.017, 2.482, 1.461(O−O)	492, 781(O−O)
O/OH	2	top	5.41		1.753, 1.932	837/597
	4	top	5.62		1.794, 1.934	749/597
H_2O_2	1	top	0.67		2.150, 1.495(O−O)	660(O−O)
OH/OH	1	top	4.92		1.922, 1.922	598
	3	top	5.21		1.929,1.938	592
O/H_2O	1	top	3.20		1.759, 2.232	849, 308
	3	top	3.86		1.784, 2.213	778, 314
OH/H_2O	2	top	3.32		1.953, 2.233	560, 303
H_2O/H_2O	1	top	1.27		2.257, 2.242	299

[a] Top: one site, with a Pt atom directly connecting to O atom; bridge, O bridges two Pt atoms. [b] AE = *E*(complex) − *E*(Pt_3,M=3) - *E*(adsorbate), separated by a slash they refer to those with and without ZPE. [c] Reference 13, full optimization using B3PW91 with LANL2DZ for Pt and 6-31G* for O implemented in Gaussian 98. [d] Reference 16. [e] Experimental.[28] [f] Reference 17, B3LYP with LANL ECP and LACVP** basis set for Pt and 6-31G** for O, Pt−Pt is fixed at 2.775 Å. [g] Reference 14, VASP (PW91-GGA/PAW) results for a four-layer Pt(111) slab.

of •OOH to Pt_3 is also reflected by a O−O vibrational frequency shift of 259 cm^{-1} in **S1** as compared with the naked •OOH (896 vs 1155 cm^{-1}), as well as by O−O bond stretching of approximately 0.1 Å (1.424 in **S1** vs 1.328 Å in •OOH).

B3LYP predicts that the distorted tetrahedron with a triplet state is the ground state for Pt_4. According to Table 2, the corresponding adsorption of OOH (**S1′**) on the Pt_4 cluster has a similar strength to that on Pt_3 (−1.67 vs −1.57 eV). The ground state of planar Pt_6 with Pt−Pt bond lengths fixed

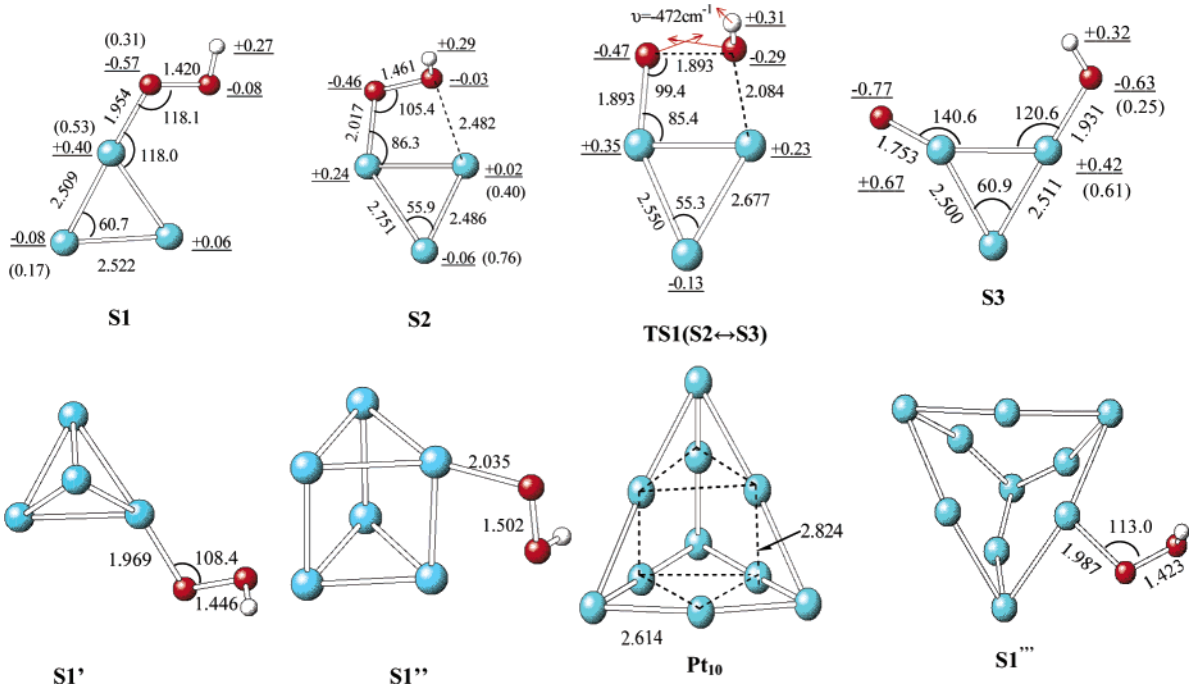


Figure 1. B3LYP optimized structures; bond lengths (in Å) and angles (°) are displayed. The underlined data are Mulliken charges, and those in parentheses are spin densities.

Table 2. Spin Multiplicity (*M*) of Complexes, Distances (R/Å) between Adsorbed Oxygen and Pt, and Adsorption Energies (ΔE/eV) of OOH and H_2O_2 on Pt Clusters Calculated with B3LYP

	Pt	Pt_2	Pt_3	Pt_4	Pt_6	Pt_{10}
			OOH			
M	2	2	2	2	2	8
R	Pt−O=1.882	Pt−O=1.941	Pt−O=1.954	Pt−O=1.969	Pt−O=2.035	Pt−O=1.987
	O−O=1.443	O−O=1.483	O−O=1.420	O−O=1.446	O−O=1.502	O−O=1.423
ΔE	−1.63	−1.42	−1.57	−1.67	−1.92	−1.18
			H_2O_2			
M	1	3	1	3	3	9
R	Pt−O=2.011	Pt1−Pt2=2.381	Pt1−Pt2=2.521	Pt1−Pt2=2.665	Pt1−Pt2=2.675	Pt1−Pt2=2.683
	O−O=1.523	Pt2−O=2.269	Pt1−O=2.173	Pt1−O=2.280	Pt1−O=2.234	Pt1−O=2.436
		O−O=1.460	O−O=1.495	O−O=1.460	O−O=1.465	O−O=1.455
				H−Pt2=2.840	H−Pt2=2.632	H−Pt2=2.575
ΔE	−0.43	−0.58	−0.67	−0.52	−0.44	−0.36

at 2.776 Å is a high spin state (multiplicity, M=7).[17] 1-fold adsorption of OOH on Pt_6 (M=6) was located yet with a lower adsorption energy of −0.86 eV. Two Pt_6 clusters, a singlet distorted octahedron and a triplet prism, were optimized, the latter having lower energy than the former by 1.02 eV. The adsorption energy of OOH on the Pt_6 prism is approximately −1.92 eV (**S1″**). The fully optimized Pt_{10} cluster shows that Pt−Pt bond lengths are different, 2.614 Å between a corner atom and an edge atom and 2.823 Å between two edge atoms. The ground state of Pt_{10} possesses high multiplicity of M = 9, and the electronic configuration for T_d-constrained Pt_{10} is $t_1^2e^2a_1t_2^3$, leading to no Jahn−Teller effect. Thus, the Pt_{10} cluster maintains T_d symmetry. The adsorption species (**S1‴**, M=8 for the ground state; adsorption energy ∼ −1.18 eV) of OOH on Pt_{10}, corresponding to **S1** of OOH on Pt_3, has also been located.

The dissociation of the superoxide radical can be catalyzed by a Pt cluster. A transition state (**TS1**) for the dissociation of **S2** has been characterized by the vibrational mode of its unique imaginary frequency (∼473 cm^{-1}) as well as by IRC analysis, as illustrated in Figure 2 that displays the electronic structure energy (relative to the reactant) against the O−O distance. The activation energy for the dissociation of **S2** into coadsorbed O and OH on Pt_3 (**S3**) is ∼0.25 eV. The dissociation of **S2** is most probably driven by the adsorption of the radical OH on Pt_3, causing a significant energy decrease of ∼1.5 eV from **S2** to **S3**. Sidik et al. obtained an activation energy of 0.06 eV for •OOH dissociation on Pt_2 with the Pt−Pt distance fixed at 2.77 Å and estimated that the activation energy increased to 0.15 eV using the same transition state structure as Pt_2−OOH over a two-layer Pt_{15}.[6] No other information on the dissociation of the peroxide radical has been reported. O and OH are quite separated at two top sites in **S3**, and the ground state is quartet (M=4) with an adsorption energy of −5.62 eV, which is rather similar to the sum of that of O (top with triplet ground state, −3.06 eV) and of OH (top with doublet ground state, −2.62 eV).

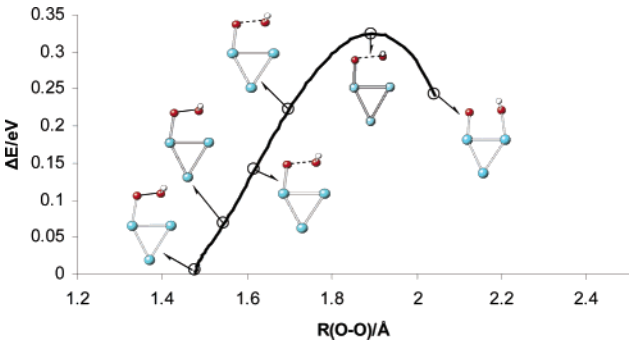


Figure 2. Relative electronic energy (in eV) against O−O distance (in Å) obtained with IRC analysis for the transition state TS1.

The preexponential factor A for the decomposition of adsorbed OOH, estimated by conventional transition-state theory, is 1.1×10^{11} s^{-1}.

3.3. H_2O_2 Adsorption and Dissociation on Pt_3. Hydrogen peroxide, H_2O_2, is usually proposed as either a product or an intermediate of the oxygen reduction reaction on Pt surfaces.[11,12] A 1-fold adsorption conformer through oxygen is located for H_2O_2 on Pt_3 (**S4,** Figure 3), but a 2-fold adsorption, i.e., bridge adsorption, was not found. Starting with a geometry close to bridge adsorption, the geometry optimization always yield two coadsorbed hydroxyls. The binding energy of H_2O_2 to Pt_3 is only ∼ −0.67 eV, which is considerably less than those of the radicals, O (−3.1), OH (−2.6), and OOH (−1.6), and the Pt−O distances of 2.173 Å are greater than those in such complexes as Pt_3−O (∼1.787 Å, top adsorption), Pt_3−OH (1.943 Å), and Pt_3−OOH (1.954 Å). Despite the weak adsorption of **S4**, as compared with naked H_2O_2, the O−O bond length is still stretched by 0.04 Å (1.495 vs 1.454 Å) with an O−O vibrational frequency shift of 273 cm^{-1} (660 vs 933 cm^{-1}). Despite the lower adsorption energies (−0.52, −0.50, and −0.36 eV), such weak adsorbed species of H_2O_2 (**S4′**, **S4″**, and **S4‴**) have also been located on Pt_4, Pt_6, and Pt_{10}, as shown in Figure 3. Panchenko et al. tried to locate the adsorption species of H_2O_2 on Pt(111), (100), and (110) surfaces, but they claimed that H_2O_2 was unstable and dissociated into two adsorbed hydroxyls.[14] Adsorbed H_2O_2 on Pt-polycrystalline has been frequently suggested as the intermediate of the ORR by experimentalists;[25] however, direct measurement or other theoretical calculations on large clusters or extended Pt surfaces have not been reported. According to Table 2, adsorption energies of H_2O_2 on the Pt_n (n=1−10) clusters

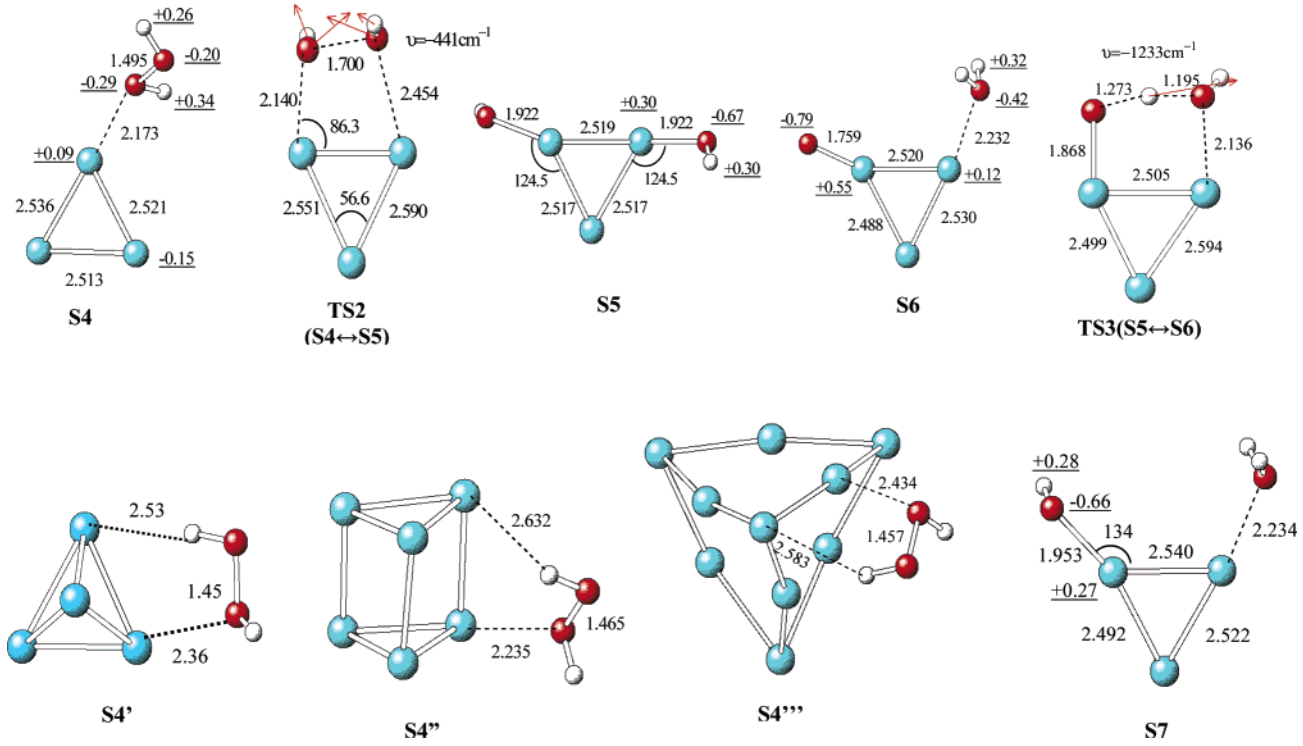


Figure 3. B3LYP optimized structures; bond lengths (in Å) and angles (°) are displayed. The underlined data are Mulliken charges.

are in the range from −0.36 to −0.67 eV. Especially, adsorption strengths of H_2O_2 on Pt_6 and Pt_{10} are very similar. Thus, it is hard to believe that the adsorption of H_2O_2 will cease as the size of the Pt cluster grows to an extended surface. However, the adsorption of H_2O_2 on Pt is still an opening question that needs to be further investigated.

A transition state (**TS2**) for the dissociation of the one-end adsorbed H_2O_2 (**S4**) into two coadsorbed hydroxyls is illustrated in Figure 3, and the activation energy is ∼0.57 eV. The O−O bond length in **TS2** is stretched to 1.7 Å. IRC analysis confirms that **TS2** connects **S4** and two coadsorbed hydroxyls on Pt_3, i.e., **S5**. **S5** is much more stable than **S4** by 2.73 eV, suggesting that the dissociation of **S4** is primarily driven by the strong adsorption of the hydroxyl radical. The preexponential factor *A* for the decomposition of **S4** estimated by transition-state theory is 1.7×10^{11} s^{-1}. The ground state of **S5** is a triplet, and its adsorption energy (−5.21 eV) is almost twice of that of OH to Pt_3, indicating that adsorption of two coadsorbed hydroxyls have a negligible cooperative effect.

Since H_2O_2 may dissociate on a Pt surface into water and oxygen,[26] and •OOH for the first reduction step of O_2 may be reduced to O and H_2O,[5] the coadsorption of O and H_2O on Pt_3 is also investigated, designated as **S6** in Figure 3. The ground state of **S6** has the same spin as the adsorbed O, triplet state, and its adsorption energy of −3.86 eV (taking O, H_2O, and triplet Pt_3 as references) is rather close to the sum of O (−3.06 eV for top adsorption) and H_2O (−0.64 eV). The thermodynamic stability of **S6** lies between **S4** (H_2O_2−Pt_3) and **S5** (2HO−Pt_3), i.e., the energy of **S6** is lower by 1.90 eV than **S4** and higher by 0.83 eV than **S5**. The proton transfer between **S5** and **S6** could be carried out via the transition state (**TS3**), in which the proton is shared by the oxygen and the hydroxyl oxygen. The activation energy from **S6** to **S5** is only ∼0.15 eV, and that of the reverse H transfer is 0.68 eV. The relatively lower energy barrier (∼0.15 eV, ∼3−4 kcal/mol) of the proton transfer from **S5** to **S6** indicates that the proton transfer can take place at room temperature.

3.4. Coadsorptions of OH/H_2O and H_2O/H_2O on Pt_3. As the possible products from the third and forth reduction steps of oxygen, coadsorbed OH/H_2O (**S7**) and H_2O/H_2O (**S8**, not shown) on Pt_3 are also optimized. Pt−O distances in **S7** are very similar to their individual complexes of OH and H_2O (1.953 vs 1.943 Å; 2.234 vs 2.254 Å), and the adsorption energy of −3.32 eV is consequently almost the sum of OH (∼2.62 eV) and H_2O (∼0.64) on Pt_3. Likewise, negligible cooperativity was found for H_2O/H_2O coadsorption, in which an adsorption energy of −1.27 eV is nearly twice that of a single H_2O on Pt_3.

3.5. Potential Energy Surface Profile of the ORR on Pt_3 Cluster. Starting with the adsorption of O_2 to Pt_3, and assuming that the adsorbed oxygen species reacts directly with a proton from the electrolyte, the potential energy surface profile for ORR has been illustrated in Figure 4 on the basis of the energies of the relevant species and available activation energies for a few key reactions. For the benefit of readers who are more familiar with electrochemical discussions of this topic, the energetics of the ORR is also given in Figure 5 referring to the standard hydrogen electrode potential as presented by Anderson and Albu.[29] We note that although the reaction mechanism may quantitatively depend on a number of factors, including electrode potential, electrolyte, and adsorbate coverage, the main features of the ORR may be qualitatively provided by cluster-based DFT calculations.

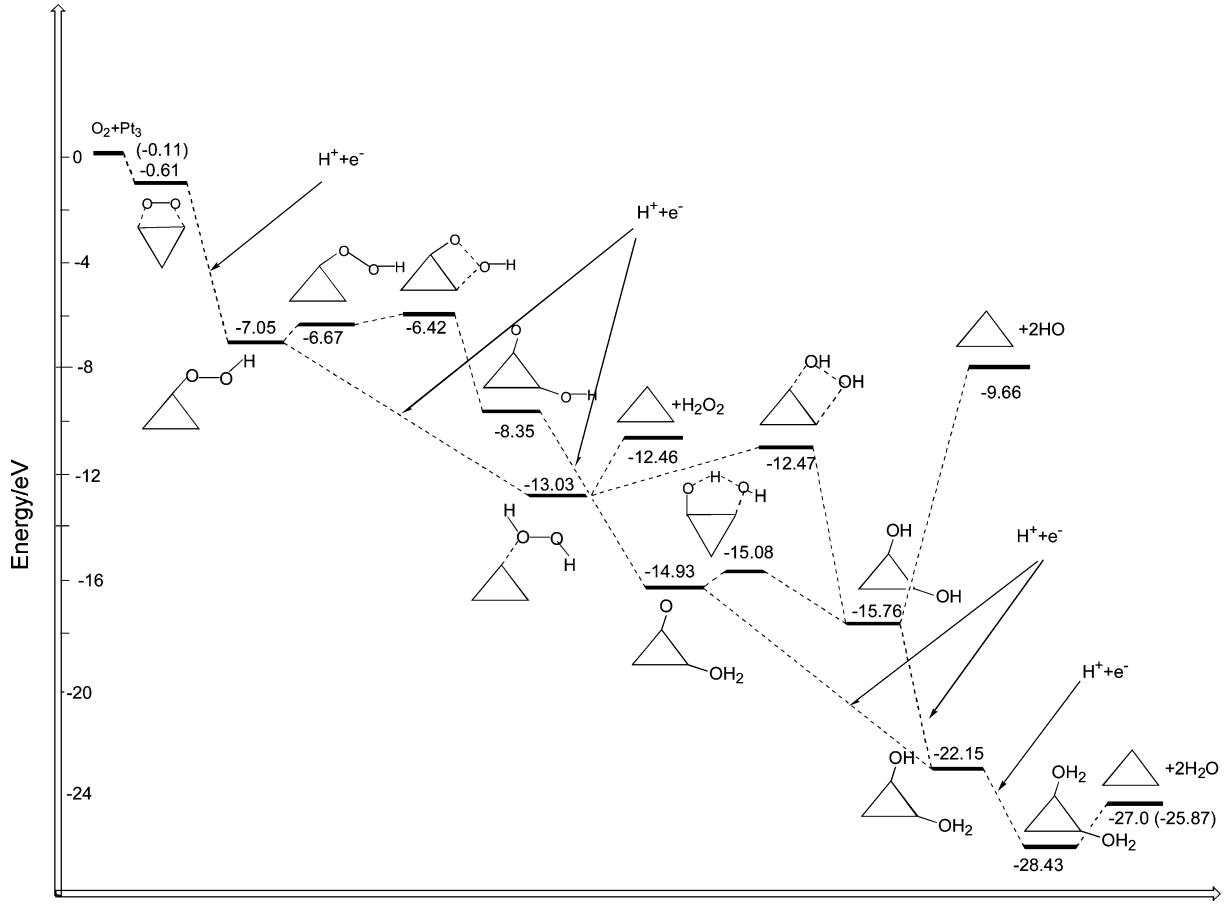
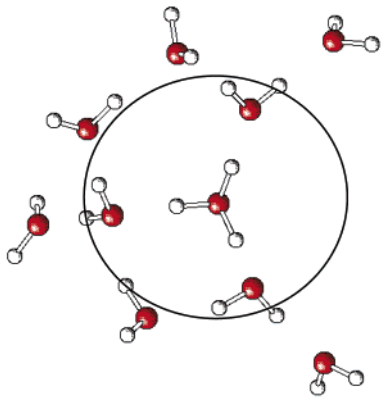


Figure 4. Potential energy surface profile for the oxygen reduction reaction: the proton was modeled by two shells of water molecules, $H^+OH_2(H_2O)_3$ $(H_2O)_6$, and the data in parentheses are Gibbs free energies.

Chart 1



Here the proton was modeled with a hydronium solvated by three water molecules, $H^+OH_2(H_2O)_3$ (model 1), and a hydronium solvated by two shells of water molecules, $H^+OH_2(H_2O)_3$ $(H_2O)_6$ (model 2, Chart 1), respectively. As shown in Table 3, the Gibbs energy variations for the $O_2 + 4H^+ + 4e^- = 2H_2O$ predicted by the two proton models are −29.8 and −25.9 eV. The former is too negative, while the latter gives a reversible potential of 1.87 V at the standard hydrogen scale. To get accurate values for the reversible potential of the ORR, further considerations such as bulk solvent and counterion effects are required. Table 3 shows that calculations with the MP2/6-311++G** method only slightly changes ΔG, as compared with B3LYP/6-311++G**. However, the effect of the basis set used at MP2/6-31G** is significant, e.g., increasing ΔG of the ORR by 2.5 eV for model 1, and MP2/6-31G** occasionally gives better predictions for the reversible potential of the ORR. Using such an increase to scale the ΔG at B3LYP/6-311++G** for model 2

$$\Delta G\,(\text{MP2/6-31G**, model 2}) \approx \Delta G(\text{B3LYP/6-311++G**, model 2}) + \Delta G(\text{MP2/6-31G**, model 1}) - \Delta G(\text{B3LYP/6-311++G**, model 1})$$

$$= -25.87 + (-27.3 + 29.83)\ \text{eV}$$

$$= -23.4\ \text{eV} \qquad (1)$$

The reversible potential of the ORR at the standard hydrogen scale yielded by eq 1 is 1.24 V, which is rather close to the experimental value (1.23 V) because of error cancellations hidden behind the model chemistry used.

Our previous CPMD simulations show that O_2 adsorbed on a Pt surface does not dissociate before the proton transfer, i.e., the formation of adsorbed OOH,[10,27] indicating that dissociation of adsorbed O_2 may have a higher activation energy than the formation of adsorbed OOH. This supports the speculation of Anderson et al.[6] Thus, it is assumed that adsorbed O_2 is reduced into adsorbed OOH (**S1**) in the first reduction step. Figure 4 shows that **S1** may isomerize into **S2,** and the latter decomposes into atomic oxygen and hydroxyl radical on Pt_3 (**S3**) with an activation barrier of 0.25 eV via a transition state (**TS1**). In the second reduction

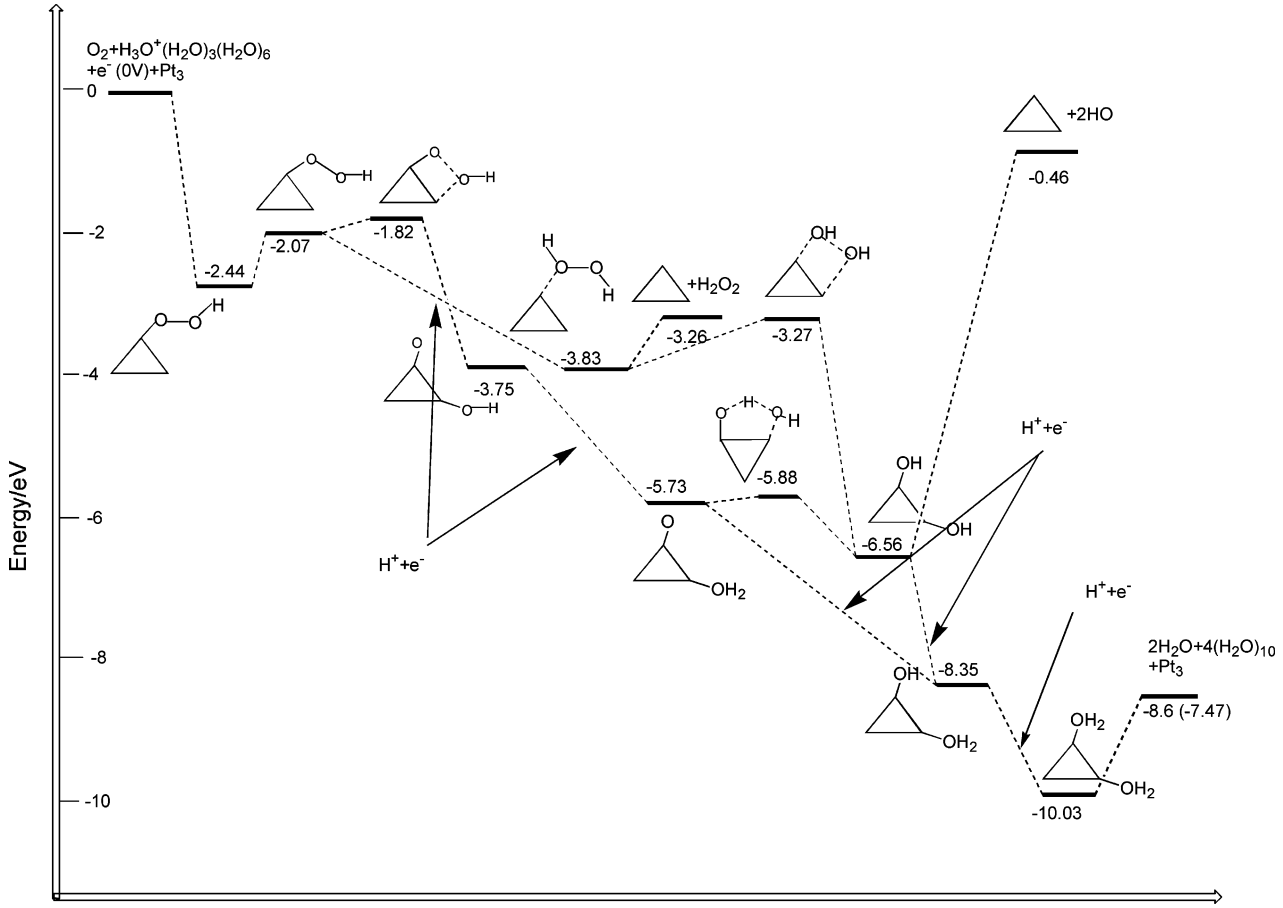


Figure 5. Potential energy surface profile for the oxygen reduction reaction at the standard hydrogen electrode potential scale: the proton was modeled by two shells of water molecules, $H^+OH_2(H_2O)_3$ $(H_2O)_6$, and the data in parentheses are Gibbs free energies.

Table 3. Reaction Energies without and with Zero Point Energy Correction (ΔE and ΔE_0, eV), Enthalpy (ΔH, eV), and Gibbs Free Energy (ΔG, eV) at 298.2 K Calculated with B3LYP

	$H_3O^+(H_2O)_3$			$H_3O^+(H_2O)_3(H_2O)_6$		
reactions	ΔE	ΔE_0	ΔG	ΔE	ΔE_0	ΔG
$O_2 + Pt_3 \rightarrow Pt_3{-}O_2$	−0.61	−0.56	−0.11	−0.61	−0.56	−0.11
$Pt_3{-}O_2 + H^+(aq) + e^- \rightarrow Pt_3{-}OOH$	−7.05	−7.03	−7.05	−6.44	−6.28	−6.06
$Pt_3{-}OOH + H^+(aq) + e^- \rightarrow Pt_3{-}O_2H_2$	−6.59	−6.57	−6.55	−5.98	−5.82	−5.56
$Pt_3{-}O_2H_2 + H^+(aq) + e^- \rightarrow HO{-}Pt_3{-}H_2O$	−9.73	−9.74	−9.74	−9.12	−8.99	−8.75
$HO{-}Pt_3{-}H_2O + H^+(aq) + e^- \rightarrow Pt_3{-}2H_2O$	−6.89	−6.86	−6.84	−6.28	−6.11	−5.85
$Pt_3{-}2H_2O \rightarrow Pt_3 + 2H_2O$	+1.46	+1.27	+0.46	+1.46	+1.27	+0.46
$O_2{+}4H^+(aq) + 4e^- \rightarrow 2H_2O$	−29.4	−29.5	−29.83	−27.0	−26.49	−25.87
$O_2{+}4H^+(aq) + 4e^- \rightarrow 2H_2O$[a]	−28.5	−28.9	−29.32			
$O_2{+}4H^+(aq) + 4e^- \rightarrow 2H_2O$[b]	−27.4	−27.3	−27.3	−25.0		−23.37[c]

[a] MP2/6-311++G**. [b] MP2/6-31G**. [c] −25.87 +(ΔG, MP2/6-31G** - ΔG, B3LYP/6-311++G**) for the model of $H_3O^+(H_2O)_3$.

step **S3** gets reduced to form coadsorbed atomic oxygen and water (**S6**). This pathway is usually referred to as a *direct* one, implying the absence of H_2O_2 in the reaction pathway. Another possibility of the second reduction step is that **S1** and/or **S2** is directly reduced by one electron to yield 1-fold adsorbed hydrogen peroxide H_2O_2 (**S4**), which can readily decompose into two coadsorbed hydroxyls (**S5**) with an activation energy barrier of 0.57 eV via the transition state (**TS2**). The pathway giving rise to H_2O_2 is usually called a *series*.

Location of one-end adsorbed hydrogen peroxide H_2O_2 (**S4**), arising from reduction of adsorbed OOH, suggests that the ORR can proceed via a *series* pathway. On the other hand, the relatively low energy barrier for the decomposition of **S2** and the high stability of **S3** in comparison with **S1** show that the end-on chemisorption precursor (**S2**) has a strong tendency to decompose before the second electron transfer. In addition, the energy of S4, ∼1.9 eV higher than S6, thermodynamically indicates that the *direct* pathway is more favorable. Therefore, although the one-end adsorbed H_2O_2 has been found, the second electron transfer step of the ORR should dominantly start with the coadsorbed O and OH (**S3**). Thus, it is theoretically validated that the O_2 reduction on a Pt(111) surface may proceed via a *parallel* pathway with the *direct* as the dominant step.

A proton transfer can occur on **S6** with an activation energy of 0.15 eV via the transition state **TS3**. Before the proton-transfer reaction, further reduction of S6 could not be excluded, which is dependent on its lifetime as well as on the competition between the proton transfer and reduction. However, intramolecular proton transfer may be more favorable than the reduction reaction involving one electron and one proton transfer.

Of the three species (**S4**, **S5**, and **S6)** from the second reduction step, **S5** (coadsorbed hydroxyls on Pt_3) is the most stable thermodynamically, followed by **S6** (coadsorbed atomic oxygen and H_2O), and **S4** (atop adsorbed H_2O_2) the least stable. H_2O_2 has quite a low adsorption energy of -0.63 eV on Pt_3, and -0.4 eV on Pt_{10}, which may enable H_2O_2 to desorb into the electrolyte phase. HO radicals are strongly adsorbed on Pt_3 with an adsorption energy of -2.61 eV per OH in **S5**. It may be speculated that OH can block active sites of Pt surface, unless OH is further reduced. Adsorbed O in S3 and S6 has the same problem due to even stronger adsorption. However, its higher reduction potential may first induce its reduction into OH.

4. Conclusions

Adsorption and decomposition of the radical OOH and H_2O_2 on Pt clusters (Pt_3, Pt_4, Pt_6, and Pt_{10}) have been extensively investigated using B3LYP type of density functional theory. Adsorption and coadsorption on Pt_3 of other relevant oxide products of the ORR have been investigated, and a potential energy surface profile for the complete reduction of oxygen in acid media has been proposed. The OOH radical has a very strong adsorption on Pt clusters. Both thermodynamics and kinetics indicate that OOH readily decomposes into O and OH before a second electron transfer takes place. Thus, although a one-end adsorbed hydrogen peroxide H_2O_2, arising from the reduction of adsorbed OOH, has also been located on Pt_3 and Pt_{10}, a favorable pathway for the second electron transfer should occur on the coadsorbed O and OH species, suggesting that O_2 reduction on a Pt surface may proceed via a *parallel* pathway: the *direct* (no H_2O_2 generated as intermediate) and *series* (H_2O_2 generated) occurring simultaneously, with the *direct* as the dominant step.

Acknowledgment. Financial support from the Department of Energy/Basic Energy Sciences, (DE-FG02-04ER15619) and from DURIP/ARO (Grant No. W911N F-04-1-0098) is gratefully acknowledged. The use of computational facilities at the NSF HPC and TeraGrid Sites: National Center for Supercomputing Applications (CHE040052) is appreciated.

References

(1) Damjanovic, A.; Brusic, V. Electrode kinetics of oxygen reduction on oxide-free platinum electrodes. *Electrochim. Acta* **1967**, *12*, 615−628.

(2) Yeager, E. B. Electrocatalysts for O_2 reduction. *Electrochim. Acta* **1984**, *29*, 1527−1537.

(3) Adzic, R. R.; Wang, J. X. Configuration and site of O_2 adsorption on the Pt(111) electrode surface. *J. Phys. Chem. B* **1998**, *102*, 8988−8993.

(4) Markovic, N. M.; Ross, P. N. Surface science studies of model fuel cell electrocatalysts. *Surf. Sci. Rep.* **2002**, *45*, 117−229.

(5) Anderson, A. B.; Albu, T. V. Catalytic effect of platinum on oxygen reduction: An ab initio model including electrode potential dependence. *J. Electrochem. Soc.* **2000**, *147*, 4229−4238.

(6) Sidik, R. A.; Anderson, A. B. Density functional theory study of O_2 electroduction when bonded to a Pt dual site. *J. Electroanal. Chem.* **2002**, *528*, 69−76.

(7) Murthi, V. S.; Urian, R. C.; Mukerjee, S. Oxygen reduction kinetics in low and medium-temperature acid environment: Correlation of water activation and surface properties in supported Pt and Pt alloy electrocatalysts. *J. Phys. Chem. B* **2004**, *108*, 11011−11023.

(8) Steele, B. C. H.; Heinzel, A. Materials for fuel-cell technologies. *Nature* **2001**, *414*, 345−352.

(9) Markovic, N. M.; Schmidt, T. J.; Stamenkovic, V.; Ross, P. N. Oxygen reduction reaction on Pt and Pt bimetallic surfaces: A selective review. *Fuel Cells* **2001**, *1*, 105−116.

(10) Wang, Y.; Balbuena, P. B. Ab initio-molecular dynamics studies of O_2 electroreduction on Pt (111): Effects of proton and electric field. *J. Phys. Chem. B* **2004**, *108*, 4376−4384.

(11) Stamenkovic, V.; Schmidt, T. J.; Ross, P. N.; Markovic, N. M. Surface composition effects in electrocatalysis: Kinetics of oxygen reduction on well-defined Pt_3Ni and Pt_3Co alloy surfaces. *J. Phys. Chem. B* **2002**, *106*, 11970−11979.

(12) Nakanishi, S.; Mukouyama, Y.; Karasumi, K.; Imanishi, A.; Furuya, N.; Nakato, Y. Appearance of oscillation through the autocatalytic mechanism by control of the atomic-level structure of electrode surfaces in electrochemical H_2O_2 reduction at Pt electrodes. *J. Phys. Chem. B* **2000**, *104*, 4181−4188.

(13) Balbuena, P. B.; Altomare, D.; Vadlamani, N.; Bingi, S.; Agapito, L. A.; Seminario, J. M. Adsorption of O, OH, and H_2O on Pt-based bimetallic clusters alloyed with Co, Cr, and Ni. *J. Phys. Chem. A* **2004**, *108*, 6378−6384.

(14) Panchenko, A.; Koper, M. T. M.; Shubina, T. E.; Mitchell, S. J.; Roduner, E. Ab initio calculations of intermediates of oxygen reduction on low-index platinum surfaces. *J. Electrochem. Soc.* **2004**, *151*, A2016−A2027.

(15) Li, T.; Balbuena, P. B. Computational studies of the interactions of oxygen with platinum clusters. *J. Phys. Chem. B* **2001**, *105*, 9943−9952.

(16) Balbuena, P. B.; Altomare, D.; Agapito, L. A.; Seminario, J. M. Adsorption of oxygen on Pt-based clusters alloyed with Co, Ni, and Cr. *J. Phys. Chem. B* **2003**, *107*, 13671−13680.

(17) Jacob, T.; Muller, R. P.; W. A. Goddard, I. Chemisorption of atomic oxygen on Pt(111) from DFT studies of Pt-clusters *J. Phys. Chem. B* **2003**, *107*, 9465−9476.

(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.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.

(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; 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. Gaussian03; Revision C.02 ed.; Gaussian, Inc.: Wallingford, CT, 2004.

(21) Wadt, W. R.; Hay, P. J. Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *J. Chem. Phys.* **1985**, *82*, 284−298.

(22) Boys, S. F.; Bernardi, F. Calculations of small molecular interactions by differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553.

(23) Majumdar, D.; Dai, D.; Balasubramanian, K. Theoretical study of the electronic states of platinum trimer (Pt_3) *J. Chem. Phys.* **2000**, *113*, 7919−7927.

(24) Michaelides, A.; Hu, P. A density functional theory study of hydroxyl and the intermediate in the water formation reaction on Pt. *J. Chem. Phys.* **2001**, *114*, 513−519.

(25) Stamenkovic, V.; Grgur, B. N.; Ross, P. N.; Markovic, N. M. Oxygen reduction reaction on Pt and Pt-bimetallic electrodes covered by CO − Mechanism of the air bleed effect with reformate. *J. Electrochem. Soc.* **2005**, *152*, A277−A282.

(26) Kim, D.-H.; Lee, B.-K.; Lee, D. S. Determination of trace anions in concentrated hydrogen peroxide by direct injection ion chromatography with conductivity detection after Pt-catalyzed on-line decomposition. *Bull. Korean Chem. Soc.* **1999**, *20*, 696−700.

(27) Wang, Y.; Balbuena, P. B. Ab initio molecular dynamics simulations of the oxygen electroreduction reaction on a Pt(111) surface in the presence of hydrated hydronium $(H_3O)+(H_2O)_2$: Direct or series pathway? *J. Phys. Chem. B*, in press.

(28) Gland, J. L. Molecular and atomic adsorption of oxygen on the Pt (111) and Pt (S)-12 (111) x (111) surfaces. *Surf. Sci.* **1980**, *93*, 487−514.

(29) Anderson, A. B.; Albu, T. V. Ab Initio Determination of Reversible Potentials and Activation Energies for Outer-Sphere Oxygen Reduction to Water and the Reverse Oxidation Reaction. *J. Am. Chem. Soc.* **1999**, *121*, 11855−11863.

CT0500794

JCTC Journal of Chemical Theory and Computation

Revisiting Hydrogen [1,5] Shifts in Cyclopentadiene and Cycloheptatriene as Bimolecular Reactions

Shinichi Yamabe,* Noriko Tsuchida, and Shoko Yamazaki

Department of Chemistry, Nara University of Education, Takabatake-cho, Nara 630-8528, Japan

Received March 11, 2005

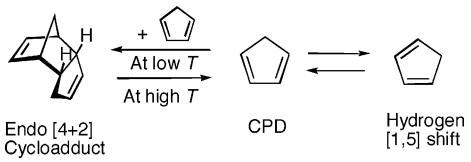
Abstract: Hydrogen [1,5] shifts are pericyclic reactions and take place typically in 1,3-pentadiene. However, because of structure restriction, the symmetry-allowed thermal reactions of 1,3-cyclopentadiene (CPD) and 1,3,5-cycloheptatriene (CHT) suffer large energy barriers, $\Delta U^{\ddagger} = +26.9$ kcal/mol and $\Delta U^{\ddagger} = +37.5$ kcal/mol by density-functional theory (B3LYP/6-31G*) calculations, respectively. This theoretical study has shown that exo [4+2] and [6+4] cycloadduct dimers involve novel hydrogen-shift channels. After hydrogen migration, one-center adducts are obtained, which undergo Cope rearrangements leading to the second one-center adducts. From the intermediates, reverse routes lead to CPD and CHT with [1,5] migrated hydrogens. A correlation between cycloadditions and [1,5] and [3,3] sigmatropic rearrangements in pericyclic reactions is also discussed.

Introduction

Pericyclic reactions were defined by Woodward and Hoffmann as reactions in which all first-order changes in bonding relationships take place concertedly.[1] Representative pericyclic reactions are cycloadditions, [1,5] and [3,3] sigmatropic shifts, cheletropic additions and eliminations, and electrocyclizations. These reactions involve π-electron conjugated systems consisting of mainly carbon and hydrogen atoms, and they are fundamental in organic reactions. The pathways of pericyclic reactions have been explained in terms of the frontier orbital theory.[2]

The Diels−Alder (DA) reaction ([4+2] cycloaddtion) and hydrogen [1,5] shift are "symmetry allowed", and for these reactions, many ab initio calculations have been reported.[3] For the (butadiene−ethylene) DA reaction, predicted transition-state (TS) geometry with the *Cs* symmetry was obtained. A [1,5] sigmatropic shift of 1,3-pentadiene was calculated, and the concerted and suprafacial hydrogen shift was confirmed.[4] Furthermore, extensive studies and calculations of the pericyclic reactions have been performed.[5] Most calculation results for those reactions support the Woodward−Hoffmann rule and the frontier molecular orbital (FMO) theory.

Scheme 1. Dual Reactivity of 1,3-Cyclopentadiene (CPD). *T* Stands for the Temperature



1,3-Cyclopentadiene (CPD) was first used for [4+2] cycloadditions by Diels and Alder.[6] Since then, CPD is the representative diene for DA reactions. Dimerization of CPD gives an endo cycloadduct, dicyclopentadiene.[7] At the same time, CPD undergoes a hydrogen [1,5] sigmatropic shift readily at room temperature (Scheme 1).[8]

Various TS structures for the hydrogen [1,5] shift in CPD have been reported.[9] So far, the two reactions in Scheme 1 have been treated separately.[3,7,10] The hydrogen [1,5] shift in a linear diene 1,3-pentadiene may occur readily[5] owing to the flexible structure of terminal methyl and methylene groups. On the other hand, the methylene C−H bond of CPD cannot be directed to the 4-position methine carbon in the C−H····C alignment owing to ring strain. That is, cleavage of the C−H bond may not accompany the bond interchange in the σ direction.

It appears curious that the hydrogen [1,5] shift of the ring strained structure in CPD has an activation energy (E_a =

* Corresponding author e-mail: yamabes@nara-edu.ac.jp.

10.1021/ct0500646 CCC: $30.25
Published on Web 08/06/2005

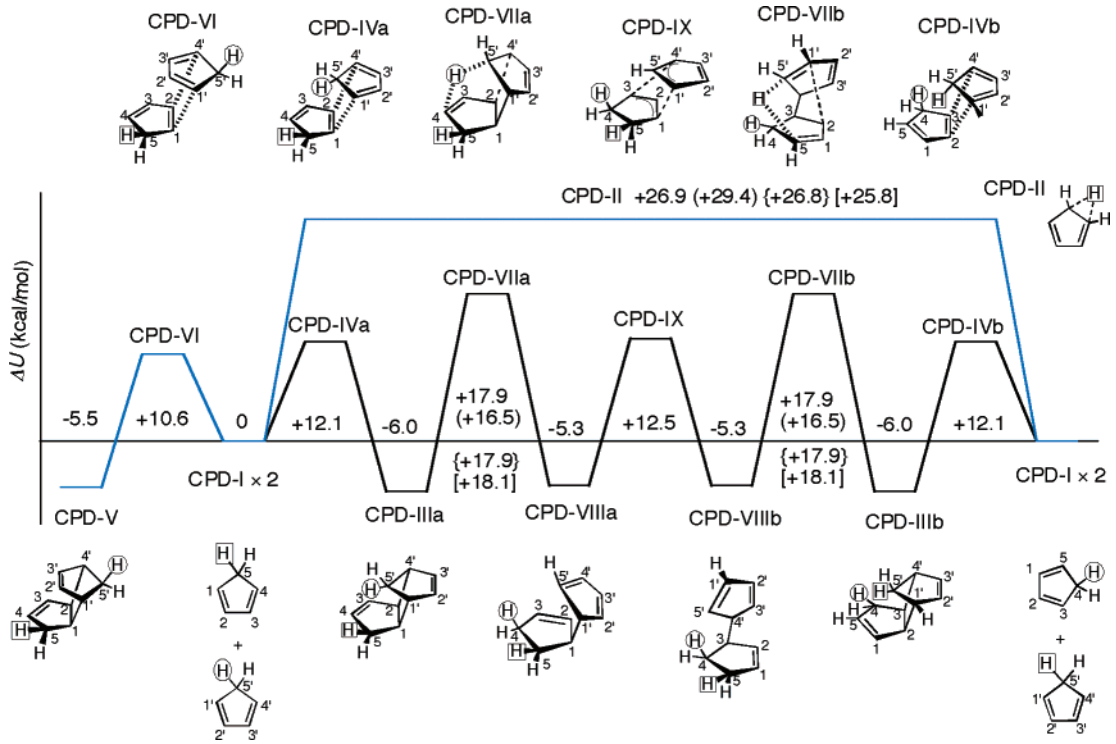


Figure 1. Changes of sums of electronic and zero-point energies (RB3LYP/6-31G*) of [4+2] cycloadditions and Cope and hydrogen [1,5] sigmatropic shifts of CPDs. Energies in parentheses were calculated by QCISD(T)/6-311G(d,p)//RB3LYP/6-31G* [QCISD(T) electronic energies and RB3LYP zero-point vibrational energies]. Those in braces { } were calculated by RB3LYP/6-31G* with the Onsager's reaction field, and those in brackets [ ] were calculated by RB3LYP/6-31G* with polarization functions on the migrating hydrogen atoms.

24.3 kcal/mol)[11] of ca. 10 kcal/mol smaller than that (E_a = 35.4 kcal/mol)[12] of the flexible structure in 1,3-pentadiene. Curiosity lies also in the computational data of activation energies. Jensen and Houk reported 34 kcal/mol[5g] for the hydrogen shift of 1,3-pentadiene, which is in good agreement with experimental data, 35.4 kcal/mol.[12] On the other hand, Okajima and Imafuku reported 28 kcal/mol for the hydrogen shift of CPD,[9d] while the experimental data gives 24.3 kcal/mol.[11] The overestimate of the calculated energy is not explicable, because RB3LYP tends to underestimate the activation energy of some proton- and hydrogen-shift reactions.[13] In this study, bimolecular processes to ensure the bond exchange have been examined for CPD, CHT, and their methyl-substituted reactants, Me-CPD and Me-CHT.

Computational Methods

The density-functional theory (DFT) calculations using the B3LYP hybrid functional[14] and the 6-31G* basis sets (RB3LYP/6-31G*) were carried out by the Gaussian 98 program package[15] in order to investigate the reaction paths of cycloadditions and rearrangements of CPD, CHT, Me-CPD, and Me-CHT. The sums of electronic and zero-point energies of all the species were obtained by vibrational analyses, and they were used as internal energies, U (T = 0 K). To check the above-mentioned underestimation of the activation energy by RB3LYP/6-31G*, QCISD(T)/6-311G-(d,p) single-point calculations were also carried out for key TS geometries.

The CPD and Me-CPD reactions are usually carried out in CCl_4 solvent, while those of CHT and Me-CHT are done in the gas phase (by the temperature over their boiling point). For the CPD and Me-CPD ones, the solvent effect was taken into account for the key species by the Onsager's reaction field approximation[16] with the dielectric constant = 2.228 (CCl_4). The solvent-effect-containing data are shown in braces { }, see Figure 1.

The p-type polarization function on the migrating hydrogen might play an important role in the calculated energies. To check the role, the function was added to the 31G (in 6-31G*) of only the migrating hydrogen. The key species were recalculated by the augmented basis set. The data of this augmented basis set are shown in square brackets [ ], see Figure 1.

Results and Discussion

A Bimolecular Process for the Hydrogen [1,5] Shift in CPD. We have considered that a bimolecular multistep reaction containing cycloaddition and hydrogen shifts can involve a suitable conformation for the bond interchange using a molecular model. These results of the energy changes are shown in Figure 1. The optimized geometries of TS structures CPD-II, CPD-IVa, CPD-VIIa, and CPD-IX are shown in Figures 2−5, respectively. The optimized energy minimum and TS geometries of CPD-I (Figure S1), CPD-IIIa,b (Figures S2 and S8), CPD-IVb (Figure S9), CPD-V (Figure S4), CPD-VI (Figure S3), CPD-VIIb (Figure S7) and CPD-VIIIa,b (Figures S5 and S6) are shown in the Supporting Information. Activation energies $\Delta U^{\ddagger}$ are calculated from the relative internal energies ΔU. As expected, the kinetically most favorable reaction is the endo [4+2] cycloaddition

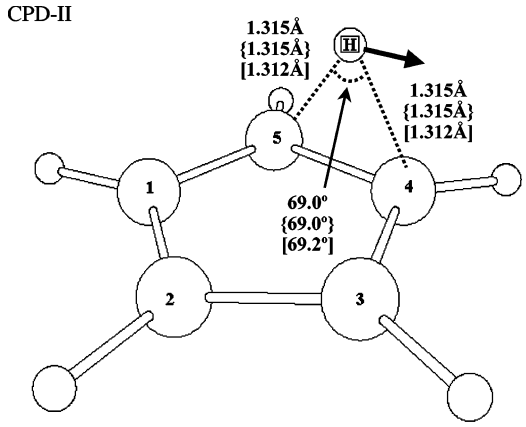


Figure 2. TS of the direct hydrogen [1,5] shift, CPD-II. $\nu^{\ddagger}$ stands for a frequency in vibrational analysis.

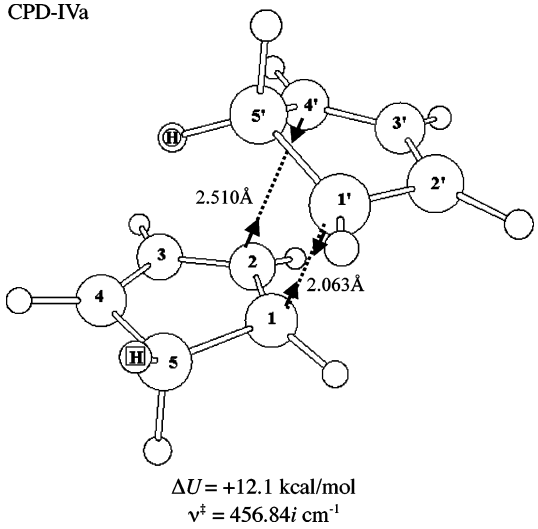
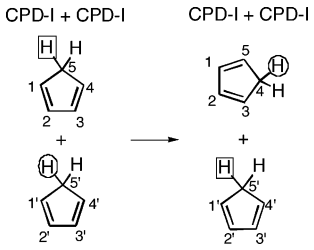


Figure 3. TS of the exo [4+2] cycloaddition, CPD-IVa. The circle and boxed hydrogens are also shown in Figure 1.

Scheme 2. Two CPD Molecules before and after the New Bimolecular Process in Figure 1



(CPD-VI; $\Delta U^{\ddagger} = +10.6$ kcal/mol) of two CPDs. However, the cycloadduct (CPD-V) is less stable than two reactants (CPD-I × 2, CPDs) in Gibbs free energy difference (Figure 6; $\Delta G = +1.4$ kcal/mol). The direct hydrogen shift (CPD-II; $\Delta U^{\ddagger} = +26.9$ kcal/mol) is the most unfavorable, and the $\Delta U^{\ddagger}$ is larger than the experimentally obtained activation energy ($E_a = 24.3$ kcal/mol).[11] The QCISD(T)/6-311G(d,p) $\Delta U^{\ddagger} = 29.4$ kcal/mol is larger than E_a. A route to the exo

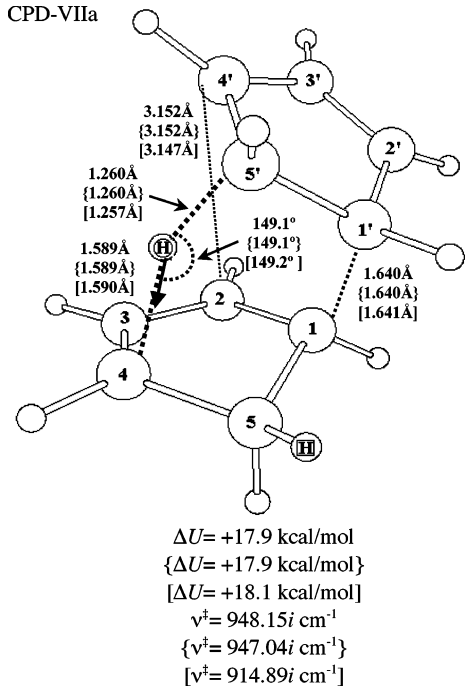


Figure 4. TS of the hydrogen shift from the exo [4+2] adduct (CPD-IIIa) to the one-center adduct intermediate (CPD-VIIIa), CPD-VIIa.

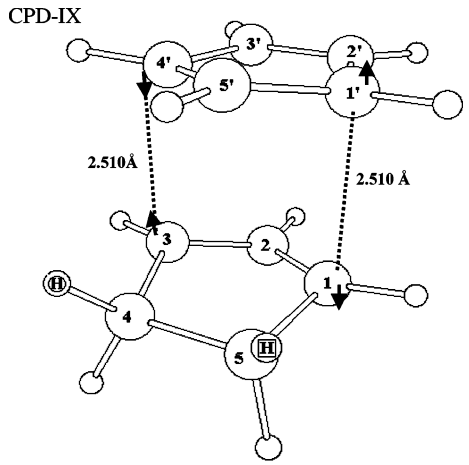
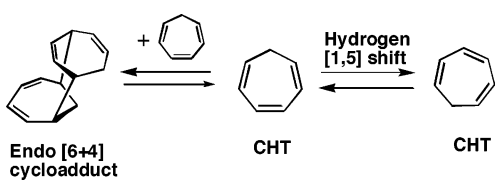


Figure 5. TS of the Cope rearrangement with *Cs* symmetry, CPD-IX.

Scheme 3. Dual Reactivity of 1,3,5-Cycloheptatriene (CHT)



[4+2] cycloadduct was also investigated. As expected, the exo [4+2] addition TS (CPD-IVa, in Figure 3) is 1.5 kcal/mol less stable than the endo [4+2] addition TS (CPD-VI). On the other hand, the exo cycloadduct (CPD-IIIa) is 0.5 kcal/mol more stable than the endo adduct (CPD-V). From CPD-IIIa, the TS of a methylene hydrogen (denoted by a

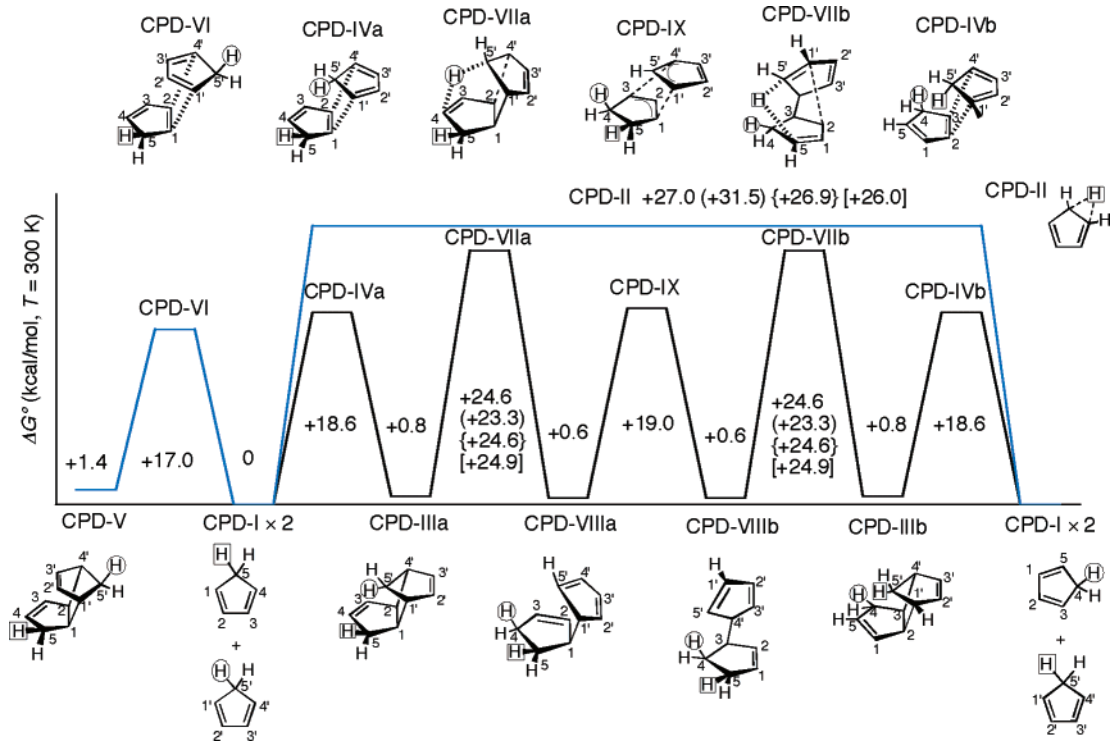


Figure 6. Changes of Gibbs free energies (RB3LYP/6-31G*, $T = 300$ K) of [4+2] cycloadditions and Cope and hydrogen [1,5] sigmatropic shifts of CPDs. Energies in parentheses were calculated by QCISD(T)/6-311G(d,p)//RB3LYP/6-31G* [QCISD(T) electronic energies and RB3LYP thermal corrections and entropies].

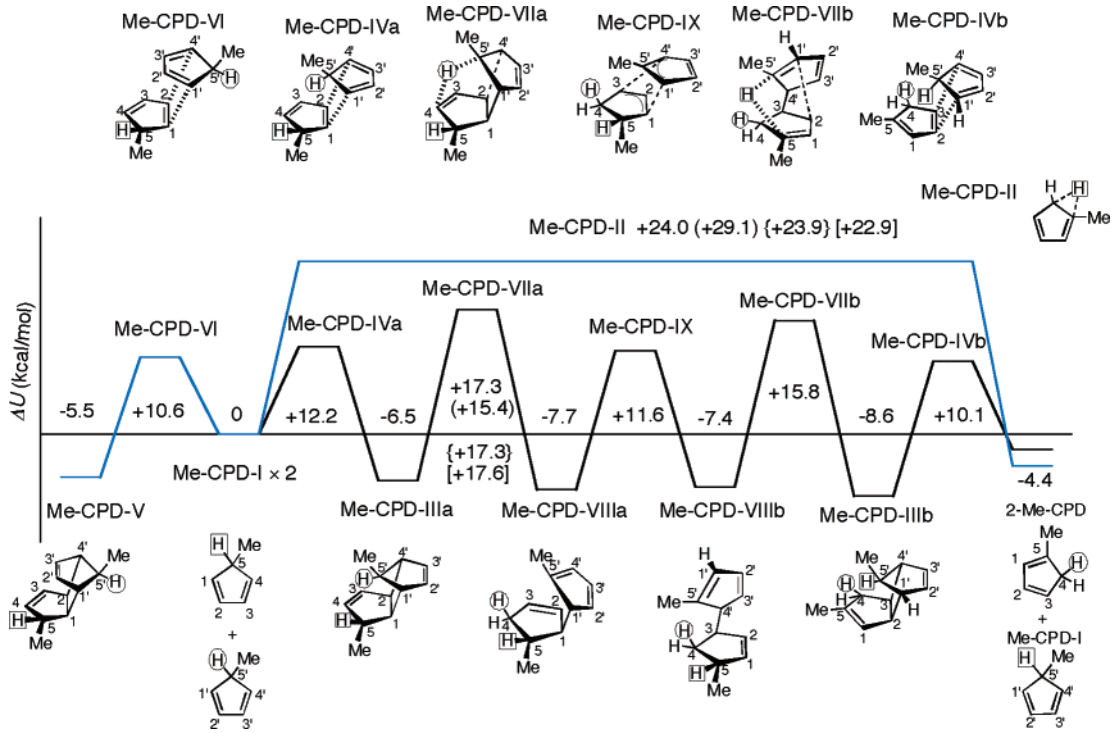


Figure 7. Changes of sums of electronic and zero-point energies (RB3LYP/6-31G*) of [4+2] cycloadditions and Cope and hydrogen [1,5] sigmatropic shifts of Me-CPD. Energies in parentheses were calculated by QCISD(T)/6-311G(d,p)//RB3LYP/6-31G* [QCISD(T) electronic energies and RB3LYP zero-point vibrational energies].

circle) migration (CPD-VIIa) was found and is shown in Figure 4. After the TS (CPD-VIIa), a one-center adduct (CPD-VIIIa) was afforded. From the adduct (CPD-VIIIa), an isomerization process (CPD-IX) was obtained and is shown in Figure 5. The process is a Cope rearrangement with *Cs* symmetry. After the Cope TS (CPD-IX), an isomer (CPD-VIIIb) that is a mirror image isomer relative to CPD-VIIIa because of the symmetry of CPD-I[17] was formed. From the one-center adduct isomer (CPD-VIIIb), a reverse route, CPD-VIIIb → CPD-VIIb → CPD-IIIb → CPD-IVb → CPD-I

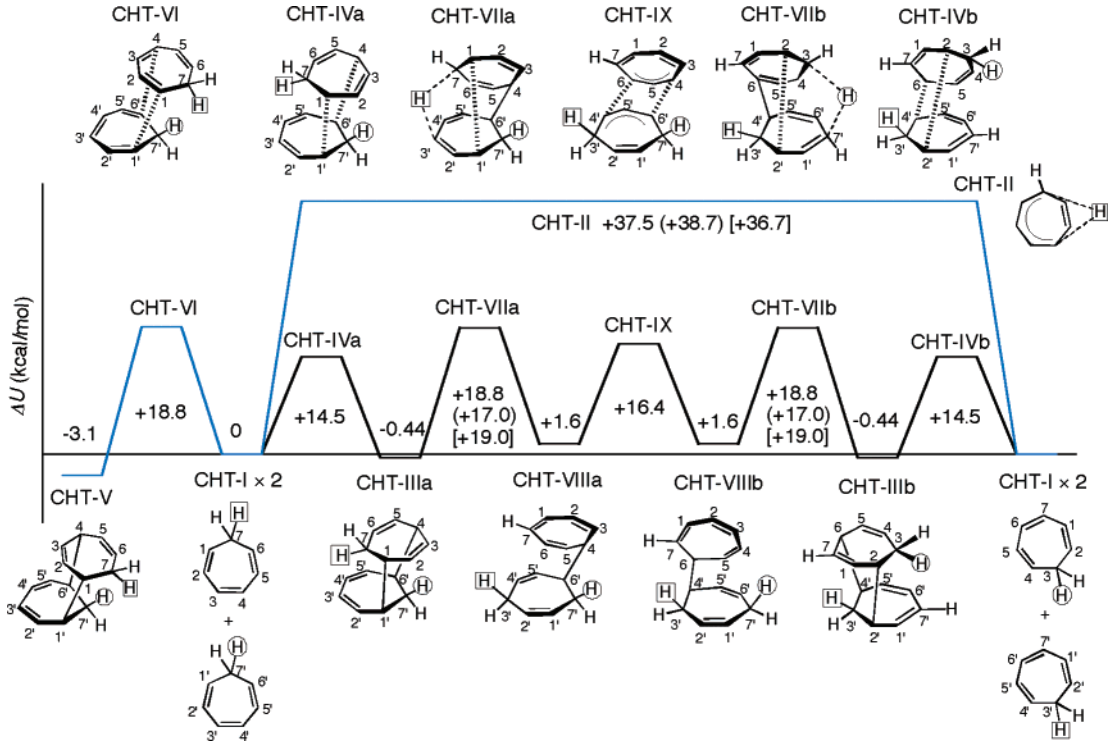


Figure 8. Changes of the sum of electronic and zero-point energies (RB3LYP/6-31G*) of [6+4] cycloadditions and Cope and hydrogen [1,5] sigmatropic shifts of CHTs. Energies in parentheses of CHT-II and CHT-VIIa,b are of QCISD(T)/6-311G(d,p)//RB3LYP/6-31G*.

+ CPD-I, is possible. As a result, a hydrogen, denoted by a circle, is [1,5] shifted (Scheme 2).

In the multistep channel, the rate-determining step is the hydrogen migration TS (CPD-VIIa or CPD-VIIb). The activation energy, $\Delta U^{\ddagger} = +17.9$ (+16.5) kcal/mol, is smaller than that of CPD-II, $\Delta U^{\ddagger} = +26.9$ (+29.4) kcal/mol. The values in parentheses are those of QCISD(T)/6-311G(d,p)//RB3LYP/6-31G*. The energy lowering due to tunnel correction[18] is 0.71 kcal/mol (CPD-II) and 0.55 kcal/mol (CPD-VIIa,b). The value ($\Delta U^{\ddagger} = +17.9$ kcal/mol) is smaller than the experimental value ($E_a = 24.3$ kcal/mol). To compare the reactivities of the unimolecular and bimolecular processes, changes of Gibbs free energies were examined and are shown in Figure 6. Although the unimolecular process is entropically more favorable than the bimolecular process, $\Delta G^{\ddagger}$ (CPD-VIIa) is smaller than $\Delta G^{\ddagger}$ (CPD-II). This result indicates that, despite the entropy loss, the present new bimolecular process is more favorable than the unimolecular process.

Next, to compare the calculation values and the experimental data further, B3LYP/6-31G* calculations of 5-methyl-1,3-cyclopentadiene (Me-CPD) were also carried out. The energy diagram is shown in Figure 7. The corresponding geometries are shown in Figures S19−S31 of the Supporting Information. Similar to CPD, the direct hydrogen shift [Me-CPD-II; $\Delta U^{\ddagger} = +24.0$ (+29.1) kcal/mol] is the most unfavorable, and the $\Delta U^{\ddagger}$ is not in agreement with the experimentally obtained activation energy ($E_a = 19.9$ kcal/mol).[19] A bimolecular route was also examined. The exo [4+2] addition TS (Me-CPD-IVa), the exo cycloadduct (Me-CPD-IIIa), the TS of a hydrogen migration (Me-CPD-VIIa), a one-center adduct (Me-CPD-VIIIa), and a Cope rearrangement TS (Me-CPD-IX) were obtained. After the Cope TS (Me-CPD-IX), a reverse route, Me-CPD-VIIIb → Me-CPD-VIIb → Me-CPD-IIIb → Me-CPD-IVb (which are not mirror image isomers because of the dissymmetry of Me-CPD-I), was obtained. Finally, 2-methyl-1,3-cyclopentadiene + Me-CPD-I were formed. The rate-determining step is the hydrogen migration TS (Me-CPD-VIIa). The activation energy, $\Delta U^{\ddagger} = +17.3$ (+15.4) kcal/mol, is in agreement with the experimentally obtained activation energy ($E_a = 19.9$ kcal/mol),[19] considering the underestimation by RB3LYP.[13] Thus, it has been shown that the hydrogen [1,5] shift of CPD takes place favorably via a bimolecular reaction.

A Bimolecular Process for the Hydrogen [1,5] Shift in CHT. 1,3,5-Cycloheptatriene (CHT) is known to undergo the hydrogen [1,5] shift at a high temperature ($T = 150$ °C).[20] In analogy with the tropone (2,4,6-cycloheptatrien-1-one)−CPD cycloaddition,[21] the dimerization of CHT is considered to be a thermally allowed [6+4] cycloaddition (Scheme 3). The 6π-electron component,[22] that is, the hexatriene moiety, is required for the reaction, because orbital lobes of FMOs (HOMO and LUMO) are largest at the triene terminal sites.

Reaction paths of cycloadditions, hydrogen [1,5] and [3,3] sigmatropic shifts, were traced. The computational energy diagram is shown in Figure 8. The TS of the direct [1,5] shift is most unstable [$\Delta U^{\ddagger} = +37.5$ (+38.7) kcal/mol; CHT-II, Figure 9]. A multistep bimolecular channel was obtained. The channel is composed of TSs of the exo [6+4] cycloaddition (CHT-IVa) in Figure 10, the hydrogen shift (CHT-VIIa) in Figure 11, a Cope shift with *Cs* symmetry (CHT-IX; Figure 12), and their mirror image isomeric reactions. Geometries of the other species, CHT-I, CHT-IIIa, CHT-VI, CHT-V, CHT-VIIIa, CHT-VIIIb, CHT-VIIb, CHT-IIIb, and CHT-IVb are shown in the Figures S10−S18 of the

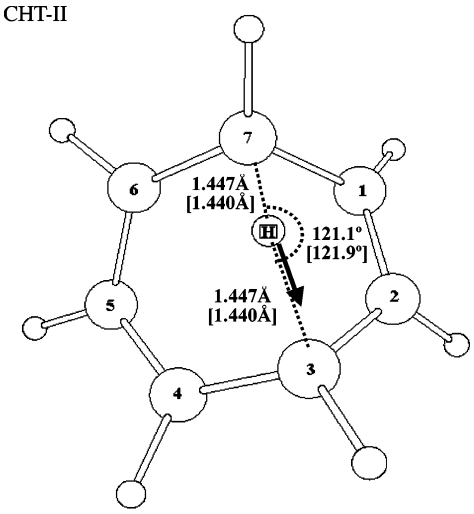


Figure 9. TS of the direct hydrogen [1,5] shift, CHT-II.

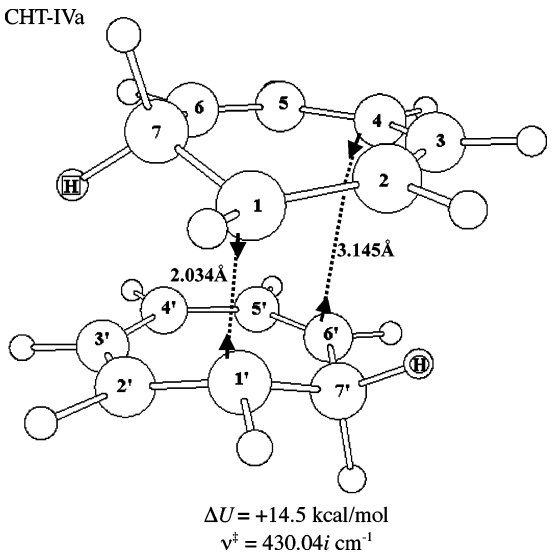


Figure 10. TS of the exo [6+4] cycloaddition, CHT-IVa. The circle and boxed hydrogens are also shown in Figure 8.

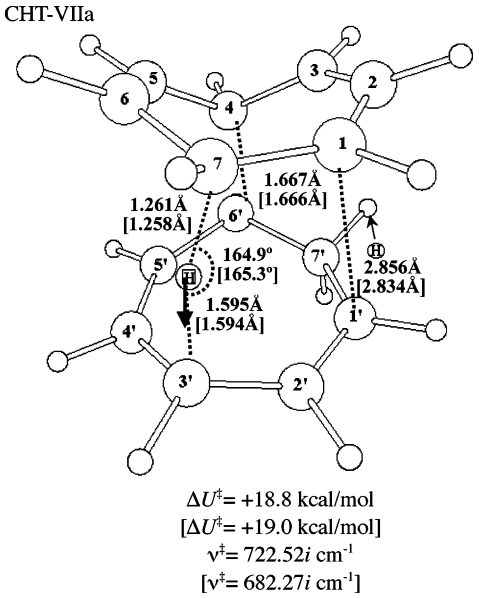


Figure 11. TS of the hydrogen shift from the exo [6+4] adduct (CHT-IIIa) to the one-center adduct intermediate (CHT-VIIIa), CHT-VIIa.

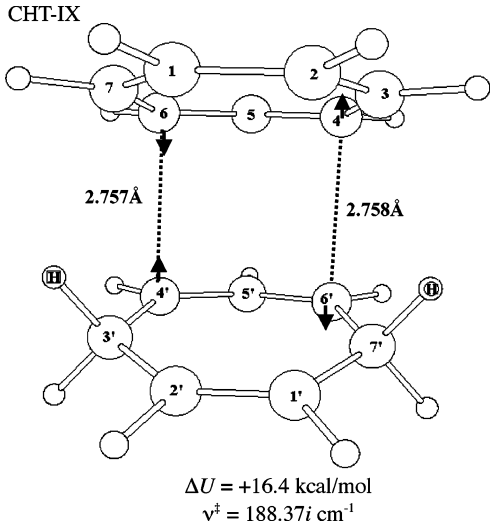


Figure 12. TS of the Cope rearrangement with *Cs* symmetry, CHT-IX.

Supporting Information. The channel leads to the double hydrogen [1,5] shifted reactants, which are located at the right edge of Figure 8 (CTH-I × 2). The rate-determining step of the channel is the hydrogen [1,5] shift [$\Delta U^{\ddagger} = +18.8$ (+17.0) kcal/mol; CHT-VIIa]. The energy lowerings by the tunnel correction are 0.98 kcal/mol (CHT-II) and 0.34 kcal/mol (CHT-VIIa,b). For a comparison containing the entropy effect, a ΔG (T = 423 K experimentally used[20]) energy diagram was depicted and is shown in Figure 13. The bimolecular process is more favorable than the unimolecular one.

B3LYP/6-31G* [QCISD/6-311G(d,p)] calculations of 7-methyl-1,3,5-cycloheptatriene (Me-CHT) were also carried out. The energy diagram of ΔU is shown in Figure 14. The corresponding geometries are shown in Figures S32−S44 in the Supporting Information. Similar to CHT, the direct hydrogen shift [Me-CHT-II; $\Delta U^{\ddagger} = +36.1$ (+37.2) kcal/mol] is the most unfavorable, and the $\Delta U^{\ddagger}$ is not in agreement with the experimentally obtained activation energy (E_a = 33.3 kcal/mol).[20b] A bimolecular route was also examined. The exo [6+4] addition TS (Me-CHT-IVa), the exo cycloadduct (Me-CHT-IIIa), the TS of a methylene hydrogen migration (Me-CHT-VIIa), a one-center adduct intermediate (Me-CHT-VIIIa), and a Cope rearrangement TS (Me-CHT-IX) were obtained. After the Cope TS (Me-CHT-IX), a reverse route, Me-CHT-VIIIb → Me-CHT-VIIb → Me-CHT-IIIb → Me-CHT-IVb, which are not mirror image isomers because of the dissymmetry of Me-CHT-I, was obtained. Finally, two 2-methyl-1,3,5-cycloheptatrienes were formed. The rate-determining step is the hydrogen migration TS (Me-CHT-VIIa). The activation energy, $\Delta U^{\ddagger} = +20.2$ (+17.3) kcal/mol, is again smaller than the experimental value (E_a = 33.3 kcal/mol). Nevertheless, the hydrogen [1,5] shift in the bimolecular route is a more favorable process in view of the comparison of the calculated energies.

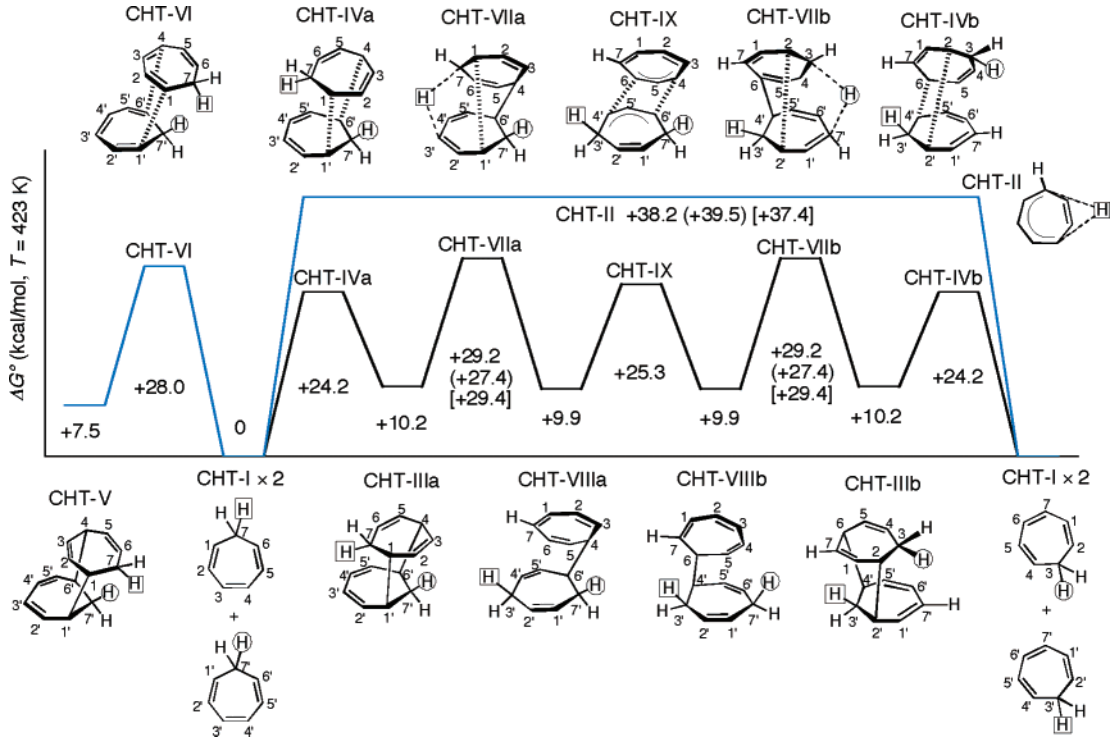


Figure 13. Changes of Gibbs free energies (RB3LYP/6-31G*, $T = 423$ K) of [6+4] cycloadditions and Cope and hydrogen [1,5] sigmatropic shifts of CHTs. Energies in parentheses were calculated by QCISD(T)/6-311G(d,p)//RB3LYP/6-31G* [QCISD(T) electronic energies and RB3LYP thermal corrections and entropies].

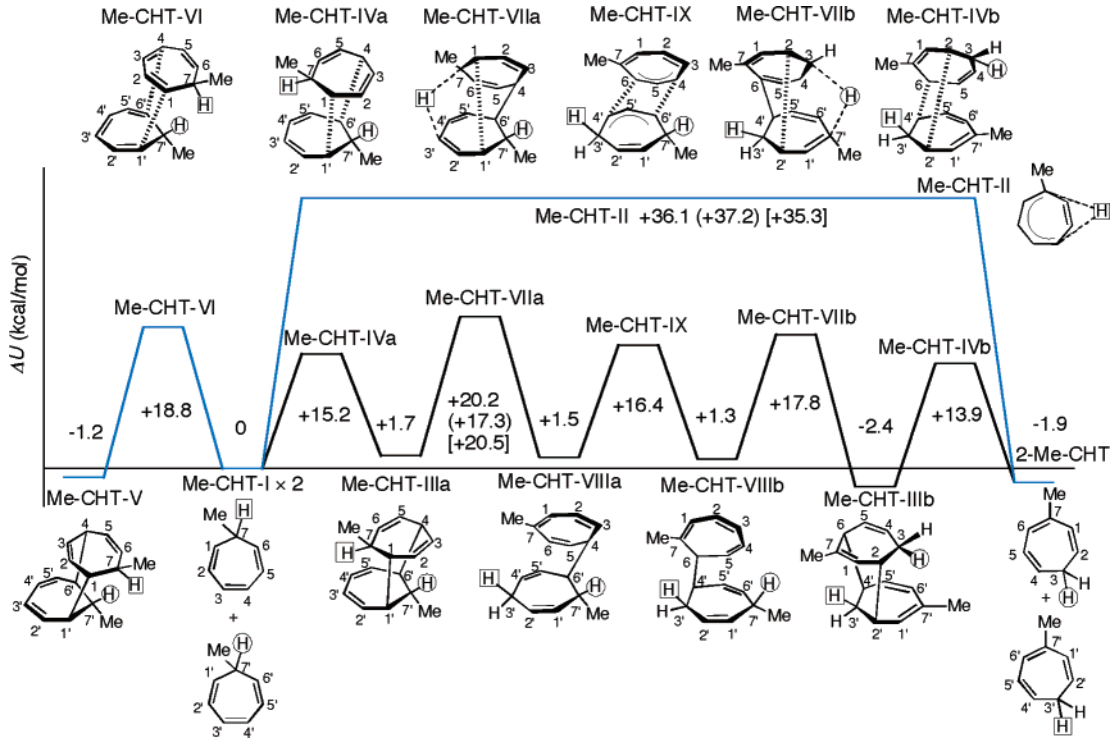
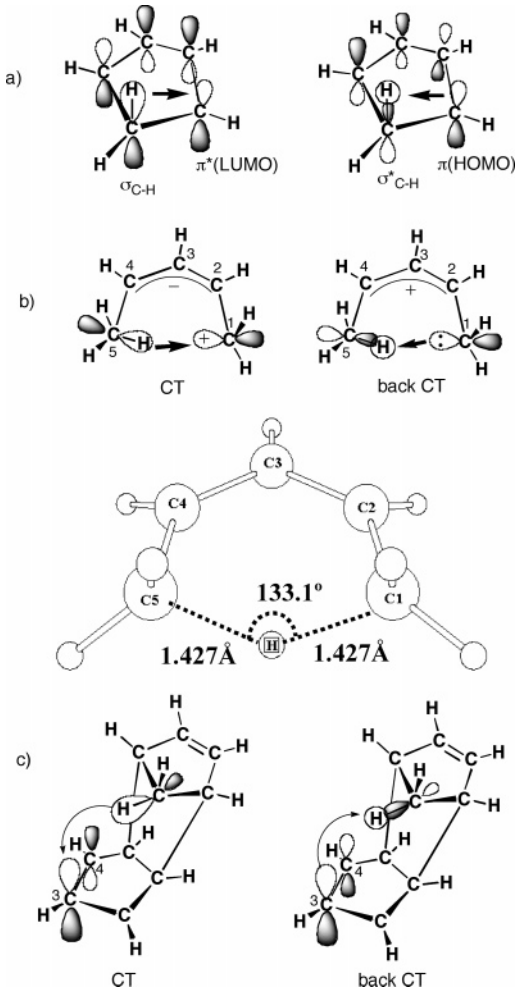


Figure 14. Changes of the sum of electronic and zero-point energies (RB3LYP/6-31G*) of [6+4] cycloadditions and Cope and hydrogen [1,5] sigmatropic shifts of Me-CHTs. Energies in parentheses of II and VIIa are of QCISD(T)/6-311G(d,p)//RB3LYP/6-31G*.

The geometries have been obtained by RB3LYP/6-31G* (gas phase) and the energies by RB3LYP/6-31G* (gas phase) and QCISD(T)/6-311G(d,p). The solvent effect has been examined for CPD and Me-CPD reactions. The data calculated by RB3LYP/6-31G* with the Onsager's reaction field are shown in braces { } in Figures 1, 2, 4, 6, and 7 and Figure S1 (Supporting Information). Since the dielectric constant of CCl_4 is small (= 2.228), the reaction field containing geometries are almost the same as those in the gas phase. Also, the energy changes are insensitive to the

Scheme 4. CT and Back CT Interactions Leading to the Hydrogen Shift. In A, Those of CPD-II Are Depicted. In B, Those of the [1,5] H Shift in 1,3-Pentadiene Are Shown. In C, Those of CPD-VIIa Are Exhibited. Through CT, the C−H Bonding Becomes Weakened and the New C−H Bond Is Formed. Also, through Back CT, the Same Bond Interchange Is Induced.



presence or absence of the effect. For instance, $\Delta U^{\ddagger} = +26.9$ kcal/mol (gas phase) is nearly equal to $\{\Delta U^{\ddagger} = +26.8$ kcal/mol (reaction field)$\}$ for CPD-II in Figure 1.

The effect of the polarization's function on the migrating hydrogen atom on the geometries and energies has been examined. The data are shown in square brackets [ ]. For the TS geometries, the effect has been found to be negligible. For instance, in Figure 4, C(4)···H distances are 1.589 Å and [1.590 Å] and C(5′)−H distances are 1.260 Å and [1.257 Å]. The energies vary by, at most, 1 kcal/mol. For example, $\Delta U^{\ddagger} = +26.9$ kcal/mol and $[\Delta U^{\ddagger} = +25.8$ kcal/mol] for CPD-II and $\Delta U^{\ddagger} = +17.9$ kcal/mol and $[\Delta U^{\ddagger} = +18.1$ kcal/mol] for CPD-VIIa in Figure 1. Thus, Onsager's reaction field and the polarization function on the migrating hydrogen atom have only a small effect on the calculated geometries and energies.

FMO Interactions for the Hydrogen Shift. In this subsection, the bimolecular process is considered in terms of the frontier-orbital interactions. Scheme 4 presents three types of the FMO interactions. In Scheme 4a, the hydrogen [1,5] shift is described by charge transfer (CT) from the sp^3 C−H bonding orbital to the diene LUMO and by back CT from the diene HOMO to the C−H antibonding vacant orbital. To cause the effective CT and back CT, large orbital overlaps are required. But, the CT in Scheme 4a suffers a small overlap owing to the nearly parallel expansion. To make the matter worse, in back CT, the overlap is almost canceled out as a result of the node of σ^*_{C-H}. Thus, the ring-strained FMO interactions are unfavorable. In Scheme 4b, CT and back CT in 1,3-pentadiene are shown. The p_π atomic orbital of the right-sided terminal methylene group may be directed to the left-sided C−H bond of the methyl group. The direction leads to the large orbital overlaps for CT and back CT, and consequently, the unimolecular [1,5] H shift is likely. In Scheme 4c, those interactions in the stacked [4+2] exo dimer of CPD, CPD-VIIa, are shown. The C−H bonding and antibonding orbitals are directed suitably to π and π^* orbitals of the C_3═C_4 bond. The σ-type orbital overlaps become feasible in the stacked form of the exo [4+2] cycloadduct. The C···H···C angle in CPD-VIIa (149.1°, Figure 4) is larger than that in 1,3-pentadiene (133.1°, Scheme 4b), and the CPD dimer provides a ground for the facile H shift. Interestingly, the sum of two C···H distances, 1.260 + 1.589 Å, in CPD-VIIa is very close to that, 1.427 + 1.427 Å, in 1,3-pentadiene (Scheme 4b).

Conclusion

Novel multistep channels for the hydrogen [1,5] shifts for CPD and CHT have been shown. The present results suggest new aspects for pericyclic reactions. The suggestion is that ring-strained methylene hydrogen may shift favorably in the [4+2] or [6+4] stacked conformation. Cycloadditions, sigmatropic shifts, and their reverse reactions work cooperatively to give hydrogen [1,5] shifts for the cycloaddition-favored reactants.

Supporting Information Available: The energy diagrams and optimized geometries not shown in Figures 1−14 for CPD, CHT, Me-CPD, and Me-CHT. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Woodward, R. B.; Hoffmann, R. *The Conservation of Orbital Symmetry*; Verlag Chemie: New York, 1970.

(2) (a) Fukui, K. *Theory of Orientation and Stereoselection*; Springer: New York, 1970. (b) Fleming, I. *Frontier Orbitals and Organic Chemical Reactions*; John Wiley & Sons: New York, 1976.

(3) Houk, K. N.; Li, Y.; Evanseck, J. D. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 682.

(4) Spangler, C. W. *Chem. Rev.* **1976**, *76*, 187.

(5) (a) Dormans, G. J. M.; Buck, H. M. *J. Mol. Struct.* **1986**, *136*, 121. (b) Dormans, G. J. M.; Buck, H. M. *J. Am. Chem. Soc.* **1986**, *108*, 3253. (c) Hess, B. A., Jr.; Schaad, L. J.; Pancir, J. *J. Am. Chem. Soc.* **1985**, *107*, 149. (d) Hess, B. A., Jr.; Schaad, L. J. *J. Am. Chem. Soc.* **1983**, *105*, 7185. (e) Dewar, M. J. S.; Merz, J. M.; Stewart, J. J. P. *J. Chem. Soc., Chem. Commun.* **1985**, 166. (f) Castenmiller, W. A. M.; Buck, H. M. *Tetrahedron* **1979**, *35*, 397. (g) Jensen, F. J.; Houk, K. N. *J. Am. Chem. Soc.* **1987**, *109*, 3139. (h) Dewar, M. J. S.; Healr, E. F.; Merz, J. M. *J. Am. Chem. Soc.* **1988**, *110*, 2666.

(6) Diels, O.; Alder, K. *Justus Liebigs Ann. Chem*. **1928**, *460*, 98.

(7) (a) Wassermann, A. *Diels−Alder Reactions*; Elsevier: London, 1965. (b) Sangwan, N. K.; Schneider, H.-J. *J. Chem. Soc., Perkin Trans. 2*. **1989**, 1223. (c) Jorgensen, W. L.; Blake, J. F.; Lim, D.; Severance, D. L. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1727. (d) Breslow, R.; Zhu, Z. *J. Am. Chem. Soc*. **1995**, *117*, 9923.

(8) Alder, K.; Ache, H. J. *Chem. Ber*. **1962**, *95*, 503; 511.

(9) (a) Kahn, S. D.; Hehre, W. J.; Rondan, N. G.; Houk, K. N. *J. Am. Chem. Soc*. **1985**, *107*, 8291. (b) Rondan, N. G.; Houk, K. N. *Tetrahedron Lett*. **1984**, *25*, 2519. (c) Bachrach, S. M. *J. Org. Chem*. **1993**, *58*, 5414. (d) Okajima, T.; Imafuku, K. *J. Org. Chem*. **2002**, *67*, 625.

(10) (a) Caramella, P.; Quadrelli, P.; Toma, L. *J. Am. Chem. Soc*. **2002**, *124*, 1130. (b) Jamróz, M.; Galka, S.; Dobrowolski, J. C. *THEOCHEM* **2003**, *634*, 225.

(11) Roth, W. R. *Tetrahedron Lett*. **1964**, 1009.

(12) Roth, W. R.; Konig, J. *Justus Liebigs Ann. Chem*. **1966**, *699*, 24.

(13) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.

(14) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648.

(15) 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.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.11.1; Gaussian, Inc.: Pittsburgh, PA, 2001.

(16) Onsager, L. *J. Am. Chem. Soc*. **1938**, *58*, 1486.

(17) The geometries of CPD-IVb, CPD-IIIb, CPD-VIIb, and CPD-VIIIb were made by changing all the signs of the Cartesian coordinates of the geometries of CPD-IVa, CPD-IIIa, CPD-VIIa, and CPD-VIIIa. The energies of CPD-IVb, CPD-IIIb, CPD-VIIb, and CPD-VIIIb are the same as those of CPD-IVa, CPD-IIIa, CPD-VIIa, and CPD-VIIIa.

(18) Wigner, E. P. *Phys. Chem*. **1933**, *B19*, 203.

(19) McLean, S.; Webster, C. J.; Rutherford, R. J. D. *Can. J. Chem*. **1969**, *47*, 1555.

(20) (a) Ter Borg, A. P.; Razenberg, E.; Kloosterziel, H. *Recl. Trav. Chim. Pays-Bas* **1965**, *84*, 1230. (b) Egger, K. W. *J. Am. Chem. Soc*. **1967**, *89*, 3688.

(21) (a) Cookson, R. C.; Drake, B. V.; Hudec, J.; Morrison, A. *J. Chem. Soc., Chem. Commun*. **1966**, 15. (b) Ito, S.; Fujise, Y.; Okuda, T.; Inoue, Y. *Bull. Chem. Soc. Jpn*. **1966**, *39*, 1351.

(22) CHT has a geometric isomer, bicyclo[4.1.0]hepta-2,4-diene (norcaradiene, NCD). NCD has a puckered conformation, which is apparently suitable for the intramolecular hydrogen shift. However, NCD cannot be a reactant for the hydrogen shift, because the cyclopropyl ring in NCD precludes the hydrogen-bridged TS structure. Cycloadditions between CHT and NCD were calculated and are less stable than the rate-determining step VIIa,b in Figure 8. A series of isomerizations of methyl-NCDs were observed. They would involve intermediates of cycloadducts of two methyl-CHTs. (a) Berson, J. A.; Willcot, M. R., III; *J. Am. Chem. Soc*. **1965**, *87*, 2751. (b) Berson, J. A.; Willcot, M. R., III; *J. Am. Chem. Soc*. **1965**, *87*, 2752.

CT0500646

JCTC Journal of Chemical Theory and Computation

Predictions of Optical Excitations in Transition-Metal Complexes with Time Dependent-Density Functional Theory: Influence of Basis Sets

Laurence Petit,[†,‡] Pascale Maldivi,[†] and Carlo Adamo*[,‡]

Laboratoire de Reconnaissance Ionique, DRFMC/LCIB (UMR_E 3 CEA-UJF), CEA-Grenoble, 17 rue des Martyrs, F-38054 Grenoble Cedex 9, France, and Laboratoire d'Électrochimie et de Chimie Analytique, CNRS UMR-7575, École Nationale Supérieure de Chimie de Paris, 11 rue P. et M. Curie, F-75231 Paris Cedex 05, France

Received March 2, 2005

Abstract: The calculation of the absorption spectra of four families of transition-metal complexes ($Ni(CO)_4$, MnO_4^-, MF_6 (M = Cr, Mo, W) and $CpM(CO)_2$ (M = Rh, Ir)) has been undertaken to unravel the influence of basis sets onto excitation energies, oscillator strengths, and assignments. Three among the most common pseudopotentials, with the corresponding valence basis sets, and two all-electron basis sets have been used for the metal center description in the framework of the time dependent Density Functional Theory (TD-DFT). Our results show that this approach does not particularly depend on the basis set used on the metal atoms. Furthermore, the chosen functional PBE0 provides transitions in good agreement with experiments, and it provides an accuracy of about 0.3 eV, comparable to that of refined post-Hartree−Fock methods.

1. Introduction

Over the past three decades, numerous computational strategies have been developed for the simulation of electronic spectra. In particular, electronic properties of transition-metal complexes have been of great interest due to the chemical relevance of such systems in many fields from medicine[1,2] to material design.[3,4] In this perspective, theoretical approaches rooted in quantum chemistry obviously present numerous advantages and have notably helped a lot in clarifying experimental spectra or in assessing solvent-related properties. Yet, available tools are often time-demanding, and, overall, they require a careful choice of the computational parameters. Both points can be a major problem when dealing with large systems of chemical interest.

Theoretical methods based on the Density Functional Theory (DFT), which combine both speed and accuracy, have thus been widely used in this field.[6] In particular, the so-called time dependent DFT (TD-DFT) approach has become a key method validated in numerous works on organic and bioorganic compounds[6−8] as well as on metal complexes.[2,9,10] Based on the linear response theory, it enables to calculate molecular properties such as polarizabilities[11,12] and excitation energies[6−10] by applying a time-dependent field to the system and analyzing its dynamic response.[13] Another main advantage of TD-DFT methods rests on their lack of preliminary assumption concerning the molecular orbitals involved in excitations. This point is of particular relevance when the nature of the excited state is unknown and cannot be clearly deduced from experiments.[9,14]

Even thought it is easily accessible, some cares, relative to the functional and basis sets used, must be taken. The first point, i.e., functional assessment, has been deeply discussed in the literature. In particular, it has been shown that large discrepancies are notably caused by the incorrect asymptotic behavior of the approximated functionals and by the self-interaction error (SIE).[15,16] Various strategies have been proposed to overcome this failure, such as asymptotically corrected potentials[17,18] or by including the Hartree−Fock potential at long-range.[19] More easily, the hybrid approach that introduces some HF exact exchange can

* Corresponding author e-mail: carlo-adamo@enscp.fr.
† DRFMC/SCIB.
‡ École Nationale Supérieure de Chimie de Paris.

10.1021/ct0500500 CCC: $30.25
Published on Web 07/19/2005

Table 1. Contraction Schemes for the All Electron and Valence Basis Sets Considered in the Present Paper

atom	LANL2DZ	SBKJC	SDD	DZVP	Watchers + f
Ni	(5s,5p,5d)/[3s,3p,2d]	(8s,8p,6d)/[4s,4p,3d]	(8s,7p,6d,1f)/[6s,5p,3d,1f]	(15s, 9p,5d)/[5s,3p,2d]	(14s,11p,6d,3f)/[8s,6p,4d,1f]
Mn	(5s,5p,5d)/[3s,3p,2d]	(8s,8p,6d)/[4s,4p,3d]	(8s,7p,6d,1f)/[6s,5p,3d,1f]	(15s, 9p,5d)/[5s,3p,2d]	(14s,11p,6d,3f)/[8s,6p,4d,1f]
Cr	(5s,5p,5d)/[3s,3p,2d]	(8s,8p,6d)/[4s,4p,3d]	(8s,7p,6d,1f)/[6s,5p,3d,1f]	(15s, 9p,5d)/[5s,3p,2d]	
Mo	(5s,6p,4d)/[3s,3p,2d]	(7s,7p,5d)/[4s,4p,3d]	(8s,7p,6d)/[6s,5p,3d]	(18s,12p,9d)/[6s,5p,3d]	
W	(8s,6p,3d)/[3s,3p,2d]	(7s,7p,5d)/[4s,4p,3d]	(8s,7p,6d)/[6s,5p,3d]		
Rh	(5s,6p,4d)/[3s,3p,2d]	(7s,7p,5d)/[4s,4p,3d]	(8s,7p,6d)/[6s,5p,3d]	(18s,12p,9d)/[6s,5p,3d]	
Ir	(8s,6p,3d)/[3s,3p,2d]	(7s,7p,5d)/[4s,4p,3d]	(8s,7p,6d)/[6s,5p,3d]		

represent an interesting alternative. Among others, the PBE0 hybrid functional[20,21,27] has many times proved its efficiency on a wide range of compounds, and it generally provides accurate results on ground and excited-state properties, including charge-transfer transitions.[6,22−26] It derivates from the Perdew-Burke-Erzenrhof PBE exchange-correlation functional,[27] but it is not fitted on any experimental data. In contrast, at the best of our knowledge, the dependence of the TD-DFT approach on basis sets has never been systematically tested for metal-containing complexes. It is nevertheless well-known that basis sets can affect molecular properties, including excited states.[28,29]

To elucidate this point, often not well explored in TD-DFT applications, we have considered in the present study several benchmark cases, combining simple structure, various ligands and metals. The chosen metal complexes have been well characterized at both experimental and theoretical levels so as to allow detailed comparisons. We have thus worked out the absorption spectra of four classes of transition-metal complexes: $Ni(CO)_4$, MnO_4^-, MF_6 (M = Cr, Mo, W) and $CpM(CO)_2$ (M = Rh, Ir). Their spectra have been computed using different pseudopotentials, selected among those commonly used in TD-DFT applications[30,31] as well as of two all-electron basis sets, especially developed for metals.

2. Computational Details

All calculations were performed with the Gaussian 03 package.[32] Optimizations were carried out at the DFT level of theory, while absorption spectra were computed as vertical electronic excitations from the ground state using the TD-DFT approach as implemented in Gaussian.[33] For both types of calculations, the PBE0 hybrid functional,[20,21] that includes 25% of exact exchange, was used. All systems were studied in their highest symmetry, that is T_d for $Ni(CO)_4$ and MnO_4^-,[34−36] O_h symmetry for MF_6 complexes (M = Cr, Mo, W)[37−41] and C_s for $CpM(CO)_2$ (M = Rh, Ir).[42] Even though the correct symmetry of CrF_6, and even its very existence, are still a source of controversy, recent studies tend to favor the O_h group.[37−39]

Three different pseudopotentials, namely those of Hay, Wadt, and co-workers (LANL2DZ[43,44]), those of Stevens and co-workers (SBKJC[45]), and those of the Stuttgard group (SDD[46,47]), with the corresponding valence basis sets are herein considered. Two all-electron basis sets were also used: the "classical" basis of Watchers,[48,49] including a *f* diffuse function, and the more recent DZVP for DFT calculations, either from Chiodo and co-workers[50] or Godbout et al.[51] All the considered basis and contraction schemes are summarized in Table 1. Whatever the complex is, remaining atoms (F, O, C, and H) were always described with the 6-31+G(2d,p) basis set as suggested by Daniel.[52]

3. Results and Discussion

We want to add few general remarks concerning our calculations before discussing them. On one hand, only the best results are discussed in details for a given system, while, for other basis sets, the sole transitions with experimental counterparts are reported in the tables. Each time, contributions to a given transition having participations below 5% are not mentioned. On the other hand, all the TD-DFT calculations have been carried out on the optimized structure with the above-mentioned criteria. We have nonetheless checked that results starting from X-ray structures were similar, with maximum discrepancies in the order of 10 nm. Since significant differences were observed only for the permanganate ion, both sets of transitions are reported in this case. On the whole, optimized geometries, reported in the Supporting Information (see Table SI), are quite close to the experimental values. Finally, we only focused on spin-allowed transitions, i.e., singlet excitations.

MF_6, M = Cr, Mo, W. We have chosen as the first tests three extensively studied hexafluoride complexes, namely CrF_6, MoF_6, and WF_6. Such systems are formally M(VI) complexes and the metal atoms have a d^0 electronic configuration. Indeed, all the valence occupied orbitals have essentially ligand character, while the first unoccupied orbitals are centered on the metal. Transitions are thereby expected to be essentially made of ligand to metal charge transfer (LMCT). Moreover, despite numerous controversies,[53−55] the octahedral geometry was adopted for the three systems. The only spin and symmetry allowed excitations are thus $^1A_{1g} \rightarrow {}^1T_{1u}$ transitions. The fluorides generate a strong crystal field that increases in going from Cr to W. The metal e_g orbitals are thus quite high in energy, and most of the absorption bands are to t_{2g} levels. As expected, CrF_6, MoF_6, and WF_6 spectra are akin in relative structure, intensities, and assignments. In particular, three allowed charge-transfer bands are recorded whatever the complex is. For MoF_6 and WF_6, they go with several dipole-forbidden transitions lying high in energy and consequently ascribed to Rydberg states.

On the whole, our results are in very good agreement with available experimental data,[56,57] as shown in Table 2. The mean absolute errors (MAE) on energies are rather low, ranging from 0.43 eV for WF6 (LANL2DZ), 0.23 eV for MoF6 (SDD), to 0.10 eV for CrF_6 (SBKJC). What is more,

Table 2. Excitation Energies (eV), Oscillator Strengths, and Corresponding Assignments Calculated on the PBE0 Structure of CrF_6, MoF_6, and WF_6[c]

PBE0			theoretical reference[a]			experiment[b]		
excitation energy	oscillator strength	assignments	excitation energy	oscillator strength	assignments	excitation energy	oscillator strength	assignments
			CrF_6 (SBKJC ECP)					
3.13	0.0018	79%($t_{1u}\rightarrow t_{2g}$) + 21%($t_{2u}\rightarrow t_{2g}$)	2.38	0.0002	($t_{1u}\rightarrow t_{2g}$) + ($t_{2u}\rightarrow t_{2g}$)	3.31	middle	
3.89	0.0272	60%($t_{2u}\rightarrow t_{2g}$) + 35%($t_{1u}\rightarrow t_{2g}$) + 5%($t_{1u}\rightarrow e_g$)	3.12	0.0138	($t_{1u}\rightarrow t_{2g}$) + ($t_{2u}\rightarrow t_{2g}$)	3.88	strong	
4.89	0.0273	76%($t_{1u}\rightarrow t_{2g}$) + 18%($t_{1u}\rightarrow e_g$) + 6%($t_{2u}\rightarrow t_{2g}$)	4.21	0.0406	($t_{1u}\rightarrow t_{2g}$) + ($t_{2u}\rightarrow t_{2g}$)	4.77	strong	
			MoF_6 (SDD)					
5.51	0.0028	85%($t_{1u}\rightarrow t_{2g}$) + 15%($t_{2u}\rightarrow t_{2g}$)	5.55	0.0243	100%($t_{1u}\rightarrow t_{2g}$)	5.90	weak	$t_{1g}\rightarrow t_{2g}$
6.03	0.0197	51%($t_{1u}\rightarrow t_{2g}$) + 49%($t_{2u}\rightarrow t_{2g}$)	6.62	0.0945	100%($t_{2u}\rightarrow t_{2g}$)	6.54	middle	$t_{1u}\rightarrow t_{2g}$
7.15	0.1618	60%($t_{1u}\rightarrow t_{2g}$) + 27%($t_{2u}\rightarrow t_{2g}$) + 13%($t_{1u}\rightarrow e_g$)	7.24	0.3549	100%($t_{1u}\rightarrow t_{2g}$)	7.12	strong	$t_{2u}\rightarrow t_{2g}$
8.71	forbidden	100%($e_g\rightarrow t_{2g}$)	8.91	forbidden	100%($t_{2g}\rightarrow t_{2g}$)	8.62	very weak	$e_g\rightarrow t_{2g}$
9.08	forbidden	100%($e_g\rightarrow t_{2g}$)	9.49	forbidden	100%($e_g\rightarrow t_{2g}$)	9.22	very weak	$e_g\rightarrow t_{2g}$
			WF_6 (LANL2DZ)					
6.72 (184.60)	0.0028	82%($t_{1u}\rightarrow t_{2g}$) + 18%($t_{2u}\rightarrow t_{2g}$)	7.20	$t_{1g}\rightarrow t_{2g}$		7.23	weak	$t_{1g}\rightarrow t_{2g}$
7.22 (171.79)	0.0280	54%($t_{1u}\rightarrow t_{2g}$) + 46%($t_{2u}\rightarrow t_{2g}$)	7.92	$t_{1u}\rightarrow t_{2g}$		8.05	middle	$t_{2u}\rightarrow t_{2g}$
8.34 (148.72)	0.2193	60%($t_{1u}\rightarrow t_{2g}$) + 29%($t_{2u}\rightarrow t_{2g}$) + 11%($t_{1u}\rightarrow e_g$)	8.68	$t_{1u}\rightarrow t_{2g}$		8.60	strong	$t_{1u}\rightarrow t_{2g}$
9.05 (136.97)	forbidden	100%($t_{2g}\rightarrow t_{2g}$)	9.07	$a_{1g}\rightarrow t_{2g}$		8.94	very weak	$t_{1u}\rightarrow t_{2g}$
10.18 (121.78)	forbidden	100%($e_g\rightarrow t_{2g}$)				10.03	very weak	$a_g\rightarrow t_{2g}$
11.11 (111.61)	forbidden	80%($t_{2g}\rightarrow t_{2g}$) + 20%($e_g\rightarrow t_{2g}$)				11.05	very weak	$e_g\rightarrow t_{2g}$

[a] Theoretical references: TD-DFT for CrF_6 (ref 60), SAC-CI for MoF_6 (ref 58), and Xα method for WF_6 (ref 59). [b] Experimental references: ref 56 for CrF_6 in N_2 matrix, ref 57 for MoF_6 and WF_6 in gaseous phase. [c] The reported results are those in better agreement with the experiments, and they have been obtained with different pseudopotentials for the metal atoms.

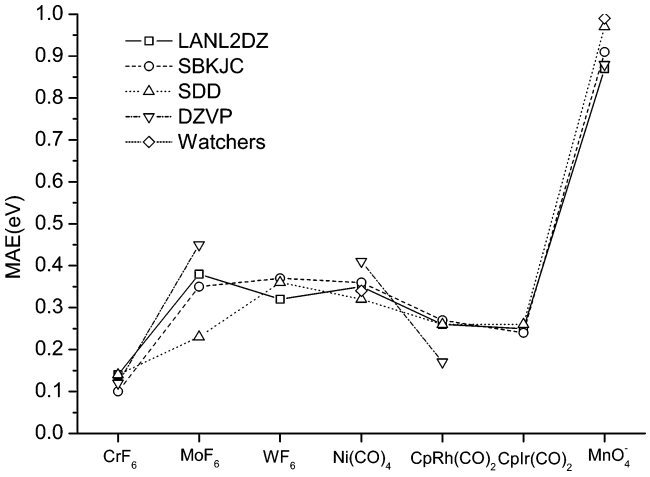


Figure 1. Mean absolute errors (MAEs, eV) for the vertical transitions of CrF_6, MoF_6, WF_6, $Ni(CO)_4$, $CpRh(CO)_2$, $CpIr(CO)_2$, and MnO_4^- complexes, computed using different pseudopotentials and related valence basis sets (LANL2DZ, SBKJC, SDD) as well as the DZVP and Watchers+f all-electron basis.

this discrepancy is almost constant from one basis set to another, as can be seen from the MAEs reported in Figure 1. MoF_6 seems however not to respect this trend since wide variations are observed from SDD (MAE of 0.23 eV) to DZVP (MAE of 0.45 eV) with intermediate MAE for LANL2DZ (0.38 eV) and SBKJC (0.35 eV). In any cases, the largest deviations (around 0.4/0.5 eV) are for the first two lowest excitations that can be assigned as LMCT transitions. These errors are nonsystematic, so that the nonsystematic trends can be clearly evidenced. Therefore it is difficult to argue that these deviations are related to the limits of the basis considered or to an intrinsic problem to the TD-DFT method when dealing with charge-transfer transitions.[16]

Intensities are well reproduced, with an increasing pattern from the first allowed band to the last one. In contrast, theoretical assignments are more questionable when compared to experimental suggestions. As illustrated by the MoF_6 complex, all our bands with the exception of the first one are consistent with the experiment, but several transitions display some interconfigurational mixing. Actually, different assignments have been proposed for the transitions of MoF_6,[23,57−59] leading to some inconsistencies. For instance, as previously revealed by Nakai and co-workers[58] the experimental assignment of the first band is arguable since $t_{1g}\rightarrow t_{2g}$ transitions are calculated at lower energies (4.91 and 5.01 eV in our work) and are forbidden (see Table SII, Supporting Information). On the other hand, beyond 7.5 eV, all the transitions are forbidden with close assignments. A rigorous analysis with respect to experiment can thus turn out to be difficult.

It is nevertheless important to highlight the accuracy of our results with respect to other theoretical calculations. With the exception of WF6 whose absorption spectrum is particularly accurate at the Xα level,[59] results on MoF_6 and CrF_6 are at least comparable or even better than in previous SAC-CI[57] and TD-DFT calculations, respectively.[60] In particular, CrF_6 needs some comments. In fact, previous calculations,[60]

Table 3. Excitation Energies (eV), Oscillator Strengths, and Corresponding Assignments Calculated with the SDD Pseudopotential on the PBE0 Optimized Structure of $Ni(CO)_4$

PBE0			CASPT2/CASSCF[a]			
excitation energy	oscillator strength	assignments	excitation energy	oscillator strength	assignments	experiment[b] excitation energ;
4.87	0.0242	82%($t_2 \rightarrow t_2$) + 18%($t_2 \rightarrow e$)	4.34	0.29	92%($t_2 \rightarrow t_2$)	4.5, 4.54
5.12	0.0760	72%($t_2 \rightarrow e$)+ 15%($t_2 \rightarrow t_2$) + 7%($t_2 \rightarrow t_1$) + 6%($e \rightarrow t_2$)	5.22	0.38	92%($t_2 \rightarrow e$)	5.4, 5.17, 5.24
5.88	0.0057	76%($e \rightarrow t_2$) + 21%($t_2 \rightarrow t_1$)	5.57	0.29	94%($e \rightarrow t_2$)	5.52
6.41	0.1700	69%($t_2 \rightarrow t_1$) + 16%($e \rightarrow t_1$) + 12%($e \rightarrow t_2$)	6.28	0.47	93%($t_2 \rightarrow t_1$)	6.0, 6.02
6.68	0.0274	100%($t_2 \rightarrow a_1$)				
7.50	0.2721	94%($e \rightarrow t_1$) + 6%($t_2 \rightarrow t_1$)	6.97	0.83	88%($e \rightarrow t_1$) + 6%($t_2 \rightarrow t_1$)	
7.65	0.0189	100%($t_2 \rightarrow t_2$)				

[a] Reference 64. [b] Reference 63 (gas phase); ref 62 (matrix); ref 61 (solvent).

carried out with a similar approach, did not achieve accurate results as can be seen from the values reported in Table 2. Depending on the used functional, the discrepancy could be attributed either to geometrical problems (LB94) or to a small basis sets (B3LYP calculations).

$Ni(CO)_4$. Nickel tetracarbonyl is a d^{10} system. No d−d transitions are therefore observed, and excitations are dominated by charge transfer. It is an interesting system in our study since three different experimental spectra are available, recorded in solution,[61] in matrix,[62] and in the gas phase.[63] Several calculations at different levels of theory have been performed subsequently.[64−68] This large amount of results allows an accurate assessment of our work. Experimentally, $Ni(CO)_4$ features four main transitions. Both solution and gas-phase spectra give a main band around 6 eV with two shoulders at 5.52 and 4.54 eV in solution and 5.4 and 4.5 eV in the gas phase. Only two transitions at 5.17 and 4.54 eV are instead reported in a matrix. These energies, listed in Table 3, are compared to our results calculated with the SDD pseudopotential for the metal center. CASPT2 values,[64] which are among the most accurate results ever calculated for $Ni(CO)_4$, are also given. We refer the reader to refs 65 and 51 for a deeper scrutiny of other theoretical methods. According to the Laporte's rule, the system, optimized in T_d symmetry, can only present $^1A_1 \rightarrow {}^1T_2$ allowed transitions. Seven of such excitations are listed in Table 3, the four first corresponding to the experimental ones. All the transitions are overestimated, and the MAE is of about 0.3 eV with a maximum deviation of 0.4 eV for the highest transition. Previous TD-DFT calculations performed with the BP functional[66] show the opposite trend (underestimation) but are nevertheless consistent with our work. As expected, the expensive CASPT2 approach is in better agreement with the experimental values, the MAE being of 0.17 eV.[64] Turning to intensities, it must be noticed that the experimental trend in solution is respected, the strongest transition being around 6.5 eV ($f = 0.170$) with the two lowest counterparts at 5.12 eV ($f = 0.076$) and 4.87 eV ($f = 0.024$). Oscillator strengths calculated by Pierloot et al.[64] behave in the same way as ours but are more intense. Nevertheless, it must be noticed that the CASPT2 method does not enable to calculate transition moments and configurations so the values reported in Table 3 originate from a CASSCF calculation.

Orbitals from HOMO-3 to LUMO+3 are calculated to have t_2 symmetry. The last occupied orbitals are essentially made of Ni 3d orbitals, but some bonding character with the carbonyls is nevertheless observed. The first LUMOs clearly show an antibonding character between the 3d(Ni) and the π^*(CO). The four first allowed transitions can be therefore ascribed to charge transfer from the 3d orbitals of nickel to the π^*(CO), resulting in a destabilization of the carbonyl. Even so, the assignment of such transitions is still a source of controversy. As previous CASPT2 (assignments made on the basis of CASSCF results, Table 3) and TD-DFT works,[64,66] we found the allowed bands I to IV to be mainly composed of $t_2 \rightarrow t_2$, $t_2 \rightarrow e$, $e \rightarrow t_2$, and $t_2 \rightarrow t_1$ transitions, respectively. Our assignments are in very good agreement with CASSCF calculations and turned out to present less interconfigurational mixing than previous TD-DFT and SAC-CI studies.[65,66] One can also notice that the excitation at 6.68 eV is not reported in other theoretical works. The CASSCF and TD-DFT calculations only quote one $e \rightarrow t_1$ transition, whereas no transition above 5.8 eV appears at the SAC-CI level. This latter excitation is herein found at 7.50 eV and is the most intense one. Actually, as underlined by van Gisbergen et al.,[66] these differences are difficult to rationalize since they strongly depend on the size of the active space considered in the CASSCF or SAC-CI calculation. In this manner, some orbitals do not appear simply because they are not included at the beginning, as some a_1 orbitals for instance.

The variations observed in going from one basis set to another are quite small. Whatever the basis set considered on the metal center is, Ni−C and C−O bonds are always strengthened by the PBE0 optimization, being on average 0.018 Å too low (see Table SI, Supporting Information). We checked, at the same level of theory with SDD pseudopotential, that this discrepancy did not significantly alter the absorption spectrum, the maximum difference being of a few nanometers. Excitation values for the considered basis sets are all listed in Table 4. Results globally indicate that the way the metal center is described hardly affects the spectrum, the pseudopotentials providing as accurate results as all-electron basis sets. Core electrons have indeed little effect on results since they do not participate in the excitation process. In contrast, the way 3d orbitals of the Ni are described can have a great influence. Furthermore, since the process consists of a charge transfer, polarization or diffuse functions should at least in principle be considered. Yet, discrepancies are herein too small to draw any conclusion.

Table 4. Allowed Transitions of $Ni(CO)_4$ (in eV) Computed with Different Pseudopotentials and All-Electron Basis Sets[a]

LANL2DZ		SBKJC		SDD		DZVP		Watchers + f	
excitation energy	oscillator strength	excitation energy	oscillator strength	excitation energy	oscillator strength	excitation energy	oscillator strength	excitation energy	oscillator strength
4.93	0.0259	4.94	0.0253	4.87	0.0242	5.04	0.0207	4.92	0.0249
5.19	0.0763	5.19	0.0742	5.12	0.0760	5.29	0.0800	5.15	0.0724
5.95	0.0035	5.96	0.0028	5.88	0.0057	6.07	0.0013	5.93	0.0026
6.45	0.1836	6.47	0.1826	6.41	0.1700	6.54	0.1868	6.42	0.1791
6.91	0.0268	6.94	0.0273	6.68	0.0274	6.99	0.0273	6.91	0.0251
7.49	0.0036	7.57	0.2922	7.50	0.2721	7.63	0.2961	7.51	0.2988

[a] Assignments being very close from one basis set to another one, they are not mentioned herein.

$CpM(CO)_2$, M = Rh, Ir. $CpM(CO)_2$ systems, M = Rh, Ir, have recently been of much interest because they are able to activate the normally unreactive C−H bond when excited.[69−71] Their absorption spectra is herein studied with three pseudopotentials (LANL2DZ, SBKJC, and SDD) and the DZVP all electron basis set for the rhodium complex. The DZVP results for $CpRh(CO)_2$ are compared with recent theoretical data at the SAC-CI level of theory[42] as well as with the few available experimental values[72] (see Table 5). Note that the SAC-CI assignments are based on our symmetry orientation, as given by the Gaussian program, so as to make comparisons easier. The optimized structures in C_s symmetry are found to have an $^1A'$ ground state, in agreement with ref 42. Whatever the basis set on the metal center is, the geometry is quite consistent with the SAC-CI one, with the exception of the d(Rh−Cp) distance that is calculated to be on average 0.16 Å too low. Likewise, frontier orbitals are similar.[42] The last occupied levels (from HOMO-4 to HOMO-1) are dominated by the 4d orbitals of rhodium, while the first unoccupied ones (from LUMO+1 to LUMO+4) present an important antibonding character between d(Rh) and π^*(CO). As a consequence, the overall good agreement with these theoretical values is not surprising. Only one significant inversion in the assignments can be observed. These orbitals are mainly involved in two MLCT transitions calculated at 4.43 eV and 4.68 eV (see Table 5). At the SAC-CI level, they are found at 4.38 and 4.45 eV, respectively, after another intense band at 4.36 eV ($f = 0.1003$). This last transition originates from a HOMO to LUMO excitation and corresponds to a charge transfer from Cp to Rh and CO.

More remarkably, in our work, three of the most intense bands lie in the range of 5.7−6.0 eV, namely at 5.75 eV (7A′, $f = 0.1079$), 5.82 eV (8A′, $f = 0.0277$), and 5.93 eV (9A′, $f = 0.1100$). None of these bands appears in the SAC-CI results because these calculations are limited to the four first A′ and A″ singlet excitations. They correspond to excitations from the HOMO toward the Cp ligand (for 8A′ and 9A′) and the carbons of carbonyls (7A′).

Results with pseudopotentials are very similar in the band positions, wavelengths, and intensities, except for the highest energies. The graph of Figure 1 well underlines this homogeneity. The three pseudopotentials behave in the same way up to 5 eV, with a small overestimation of excitation energies of about 0.1−0.2 eV with respect to the DZVP values. For instance, the MLCT transitions with LANL2DZ pseudopotential are found to be at 4.60 and 4.79 eV instead of 4.43 and 4.68 eV. Intensities and orbitals are also akin to the SAC-CI ones. In contrast, when going to high energies, discrepancies are more marked, with states inversions, decreases in intensities, and changes in assignments. More in details, 7A′, 8A′, and 9A′ intense states calculated with the DZVP basis set appear to be weaker, whereas forbidden A″ excitations become allowed. Such differences are quite puzzling. Actually, in Hu et al.[42] SAC-CI study, metals are described by the Hay−Wadt relativistic effective pseudopotential[73] that has almost the same characteristics as ours (28 core electrons, (5s,6p,4d) → [3s,3p,2d] for Rh). The observed difference between our results with pseudopotentials and the SAC-CI ones is thus a matter of methodology. When using the all-electron DZVP basis set, results become surprisingly better, whereas no relativistic effect is taken into account. Presumably, this behavior depends on error compensation.

The calculated excitation energies, oscillator strengths, and assignments for the $CpIr(CO)_2$ complex with LANL2DZ, SBKJC, and SDD pseudopotentials are shown in Table 6. On the whole, they are very close to the ones with rhodium, except for high energies transitions and for small differences in assignments. Yet, since the DZVP basis set is not available, they are not more accounted for herein. One should notice that the comparison with SAC-CI results is not complete because of problems of assignments.

$MnO_4^{-\cdot}$. The permanganate ion is a very challenging system because it involves both configuration mixing in excitations and important correlation effects. For this reason, numerous theoretical works have been dedicated to the study of its absorption spectrum with a wide range of methods, from ΔSCF[74] to CI(SD).[75] Yet, lots of inconsistencies in energies and assignments appear. A CASPT2 multiconfigurational treatment is difficult since a large active space would be necessary. A SAC-CI study[76] has nevertheless managed to calculate energies in quite good agreement with experimental data. Our results, based on the optimized tetrahedral structure with LANL2DZ pseudopotential for singlet excitation energies, oscillator strengths, and assignments, are listed in Table 7. They are compared to the experimental values recorded by Holt and co-workers.[77] As expected for a d^0 complex, we found the last occupied levels to be mainly composed of p(O) orbitals, while the first unoccupied levels mix 3d orbitals of manganese and p orbitals of oxygen. In this manner, the allowed bands correspond to LMCT bands. In agreement with other theoretical works,[60,66,76] the lowest allowed band is ascribed to a $t_1 \rightarrow e$ transition at 3.08 eV.

Table 5. Excitation Energies (eV), Oscillator Strengths, and Assignments Calculated with the DZVP Basis Set on the PBE0 Optimized Structure of $CpRh(CO)_2$

PBE0				SAC-CI[a]				
state	excitation energy	oscillator strength	assignments	state	excitation energy	oscillator strength	assignments	experiment[b] excitation energy
1A″	2.73	forbidden	100%(34a′→21a″)	1A″	2.71	forbidden	57%(34a′→36a′) + 29%(34a′→35a′) + 7%(19a″→36a′)	2.93
2A″	3.55	0.0050	100%(33a′→21a″)	2A″	3.22	0.0081	64%(33a′→36a′) + 25%(33a′→35a′)	
1A′	3.93	forbidden	95%(34a′→36a′) + 5%(19a″→21a″)	1A′	3.95	forbidden	61%(20a″→36a′) + 24%(20a″→35a′) + 7%(34a′→28a″)	
2A′	4.36	0.1003	86%(34a′→35a′) + 9%(20a″→21a″) 5%(32a′→36a′)	4A′	4.57	0.1671	67%(34a′→21a″) + 24%(34a′→22a″) + 9%(34a′→23a″)	
3A′	4.43	0.0021	79%(19a″→21a″) + 14%(33a′→35a′)	2A′	4.38	0.0116	31%(33a′→22a″) + 29%(33a′→21a″) + 24%(34a′→23a″) + 16%(34a′→22a″)	4.34
3A″	4.59	forbidden	97%(34a′→23a″)					
4A′	4.68	0.0095	89%(33a′→35a′) + 11%(19a″→21a″)	3A′	4.45	0.0359	34%(33a′→22a″) + 34%(33a′→21a″) + 19%(34a′→23a″) + 13%(33a′→23a″)	
4A″	4.90	0.0227	66%(34a′→22a″) + 34%(32a′→21a″)					
5A′	4.97	0.0007	97%(33a′→36a′)					
5A″	5.14	0.0147	65%(32a′→21a″) + 24%(34a′→22a″) + 7%(32a′→23a″)	3A″	4.84	0.0045	61%(32a′→36a′) + 24%(32a′→35a′)	
6A′	5.20	forbidden	100%(34a′→37a′)					
6A″	5.48	forbidden	100%(33a′→22a″)					
7A″	5.65	0.0051	49%(33a′→23a″) + 31%(31a′→21a″) + 11%(34a′→24a″)	4A″	5.86	0.0012	28%(34a′→36a′) + 28%(19a″→36a′) + 24%(19a″→35a′) + 20%(34a′→35a′)	
8A″	5.65	0.0005	53%(33a′→23a″) + 32%(31a′→21a″) + 12%(34a′→24a″)					
7A′	5.75	0.1079	70%(34a′→38a′) + 17%(20a″→21a″) + 9%(34a′→39a′)					
9A″	5.75	forbidden	70%(34a′→24a″) + 9%(34a′→25a″) + 8%(19a″→35a′) + 8%(31a′→21a″)					
10A″	5.79	0.0014	59% (19a″→35a′) + 17%(34a′→25a″) + 11%(31a′→21a″) + 8%(34a′→24a″)					
8A′	5.82	0.0277	54%(34a′→39a′) + 28%(34a′→38a′) + 14%(34a′→40a′)					
11A″	5.88	forbidden	79%(19a″→36a′) + 12%(31a′→21a″) +					
9A′	5.93 (209.23)	0.1100	43%(34a′→39a′) + 26%(34a′→40a′) + 23%(20a″→21a″) + 6%(34a′→38a′)					

[a] Reference 42. [b] In decalin, ref 72.

This energy is upshifted of 0.8 eV with respect to the experimental excitation. A similar trend is found for the other bands, the largest difference being about 1 eV. In contrast, oscillator strengths behave in the same way as the experimental spectrum. The third band is indeed the most intense, with another first strong excitation and a weak one at 4.31

Table 6. Excitation Energies (eV), Oscillator Strengths, and Assignments Calculated with LANL2DZ Pseudopotential on the Optimized Structure of $CpIr(CO)_2$[b]

PBE0				SAC-CI[a]			
state	excitation energy	oscillator strength	assignments	state	excitation energy	oscillator strength	assignments
1A″	3.57	forbidden	100%(24a′→17a″)	1A″	3.30	forbidden	50%(23a′→17a″) + 50%(23a′→25a′)
2A″	4.40	0.0042	100%(23a′→17a″)	2A″	3.84	0.0073	53%(16a″→25a′) + 47%(16a″→17a″)
1A′	4.43	forbidden	78%(24a′→25a′) + 9%(24a′→27a′) + 7%(24a′→26a′)	1A′	4.55	0.0002	38%(23a′→27a′) + 25%(15a″→25a′) + 19%(15a″→17a″)
2A′	4.97	0.1318	82%(24a′→26a′) + 9%(24a′→25a′)	3A′	5.00	0.2132	73% (23a′→24a′) + 27%(23a′→18a″)
3A″	5.04	0.0003	54%(24a′→19a″) + 37%(24a′→20a″) + 9%(24a′→18a″)	3A″	5.15	0.0001	
4A″	5.18	0.0094	83%(24a′→18a″) + 7%(22a′→17a″) + 6%(24a′ →19a″)				
3A′	5.19	0.0064	64%(23a′→26a′) + 19%(16a″→17a″) + 8%(23a′→25a′) + 5%(23a′ →28a′)	2A′	4.93	0.0335	54%(16a″→18a″) + 46%(16a″→24a′)
4A′	5.28	0.0026	87%(24a′→27a′) + 9%(24a′→25a′)				
5A′	5.39	0.0015	76%(16a″→17a″) + 12%(23a′→25a′) + 9%(23a′→26a′)	4A′	5.13	0.0140	37%(15a″→25a′) + 34%(15a″→17a″) + 15%(23a′→26a′)
6A′	5.44	0.0053	68%(23a′→25a′) + 15%(23a′→26a′) + 10%(23a′→27a′)				
5A″	5.64	0.0001	100%(23a′→18a″)				
6A″	5.78	forbidden	60%(24a′→20a″) + 40%(24a′→19a″)				
7A′	5.81	0.0007	100%(24a′→28a′)				
8A′	5.83	0.0104	88%(24a′→29a′) + 12%(24a′→30a′)				
9A′	5.89	0.0463	83%(24a′→30a′) + 17%(24a′→29a′)				
7A″	6.00	0.0335	68%(24a′→21a″) + 18%(24a′→22a″) + 10%(22a′→17a″)				
8A″	6.03	0.0058	47%(23a′→19a″) + 39%(23a′→20a″) + 11%(22a′→17a″)				
9A″	6.20	0.0035	41%(22a′→17a″) + 37%(16a″→26a′) + 7%(23a′→20a″) + 6%(16a″→25a′)				
10A′	6.27	0.0034	75%(24a′→31a′) + 20%(23a′→27a′)				
11A′	6.29	0.0080	58%(23a′→27a′) + 27%(24a′→31a′) + 6%(23a′→21a″) + 5%(15a″→17a″)				

[a] Reference 42. [b] Notice that the SAC-CI assignments are not complete, as reported in ref 39.

Table 7. Excitation Energies (eV), Oscillator Strengths, and Assignments Calculated with the LANL2DZ Pseudopotential on the Optimized Structure of MnO_4^-

TD-DFT/PBE0			SAC-CI[a]			experiment[b]		
excitation energy	oscillator strength	assignments	excitation energy	oscillator strength	assignments	excitation energy	oscillator strength	assignments
3.08	0.0095	89%(t_1→e) + 11%(t_2→e)	2.57	0.0202	t_1→e	2.27	strong (0.03)	t_1→e
4.31	0.0015	62%(t_2→e) + 38%(t_1→e)	3.58	0.0045	t_1→t_2	3.47	weak	t_1→t_2
4.84	0.0105	49%(t_1→t_2) + 25%(t_2→e) + 16%(t_2→t_2) + 8%(a_1→t_2)	3.72	0.0136	t_2→e	3.99	strong (0.07)	t_2→e
6.42	0.0016	76%(t_1→t_2) + 17%(t_2→e) + 7%(a_1→t_2)	5.82	0.0022	t_2→t_2 + a_1→t_2	5.45	strong	t_2→t_2
6.66	0.0017	50%(t_2→e) + 34%(a_1→t_2) + 16%(t_2→t_2)						

[a] Reference 76. [b] Reference 77.

eV. One should notice that a fifth band around 6.5−7.0 eV is found but is neither reported in experimental records nor in other theoretical calculations.

Although they are consistent with the global behavior of the experimental spectrum, our results seem quite inaccurate. Furthermore, they are almost unaffected by the basis sets, as can be seen from the data of Table 8. Even basis sets including diffuse functions (SDD and Watchers) give large errors. Previous TD-DFT works give similar inaccuracy.[60,66] Different explanations have been proposed for such a failure, but the multiconfigurational nature of MnO_4^- seems the main cause.[60] This is well illustrated by the SAC-CI calculations,[76] whose results become more and more reliable as the active space increases. The energies listed in Table 7 come from an active space composed of 12 occupied and 41 empty orbitals. They are quite close to the experimental values, with a mean error of 0.26 eV.

To separate electronic from geometrical effects, we have carried out the same calculations on the experimental geometry (d_{Mn-O} = 1.629 Å). The results reported in Table

Table 8. Main Vertical Transitions (eV) of MnO_4^- with Different Basis Sets and Pseudopotentials[a]

experimental structure			optimized structure		
excitation energy	oscillator strength	assignments	excitation energy	oscillator strength	assignments
		LANL2DZ			
2.77	0.0077	$t_1 \rightarrow e$	3.08	0.0095	$t_1 \rightarrow e$
3.89	0.0007	$t_2 \rightarrow e$	4.31	0.0015	$t_2 \rightarrow e$; $t_1 \rightarrow t_2$
4.41	0.0099	$t_1 \rightarrow t_2$; $t_2 \rightarrow e$	4.84	0.0105	$t_1 \rightarrow t_2$
6.09	0.0044	$t_2 \rightarrow e$; $a_1 \rightarrow t_2$	6.42	0.0016	$t_1 \rightarrow t_2$
6.45	0.0041	$t_1 \rightarrow t_2$	6.66	0.0017	$t_2 \rightarrow e$; $a_1 \rightarrow t_2$
		SBKJC			
2.82	0.0081	$t_1 \rightarrow e$	3.08	0.0097	$t_1 \rightarrow e$
3.95	0.0007	$t_2 \rightarrow e$	4.30	0.0015	$t_2 \rightarrow e$
4.46	0.0106	$t_2 \rightarrow e$	4.83	0.0112	$t_1 \rightarrow t_2$
6.16	0.0055	$a_1 \rightarrow t_2$	6.62	0.0045	$a_1 \rightarrow t_2$; $t_2 \rightarrow e$
			7.28	0.0281	$t_1 \rightarrow t_2$
		SDD			
2.80	0.0077	$t_1 \rightarrow e$	3.13	0.0096	$t_1 \rightarrow e$
3.91	0.0009	$t_2 \rightarrow e$	4.35	0.0021	$t_2 \rightarrow e$
4.42	0.0100	$t_1 \rightarrow t_2$; $t_2 \rightarrow e$	4.89	0.0104	$t_1 \rightarrow t_2$
6.13	0.0079	$a_1 \rightarrow t_2$	6.68	0.0063	$t_2 \rightarrow e$
6.75	0.0051	$t_2 \rightarrow a_1$	6.79	0.0010	$a_1 \rightarrow a_1$; $t_2 \rightarrow a_1$
		DZVP			
2.83	0.0081	$t_1 \rightarrow e$	3.06	0.0095	$t_1 \rightarrow e$
3.94	0.0007	$t_2 \rightarrow e$	4.25	0.0014	$t_2 \rightarrow e$
4.47	0.0108	$t_2 \rightarrow e$	4.80	0.0115	$t_2 \rightarrow e$; $t_1 \rightarrow t_2$
6.15	0.0045	$t_2 \rightarrow e$; $a_1 \rightarrow t_2$	6.59	0.0035	$t_2 \rightarrow e$;$a_1 \rightarrow t_2$
7.31	0.0299	$t_2 \rightarrow a_1$	7.80	0.1575	$e \rightarrow t_2$; $a_1 \rightarrow t_2$
		Watchers + f			
2.84	0.0080	$t_1 \rightarrow e$	3.15	0.0098	$t_1 \rightarrow e$
3.96	0.0010	$t_2 \rightarrow e$	4.36	0.0023	$t_2 \rightarrow e$
4.48	0.105	$t_2 \rightarrow e$	4.92	0.0109	$t_1 \rightarrow t_2$
6.16	0.0048	$a_1 \rightarrow t_2$	6.71	0.0034	$t_2 \rightarrow e$;$a_1 \rightarrow t_2$
7.28	0.0498	$e \rightarrow t_2$	7.73	0.1242	$a_1 \rightarrow t_2$

[a] Both results calculated with the experimental or with the PBE0 optimized structure are listed.

8 show a better agreement with the experimental values, even if the gap is still evident.

Finally, to verify the basis set effects on this difficult case, the only charged species in our set, we have carried out some calculations using the large 6-311++G(3df) basis for O and the LANL2DZ pseudopotential for Mn. Significant variations (around 0.2 eV) have been found, all the transition decreasing of about 0.2 eV. Nevertheless, the errors are still larger than in the other systems, thus clearly showing the predominant role of the multideterminant effects.

To sum up, the permanganate ion is a very complex system. The analysis of results is particularly difficult since values are altered by several parameters. Calculations on the X-ray structure point out the major role of correlation on the structural and spectroscopic parameters as well as the inherent defaults of the TD-DFT approach. When considering optimized structures, these inaccuracies are strengthened by geometrical problems.

4. Conclusion

We have investigated the optical excitations of four classes of transition-metal complexes in the framework of the time dependent DFT with the PBE0 hybrid functional. For each system, excited states, oscillator strengths, and configurations have been analyzed in detail and compared to experimental and theoretical results so as to assess the effect of different basis sets (DZVP and Watchers + f) and pseudopotentials (LANL2DZ, SBKJC, and SDD) for the description of the metal.

The accurate calculation of charge transfer excitations remains a difficult task even though the combination of TD-DFT with hybrid functionals provides an satisfactory avenue to do so. Our results globally feature an absolute mean error in the order of 0.3−0.4 eV for most compounds. Largest discrepancies (up to 0.9 eV) have been found for the permanganate ion. In this case multiconfigurational effects underline a severe limit of the TD-DFT approach. All the same, these values are in the range of what is usually expected with TD-DFT.[72] Corresponding intensities are much less affected and reproduce very well the experimental trends.

The effect of the basis set of the metal atoms turns out to be much more limited, the pseudopotentials being a more precise and convenient alternative to the all electron basis, even for first-row transition metals. Pseudopotentials can hence be recommanded within the TD-DFT/PBE0 route and

represents a useful and fast tool for scientists for clarifying excitation processes of transition-metal complexes.

Supporting Information Available: Geometrical parameters (distances in Å and angles in deg) for optimized structures of CrF_6, MoF_6, WF_6, $Ni(CO)_4$, $CpRh(CO)_2$, $CpIr(CO)_2$, and MnO_4^- (Table SI), excitation energies, oscillator strengths, and main configurations calculated on the PBE0 optimized structure of CrF_6 (SBKJC), MoF_6 (SDD), and WF_6 (LANL2DZ) (Table SII), and excitation energies, oscillator strengths, and main configurations calculated with the SDD pseudopotential on the PBE0 optimized structure of $Ni(CO)_4$ (Table SIII). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Neese, F.; Zaleski, J. M.; Zaleski, K. L.; Solomon, E. I. *J. Am. Chem. Soc*. **2000**, *122*, 11703.

(2) Petit, L.; Adamo, C.; Russo, N. *J. Phys. Chem. B* **2005**, *109*, 12214.

(3) Ciofini, I.; Daul, C. A.; Adamo, C. *J. Phys. Chem. A* **2003**, *107*, 11182.

(4) Koepke, C.; Wisniewski, K.; Grinberg, M.; Russell, D. L.; Holliday, K.; Beall, G. H. *J. Lumin.* **1998**, *78*, 135.

(5) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, Germany, 2000.

(6) Adamo, C.; Barone, V. *Chem. Phys. Lett*. **2000**, *330*, 152.

(7) Parac, M.; Grimme, S. *J. Phys. Chem. A* **2003**, *106*, 6844.

(8) Yamaguchi, Y.; Yokoyama, S.; Mashiko, S. *J. Chem. Phys*. **2002**, *116*, 6541.

(9) Ciofini, I.; Laine, P. P.; Bedioui, F.; Adamo, C. *J. Am. Chem. Soc*. **2004**, *126*, 10763.

(10) Cavillot, V.; Champagne, B. *Chem. Phys. Lett*. **2002**, *354*, 449.

(11) Van Caillie, C.; Amos, R. D. *Chem. Phys. Lett.* **1998**, *291*, 71.

(12) Adamo, C.; Cossi, M.; Scalmani, G.; Barone, V. *Chem. Phys. Lett.* **1999**, 307, 265.

(13) Casida, M. E. In *Recent Advances in Density Functional Methods, Part I*; Chong, D. P., Ed.; World Scientific: Singapore, 1995.

(14) Lainé, P. P.; Ciofini, I.; Ochsenbein, P.; Amouyal, E.; Adamo, C.; Bedioui, F. *Chem. Eur. J*. **2005**, *11*, 3711

(15) Tozer, D. J.; Handy, N. C. *Phys. Chem. Chem. Phys*. **2000**, *2*, 2117.

(16) Dreuw A.; Head-Gordon, M. *J. Am. Chem. Soc*. **2004**, *126*, 4007.

(17) van Leeuwen, R; Baerends, E. J. *Phys. Rev. A* **1994**, 49, 2421

(18) Gritsenko, O.; Baerends, J. *Chem. Phys.* **2004**, *121*, 655.

(19) Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 10180.

(20) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys*. **1999**, *109*, 911.

(21) Adamo, C.; Barone, V. *J. Chem. Phys*. **1999**, *110*, 6158.

(22) Adamo, C.; Barone, V. *Theor. Chem. Acc*. **2000**, *105*, 169.

(23) Aquilante, F.; Cossi, M.; Crescenzi, O.; Scalmani, G.; Barone, V. *Mol. Phys*. **2003**, *101*, 1945.

(24) Adamo, C.; Scuseria, G. E.; Barone, V. *J. Chem. Phys*. **1999**, *111*, 2889.

(25) Adamo, C.; Barone, V. *Chem. Phys. Lett*. **1999,** *314*, 152.

(26) Bernasconi, L.; Blumberger, J.; Sprik, M.; Vuilleumier, R. *J. Chem. Phys*. **2004**, *121* 11885.

(27) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett*. **1996**, *77*, 3865, **1997**, *78*, 1396.

(28) Wiberg, K. B.; De Oliveira, A. E.; Trucks, G. *J. Phys. Chem. A* **2002**, *106*, 4192.

(29) Broclawok, E.; Borowski, T. *Chem. Phys. Lett.* **2001**, *339*, 433.

(30) Wiberg, K. B. *J. Comput. Chem*. **2004**, *25*, 1342.

(31) Monat, J.; Rodriguez, J. H.; McCusker, J. K. *J. Phys. Chem. A.* **2002**, *106*, 7399.

(32) Gaussian 03, Revision B. 05. 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.

(33) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys*. **1998**, *109*, 8128.

(34) Hedberg, L.; Iijima, T.; Hedberg, K. *J. Chem. Phys*. **1979**, *70*, 3224.

(35) Brockway, L. O.; Cross, P. C. *J. Chem. Phys*. **1935**, *3*, 828.

(36) Palenik, G. J. *Inorg. Chem*. **1967**, *6*, 503.

(37) Russo, T. V.; Martin, R. L.; Hay, P. J. *J. Chem. Phys*. **1995**, *102*, 8023.

(38) Gillespie, R. J.; Bytheway, I.; Tang, T. H.; Bader, R. W. *Inorg. Chem*. **1996**, *35*, 3954.

(39) Quinones, G. S.; Hägele, G.; Seppelt, K. *Chem. Eur. J*. **2004**, *10*, 4755.

(40) Marsden, C. J.; Moncrieff, D.; Quelch, G. E. *J. Phys. Chem*. **1994**, *98*, 2038.

(41) Tanpipat, N.; Baker, J. *J. Phys. Chem*. **1996**, *100*, 19818.

(42) Hu, Z.; Boyd, R. J.; Nakatsuji, H. *J. Am. Chem. Soc*. **2002**, *124*, 2664.

(43) Hay, P. J.; Wadt, W. R. *J. Chem. Phys*. **1985**, *82*, 284.

(44) Ortiz, J. V.; Hay, P. J.; Martin, R. L. *J. Am. Chem. Soc.* **1992**, *114*, 2736.

(45) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612.

(46) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866.

(47) Andrae, D.; Haeussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123.

(48) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033.

(49) Bauschlicher, C. W.; Langhoff, S. R., Jr.; Barnes, L. A. *J. Chem. Phys.* **1989**, *91*, 2399.

(50) Chiodo, N. S.; Russo, N.; Sicilia, E. *J. Comput. Chem.* **2005**, *26*, 175.

(51) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, 70, 560.

(52) Daniel, C. *Coord. Chem. Rev.* **2003**, *238*−*239*, 143.

(53) Marsden, C. J.; Wolynec, P. P. *Inorg. Chem.* **1991**, *30*, 1682.

(54) Hope, E. G.; Levason, W.; Ogden, J. S. *Inorg. Chem.* **1991**, *30*, 4873.

(55) Pierloot, K.; Roos, B. O. *Inorg. Chem.* **1992**, *31*, 5353.

(56) Hope, E. G.; Jones, P. J.; Levason, W.; Ogden, J. S.; Tajik, M.; Turff, J. W. *J. Chem. Soc., Dalton Trans.* **1985**, 1443.

(57) McDiarmid, R. *J. Chem. Phys.* **1974**, *61*, 3333.

(58) Nakai, H.; Morita, H.; Tomasello, P.; Nakatsuji, H. *J. Phys. Chem. A* **1998**, *102*, 2033.

(59) Bloor, J. E.; Sherrod, R. E. *J. Am. Chem. Soc.* **1980**, *102*, 4333.

(60) Boulet, P.; Chermette, H.; Daul, C.; Gilardoni, F.; Rogemont, F.; Weber, J.; Zuber, G. *J. Phys. Chem. A* **2001**, *105*, 885.

(61) Schreiner, A. F.; Brown, T. L. *J. Am. Chem. Soc.* **1968**, *90*, 3366.

(62) Lever, A. B. P.; Ozin, G. A.; Halan, A. J. L.; Power, W. J.; Gray, H. B. *Inorg. Chem.* **1979**, *18*, 2088.

(63) Kotzian, M.; Rösch, N.; Schröder, H.; Zerner, M. C. *J. Am. Chem. Soc.* **1989**, *111*, 7687.

(64) Pierloot, K.; Tsokos, E.; Vanquickenborne, L. G. *J. Phys. Chem.* **1996**, *100*, 16545.

(65) Hada, M.; Imai, Y.; Hidaka, M.; Nakatsuji, H. *J. Phys. Chem.* **1995**, *103*, 6993.

(66) Gisbergen, S. J. A.; Groeneveld, J. A.; Rosa, A.; Snijders, J. G.; Baerends, E. J. *J. Phys. Chem. A* **1999**, *103*, 6835.

(67) Rösch, N.; Jörg, H.; Kotzian, M. *J. Chem. Phys.* **1987**, *86*, 4038.

(68) Nooijen, M.; Lotrich, V. *J. Chem. Phys.* **2000**, *113*, 494.

(69) Dunwoody, N.; Lees, A. J. *Organometallics* **1997**, *16*, 5770.

(70) Hoyano, J. K.; Graham, W. A. *J. Am. Chem. Soc.* **1982**, *104*, 3723.

(71) Hoyano, J. K.; McMaster, A. D.; Graham, W. A. *J. Am. Chem. Soc.* **1983**, *105*, 7190.

(72) Purwoko, A. A.; Drolet, D. P.; Lees, A. J. *J. Organomet. Chem.* **1995**, *504*, 107.

(73) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270.

(74) Atananov, M.; Brunold, T. C.; Güdel, H. U.; Daul, D. *Inorg. Chem.* **1998**, *37*, 4789.

(75) Johansen, H.; Rettrup, S. *Mol. Phys.* **1983**, *49*, 1209.

(76) H. Nakai, H.; Ohmori, Y.; Nakatsuji, H. *J. Chem. Phys.* **1991**, *95*, 8287.

(77) Holt, L.; Balhausen, C. *Theor. Chim. Acta* **1967**, *7*, 313.

CT0500500

JCTC Journal of Chemical Theory and Computation

Molecular Dynamics Study of a Surfactant Monolayer Adsorbed at the Air/Water Interface

Jnanojjal Chanda and Sanjoy Bandyopadhyay*

Molecular Modeling Laboratory, Department of Chemistry, Indian Institute of Technology, Kharagpur - 721302, India

Received February 3, 2005

Abstract: A constant volume and temperature (NVT) molecular dynamics (MD) simulation has been carried out to investigate the properties of a monolayer of monododecyl hexaethylene glycol ($C_{12}E_6$) adsorbed at the air/water interface at a surface coverage corresponding to that at its critical micelle concentration (55 Å^2 per molecule). The simulated results have been found to agree reasonably well with available experimental data and with other simulation studies. The study shows that the long polar headgroups of the surfactants are more tilted toward the aqueous layer due to strong interaction between them and water. It has been shown that the surfactant monolayer strongly influences the translational and rotational mobility of interfacial water molecules. A drastic change in the dipolar reorientational motion of water molecules in the aqueous layer is observed with a small variation of distance from the surfactant headgroups.

1. Introduction

The study of the structure and dynamics of organized surfactant assemblies at different interfaces and in bulk solutions is of great importance in various industrial processes, such as detergency, oil recovery, purification, food processing, paints, lubrication, and so forth.[1] Much successful work has been done to understand the properties of surfactant aggregates in solutions,[2−4] but at interfaces, proper systematic studies of adsorbed films are still lacking. This is mainly due to the general lack of suitable experimental techniques capable of probing wet interfaces.

In this work we have focused on the surfactant $C_{12}E_6$, which is a member of an important group of nonionic surfactants, known as the C_mE_n family. These are monoalkyl ethers of the form $C_mH_{2m+1}(OC_2H_4)_nOH$, with a long headgroup consisting of a hydroxy (OH) group followed by a variable number of polar OC_2H_4 (E) groups and a nonpolar C_mH_{2m+1} tail, also of variable length. Polyoxyethylene surfactants (which share the E_n headgroup but may have more complicated hydrophobic tail structures) are the most important nonionic surfactants in commercial use.[1] In particular, the C_mE_n family of surfactants are used in the detergency, cosmetics, and pharmaceutical industries. These surfactants are also highly important from a theoretical point of view, since systematic variation of the values of *m* and *n* alters the properties of the surfactant toward hydrophobicity or hydrophilicity and thus can provide valuable insight into the mechanisms governing self-assembly phenomena in these systems.

A large number of experimental studies have been reported on the properties of different surfactant layers adsorbed at vapor/liquid, liquid/liquid, and liquid/solid interfaces. Richmond and co-workers[5] have studied the structure and orientation of interfacial water molecules using vibrational sum frequency spectroscopy on both cationic and anionic surfactant monolayers at the air/water interface. They have also studied in detail the conformational order of different surfactant monolayers adsorbed at the water/carbon tetrachloride (CCl_4) interface at different surface coverages.[6,7] Grubb et al.[8] have compared the orientation of surfactant molecules at air/water, water/decane, and water/CCl_4 interfaces using second harmonic generation studies. Recently, time-resolved quasielastic laser scattering measurement has been applied to study the dynamics and collective properties of anionic surfactant sodium dodecyl sulfate (SDS) and neutral surfactant Triton X-100 adsorbed at water/nitrobenzene interface.[9] Fluorescence[10] and resonance Raman spec-

* Corresponding author phone: 91-3222-283344; fax: 91-3222-255303; e-mail: sanjoy@chem.iitkgp.ernet.in.

10.1021/ct050019y CCC: $30.25
Published on Web 07/15/2005

tra[11] have also been used in earlier studies on different surfactant monolayers adsorbed at liquid/liquid interfaces. Recently, Linse and co-workers[12] have studied adsorption of several nonionic surfactants at liquid/vapor and solid/liquid interfaces using surface tension and ellipsometric measurements. Only recently, the relatively new technique of neutron reflection has been successfully employed by Thomas and co-workers[13−17] to study the properties of various surfactant layers, including C_mE_n type of surfactants adsorbed at air/liquid, liquid/liquid, and liquid/solid interfaces. Neutron experiments are capable of providing accurate information on the density profiles of the surfactant at an interface. Neutron reflection measurements with isotopic substitution can provide information on structural characteristics of different fragments of the molecules comprising the monolayer.[13,14] Thomas and co-workers[13,14] have studied in detail the adsorption of a series of monododecyl ether surfactants, $C_{12}E_n$, at the air/water interface by varying the number of oxyethylene headgroups between $n = 1$ to $n = 12$. The structure of adsorbed layers were determined by increasing the concentration of the surfactants from a low value to the critical micelle concentration (cmc). It is observed that the alkyl chain thickness remained constant throughout the series, and the value of structural parameters indicate a large average tilt of the surfactant molecules away from the surface normal. These experiments also indicate that the surfactant layers are roughened at the interface with significant intermixing of alkyl chains and oxyethylene headgroups.

The time scales associated with the adsorption of surfactants leading to the formation of a monolayer at the air/water interface or the exchange of monomers between the monolayer and those in the bulk solution are in the range of micro- to milliseconds and therefore are beyond the scope of current generation atomistic molecular dynamics (MD) simulations. However, it is possible to study the microscopic properties of a monolayer already adsorbed at an interface within a reasonable time. This approach has been employed in several simulation studies.[18−32] Tarek et al.[18] have carried out MD studies of cationic surfactant cetyltrimethylammonium bromide (CTAB) adsorbed at the air/water interface at two different concentrations. The simulated density profiles were found to be in good agreement with neutron reflectivity data. Wijmans and Linse[22] have used the Monte Carlo method to study the self-assembly of nonionic surfactants at a hydrophilic surface. Berkowitz and co-workers[19,23,24] have studied in detail the monolayers of ionic surfactants adsorbed at air/water and water/CCl_4 interfaces. They observed a significant difference in the molecular configuration of the surfactants adsorbed at the two interfaces.[19] It was also shown that the polarity of a surfactant has strong influence on the structure and dynamics of water near the interface.[23] Kuhn and Rehage[25−27] have studied in detail the orientation and dynamic properties of monododecyl pentaethylene glycol ($C_{12}E_5$) monolayer adsorbed at air/water and oil/water interfaces. Recently, Rossky and co-workers[28,29] have reported atomistic MD simulation studies on fluorocarbon surfactant monolayers at CO_2/water interface. They noticed that water penetrated the fluorocarbon surfactants to a lesser extent compared to the hydrocarbon surfactants. We have recently studied monolayers of monododecyl diethylene glycol ($C_{12}E_2$) as well as AOT adsorbed at the air/water interface using atomistic MD simulations.[20,21] Only recently, Smit and co-workers[30,31] have used the dissipative particle dynamics method to study the properties of surfactant monolayers at an oil/water interface. They have investigated the effect of the surfactant structure on the bending moduli of the adsorbed monolayers. MD simulations of monolayers of anionic−nonionic surfactant mixtures at a liquid/liquid interface has also been reported recently.[32]

In this work, the microscopic properties of monolayers of monododecyl hexaethylene glycol, $C_{12}E_6$, adsorbed at the air/water interface at a surface coverage corresponding to its cmc (55 Å^2 per molecule) have been studied using MD simulation techniques. The structure and dynamics of the adsorbed monolayer films as well as that of the interfacial water molecules have been investigated in detail at an atomic level resolution. We have organized the article as follows. In the next section we discuss the setup of the system, and the simulation methods employed. This is followed by the results obtained from our investigations and their interpretation. In the last section we summarize the important findings obtained from our study.

2. System Setup and Simulation Details

To study a surfactant monolayer adsorbed at an interface, one has to carefully handle the long-range electrostatic interactions along with a proper choice of boundary conditions. In accordance with our earlier studies,[20,21] we set up the system by placing two monolayers on opposite sides of a slab of aqueous solution, which is thick enough for the two monolayers to remain effectively isolated from each other.

The initial configuration of our constant temperature and volume (NVT) simulation system was set up by arranging a uniform monolayer of 64 surfactants with their hydroxy groups on an appropriate 8×8 square lattice in the *xy* plane with the hydrocarbon chains directed perpendicular to the lattice plane in extended configurations. The lattice constants were chosen to give the surface area per molecule of 55 Å^2, corresponding to the experimentally observed value for adsorption at the air/water interface at cmc.[33] Then two such Langmuir type monolayers were placed, with their hydroxy groups solvated, in the *xy* plane of a roughly 35 Å thick slab of water molecules. The thickness of the aqueous layer was large enough to give a distinct region of bulk solution in the middle of the simulation cell. The overall system contained 128 surfactants and 4133 water molecules. The dimension of the simulation cell in the *x* and *y* directions was 59.4 Å, while the *z* dimension was kept large at ∼200 Å. It is assumed that such large *z* dimension results in negligible interactions between periodic replicas in the *z*-direction.

Initially the surfactant headgroups and the water molecules were kept frozen and a short MD run of 10 ps was performed at a high temperature of 1000 K. This is done to randomize the conformations of the hydrocarbon chains of the surfactants. The surfactant headgroups and the water molecules were then unfrozen, and the temperature of the system was

lowered to 10 K. The temperature was then slowly increased to 298 K over next 50 ps. The resulting configuration was then equilibrated at constant temperature ($T = 298$ K) and volume (NVT) for about 2 ns. This was followed by a production run of another 3.5 ns duration. The trajectories were stored during the production phase of the simulation for subsequent analysis.

The simulations utilized the Nosé-Hoover chain thermostat extended system method[34] as implemented in the PINY-MD simulation code.[35] A recently developed reversible multiple time step algorithm, RESPA,[34] allowed us to employ a 4 fs MD time step. This was achieved using a three-stage force decomposition into intramolecular forces (torsion/bend-bond), short-range intermolecular forces (a 7.0 Å RESPA cutoff distance), and long-range intermolecular forces. Electrostatic interactions were calculated by using the particle mesh Ewald (PME) method.[36] The PME and RESPA were combined following the method suggested by Procacci et al.[37,38] The minimum image convention[39] was employed to calculate the Lennard-Jones interactions and the real-space part of the Ewald sum using a spherical truncation of 7 and 10 Å, respectively, for the short- and the long-range parts of the RESPA decomposition. 'SHAKE/ROLL' and 'RATTLE/ROLL' methods[34] were implemented to constrain all bonds involving hydrogen atoms to their equilibrium values.

The intermolecular potential model was based on pairwise additive site−site electrostatic and Lennard-Jones contributions. The rigid three-site SPC/E model[40] was employed for water. The CH_3 and CH_2 groups of the surfactants were treated as united atoms (i.e. these groups were represented by single interaction sites). The potential parameters for the alkane chain groups were taken from the work of Martin and Siepmann,[41] while the oxyethylene groups were modeled using the OPLS parameters.[42,43] The surfactant chains were made flexible by including bond stretching, bending, and torsion interactions.

3. Results and Discussion

3.1. Structural Properties. The configurations of the system at the beginning (a) and at the end of the simulation (b) are displayed in Figure 1. It is clear from the figure that the monolayer lost its initial ordering during the nanosecond time scale of the simulation. The important notable feature from the figure is that, as the simulation progressed, a significant roughness developed near the surfactant headgroup region of the interface. This is in contrast to an almost flat interface at the beginning. The structural properties of the monolayers can be obtained by studying various distribution functions. In Figure 2(a) we display the average number density profiles (NDP) of different components of the surfactant chains from the center of the simulation cell in a direction normal to the plane of the interface (z). The different components of the surfactants for which the NDPs are computed separately include the terminal hydroxy group (OH), the oxyethylene groups (E), and the hydrocarbon chain atoms (C). The distribution for the water molecules is also computed and plotted. The primary features of the plot are similar to the well-known picture of the surfactant organization at an interface.[18−21,32,44] The surfactant headgroups are hydrated and reside at the interface, while the hydrocarbon tails are excluded from water. Only a fraction of water penetrated within the hydrocarbon tail part of the surfactants. Distribution of a few selected oxyethylene groups (E1, E3, and E6) as displayed in Figure 2(b) shows clearly that almost the entire oxyethylene headgroup region is hydrated. This suggests that the polar nature of the long oxyethylene headgroups helps water molecules penetrate the monolayer film. Finally, the symmetric nature of the density profiles of the two monolayers indicates a well-equilibrated system with an identical overall structure of the two monolayers. The presence of an approximately 25 Å thick slab of bulk water between the two layers indicates that they are independent and have no influence on each other. Further insight into the structure of the surfactant monolayers was obtained from the probability distribution function of the oxyethylene groups (E1−E6) of the surfactants with respect to the normal (z) to the plane of the interface. This is shown in Figure 3 separately for the two monolayers. Again with regard to their average structure the two monolayers are very similar. The relatively broader distributions of the first two oxyethylene groups (E1 and E2) indicate the roughness of the monolayer structure at the interface.

We have calculated the thickness of the adsorbed monolayers from the simulated MD trajectory. The time evolution of the overall thickness (d_s) averaged over both the monolayers is shown in Figure 4. The figure also displays separately the time evolution of the thickness of the oxyethylene (E1−E6) headgroup part (d_h) and the dodecyl hydrocarbon chain (d_t). The presence of sharp oscillations in the plots again indicates the roughness of the adsorbed monolayers. The average thickness values as obtained from our calculations are listed in Table 1. For comparison, the corresponding values obtained from neutron reflection studies[14,33] are also listed in the table. Considering the approximations involved in the models used to obtain the area per molecule and the widths of the profiles from neutron reflectivity measurements, our results agree reasonably well with the experimental data. The accuracy of the force field employed might also be responsible to some extent for the deviations noticed, particularly for the thickness of the hydrophilic headgroup moiety. However, the most interesting feature to note from this result is that the sum of the thicknesses of the two constituent parts of the surfactant molecule is greater than its overall thickness. This suggests that there is an extensive mixing of the head and tail parts of the surfactant chains. The extent of such mixing is in accordance with the density distribution as shown in Figure 2. There appears to be no distinction between the two parts of the surfactants in the adsorbed layer. A similar behavior was observed recently for surfactants with much smaller headgroups ($C_{12}E_2$) adsorbed at an interface.[20]

3.2. Surfactant Orientations. To investigate the orientation of head and tail parts of the surfactants at the interface, we have calculated the orientation of the head and the hydrocarbon tail vectors with respect to the normal (z) to the plane of the interface. The head vector is defined as the vector connecting the hydroxy group and the oxygen atom

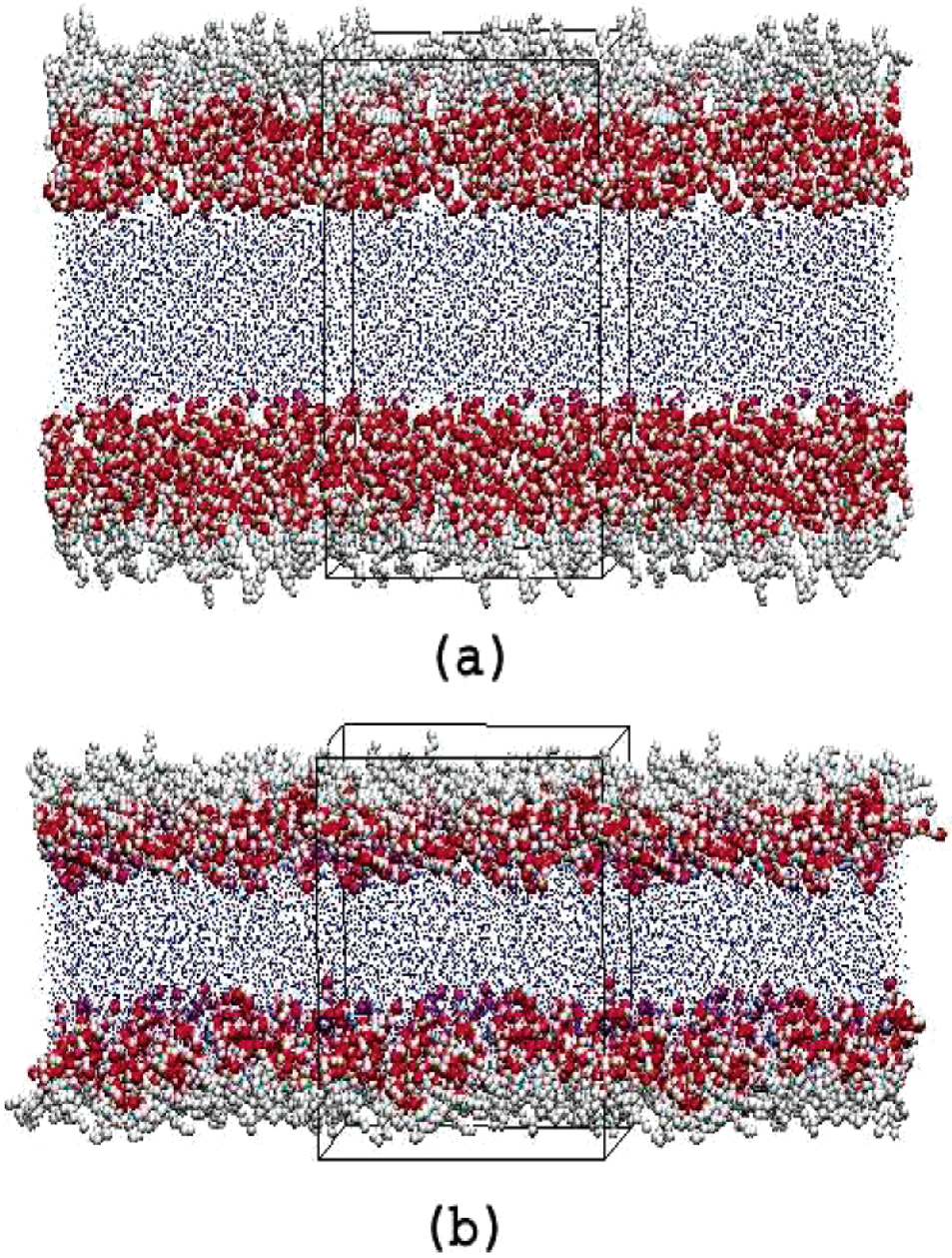


Figure 1. Snapshots of the configuration of the system near the beginning (a) and at the end (b) of the simulation. The oxygen atoms, methyl (CH_3), and methylene (CH_2) groups of the surfactant molecules are drawn as spheres, while only the oxygen atoms of the water molecules are drawn as dots. The atom coloring scheme is O (surfactant), red; CH_3/CH_2, gray; and O (water), blue. For visual clarity, the system is replicated once on both sides of the central simulation cell and the dimension of the cell is marked arbitrarily in the direction normal to the plane of the interface (i.e., along the *z* direction).

of the last oxyethylene group (E6). The vector that connects the first methylene group (CH_2) and the terminal methyl group (CH_3) of the dodecyl chain is defined as the tail vector. The probability distribution, $\rho(\theta)$, of the tilt angle (θ) between the head or tail vector with respect to the normal to the interface (z) is displayed in Figure 5. The distributions are calculated by taking an average over both the monolayers. It is clear from the figure that both the head and tail parts of the surfactant molecules are significantly tilted from the normal (z). The average tilt angle values of the head and tail vectors have been found to be 56.8° and 54°, respectively. This average value of the tilt angle of the dodecyl hydrocarbon chain is comparable with the reported experimental value of 45°.[14] However, the most noticeable feature of the distribution is the presence of a significant fraction (∼13%) of the head vectors oriented with an angle of $\theta \geq 90°$. This suggests that some of the head vectors are reoriented at the interface. Such a reorientation occurs because of a strong interaction between the water molecules and the polar oxyethylene headgroups of the surfactants. Similar orientational bahavior was observed recently with adsorbed $C_{12}E_2$ surfactant monolayers.[20]

3.3. Surfactant and Water Dynamics. To obtain further insight into the microscopic properties of the adsorbed surfactant monolayer, it is crucial to investigate the dynamics of the surfactant molecules as well as the water around them.

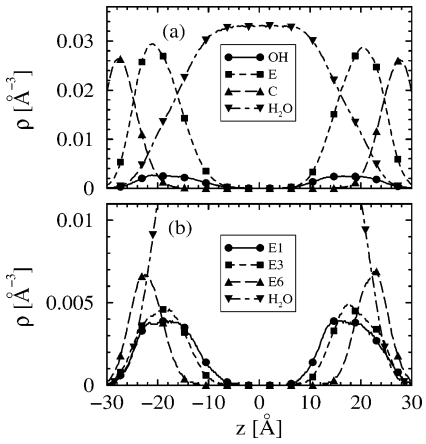


Figure 2. (a) Number density profiles of the hydroxy group (OH), oxyethylene groups (E), hydrocarbon chain atoms (C), and water molecules, measured with respect to the center of the simulation cell in the direction normal to the plane of the interface (i.e., along the *z* direction). The density profiles of a few selected oxyethylene groups are shown in (b). The oxyethylene E groups are numbered starting from the terminal OH group of the surfactant molecule.

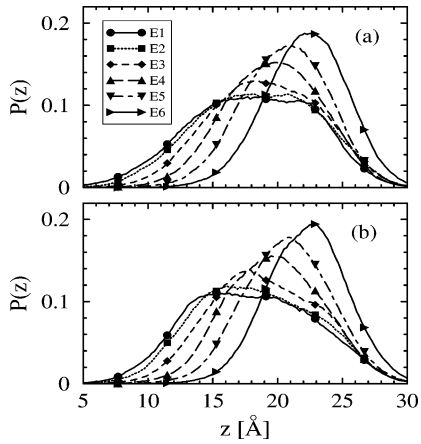


Figure 3. The distribution of the individual oxyethylene groups (E1−E6) of the surfactant molecules along the *z*-direction. The distributions for the two layers are drawn separately in (a) and (b). The numbering scheme of the E groups is the same as in Figure 2.

The dynamical behavior of the surfactants and water molecules are discussed in this section.

3.3.1. Surfactant Dynamics. The translational mobility of the surfactant molecules can be investigated by measuring the mean square displacements (MSD) of the center of mass of the surfactant chains from the MD trajectory. To investigate whether the adsorbed surfactant molecules exhibit any anisotropy in their translational mobility, we have separately measured the center of mass MSDs in the plane of the interface (i.e., *xy* plane) and in the direction normal to it (i.e., along the *z* direction). These are displayed in Figure 6. We notice from the figure that the in-plane motion of the surfactant molecules is higher than the out-of-plane motion. Slightly higher in-plane mobility arises from the lateral rattling motion of the surfactant molecules, which dominates the time scale over which the calculations are carried out.

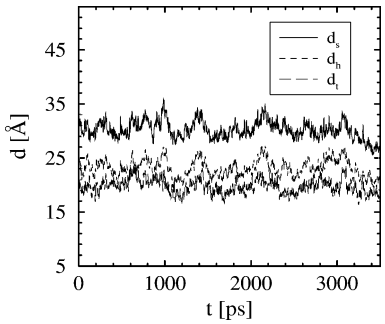


Figure 4. Time evolution of the thickness of the monolayer during the last 3.5 ns of the simulation. The thickness of the whole surfactant chain (d_s) is drawn as a solid line, while that of the oxyethylene headgroups (d_h) and the hydrocarbon tail (d_t) are drawn as broken lines. The calculation is done by averaging over both the monolayers.

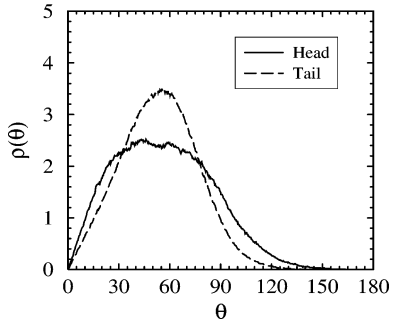


Figure 5. Distribution of the tilt angle, θ (in degrees), of the oxyethylene headgroup (solid line) and the hydrocarbon tail (dashed line) with respect to the normal to the interface, *z*.

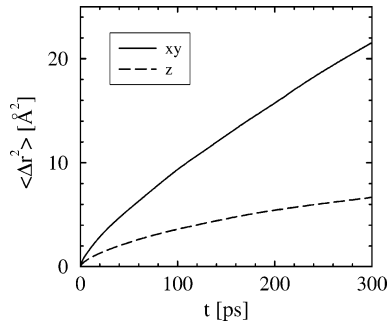


Figure 6. Time evolution of the in-plane (i.e., in *xy* plane) and the out-of-plane (i.e., along *z*) mean square displacements of the center of mass of the surfactant molecules.

Table 1: Comparison of MD and Experimental Values[14,33] of the Oxyethylene Group Thickness (d_h), Dodecyl Chain Thickness (d_t), and the Overall Thickness (d_s) of $C_{12}E_6$ Monolayers Adsorbed at the Air/Water Interface at a Surface Coverage of 55 Å^2 Per Molecule

quantity (Å)	MD	experiment
d_h	23.2 ± 1.0	19.5 ± 1.0
d_t	19.7 ± 0.9	19.0 ± 1.0
d_s	30.3 ± 1.3	26.5 ± 2.0

The out-of-plane protrusion motion of the surfactants is more restricted. Such restricted dynamics of the surfactant molecules is a signature of anomalous sublinear diffusion in heterogeneous anisotropic or confined systems.[45,46] To

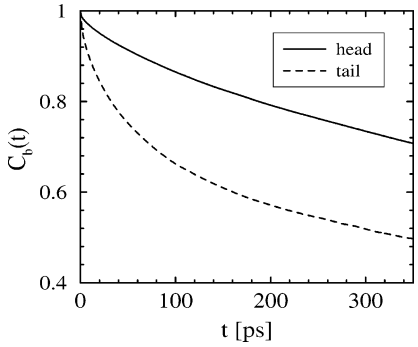


Figure 7. Reorientational time autocorrelation functions for the oxyethylene head vectors and the hydrocarbon tail vectors of the surfactant molecules.

investigate the extent of anomalous diffusion, if any, the curves in Figure 6 are fitted to a law

$$\langle \Delta r^2 \rangle \simeq t^{\alpha} \tag{1}$$

where the exponent α is expected to be smaller than one. The estimated values of α have been found to be 0.53 and 0.32 for the in-plane and out-of-plane motions, respectively. Such small α values clearly show the presence of significant anomaly in the diffusion behavior of surfactant molecules within the time scale of our simulation.

The rotational motion of the surfactant molecules has been investigated by measuring the reorientational dynamics of the head and tail vectors (as defined earlier) of the individual monomers. The reorientational motion has been studied by measuring the time correlation function (TCF), $C_b(t)$, defined as

$$C_b(t) = \frac{\langle \hat{b}_i(t+\tau).\hat{b}_i(\tau)\rangle}{\langle \hat{b}_i(\tau).\hat{b}_i(\tau)\rangle} \tag{2}$$

where $\hat{b}_i(t)$ represents the unit vector corresponding to the head or tail of the ith surfactant molecule at time t, and the angular brackets denote averaging over the surfactant molecules and over initial times τ. The variation of $C_b(t)$ against time has been displayed in Figure 7. It is clearly evident from the figure that the headgroups of the surfactant molecules reorient much slowly than the hydrocarbon tails. Such slow relaxation of $C_b(t)$ for the headgroups arises from the strong interaction between the interfacial water molecules and the oxyethylene groups, as mentioned before. The hydrocarbon tails of the surfactants being away from the interface are more flexible in nature and thus reorient faster.

3.3.2. Water Dynamics. The dynamics of water near the interface plays a crucial role in determining the behavior of surfactant aggregates. The water dynamics near the surface of organized molecular assemblies such as micelles[47−49] and heterogeneous macromolecules such as proteins[50] is well studied. However, the dynamics of water near the surface of an adsorbed surfactant film has not been explored much. In this section we study the translational and reorientational motion of water near the interface.

3.3.3. Translational Motion of the Interfacial Water. The mean square displacements of water molecules at different distances from the interface have been calculated. To be

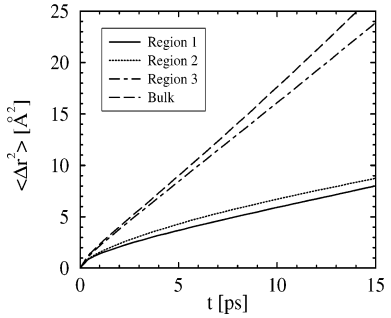


Figure 8. Mean square displacement (MSD) of the water molecules in different regions of the aqueous layer. Region 1 comprises water molecules which are within 4 Å from the hydroxy group oxygen atoms, those which are within 4−7 Å are in region 2, while those which are beyond 7 Å are said to be in region 3. The MSD of water molecules in pure bulk water is also shown for comparison.

Table 2: Diffusion Coefficient of the Water Molecules in Different Regions as Obtained from Their Mean Square Displacements[a]

region	D (10^{-5} cm^2 s^{-1})
region 1 (0−4 Å)	0.72
region 2 (4−7 Å)	0.73
region 3 (>7 Å)	2.57
bulk water	2.84

[a] The value for bulk water is also listed for comparison.

specific, we have defined three hydration regions. Water molecules which are within a distance of 4 Å from the hydroxy group oxygen atoms of the surfactant headgroups form region 1. This essentially corresponds to the first hydration layer with respect to the hydroxy group oxygen atoms of the surfactants. Region 2 comprises water molecules within 4−7 Å from the hydroxy group oxygen atoms, while those beyond 7 Å are considered to be in region 3. The distances are measured with respect to both the monolayers by tagging the water molecules at different time origins. The plots are displayed in Figure 8. For comparison, we have also displayed the same for pure bulk water, which was obtained from a MD simulation of pure SPC/E water at room temperature. It is evident from the figure that at a distance close to the surfactant headgroups (region 1) the translational mobility of the water molecules is significantly restricted. The mobility of water molecules increases as the distance from the surfactant headgroups is increased. Within a short distance of approximately 7 Å from the interface (region 3), the mobility of the water molecules approaches close to that of pure bulk water. Such restricted mobility of water near the interface of organized molecular assemblies, such as micelles as well as surfactant aggregates at an interface, has been studies recently.[21,47−49] This might arise because of the strong interaction between the hydrophilic oxyethylene headgroups of the surfactants and the surrounding water molecules. The diffusion coefficients of water molecules in the three regions have been calculated from a linear fit to the corresponding mean square displacements, which are listed in Table 2. The data show that water in regions 1 and 2 are much less mobile than those in region 3 and pure bulk water. In fact, the water in region 3 exhibits three times faster

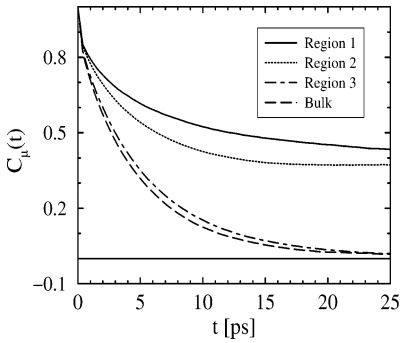


Figure 9. Reorientational time correlation function (TCF) of the water dipoles, $C_\mu(t)$, for the water molecules in three regions of the aqueous layer. The definitions of the regions are same as in Figure 8. The TCF for pure bulk water is also shown for comparison.

dynamics than those in region 1. Such an approach to calculate the diffusion coefficients of interfacial water molecules from MD simulations has been used earlier.[21,49] It may be noted that the absolute values of the calculated diffusion coefficients may not be very authentic in this case. This is because the surfactant monolayers provide a heterogeneous anisotropic environment for the water molecules, which is not strictly three-dimensional.[45,46] However, here we are more interested in comparing the relative diffusion of water in different regions close to the interface, rather than measuring the absolute diffusion coefficient values.

3.3.4. Orientational Motion of the Interfacial Water. It is expected that the rotational motion of the water molecules close to the surfactants at the interface will also be affected. The rotational motion of water can be investigated by measuring the reorientational dynamics of its electrical dipole $\vec{\mu}$, defined as the vector connecting the oxygen atom of the water molecule to the center of the line connecting the two hydrogen atoms. The time evolution of $\vec{\mu}$ can be estimated by measuring the dipole−dipole time correlation function (TCF), defined as

$$C_\mu(t) = \frac{\langle \hat{\mu}_i(t+\tau).\hat{\mu}_i(\tau)\rangle}{\langle \hat{\mu}_i(\tau).\hat{\mu}_i(\tau)\rangle} \qquad (3)$$

where $\hat{\mu}_i(t)$ is the unit dipole moment vector of the ith water molecule at a time t, and the angular brackets denote averaging over the tagged water molecules and over initial times τ. Again, we restrict ourselves to the reorientational motion of water molecules in the three regions as discussed before. The correlation functions were calculated by averaging over these water molecules only, which are shown in Figure 9. For comparison, the decay curve for pure bulk water is also displayed. It is clear from the figure that the relaxation curves decay nonexponentially, and for water close to the surfactants (regions 1 and 2) they do not decay to zero. This shows that the presence of the surfactants significantly restricts the reorientational motion of water. However, similar to translational mobility, the rotational motion of the water molecules too approach that of pure bulk water within a short distance of about 7 Å (region 3). The slowing down of the rotational dynamics of interfacial water has been shown recently in an adsorbed monolayer of anionic surfactants, AOT.[21] Thus the present investigation in conjunction with our earlier work establishes the fact that the restricted dynamics of the interfacial water in an adsorbed surfactant monolayer is independent of the type of the surfactant molecules. To obtain a quantitative estimation of the time scales associated with the reorientational motion of the water molecules from the dipolar correlation functions, we have fitted the correlation functions with multiexponentials. It is a common practice to use multiexponentials, as one can then directly obtain time constants associated with different motions.[21,47] These time constants can be assigned to different relaxation processes of the system. Here we have used a sum of three exponentials to fit the data for each of the three TCFs. The parameters for best fit are shown in Table 3. The average reorientational time constant ($\langle\tau_\mu\rangle$) values clearly indicate that the rotational motion of water close to the surfactant headgroups is severely restricted. The slow decay for water close to the interface arises because of their strong interaction and formation of hydrogen bonds with the polar oxyethylene headgroups of the surfactants. As the dynamics of the individual surfactant chains in the adsorbed monolayer is much slower compared to the time scale of water dynamics, such strong surfactant−water interaction slows down the reorientational motion of these water molecules *bound* to the interface. Interestingly, an extraordinary slow decay with a long time component (nearly flat decay curve), which we could not determine because of the limitation of the time scale of our simulation has been observed for water molecules in region 2. It has been found that the unusually slow decay for water in region 2 arises due to the presence of a small fraction of water molecules which are pulled into the monolayer by the long polar oxyethylene headgroups and remain trapped for long times. It may be noted that the $\langle\tau_\mu\rangle$ value for water in region 3 approaches that for pure bulk water and is about 13 times smaller than that for water in region 1. Such drastic changes in the reorientational motion of water molecules with a slight variation of distance from the surfactant headgroups of an adsorbed monolayer has also been observed recently.[21]

Table 3: Multiexponential Fitting Parameters for the Dipolar Time Correlation Functions of Water Molecules in Different Regions[a]

region	time constant (ps)	amplitude (%)	$\langle\tau_\mu\rangle$ (ps)
region 1 (0−4 Å)	0.24	14.5	66.13
	4.57	32.7	
	122.26	52.8	
region 2 (4−7 Å)	0.19	13.0	-
	4.73	50.5	
	>500	36.5	
region 3 (>7 Å)	0.18	12.5	5.12
	4.29	54.4	
	8.38	33.1	
bulk water	0.22	13.8	4.73
	4.68	80.6	
	16.63	5.6	

[a] $\langle\tau_\mu\rangle$ is the average time constant. Corresponding parameters for bulk water are also listed for comparison.

4. Conclusion

In this article, we have presented results obtained from an atomistic MD simulation study of the structure and dynamical properties of adsorbed monolayers of the nonionic surfactant $C_{12}E_6$ adsorbed at the air/water interface. The simulation has been performed at a constant volume and at room temperature with a surface coverage of 55 Å^2 per molecule, which corresponds to the surface coverage of the surfactant at cmc. The structural properties of the adsorbed monolayer, such as the number density profiles of different components normal to the plane of the interface (i.e., along the *z* direction) and the thickness of the monolayer (d_h, d_t), are calculated. It is observed that significant roughness developed at the interface during the time scale of the simulation with extensive mixing of the head and tail segments of the surfactant chains. The calculated thickness of the adsorbed layer has been found to agree reasonably well with experimental results.[14,33] It is found that there is a significant fraction of the surfactants with their headgroups reoriented within the aqueous layer close to the interface. This occurs due to strong interaction between the long polar headgroups of the surfactants and the interfacial water molecules. Such unique orientational behavior was observed earlier for this class of surfactants with smaller headgroups, namely, $C_{12}E_2$.[20] Such a strong interaction with water is also reflected in slower reorientational motion of the surfactant headgroups. We observed significant anomaly in the diffusion behavior of the surfactant molecules within the time scale of the simulation.

The dynamics of the interfacial water molecules are also investigated in detail. Both translational and reorientational motions of the water molecules have been found to be restricted near the interface. An important observation from this study has been the drastic increase in both translational and rotational motion of water molecules with a small increase in distance from the surfactant headgroups. We noticed that the average relaxation time for the reorientational motion of water in different regions of the aqueous layer can drastically vary with distance from the surfactant headgroups.

It is important to obtain a further microscopic level understanding of the interaction between the surfactants and the interfacial water molecules. In particular, it would be interesting to study the formation and breaking of hydrogen bonds and their lifetimes between the water molecules and the surfactant headgroups. This will help obtaining a greater molecular level understanding of the properties of surfactant monolayers adsorbed at an interface. Some of these aspects are presently under investigation in our laboratory.

Acknowledgment. This work was supported in part by generous grants from the Council of Scientific and Industrial Research (CSIR) and the Department of Science and Technology (DST), Government of India. One of us (J.C.) thanks Sudip Chakraborty for many useful discussions.

References

(1) Laughlin, R. G. *The Aqueous Phase Behavior of Surfactants*; Academic Press: New York, 1994.

(2) Funari, S. S.; Rapp, G. *J. Phys. Chem. B* **1997**, *101*, 732−739.

(3) Briganti, G.; Segre, A. L.; Capitani, D.; Casieri, C.; Mesa, C. L. *J. Phys. Chem. B* **1999**, *103*, 825−830.

(4) Capitani, D.; Casieri, C.; Briganti, G.; Mesa, C. L.; Segre, A. L. *J. Phys. Chem. B* **1999**, *103*, 6088−6095.

(5) Gragson, D. E.; McCarty, B. M.; Richmond, G. L. *J. Phys. Chem.* **1996**, *100*, 14272−14275.

(6) Conboy, J. C.; Messmer, M. C.; Richmond, G. L. *J. Phys. Chem. B* **1997**, *101*, 6724−6733.

(7) Conboy, J. C.; Messmer, M. C.; Richmond, G. L. *Langmuir* **1998**, *14*, 6722−6727.

(8) Grubb, S. G.; Kim, M. W.; Raising, T.; Shen, Y. R. *Langmuir* **1988**, *4*, 452−454.

(9) Zhang, Z. H.; Tsuyumoto, I.; Kitamori, T.; Sawada, T. *J. Phys. Chem. B* **1998**, *102*, 10284−10287.

(10) Piasecki, D. A.; Wirth, M. J. *J. Phys. Chem.* **1993**, *97*, 7700−7705.

(11) Tian, Y.; Umemura, J.; Takenaka, T.; Kunitake, T. *Langmuir* **1988**, *4*, 1064−1066.

(12) Kjellin, U. R. M.; Claesson, P. M.; Linse, P. *Langmuir* **2002**, *18*, 6745−6753.

(13) Lu, J. R.; Li, Z. X.; Thomas, R. K.; Staples, E. J.; Thompson, L.; Tucker, I.; Penfold, J. *J. Phys. Chem.* **1994**, *98*, 6559−6567.

(14) Lu, J. R.; Su, T. J.; Li, Z. X.; Thomas, R. K.; Staples, E. J.; Tucker, I.; Penfold, J. *J. Phys. Chem. B* **1997**, *101*, 10332−10339.

(15) Penfold, J.; Staples, E. J.; Tucker, I.; Thomas, R. K. *J. Colloid Interface Sci.* **1998**, *201*, 223−232.

(16) Lu, J. R.; Li, Z. X.; Thomas, R. K.; Binks, B. P.; Crichton, D.; Fletcher, P. D. I.; McNab, J. R.; Penfold, J. *J. Phys. Chem. B* **1998**, *102*, 5785−5793.

(17) Penfold, J.; Thomas, R. K. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2648−2652.

(18) Tarek, M.; Tobias, D. J.; Klein, M. L. *J. Phys. Chem.* **1995**, *99*, 1393−1402.

(19) Schweighofer, K. J.; Essmann, U.; Berkowitz, M. *J. Phys. Chem. B* **1997**, *101*, 3793−3799.

(20) Bandyopadhyay, S.; Chanda, J. *Langmuir* **2003**, *19*, 10443−10448.

(21) Chanda, J.; Chakraborty, S.; Bandyopadhyay, S. *J. Phys. Chem. B* **2005**, *109*, 471−479.

(22) Wijmans, C. M.; Linse, P. *J. Phys. Chem.* **1996**, *100*, 12583−12591.

(23) Schweighofer, K. J.; Essmann, U.; Berkowitz, M. *J. Phys. Chem. B* **1997**, *101*, 10775−10780.

(24) Dominguez, H.; Berkowitz, M. L. *J. Phys. Chem. B* **2000**, *104*, 5302−5308.

(25) Kuhn, H.; Rehage, H. *J. Phys. Chem. B* **1999**, *103*, 8493−8501.

(26) Kuhn, H.; Rehage, H. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1023−1028.

(27) Kuhn, H.; Rehage, H. *Colloid Polym. Sci.* **2000**, *278*, 114−118.

(28) da Rocha, S. R. P.; Johnston, K. P.; Rossky, P. J. *J. Phys. Chem. B* **2002**, *106*, 13250−13261.

(29) Stone, M. T.; da Rocha, S. R. P.; Rossky, P. J.; Johnston, K. P. *J. Phys. Chem. B* **2003**, *107*, 10185−10192.

(30) Rekvig, L.; Hafskjold, B.; Smit, B. *J. Chem. Phys.* **2004**, *120*, 4897−4905.

(31) Rekvig, L.; Hafskjold, B.; Smit, B. *Phys. Rev. Lett.* **2004**, *92*, 116101.

(32) Dominguez, H. *J. Phys. Chem. B* **2002**, *106*, 5915−5924.

(33) Lu, J. R.; Li, Z. X.; Thomas, R. K.; Staples, E. J.; Tucker, I.; Penfold, J. *J. Phys. Chem.* **1993**, *97*, 8012−8020.

(34) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117−1157.

(35) Tuckerman, M. E.; Yarne, D. A.; Samuelson, S. O.; Hughs, A. L.; Martyna, G. J. *Comput. Phys. Commun.* **2000**, *128*, 333−376.

(36) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(37) Procacci, P.; Darden, T.; Marchi, M. *J. Phys. Chem.* **1996**, *100*, 10464−10468.

(38) Procacci, P.; Marchi, M.; Martyna, G. J. *J. Chem. Phys.* **1998**, *108*, 8799−8803.

(39) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids;* Clarendon: Oxford, 1987.

(40) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.

(41) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.

(42) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276−1284.

(43) Briggs, J. M.; Matsui, T.; Jorgensen, W. L. *J. Comput. Chem.* **1990**, *11*, 958−971.

(44) Tu, K.; Tobias, D. J.; Blasie, J. K.; Klein, M. L. *Biophys. J.* **1996**,*70*, 595−608.

(45) Liu, P.; Harder, E.; Berne, B. J. *J. Phys. Chem. B* **2004**, *108*, 6595−6602.

(46) Christensen, M.; Pedersen, J. B. *J. Chem. Phys.* **2003**, *119*, 5171−5175.

(47) Balasubramanian, S.; Bagchi, B. *J. Phys. Chem. B* **2002**, *106*, 3668−3672.

(48) Balasubramanian, S.; Pal, S.; Bagchi, B. *Phys. Rev. Lett.* **2002**, *89*, 115505.

(49) Faeder, J.; Ladanyi, B. M. *J. Phys. Chem. B* **2000**, *104*, 1033−1046.

(50) Bandyopadhyay, S.; Chakraborty, S.; Balasubramanian, S.; Pal, S.; Bagchi, B. *J. Phys. Chem. B* **2004**, *108*, 12608−12616.

CT050019Y

JCTC Journal of Chemical Theory and Computation

The Interaction of Coinage Metal Clusters with the MgO(100) Surface

Giovanni Barcaro and Alessandro Fortunelli*

Molecular Modeling Laboratory, Istituto per i Processi Chimico-Fisici (IPCF) del C. N. R., via V. Alfieri 1, 56010 Ghezzano (PI), Italy

Received March 22, 2005

Abstract: The results of a systematic study of the interaction of small coinage metal clusters (M_n, $n = 1-3$) and extended deposition (one and two MLs; ML = monolayer) with the regular and locally defected (F_s center and divacancy) neutral MgO(100) surface are presented. The calculations have been performed at the DFT level employing plane waves as a basis set and using a gradient-corrected exchange-correlation functional (PW91). The adhesion energy along the group follows a trend that can be rationalized in terms of the strength and "stickiness" of the metallic bond, electrostatic polarization effects, and chemical interactions. Coinage metal dimers and trimers are absorbed on the regular surface in an upright position with little modification with respect to the gas-phase structure and can easily diffuse from site to site (in the case of trimers, also because of their fluxional character). In the case of extended deposition, the adhesion energy increases when passing from one to two MLs because of a "metal-on-top" stabilization mechanism. Neutral localized defects on the surface such as the F_s center (generated by a missing O atom) and the double vacancy (generated by a missing MgO dimer) act as strong trapping centers for small clusters and remarkably increase the adhesion of metal slabs to the surface in the case of extended deposition. At variance with the F_s center, the double vacancy induces a strong structural and energetic modification of the surrounding oxide lattice, varying also as a function of the metal deposition. A peculiar structural rearrangement consisting of the segregation of the metal slab into "islands" on the surface is observed in the case of one ML Cu and (to a lesser extent) Au.

1. Introduction

Transition metal nanoclusters have gained increasing attention in science and application in the past several years as a result of the observation of properties that are unique to the nanoscale domain.[1−4] From a technological point of view, the process of stabilization of the metal clusters by coating with surfactants or through absorption on a substrate is an essential step to exploit the many potential applications of these materials. For this reason, a considerable experimental and theoretical effort has been dedicated to the study of the interaction and structural modifications induced in the metal aggregate by the presence of a supporting substrate[5−29] or surfactant and solvent molecules.[33−35] Moreover, the experimental conditions of the stabilization process can be tuned in order to induce the formation of supraorganized structures of single nanocluster units.[4,34−36] From a theoretical point of view, the characterization of transition metal clusters absorbed upon a supporting substrate has mainly concentrated on the study of static structures and has rarely dealt with the mechanism of nucleation and growth (see, however, refs 29−32, 100). Furthermore, despite the many contributions to this field, previous theoretical work lacks a systematic homogeneous comparison between elements belonging to the same group. The aim of the present paper is to provide such a comparison for the metals belonging to the group IB and the MgO(100) surface. Attention will be concentrated on the interaction of small coinage metal clusters and extended

* Corresponding author. Tel: +39-050-3152447, e-mail: fortunelli@ipcf.cnr.it.

10.1021/ct050073e CCC: $30.25
Published on Web 08/12/2005

deposition with the regular or localized defected neutral MgO(100) surface. These systems present several elements of interest. On one hand, nanoclusters formed by coinage metals (Cu, Ag, and Au), both mono- and bimetallic, exhibit interesting optical, electronic transport, and catalytic properties, both in the gas phase and when stabilized by a substrate or by surfactant molecules.[33,37−44] On the other hand, the MgO(100) surface both is experimentally well-characterized and has technological (e.g., catalytic) applications.[3,45] Moreover, its theoretical description is simplified by the fact that this surface does not present the complications associated with surface reconstruction, MgO being a simple ionic solid and the (100) surface being a nonpolar one. In discussing the results, we will highlight analogies and differences in the behaviors of the coinage metals in order to extract information on both static and dynamic properties of the metal/MgO interaction. These results can form the basis for (a) a comparison with experimental data on the growth of coinage metal clusters and thin films, (b) building semiempirical metal/oxide interaction potential, which can be used in the simulation of actual growth processes, and (c) improving other simplified (for example, thermodynamic) theoretical models.

Previous Work. During the past few years, several theoretical approaches have been used to study transition metal nanoclusters. First principles calculations have been performed on small metal aggregates in order to determine the minimum energy structure and structure/property relationship as a function of cluster size.[46−55] Numerical procedures based on genetic algorithms for single and bimetallic structures using atom−atom pair potentials[56] or many-body potentials[56−64] have been used to investigate a broader range of cluster sizes. Focusing on transition metal clusters supported on magnesia surfaces, in the past 10 years, a wealth of results has been collected on the interaction of single metal atoms on the regular and defected sites of the MgO(100) surface.[65−69] The extended and localized defected sites of the MgO surface have attracted remarkable attention because of their intrinsic chemical and physical properties[70−84] and because they can act as trapping centers for metal atoms absorbed on the oxide surface.[67,85] The results of such calculations have demonstrated that the choice of the model influences the description of the system[86] and that the metal/surface interaction changes primarily as a function of the electronic configuration and the dimensions of the metal species considered. For example, the coinage metals, which are characterized by a filled d shell plus an outer, unpaired s electron, are weakly bound to the surface (with adhesion energy less of 1 eV) because of the Pauli repulsion with the electron density of the oxide.[67] The catalytic properties of systems formed by a single atom or small clusters absorbed on the regular and defected MgO surfaces have been studied by monitoring the absorption of probe molecules on the metal atoms.[39,78,87−92] The real-time monitoring of cluster growth on MgO offered by GISAXS[9] has further promoted the study of the minimum structure configurations of metal clusters and extended deposition on regular and defected MgO(100) surfaces in order to extract information on the mechanism of nucleation and growth and on the modifications in structure and properties induced in the metal aggregate by absorption on the oxide surface.[42,65,93−102] In particular, the study of the small clusters and extended deposition of coinage metals (Cu and Ag) has been performed through standard DFT calculations,[103−109] Car Parrinello calculations,[29−32] and a thermodynamic model describing the metal/oxide system in terms of a solid solution.[83,110] The general conclusion is that the Cu and Ag interactions with the regular surface are weak. Small clusters are absorbed in an upright position with negligible modifications with respect to the gas-phase structure and are able to diffuse over the surface through leapfrog, twisting, and rolling movements. In the case of extended deposition, there is an increase of the adhesion energy to the surface when passing from one to two MLs, thanks to a polarization stabilization mechanism. Classical molecular dynamics calculations of the growth process have, until now, been performed only for Pd clusters absorbed on the regular MgO surface[111−113] thanks to the development of a specific metal/oxide potential, fitted on the results of first-principle calculations. An analogous potential is still lacking for the coinage metal clusters. One of our goals is to provide the theoretical input data for the construction of such a potential, accounting not only for the interaction with the regular surface sites but also for the interaction with localized neutral vacancies.

2. Computational Details

All the calculations reported here were performed at the DFT level employing the PWscf (Plane-Wave Self-Consistent Field) computational code[114]. For each atom, the interaction between the outer-shell electrons and the positive nucleus is shielded by a pseudopotential accounting for the influence of the inner-shell electrons. All the calculations were performed spin-unrestricted. The PW91 exchange-correlation functional was employed,[115] which is a gradient-corrected density functional. From previous experience, we do not expect qualitative changes when using other gradient-corrected density functionals. However, it has been shown (see, e.g., ref 86) that the use of hybrid density functionals produces an overall (roughly constant) reduction in the metal/surface interaction energies.

The geometry optimizations were stopped when maximum force on the atoms was less than 4×10^{-4} au. All of the configurations reported here were checked to insure that they were local minima in the energy hypersurface. Few saddle point configurations can be found among those reported in the Supporting Information.

The dimensions of the plane-wave basis set are determined by the kinetic energy cutoff, which was fixed at the value of 40 Ry for all of the calculations. The *k*-point sampling of the Brillouin zone was changed according to the conductive properties of the metal deposition absorbed on the oxide. In the case of the deposition of one and two ML metallic slabs, which are conductive (small energy gap between the HOMO and LUMO energy), it was necessary to increase the number of *k* points per unit cell and to apply a procedure of Gaussian smearing (a value of 0.002 Ry was chosen as the smearing parameter for every sampled point of the Brillouin zone).

The dimensions of the unit cell were modified according to the metal deposition, as specified in detail in the following sections. A common aspect of all the different cells employed is that the oxide surface was built by positioning the Mg and O atoms at the equilibrium lattice positions of the MgO rock-salt structure (isomorphic with NaCl) and cutting this bulk structure along 100 directions. The O atoms on the surface are surrounded by four Mg surface atoms plus one Mg bulk atom and vice versa. The distance between the two opposite surface layers of the oxide slab is about 10 Å. The lattice constant of the oxide corresponded to the experimental value of 4.208 Å (implying a Mg−O distance of 2.104 Å). When the oxide bulk structure is relaxed at the DFT level, the equilibrium lattice constant is overestimated by about 2%. However, the metal/surface interaction energies do not qualitatively vary when using the DFT-optimized lattice constant (unpublished results).

In the calculations involving the regular (undefective) oxide surface, the MgO structure was left frozen and only the atomic coordinates of the metallic atoms absorbed on the surface were optimized. In the case of the defected surfaces, we have followed two different approaches. In the case of a single F_s vacancy, we have relaxed the oxide lattice around the defect in the absence of metal atoms; such a structure has then been frozen, allowing only the relaxation of the metal atoms absorbed upon it. We have checked in test cases that there is a negligible difference between the results obtained following this approach and the results obtained allowing a further relaxation of the oxide surface upon metal deposition. In the case of the double vacancy (which can be thought of as a desorption of a neutral MgO dimer in the 100 direction), the relaxation plays a very important role,[100] and to obtain an accurate estimate of the metal absorption energies, we allowed a complete relaxation of the atoms of the defected oxide and all the metal atoms. For this reason, the calculations involving the double vacancy are, computationally, rather intensive.

3. Results and Discussions

This section is divided into three subsections: absorption (I) on the regular terrace of the MgO(100) surface, (II) on a neutral oxygen vacancy (F_s center), and (III) on a neutral divacancy

Each of these subsections will separately treat the case of the absorption of (A) isolated metal atoms, (B) metal dimers, (C) metal trimers, and (D) extended depositions (one and two MLs)

For each system, we report four values of energy:

(1) The Absorption Energy (E_{int}). This quantity is calculated by subtracting the ground-state energy of the isolated oxide and the isolated metal atoms from the value of the total energy of the system. The absorption energy corresponds to the energy gain achieved by moving the single cluster constituents from infinite distance to the equilibrium distance characterizing a particular configuration.

(2) The Adhesion Energy (E_{adh}). This quantity is calculated by subtracting the energy of the oxide surface and that of the metal cluster, both frozen in the interacting configuration, from the value of the total energy of the system.

(3) The Oxide Surface Distortion Energy (ΔE_{MgO}). This quantity corresponds to the "distortion energy" of the oxide surface. It is calculated by subtracting the energy of the oxide surface in the interacting configuration from the ground-state energy of the surface.

(4) The Metal Cluster Distortion Energy (ΔE_{met}). A quantity analogous to the previous one but for the metal cluster.

We report only a selection of the configurations that have been considered. Additional data can be found in the Supporting Information.

I. Interaction of Coinage Metal Clusters with the Regular Terrace of the MgO(100) Surface. As anticipated before, for this system, no relaxation of the MgO surface was allowed ($\Delta E_{MgO} = 0$). This is consistent with the experimental observation that this surface does not reconstruct. We optimized only the coordinates of the metal atoms because the surface structure modification induced by the absorbed metal is negligible.

I.A. Isolated Adatoms. The absorption of the isolated adatoms has been studied modeling the (100) terrace by a three-layer slab, each containing four Mg atoms and four O atoms (a five-layer-slab model gave essentially identical results). The minimum distance between metal atoms in neighboring unit cells was larger than 6 Å; this distance is sufficiently large to make the interaction between the metal atoms negligible.

Three characteristic sites of absorption have been considered: (1) the O^{2-} ion, (2) the Mg^{2+} ion, and (3) the hollow site (see Figure 1a). The results of the interaction energies are reported in Table 1. In Figure 1 of the Supporting Information, the absorption energies on the O^{2-} ion are reported for the three metals as functions of the metal−surface distance. The interaction is stronger for Cu and Au atoms, while it is weaker for Ag;[68] all three metals interact more strongly with the O^{2-} ion site of the surface, while the Mg^{2+} ion is the site characterized by the smallest value of adhesion energy.

The tendency of absorbing atop the O^{2-} site is a common trend among transition metal atoms.[32,65,66] The stabilizing contributions of the interaction can be of three types: (1) chemical bonding, (2) electrostatic stabilization, and (3) dispersion interactions. The Pauli repulsion between the metal cloud and the electronic density of the surface represents, instead, the mainly destabilizing contribution. This repulsion is certainly stronger for the coinage metal atoms that are characterized by a diffuse unpaired s electron, and this gives a rough explanation of the smaller adhesion energies of Cu, Ag, and Au with respect to other transition metal atoms such as Pd or Ni.[67] It is still under debate whether chemical bonding plays a significant role in the adhesion of the coinage metals; further studies need to be carried out to clarify this point.

I.B. Dimers. The formation of a dimer can be viewed as the first step in the process of the growth of metal clusters on the oxide surface. The unit cell was chosen, consisting of a three-layer MgO (each layer containing nine O atoms and nine Mg atoms) and two metal atoms. Two possible sites of absorption on the regular surface have been considered

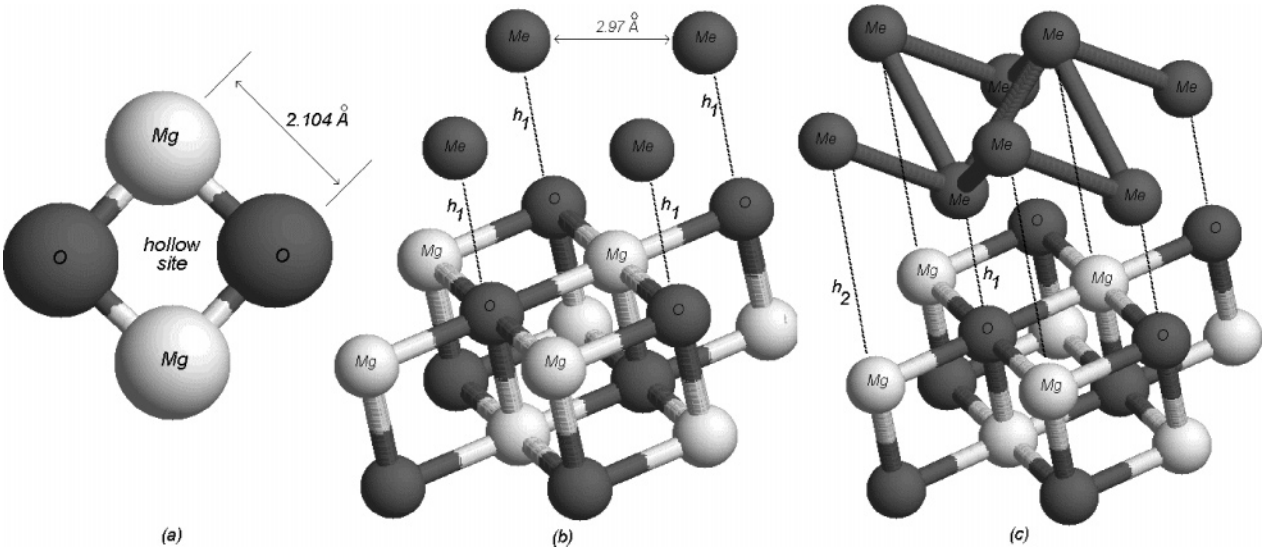


Figure 1. (a) Unit cell for a monolayer MgO slab. (b) One ML metal slab geometry absorption. (c) Two ML metal slab geometry absorption.

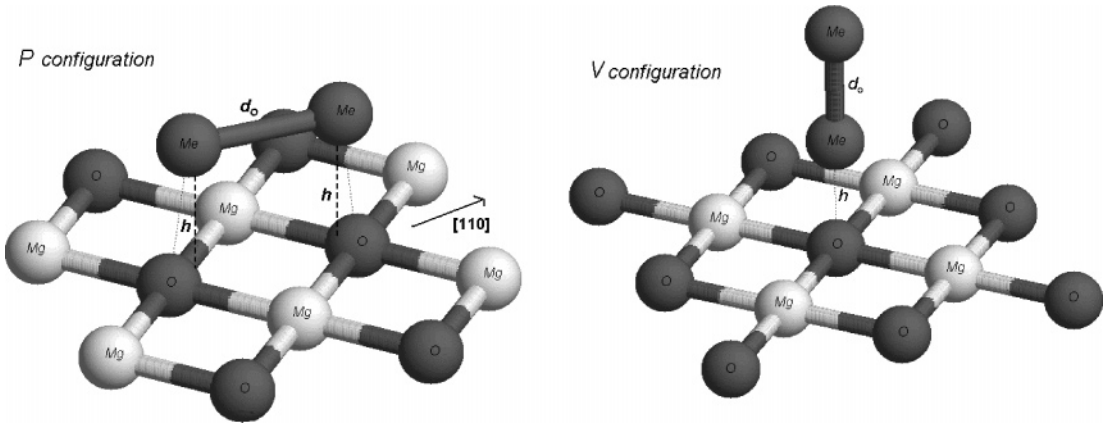


Figure 2. Dimer configurations on the regular MgO(100) surface.

Table 1. Interaction Energies and Distances Characterizing the Isolated Atom Absorption on the Regular MgO(100) Surface

adatoms		O site	Mg site	hollow site
Cu	E_{adh} (eV)	0.76	0.26	0.43
	d_o (Å)	2.1	2.8	2.3
Ag	E_{adh} (eV)	0.40	0.22	0.32
	d_o (Å)	2.5	2.9	2.6
Au	E_{adh} (eV)	0.87	0.49	0.67
	d_o (Å)	2.3	2.7	2.4

Table 2. (A) Characteristics of the Coinage Metal Dimers in the Gas Phase and (B) Geometric Parameters Characterizing the Geometric Configurations Shown in Figure 2

A. Characteristics of the Coinage Metal Dimers

	Cu_2	Ag_2	Au_2
binding energy (eV)	2.09	1.72	2.33
equilibrium bond length (Å)	2.25	2.59	2.52

B. Geometric Parameters

		d_0 (Å)	h (Å)	E_{int} (eV)	E_{adh} (eV)	ΔE_{met} (eV)
Cu_2	P	2.31	2.49	2.96	0.89	0.02
	V	2.26	2.16	3.24	1.15	0.00
Ag_2	P	2.61	2.79	2.18	0.46	0.00
	V	2.59	2.41	2.39	0.68	0.01
Au_2	P	2.56	2.63	2.92	0.61	0.02
	V	2.53	2.31	3.75	1.42	0.00

as interesting for a study of growth and diffusion (see Figure 2). The optimized values for the geometric parameters and the interaction energies are reported in Table 2b, while in Table 2a, the properties of the gas-phase dimers are reported.

The coinage metal dimers exhibit a similar behavior when absorbed on the regular surface. The values of the adhesion energy follow the same order characterizing the isolated adatoms: Au > Cu > Ag. The most stable geometry corresponds to the V configuration, in which the dimer is perpendicular to the surface atop the O^{2-} ion site. For Cu, this is in agreement with the results of refs 29−32, although we do not find any tilt angle with respect to the surface because we perform a symmetry-constrained relaxation along the surface normal direction. Although only one metal atom interacts with the surface, the adhesion energy of the dimer is *larger* than the value characterizing the isolated adatom.

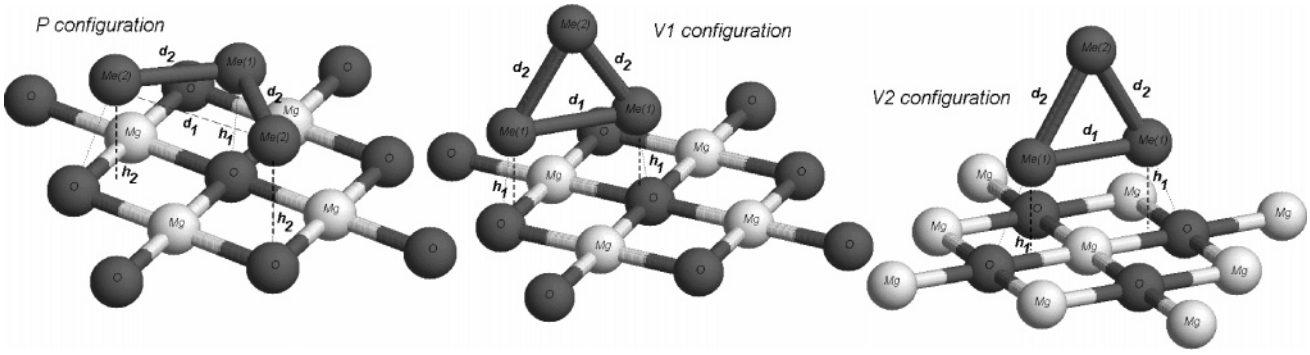


Figure 3. Trimer configurations on the regular surface. The last configuration on the right corresponds to the relaxation of the Au trimer on the F_s vacancy.

This result can be viewed as a first manifestation of the tendency of a vertical 3D growth of the coinage metal deposition on the oxide surface[29] and can be explained in terms of a stabilizing interaction between the dimer electron density and the field of the ionic surface.[103] In other words, the presence of metal atoms *above* those directly interacting with the surface *increases* the adhesion energy. This is what we call a "metal-on-top" stabilization mechanism. In the P configuration, the metal atoms can interact with two different O^{2-} ions, but the interaction energy is reduced because of the work needed to stretch the intermetal distance from the equilibrium value characterizing the isolated molecule. As a consequence, an almost identical intermetal distance is retained in all configurations considered and very small values of ΔE_{met} are observed.

To describe the diffusion process of the dimers on the surface, a leapfrog mechanism has been proposed for Cu, passing through the P configuration[29] with an activation energy barrier of 0.04 eV, although we find a larger energy difference between the V and P configurations.

It is interesting to note that, at variance with the coinage metal dimers, the most stable configuration for Pd_2 is the P one,[96] with the dimer parallel to the surface in a quasi-epitaxial geometry. This is due to a weaker metallic bond (especially noteworthy for the Pd_2 dimer) and a metal-on-top effect (probably because of a weaker polarizability of the d electrons with respect to the outer s electron).

I.C. Trimers. The chosen configurations are depicted in Figure 3. The unit cell was chosen, consisting of a three-layer MgO (each layer containing nine O atoms and nine Mg atoms) and three metal atoms. The values of the geometric parameters and of the interaction energies for each configuration are reported in Table 3b.

The trimers of the coinage metals are characterized by an s^3 electronic configuration, which produces a Jahn−Teller distortion and a consequent symmetry breaking from an equilateral triangle to an isosceles triangle. Looking at the optimized geometries of the three metals, we note that the distance between the central atom and the two lateral atoms is optimized at almost the equilibrium value of the gas-phase structure (reported in Table 3a), while the distance between the two lateral atoms is optimized in order to achieve the best interaction with the surface. Because of the Jahn−Teller effect, stretching the distance of the two lateral atoms does not imply any appreciable destabilization of the metal cluster.

Table 3. (A) Binding Energies and Geometric Parameters of the Coinage Metal Trimers in the Gas Phase and (B) Values of the Geometric Parameters Reported in Figure 4

A. Binding Energies and Geometric Parameters

	Cu_3	Ag_3	Au_3
binding energy (eV)	3.40	2.59	3.60
bond length d_1 (Å)	2.61	3.09	2.90
bond length d_2 (Å)	2.32	2.66	2.61

B. Values of Geometric Parameters

		d_1 (Å)	d_2 (Å)	h_1 (Å)	h_2 (Å)	E_{int} (eV)	E_{adh} (eV)	ΔE_{met} (eV)
	P	3.77	2.37	2.25	2.17	4.50	1.29	0.19
Cu_3	V1	2.61	2.33	2.10		5.00	1.60	0.00
	V2	3.50	2.36	2.06.		4.85	1.62	0.17
	P	4.33	2.69	2.73	2.67	3.27	0.72	0.04
Ag_3	V1	3.08	2.66	2.43		3.51	0.92	0.00
	V2	4.00	2.66	2.51		3.45	0.90	0.04
	P							
Au_3	V1	2.91	2.62	2.32		5.15	1.56	0.00
	V2	3.38	2.60	2.29		5.28	1.76	0.08

We refer to this behavior as *fluxional* (see the small values of ΔE_{met}). The larger values of ΔE_{met} for Cu are due to its shorter intermetal distances, which have to be stretched more to achieve a good interaction with the stabilizing sites of the surface.

As for the dimers, the preferred configurations of the trimers are the ones in which the cluster is perpendicular to the surface,[29] because of the metal-on-top stabilization mechanism. It can be noted that in the Au case, this effect is so strong that the P configuration relaxes toward the V2 configuration. These results are in fair agreement with those of ref 29 and suggest that the diffusion of the coinage trimers can take place through movements of "twisting" between the V1 and the V2 configurations, with low activation energy barriers especially for Ag and Au (slightly larger for Cu).

I.D. Extended Deposition. We built a unit cell assuming a pseudomorphic growth of the metal slab upon the oxide surface,[102,103] that is, positioning each metal atom above an O^{2-} (see Figure 1b), a Mg^{2+}, or a hollow site. Such a configuration corresponds to the growth of the (100) metal surface above the (100) oxide surface. Although the surface energy of the (100) surface is larger than the surface energy of the (111) surface for all the metals considered, the simplest metal configuration compatible with the symmetry of the

Table 4. Results for the Absorption of One and Two MLs on the Regular Surface[a]

		one ML			two ML		
		O site	Mg site	hollow site	O site	Mg site	hollow site
Cu	E_{int} (eV)	1.95	1.67	1.75	2.75	2.52	2.57
	E_{adh} (eV)	0.32	0.045	0.13	0.50	0.041	0.15
	h_1 (Å)	2.3	3.4	2.7	2.2	3.4	2.6
	h_2 (Å)				3.4	4.6	3.8
Ag	E_{int} (eV)	1.80	1.67	1.72	2.12	2.03	2.05
	E_{adh} (eV)	0.16	0.036	0.083	0.24	0.066	0.10
	h_1 (Å)	2.8	3.4	3.0	2.6	3.9	2.9
	h_2 (Å)				4.6	5.9	4.9
Au	E_{int} (eV)	2.26	2.20	2.24	2.66	2.59	2.62
	E_{adh} (eV)	0.15	0.095	0.13	0.22	0.084	0.15
	h_1 (Å)	2.9	3.1	2.9	2.8	3.3	2.9
	h_2 (Å)				4.8	5.3	4.9

[a] h_1 and h_2 respectively correspond to the heights of the first and the second metal layers on the MgO surface.

oxide cell is the one adopted here. The strict constraint of this assumption is that the first neighbor distance between the metal atoms of the slab is fixed at the value 2.975 Å. The chosen cell consists of three MgO layers, each containing only two Mg atoms, two O atoms, and two (four) metal atoms in the case of one (two) ML absorption (a five-layer slab model gave essentially identical results). In the two ML case, the atoms of the second metal slab were positioned in the interstitial sites of the first slab in such a way as to build the *fcc* structure of the bulk metals.

One ML. The results of one ML absorption are reported in Table 4, in which all energies are per atom. The values of ΔE_{met} are not reported because we did not perform a calculation by relaxing the reticular lattice parameter of the isolated metal ML.

From the results, which are in agreement with those reported in ref 103, we can derive the following observations:

(1) The order of stability of the three possible sites of absorption is the same as that found for the isolated atoms; the best configuration corresponds to all metal atoms on top of the O atoms on the surface.

(2) For all of the three system considered, there is a decrease of the adhesion energy per atom with respect to the case of the isolated atoms and a corresponding increase of the equilibrium distance between the metallic slab and the oxide surface.

(3) The absolute value of the adhesion energy decrease is remarkably large for Au on top of the O atoms; as a consequence, the stability order among the three metals is no longer Au > Cu > Ag (as for the isolated adatoms) but Cu > Ag ≈ Au.

To explain this behavior, two physical factors have to be taken into account: (1) the mismatch between the metal lattice constant and the reticular constant characterizing the pseudomorphic growth and (2) the strength of the metallic bond. The weak adhesion of the Au ML to the surface is due to the small mismatch between the reticular constant of bulk gold (2.885 Å) and the lattice parameter imposed by the pseudomorphic growth (2.975 Å), which implies a strong metal bond and, thus, a decreased interaction energy with the surface. Cu is characterized by a metal bond of strength comparable to Au, but the small value of the reticular constant of bulk copper corresponds to a large mismatch (about 16%) between the metal and the oxide lattice. The adhesion with the surface, thus, overcomes the metal binding energy.

Note, however, that these considerations neglect the change in the nearest-neighbor distances due to reduced coordination (from 12 to 4) in the one ML case. This determines a contraction of the lattice parameter,[57] consequently increasing the mismatch with the reticular constant of the oxide.

Two ML. The results for these systems are summarized in Table 4. Note that the adhesion energies are calculated normalizing the total interaction energy to the number of atoms in contact with the oxide surface (as we did in the case of the monolayer). [This is not a universally accepted convention, and some authors prefer to normalize the adhesion energy to the number of total atoms (considering those involved in the direct interaction with the surface and also the atoms in the second layer).] The general behavior is very similar to the one ML case: the order of stability among the three metals is Cu > Ag ≈ Au, and so forth. In passing from one to two ML deposition, however, one observes an increase in the adhesion energy (in a remarkable way, for the absorption atop the O^{2-} site). This increase in the adhesion energy has already been observed for Cu and Ag extended deposition[103] and explained in terms of the enhanced electrostatic polarization of the metallic electronic cloud in the surface field of the oxide, what we call a "metal-on-top" effect, in excellent agreement with the stabilization of the upright configurations reported in Sections I.B and I.C for the dimers and trimers. This happens despite the fact that the increase in the coordination number of the metal (from 4−8) increases its lattice constant, thus decreasing the mismatch with the lattice constant of the oxide and, consequently, enhancing the metallic bond in the slab.

We conclude this section by stressing that, when treating isolated atoms, among the coinage, we can distinguish between Cu and Au from one side and Ag on the other: Cu and Au are more strongly bound to the surface, while the interaction of Ag is much weaker. On the contrary, when passing to the extended deposition of one and two MLs, we observe a leveling of the adhesion energies among the three metals, with a pronounced resemblance between Ag and Au with respect to Cu.

II. Interaction of Coinage Metal Clusters with the an F_s Defect of the MgO(100) Surface. The F_s defect is formed when a neutral O atom leaves the MgO(100) surface but the two electrons carrying the negative charge of the O atom are kept in the vacancy. As the F_s vacancy constitutes a good candidate as a center of nucleation for the absorption of several transition metals on the oxide surface, this kind of defect has been extensively studied both experimentally and theoretically in recent years.

We considered the same unit cells used for the absorption on the regular surface: a three-layer MgO, each layer containing four or nine Mg and O atoms, depending on the size of the adsorbed metal cluster. Selected calculations using larger cells have also been performed, producing essentially comparable results.

Table 5. Adatoms Absorbed Atop the F_s Vacancy

adatoms	Cu	Ag	Au
$E_{adh} = E_{int}$ (eV)	1.76	1.59	3.04
h_o (Å)	1.80	2.00	1.81

Table 6. Dimer Geometries and Interaction Energies of the Coinage Dimers on the F_s Defect

		P	V		P	V
Cu_2	d_0 (Å)	2.33	2.34	E_{int} (eV)	4.13	4.16
	h_1 (Å)	1.55	1.68	E_{adh} (eV)	2.05	2.09
	h_2 (Å)	2.86	4.02	ΔE_{met} (eV)	0.01	0.02
Ag_2	d_0 (Å)	2.65	2.64	E_{int} (eV)	3.59	3.62
	h_1 (Å)	1.80	1.82	E_{adh} (eV)	1.89	1.92
	h_2 (Å)	3.91	4.46	ΔE_{met} (eV)	0.01	0.01
Au_2	d_0 (Å)	2.59	2.59	E_{int} (eV)		6.05
	h_1 (Å)	1.60	1.60	E_{adh} (eV)		3.75
	h_2 (Å)	4.19	4.19	ΔE_{met} (eV)		0.03

As anticipated before, the reconstruction of the MgO lattice around the F_s vacancy is negligible both in the absence and in the presence of metal absorption, in agreement with previous studies.[67,70] For this reason, we did not perform any relaxation of the MgO surface ($\Delta E_{MgO} = 0$).

II.A. Isolated Adatoms. The single atoms are trapped atop the F_s center. The values of the interaction energy and of the equilibrium distance are reported in Table 5 and are in fair agreement with the results reported in ref 67.

The order among the three metals is same as that for the adsorption on the regular surface. The atoms can get closer to the vacancy on the surface because of the decrease of the Pauli repulsion consequent to the removal of the O atom, thus experiencing a stronger electric field, with a consequent increase of the polarization interaction. The relative increase of the interaction energy is larger in the cases of Ag and Au with respect to Cu, probably because of chemical bond effects.

II.B. Dimers. We considered the same configurations used for the absorption on the regular surface (see Figure 2). In Table 6, the geometric parameters and the energies for each configuration are reported.

The results for the P configuration of Au_2 is not reported because it relaxes toward the perpendicular configuration atop the vacancy (V configuration).

As for the perfect surface, the intermetal distances are only slightly elongated with respect to the equilibrium gas-phase values; the distortion energies are consequently very small. Also, the differences in adhesion and interaction energies between the P and V configurations are very small; this can be explained by noting that the P configuration is remarkably tilted with respect to the surface, so the dimers in the two configurations probably experience a similar field.

The interaction mechanism is the same as that for the adsorption of a single atom on the F_s center. Only, the decrease in equilibrium distances and, thus, the stabilizing effect are still more pronounced, as a result of the metal-on-top mechanism.

II.C. Trimers. We chose two configurations: one is similar to V1 from Figure 3, in which the F_s vacancy is under a lateral metal atom, and the other one is a linear epitaxial configuration, in which the F_s vacancy is under one of the lateral atoms. The geometric parameters and the energies characterizing the absorption of the coinage metal trimers on the F_s vacancy are reported in Tables 7 and 8, from which it can be seen that an appreciable rearrangement of the metal atoms not directly bound to the defect occurs. The dominant contribution to the adhesion energy is given by the interaction of the metal atom atop the defect. As a result of the fluxionality of the trimers and the interplay between the metal-on-top effect and the direct metal−surface interaction, there exists a large set of different configurations very close in energy, among which we selected two representative ones, without being certain to have singled out the absolute energy minimum. As before, the energy differences are slightly larger for Cu, because of its reduced fluxionality. The order of stability among the three coinage metals is the same as that found in the previous sections. The adhesion energies for all three metals are still larger than for the dimers, as fluxionality allows the trimers to deform in such a way as to better interact with the surface and exploit the metal-on-top effect or the direct metal−surface interaction.

Table 7. Geometric Parameters for the Coinage Trimers Absorbed on the F_s Vacancy[a]

		d_{1-2} (Å)	d_{2-3} (Å)	d_{1-3} (Å)	h_1 (Å)	h_2 (Å)	h_3 (Å)
Cu_3	V1	2.58	2.36	2.40	1.68	2.17	3.90
	L1	2.47	2.37	4.80	1.91	2.48	2.43
Ag_3	V1	2.97	2.73	2.70	1.80	2.67	4.40
	L1	2.73	2.71	5.30	1.92	3.04	2.92
Au_3	V1	2.80	2.69	2.69	1.57	2.85	4.25
	L1	2.67	2.62	4.68	1.66	3.37	2.65

[a] The atom labeled 1 is the one absorbed atop the vacancy.

Table 8. Energies Involved in the Trimer Absorption on the F_s Vacancy

		E_{int} (eV)	E_{adh} (eV)	ΔE_{met} (eV)
Cu3	V1	6.12	2.75	0.03
	L1	5.58	2.48	0.31
Ag3	V1	4.96	2.39	0.02
	L1	4.78	2.25	0.07
Au3	V1	7.80	4.24	0.04
	L1	7.70	4.09	0.01

The main conclusion that we can extract from the results of the absorption of atoms, dimers, and trimers on the oxygen vacancy is that the defect acts as a strong trapping center for all the coinage metals. The differences in adhesion between several configurations are small with respect to the absolute values of the absorption energies found in the calculations. Every possible mechanism of diffusion of small clusters from the site of the vacancy to the neighboring regular sites is strongly disfavored; the more it is disfavored, the larger the metal cluster.

II.D. Extended Depositions. One ML. The mean values of the adhesion and adsorption energies are evaluated by calculating the total absolute value of the respective energies and dividing by the number of the atoms in contact with the surface (four both for one and two MLs). The word "mean" refers to the fact that the single atoms experience a different

Table 9. One and Two MLs Absorbed on the F_s Vacancy

	one ML			two ML		
	Cu	Ag	Au	Cu	Ag	Au
$\langle E_{int} \rangle$ (eV)	2.14	1.99	2.72	2.84	2.25	2.85
$\langle E_{adh} \rangle$ (eV)	0.54	0.41	0.84	0.68	0.47	0.79
$\langle \Delta E_{met} \rangle$ (eV)	0.03	0.06	0.20	0.01	0.01	0.09
h_1 (Å)[a]	1.79	2.00	1.71	1.83	2.31	1.80
h_2 (Å)[b]	2.41	2.84	2.90	2.17	2.74	2.69
h_3 (Å)[c]	2.54	2.96	2.99	2.18	2.77	2.73

[a] Distance of the metal atom atop the vacancy. [b] Distance of the metal atom atop an O ion at 2.97 Å away from the vacancy. [c] Distance of the metal atom atop an O ion at 4.21 Å away from the vacancy.

Table 10. Coinage Dimers on the Double Vacancy

	Cu_2	Ag_2	Au_2		Cu_2	Ag_2	Au_2
d_o (Å)	2.43	2.68	2.69	E_{int} (eV)	5.51	4.23	5.21
h_1 (Å)[a]	0.191	0.525	0.467	E_{adh} (eV)	4.56	3.21	3.89
h_2 (Å)[b]	1.59	2.10	2.04	ΔE_{met} (eV)	0.11	0.02	0.12
r_{1-Mg} (Å)	0.089	0.276	0.370	ΔE_{MgO} (eV)	1.03	0.68	0.89
r_{2-O} (Å)	0.028	0.336	0.442	ΔE_{relax} (eV)	0.86	0.60	0.73

[a] Height of the atom atop the Mg^{2+} empty site. [b] Height of the atom atop the O^{2-} empty site.

interaction with the surface. The sites of absorption are four: one atop the vacancy, two on regular O^{2-} ions at a distance of 2.97 Å from the defect, and one on a regular O^{2-} at a distance 4.21 Å from the defect. The results are reported in Table 9. The values reported as $\langle \Delta E_{met} \rangle$ correspond to the energy increase due to the corrugation of the metal slab.

The final relaxed topology of the metallic slab is characterized by a corrugation due to the different heights of the metal atoms upon the four different sites of the surface. The atom directly atop the vacancy lies almost at the equilibrium height characterizing the isolated adatom, which suggests that this atom is practically not involved in the metallic bond with the other metal atoms. By using the data of Table 4, we can give an estimate of the interaction energy of the lower metal atom with the vacancy: supposing that the three metal atoms on the regular sites interact with the surface with the same amount of adhesion energy characterizing the one ML on the regular surface, we get

- Cu: 1.44 eV (real value of the adatom: 1.76 eV)
- Ag: 1.28 eV (real value of the adatom: 1.59 eV)
- Au: 3.02 eV (real value of the adatom: 3.04 eV)

In the one ML system, the interaction of the lower atom with the vacancy is decreased by about 20% in the cases of Cu and Ag and only less than 1% in the case of Au.

Two ML. When studying the absorption of a second monolayer absorbed on the first one, we do not find any corrugation of the second metal layer. The results are reported in Table 9.

It can be noted that for Cu and Ag, we observe an increase of the adhesion energy and a corresponding decrease of the distance of the metal atoms on the regular sites from the surface. On the contrary, Au exhibits a decrease of the adhesion energy; the stabilization due to the metal-on-top effect does not manage to compensate for the loss of the metal bond energy due to the corrugation of the first layer.

III. Interaction of Coinage Metal Clusters with the a Double Vacancy of the MgO(100) Surface. Another surface defect that has attracted considerable interest in recent years is the double vacancy.[100] This is a neutral defect and corresponds to the removal of an MgO dimer in the 100 direction of the regular oxide terrace. Its surface concentration is estimated to be rather high.[7,72] The difference with respect to the F_s defect, which, as we will see, has important consequences in view of the nucleation process, is that the removal of an MgO dimer induces an appreciable modification in the structure of the oxide slab. In fact, allowing relaxation of the defected surface results in an increase of the dimensions of the cavity with a corresponding stabilization energy of about 2.90 eV. The distances between the prospicient atoms at the border of the cavity increase by up to 10% with respect to the nonrelaxed values (see Figure 4a, where some characteristic distances between the borders of the cavity are reported).

As anticipated above, we always relaxed all the geometric parameters upon absorption. The unit cell of the MgO slab is made of three layers containing a total of 52 atoms. This is the largest unit cell we have considered, and this, together with the need of optimizing the degrees of freedom also corresponding to the MgO surface, means that the calculations involving the divacancy are the most computationally intensive.

III.A. Isolated Atoms. The results of the absorption of single coinage metal atoms on the double vacancy are reported in Table 11. In the table, E_{int}, E_{adh}, ΔE_{MgO} are the quantities defined at the beginning of Section 3; ΔE_{relax} is a new quantity defined as the energy gain obtained by performing a relaxation of the cluster coordinates, keeping the coordinates of the lattice frozen in the position characterizing the relaxed isolated oxide, and it is useful in checking the validity of the quadratic expansion of the interaction energy (see below).

From the analysis of the data from Table 11, two common features can be derived: (1) All the metals adsorb near the site left empty by the removal of the Mg atom; all the adsorbed atoms go deeply into the cavity reaching almost the same height as the Mg and O atoms of the surface oxide layer. (2) The interaction energies with the divacancy are of the same order of magnitude as the energies characterizing the interaction with the F_s defect, suggesting that this defect is another good candidate as a nucleation site on the MgO(100) terrace.[100]

Some quantitative differences can be evinced through a more detailed analysis: Cu has the smallest size and, thus, it gets closer than Ag or Au to the empty site of the Mg^{2+} ion. In such a way, it can interact with the O^{2-} ion beneath it and also with the other three O^{2-} ions at the borders of the vacancy. On the other hand, Ag and Au are too big to get close to the empty site of the Mg^{2+} ion and move toward the center of the vacancy, still keeping an appreciable interaction with the four O^{2-} surrounding ions. These considerations are confirmed by the amount of the MgO lattice "back relaxation", that is, the distortion energy of the lattice in the configuration optimally interacting with the metal atom with respect to the isolated relaxed defect,

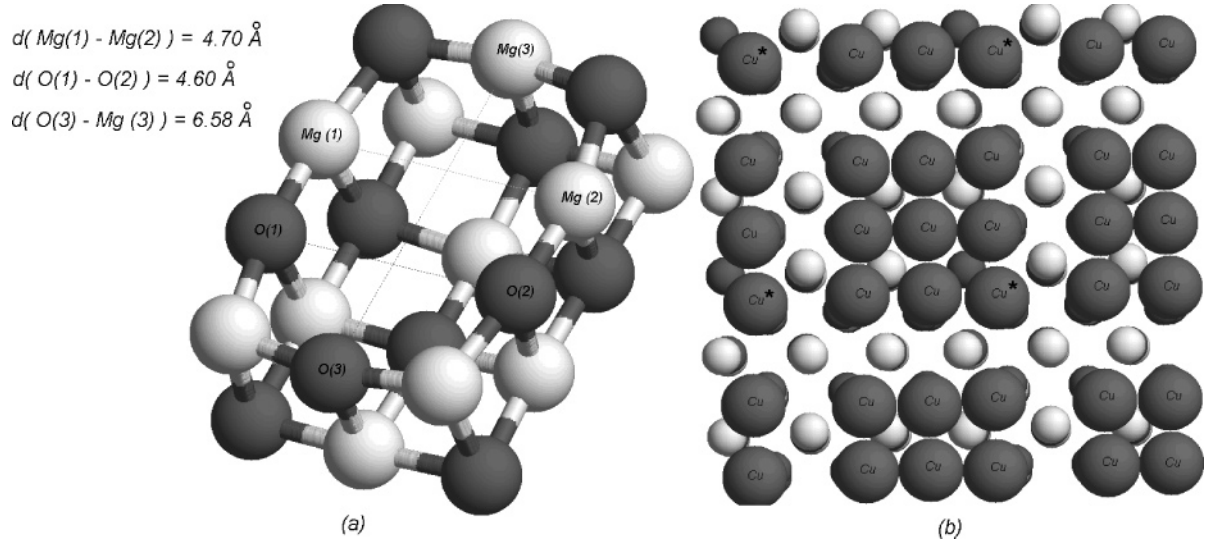


Figure 4. (a) Relaxed geometry of the oxide surface around the divacancy. (b) Cu_9 island formation; the Cu* atoms are the ones absorbed upon the vacancy.

Table 11. Interaction Energies and Characteristic Distances of the Isolated Adatoms on the Double Vacancy

adatoms	Cu	Ag	Au
E_{int} (eV)	2.83	1.95	2.34
E_{adh} (eV)	3.94	2.47	2.90
ΔE_{MgO} (eV)	1.11	0.522	0.568
ΔE_{relax} (eV)	0.778	0.454	0.481
h_o (Å)	0.189	0.547	0.501
distance from Mg site (Å)	0.176	0.585	0.781

expressed as ΔE_{MgO} in Table 11. In the case of Cu, the relaxed oxide lattice *closes* around the metal atom (ΔE_{MgO} has a value of 1.11 eV) much more than for Ag and Au. Finally, it can be noted that the values of ΔE_{relax} are roughly proportional to the values of ΔE_{MgO}, in agreement with a quadratic expansion of the interaction energy (see, e.g., ref 116).

III.B. Dimers. Whatever the starting configuration, all the coinage metal dimers relaxed to a configuration with the two atoms of the molecule atop the two empty sites of the vacancy. This behavior is analogous to the absorption of the Pd dimer on the double vacancy.[100] The remarkable stability of this configuration is due to the fact that the two atoms can interact with the three O^{2-} ions and the three Mg^{2+} ions at the border of the cavity; both atoms of the metal cluster can get close to the surface thanks to the big dimensions of the cavity. Our results are reported in Table 10.

The order of stability among the coinage metals is the same as that for the absorption of the single adatom; furthermore, also in this case, the smaller dimensions of the Cu dimer favor an enhanced backward movement of the oxide lattice around the cluster absorbed (see the values of ΔE_{MgO}). Again, the gain in interaction energy due to the relaxation of the oxide lattice (i.e., the values of ΔE_{relax}) are roughly proportional to ΔE_{MgO}, in fair agreement with the quadratic expansion theorem, showing that the system is roughly in the linear regime. This is further confirmed by the fact that the value of ΔE_{relax} is approximately proportional to the corresponding adhesion energy (ΔE_{adh}).

Table 12. Coinage Trimer Geometries on the Double Vacancy

	Cu_3		Ag_3		Au_3	
	A	B	A	B	A	B
d_{1-2} (Å)	2.49	2.44	2.68	2.70	2.63	2.64
d_{2-3} (Å)	2.47	2.61	2.76	2.82	2.65	2.61
d_{1-3} (Å)	2.38	2.39	2.80	2.79	2.81	2.91
h_1 (Å)[a]	0.217	0.255	0.789	0.808	1.13	1.15
h_2 (Å)[b]	1.56	1.67	2.17	2.17	1.82	1.84
h_3 (Å)	2.68	2.05	3.55	2.92	3.73	3.10

[a] Height of the atom atop the Mg^{2+} empty site. [b] Height of the atom atop the O^{2-} empty site

Table 13. Energies for the Trimers on the Double Vacancy

	Cu_3		Ag_3		Au_3	
	A	B	A	B	A	B
E_{int} (eV)	6.90	7.09	5.19	5.20	6.74	6.82
E_{adh} (eV)	4.65	5.00	3.23	3.28	4.11	4.26
ΔE_{met} (eV)	0.06	0.07	0.03	0.02	0.01	0.005
ΔE_{MgO} (eV)	1.09	1.24	0.60	0.65	0.96	1.03

III.C. Trimers. The geometries we considered are with two metal atoms on top of the two sites of the vacancy and the third metal atom either atop the other two (A), not in direct contact with the surface, or atop the O^{2-}(2) ion at the border of the cavity (B) (see Figure 4a for the nomenclature of the surface sites). The results of the calculations are reported in Tables 12 and 13.

The results for the three metals do not qualitatively differ very much: the two selected configurations are characterized by very similar values of the absorption energy, the distortion energies of the metallic clusters are quite small, and among the three coinage metals, the largest values are observed in the case of Cu, in agreement with the results of Sections I.C and II.C. As in the case of the interactions of the single atoms and dimers, the values of the distortion energy of the surface lattice are largest in the case of the absorption of Cu clusters

Table 14. One and Two MLs Absorbed on the Double Vacancy

	1 ML			2 ML		
	Cu	Ag	Au	Cu	Ag	Au
$\langle E_{int} \rangle$ (eV)	2.19	1.81	2.31	2.85	2.18	2.73
$\langle E_{adh} \rangle$ (eV)	0.47	0.25	0.31	0.78	0.32	0.50
$\langle \Delta E_{met} \rangle$ (eV)	−0.17	0.05	0.02	−0.02	−0.03	0.02
ΔE_{MgO} (eV)	0.70	0.29	0.67	1.37	0.67	1.00

[a] Distance of the metal atom atop the vacancy. [b] Distance of the metal atom atop an O ion at 2.97 Å away from the vacancy. [c] Distance of the metal atom atop an O ion at 4.21 Å away from the vacancy.

(because of the smaller dimensions of the Cu atom) and smallest in the case of Ag.

III.D. Extended Deposition. These results are reported in Table 14. The study of one (two) ML absorption above the double-vacancy defected MgO surface involves a heavy calculation implying the full relaxation of all the 61 (70) atoms per unit cell. For all three metals considered, a spin-restricted ground state was found. The chosen geometry is a pseudomorphic one with each metal atom atop a different O^{2-} ion of the surface layer. The particularity of this configuration is that one metal atom is put on top of the site of the vacancy with the weaker interaction (the empty O site).

The results of the calculations are discussed separately for the three coinage metals because of their peculiar behavior. Starting with Ag, for the which the discussion is simpler, we note that Ag is characterized by a good match between the lattice constant of the metallic layer and that of the oxide support. Moreover, Ag has a weaker interaction energy with the surface (both regular and defected) and a weaker metallic bond with respect to Au and Cu. The small value of ΔE_{MgO} (only 0.29 eV for one ML) and the fact that the metal atoms essentially maintain their starting positions suggest that the layer does not interact very much, even with the defected surface. In particular, the Ag atom atop the vacancy relaxes by simply decreasing its distance from the surface. The adhesion energy is quite low: 2.25 eV for nine atoms. If we take a value of 0.16 eV as the interaction energy of the eight atoms on the regular sites (the value characterizing the interaction with the regular surface), we obtain by difference that the atom atop the defect contributes about 0.9−1.0 eV, a value appreciably smaller than that for the absorption of the single atom on the divacancy (see Table 11).

Au, too, is characterized by a good match between the lattice constants of the metal and that of the oxide. However, Au interacts more strongly than Ag with the regular and defected surface sites of the surface and, at the same time, has a stronger metallic bond. The Au monolayer, thus, presents a more pronounced corrugation; in particular, the atom atop the empty O site remarkably decreases its distance from the surface (with a consequent weakening of its metal bonding), while the three Au atoms at the border of the vacancy move slightly inward. What we observe is a segregation of the metal slab with the formation of islands of nine atoms with the divacancy placed at one edge of the square (see Figure 4b for the case of Cu, in which this effect is even more pronounced). The stronger interaction with the surface is reflected in the larger value of the lattice distortion energy with respect to the Ag case.

The Cu monolayer exhibits the same behavior encountered in the Au case but in a more pronounced way. The larger mismatch between the Cu lattice constant and the surface lattice parameter and the stronger interaction of the Cu atom with the divacancy, in fact, determine a strong segregation of the metal slab with the formation of nine-atom islands (see Figure 4b). From the data of ΔE_{met} in Table 14, we observe that the formation of islands is accompanied by a strengthening of the metallic bond with respect to the pseudomorphic configuration. Segregation is also expected to occur in the absence of the divacancy defect, but it is significant that the defect places itself at a border of the metal islands; this suggests a peripheral growth of the metal cluster with respect to the double vacancy (at variance with the F_s center case, in which the cluster is expected to grow all around the defect).

The absorption of the second metal layer above the first one determines an increase of the values of ΔE_{MgO} and E_{adh} for all the three coinage metals, essentially because of the metal-on-top mechanism. In the case of Ag, the deposition of the second layer does not affect the geometry of the metal slab and the increase in adhesion energy is on the same order of magnitude as in the case of absorption on the regular surface and on the F_s defect. In the case of Cu and Au, the presence of the second metal layer above the first one prevents the segregation observed in the case of the single monolayer, thus increasing the direct interaction of the first metal slab with the surface and the metal-on-top effect. The increase in adhesion energy is consequently larger than in the case of absorption on the regular terrace and that on the single vacancy.

4. Conclusions

The results of a systematic study of the interaction of small coinage metal clusters (M_n, $n = 1-3$) and extended deposition (one and two ML) with the regular and locally defected (F_s center and divacancy) neutral MgO(100) surface are presented. These results can be summarized as follows, underlining analogies and differences between the three coinage metals.

In agreement with previous work,[65−67] the most stable site of nucleation on the regular surface for all the isolated adatoms is the O^{2-} ion. The interaction energy is larger in the case of Cu and Au with respect to Ag. This order of interaction energies is found to hold for nearly all the systems considered. In the dimers, the metal chemical bond is stronger than the interaction with the surface and the "sticky" character of the metallic bond prevents the stretching of the metal−metal distance from the equilibrium value characterizing the isolated dimers. The most stable configuration is perpendicular to the surface. Despite the fact that only one atom interacts with the surface, the adhesion energy of the cluster is larger than that for the single adatom, which suggests that the presence of metal atoms *on-top* of those directly interacting with the surface *increases* the adhesion energy ("metal-on-top" stabilization mechanism). A peculiar characteristic of the coinage metal trimers is their fluxional

behavior, originating from the Jahn−Teller symmetry breaking due to the electronic configuration of these molecules. As in the case of the dimers, the configurations perpendicular to the surface are the most stable ones ("metal-on-top" stabilization mechanism). Nevertheless, thanks to their fluxionality, the trimers manage to optimize both the metallic bond and the adhesion with the surface in several different configurations. The consequently small energy differences between the corresponding local minima allow the diffusion of the clusters on the surface even at rather low temperatures. This is particularly true for Ag and Au, which are characterized by longer intermetal bond lengths with respect to Cu and a, thus, smaller mismatch with the oxide lattice parameter.

For the absorption of extended pseudomorphic deposition (one and two ML), Ag[83,103,104] and Au are characterized by a good match between the bulk metal lattice constant and the lattice constant of the oxide pattern. The strength of the metallic bond then determines a decrease of the adhesion energy with the surface in passing from small clusters to extended deposition, especially in the case of Au, for the which the metallic bond is stronger and "stickier".[57] Instead, the mismatch between the lattice parameters of the Cu overlayer and the surface determines a smaller decrease of the adhesion energy with the surface.[83,103,104] For all three coinage metals, the thickening of the metal slab from one to two MLs determines an increase of the adhesion energy per atom in agreement with the "metal-on-top" stabilization mechanism.

In the case of the F_s center and the double vacancy, it is found that these defects act as strong trapping centers for small clusters and strongly increase the adhesion of metal slabs to the surface in the case of extended deposition.[67,108] The main difference between the two defects is that the presence of the double vacancy induces a strong structural and energetic modification in the surrounding oxide lattice;[67] moreover, such a structural relaxation changes according to the extent of the metal deposition upon the surface. On the contrary, the presence of the single vacancy induces negligible structural and energetic modifications of the oxide lattice both in the absence and in the presence of the metal deposition.

In the case of the single vacancy, the metal atoms can get closer to the surface because of the decreased Pauli repulsion consequent to the removal of the O atom.[103] The adhesion energy is stronger for Au than for Cu and Ag; since the dimensions and the polarizability of the gold atom are similar to those of Ag, this suggests that chemical bonding effects are important for Au. The most stable configurations of dimers and trimers are, again, those in which the cluster is perpendicular to the surface, with one of the lower atoms directly atop the vacancy. In the case of extended deposition, the atom atop the vacancy is weakly involved in the metallic bond. In the case of the two ML absorption, we observe an increase of the adhesion energy in the case of Cu and Ag but not for Au, for which the distortion of the metal structure overwhelms the stabilization as a result of the thickening of the metal slab.

In the case of the double vacancy, the metal atoms are free to choose between the two possible sites of the vacancy (the empty Mg site or the empty O site). Ag and Au, which are bigger than Cu, move toward the center of the cavity in order to reduce the Pauli repulsion with the O^{2-} ions at the border. Cu, instead, absorbs near the empty Mg site and, consequently, interacts much more strongly with the surface. In the case of the dimers and trimers, two metal atoms occupy the two empty sites of the surface. In the case of extended deposition, the pseudomorphic growth is disfavored by the fact that the metal atom directly above the cavity is atop the O empty site. For Cu and Au, this, together with the mismatch between the Cu lattice parameter of the metal and that of the oxide, induces the formation of segregated metal islands with the cavity at one of the corners. This effect is decreased in the two ML case, in which the adhesion energy increases appreciably for all three coinage metals.

To briefly summarize, we find that coinage metals can easily diffuse on the perfect MgO(100) surface, with small clusters diffusing even faster than single atoms, to be strongly trapped at local defects, such as the F_s center and the divacancy, which can, thus, act as nucleation centers. The presence of metal atoms above those directly interacting with the surface increases the adhesion energy, according to a "metal-on-top" mechanism, for both perfect and defected oxide. A peculiar structural relaxation is found for the divacancy, which suggests a "peripheral" growth of the metal clusters.

Acknowledgment. We thank Gianfranco Pacchioni (Milano, Italy) and Riccardo Ferrando (Genova, Italy) for useful discussions. We acknowledge financial support from the Italian CNR for the project "(Supra-)Self-Assemblies of Transition Metal Nanoclusters" within the framework of the ESF EUROCORES SONS and from the European Community Sixth Framework Programme for the STREP project "Growth and Supra-Organization of Transition and Noble Metal Nanoclusters" (Contract NMP4-CT-2004-001594). A.F. acknowledges the Italian INSTM for a grant at the CINECA supercomputing center. (After the manuscript was submitted, we learned of a related work on Au small clusters absorbed on regular and defected MgO surfaces[117]).

Acknowledgment. This section tagged Supporting Information

Supporting Information Available: Additional data on configurations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Aiken, J. D., III; Finke, R. G. *J. Mol. Catal. A* **1999**, *145*, 1−44.

(2) Jensen, P. *Rev. Mod. Phys*. **1999**, *71*, 1695−1735.

(3) Henry, C. R. *Surf. Sci*. **1998**, *31*, 235−325.

(4) Pileni, M.-P. *J. Phys. Chem. B* **2001**, *105*, 3358−3371.

(5) Henry, C. R.; Meunier, M. *Mater. Sci Eng*., A **1996**, *A217/218*, 239−243.

(6) Leroy, F.; Revenant, C.; Renaud, G.; Lazzari, R. *Appl. Surf. Sci.* **2004**, *238*, 233−237.

(7) Revenant, C.; Leroy, F.; Lazzari, R.; Renaud, G.; Henry, C. R. *Phys. Rev. B* **2004**, *69*, 035411.

(8) Renaud, G. *Surf. Sci. Rep*. **1998**, *32*, 1−90.

(9) Renaud, G. *Science* **2003**, *300*, 1416−1419.

(10) Thornton, G. *Science* **2003**, *300*, 1378−1379.

(11) Hansen, K. H.; Ferrero, S.; Henry, C. R. *Appl. Surf. Sci.* **2004**, *226*, 167−172.

(12) Becker, C.; Rosenhahn, A.; Wiltner, A.; Von Bergmann, K.; Schneider, J.; Pervan, P.; Milun, M.; Kralj, M.; Wandelt, K. *New J. Phys*. **2004**, *4*, 75.1−75.15.

(13) Degen, S.; Becker, C.; Wandelt, K. *Faraday Discuss*. **2004**, *125*, 343−356.

(14) Becker, C.; Von Bergmann, K.; Rosenhahn, A.; Schneider, J.; Wandelt, K. *Surf. Sci.* **2001**, *486*, L443−L448.

(15) Wiltner, A.; Rosenhahn, A.; Schneider, J.; Becker, C.; Pervan, P.; Milun, M.; Kralj, M.; Wandelt, K. *Thin Solid Films* **2001**, *400*, 71−75.

(16) Xu, C.; Oh, W. S.; Liu, G.; Kim, D. Y.; Goodman, D. W. *J. Vac. Sci. Technol., A* **1997**, *15*, 1261−1268.

(17) Yoon, B.; Luedtke, W. D.; Gao, J.; Landman, U. *J. Phys. Chem. B* **2003**, *107*, 5882−5891.

(18) Bréchignac, C.; Cahuzac, P.; Carlier, F.; Colliex, C.; Leroux, J.; Masson, A.; Yoon, B.; Landman, U. *Phys. Rev. Lett.* **2002**, *88*, 196103.

(19) Moseler, M.; Hakkinen, H.; Landman, U. *Phys. Rev. Lett*. **2002**, *89*, 176103.

(20) Campbell, C. T.; Starr, D. *J. Am. Chem. Soc*. **2002**, *124*, 9212−9218.

(21) Kim, Y. D.; Stultz, J.; Wei, T.; Goodman, D. W. *J. Phys. Chem. B* **2002**, *106*, 6827−6830.

(22) Larsen, J. H.; Ranney, J. T.; Starr, D. E.; Musgrove, J. E.; Campbell, C. T. *Phys. Rev. B* **2001**, *63*, 195410.

(23) Suzuki, T.; Hishita, S.; Oyoshi, K.; Souda, R. *Surf. Sci.* **1999**, *442*, 291−299.

(24) Schaffner, M. H.; Patthey, F.; Schneider, W. D. *Surf. Sci.* **1998**, *417*, 159−167.

(25) Barbier, A.; Renaud, G.; Jupille, J. *Surf. Sci*. **2000**, *454*−*456*, 979−983.

(26) Stracke, P.; Krischok, S.; Kempter, V. *Surf. Sci.* **2001**, *473*, 86−96.

(27) Lagarde, P.; Colonna, S.; Flank, A. M.; Jupille, J. *Surf. Sci.* **2003**, *524*, 102−112.

(28) Abbet, S.; Ferrari, A. M.; Giordano, L.; Pacchioni, G.; Hakkinen, H.; Landman, U.; Heiz, U. *Surf. Sci.* **2002**, *514*, 249−255.

(29) Musolino, V.; Selloni, A.; Car, R. *Phys. Rev. Lett*. **1999**, *83*, 3242−3245.

(30) Musolino, V.; Selloni, A.; Car, R. *J. Chem Phys*. **1998**, *108*, 5044−5054.

(31) Musolino, V.; Del Corso, A.; Selloni, A. *Phys. Rev. Lett*. **1999**, *83*, 2761−2764.

(32) Musolino, V.; Selloni, A.; Car, R. *Surf. Sci.* **1998**, *402*−*404*, 413−417.

(33) Hakkinen, H.; Barnett, R. N.; Landman, U. *Phys. Rev. Lett*. **1999**, *82*, 3264−3267.

(34) Landman, U.; Luedtke, W. D. *Faraday Discuss.* **2004**, *125*, 1−22.

(35) Pileni, M.-P. *C. R. Chimie* **2003**, *6*, 965−978.

(36) Pileni, M. P.; Lalatonne, Y.; Ingert, D.; Lisiecki, I.; Coutry, A. *Faraday Discuss*. **2004**, *125*, 251−264.

(37) Shibata, T. *J. Am. Chem. Soc* **2002**, *124*, 11989−11996.

(38) Kreibig, U.; Vollmer, M. *Optical Properties of Metal Clusters*; Springer: Berlin, 1995.

(39) Sanchez, A.; Abbet, S.; Heiz, U.; Schneider, W. D.; Hakkinen, H.; Barnett, R. N.; Landman, U. *J. Phys. Chem. A* **1999**, *103*, 9573−9578.

(40) Hakkinen, H.; Landman, U. *J. Am. Chem. Soc*. **2001**, *123*, 9704−9705.

(41) Socaciu, L. D.; Hagen, J.; Bernhardt, T. M.; Woste, L.; Heiz, U.; Hakkinen, H.; Landman, U. *J. Am. Chem. Soc*. **2003**, *125*, 10437−10445.

(42) Del Vitto, A.; Sousa, C.; Illas, F.; Pacchioni, G. *J. Chem. Phys*. **2004**, *121*, 7457−7466.

(43) Garzon, I. L.; Beltran, M. R.; Gonzalez, G.; Gutierrez-Gonzales, I.; Michaelian, K.; Reyes-Nava, J. A.; Rodriguez-Hernandez, J. I. *Eur. Phys. J. D* **2003**, *24*, 105−109.

(44) Roman-Velasquez, C. E.; Noguez, C.; Garzon, I. L. *J. Phys. Chem. B*, **2003**, *107*, 12035−12038.

(45) Robach, O.; Renaud, G.; Barbier, A. *Surf. Sci*. **1998**, *401*, 227−235.

(46) *Progress in Experimental and Theoretical Studies of Clusters*; Kondow, T., Mafuné, F., Eds.; World Scientific: New York, 2003.

(47) Bravo-Perez, G.; Garzon, I. L.; Novaro, O. *THEOCHEM* **1999**, *493*, 225−231.

(48) Aprà, E.; Fortunelli, A. *J. Phys. Chem. A* **2003**, *107*, 2934−2942.

(49) Fortunelli, A.; Velasco, A. M. *THEOCHEM* **2002**, *586*, 17−27.

(50) Aprà, E.; Fortunelli, A. *THEOCHEM* **2000**, *501*−*502*, 251−259.

(51) Fortunelli, A. *THEOCHEM* **1999**, *493*, 233−240.

(52) Yannouleas, C.; Landman, U.; Herlert, A.; Schweikhard, L. *Phys. Rev. Lett*. **2001**, *86*, 2996−2999.

(53) Yoon, B.; Hakkinen, H.; Landman, U. *J. Phys. Chem. A* **2003**, *107*, 4066−4071.

(54) Hakkinen, H.; Landman, U. *Phys. Rev. B* **2000**, *62*, R2287−R2290.

(55) Cleveland, C. L.; Luedtke, W. D.; Landman, U. *Phys. Rev. B*, **1999**, *60*, 5065−507.7

(56) Wolf, M. D.; Landman, U. *J. Phys. Chem. A* **1998**, *102*, 6129−6137.

(57) Baletto, F.; Ferrando, R.; Fortunelli, A.; Montalenti, F.; Mottet, C. *J. Chem. Phys.* **2002**, *116*, 3856−3863.

(58) Baletto, F.; Mottet, C.; Ferrando, R. *Phys. Rev. Lett*. **2000**, *84*, 5544−5547.

(59) Baletto, F.; Mottet, C.; Ferrando, R. *Phys. Rev. B* **2002**, *66*, 1−11.

(60) Rossi, G.; Mottet, C.; Fortunelli, A.; Baletto, F.; Ferrando, R. *Phys. Rev. Lett*. **2004**, *93*, 105503.

(61) Garzon, I. L.; Kaplan, I. G.; Santamaria, R.; Novaro, O. *J. Chem. Phys*. **1998**, *109*, 2176−2184.

(62) Garzon, I. L.; Michaelian, K.; Beltran, M. R.; Posada-Amarillas, A.; Ordejon, P.; Artacho, E.; Sanchez-Portal, D.; Soler, J. M. *Phys. Rev. Lett.* **1998**, *81*, 1600−1603.

(63) Garzon, I. L.; Michaelian, K.; Beltran, M. R.; Posada-Amarillas, A.; Ordejon, P.; Artacho, E.; Sanchez-Portal, D.; Soler, J. M. *Eur. Phys. J. D* **1999**, *9*, 211−215.

(64) Michaelian, K.; Rendon, N.; Garzon, I. L. *Phys. Rev. B* **1999**, *60*, 2000−2010.

(65) Yudanov, I. V.; Vent, S.; Neyman, K.; Pacchioni, G.; Rosch, N. *Chem. Phys. Lett*. **1997**, *275*, 245−252.

(66) Neyman, K. M.; Vent, S.; Pacchioni, G.; Rosch, N. *Nuovo Cimento* **1997**, *19*, 1743−1748.

(67) Matveev, A. V.; Neyman, K. M.; Yudanov, I. V.; Rosch, N. *Surf. Sci*. **1999**, *426*, 123−139.

(68) Neyman, K. M., Inntam, C.; Nasluzov, V. A.; Kosarev, R.; Rosch, N. *Appl. Phys. A* **2004**, *78*, 823−828.

(69) Markovits, A.; Skalli, M. K.; Minot, C.; Pacchioni, G.; Lopez, N.; Illas, F. *J. Chem. Phys.* **2001**, *115*, 8172−8177.

(70) Menetrey, M.; Markovits, A.; Minot, C.; Del Vitto, A.; Pacchioni, G. *Surf. Sci*. **2004**, *549*, 294−304.

(71) Chiesa, M.; Paganini, M. C.; Giamello, E.; Di Valentin, C.; Pacchioni, G. *Angew. Chem., Int. Ed.* **2003**, *42*, 1759−1761.

(72) Pacchioni, G. *ChemPhysChem* **2003**, *4*, 1041−1047.

(73) Dominguez-Ariza, D.; Sousa, C.; Illas, F.; Ricci, D.; Pacchioni, G *Phys. Rev. B* **2003**, *68*, 054101.

(74) Ricci, D.; Di Valentin, C.; Pacchioni, G.; Sushko, P. V.; Shluger, A. L.; Giamello, E. *J. Am. Chem. Soc*. **2003**, *125*, 738−747.

(75) Ricci, D.; Pacchioni, G.; Sushko, P. V.; Shluger, A. L. *J. Chem. Phys*. **2002**, *117*, 2844−2851.

(76) Soave, R.; Ferrari, A. M.; Pacchioni, G. *J. Phys. Chem. B* **2001**, *105*, 9798−9804.

(77) Pacchioni, G. *Solid State Sci.* **2000**, 2, 161−179.

(78) Pacchioni, G.; Ferrari, A. M. *Catal. Today* **1999**, *50*, 533−540.

(79) Sousa, C.; Pacchioni, G.; Illas, F. *Surf. Sci*. **1999**, *429*, 217−228.

(80) Pacchioni, G.; Pescarmona, P. *Surf. Sci.* **1998**, *412/413*, 657−671.

(81) Pacchioni, G.; Ferrari, A. M.; Ieranò, G. *Faraday Discuss*. **1997**, *106*, 155−172.

(82) Giamello, E.; Paganini, M. C.; Murphy, D. M.; Ferrari, A. M.; Pacchioni, G. *J. Phys. Chem. B* **1997**, *101*, 971−982.

(83) Fucks, D.; Dorfman, S.; Kotomin, E. A.; Zhukovskii, Y. F.; Stoneham, M. *Phys. Rev. Lett*. **2000**, *85* (20), 4333−4336.

(84) Kuzovkov, V. N.; Popov, A. I.; Kotomin, E. A.; Monge, M. A.; Gonzalez, R.; Chen, Y. *Phys. Rev. B* **2001**, *64*, 1−5.

(85) Giordano, L.; Goniakowski, J.; Pacchioni, G. *Phys. Rev. B* **2001**, *75*, 1−9.

(86) Lopez, N.; Illas, F.; Rosch, N.; Pacchioni, G. *J. Chem. Phys*. **1999**, *110*, 4873−4879.

(87) Worz, A. S.; Judai, K.; Abbet, S.; Antonietti, J. M.; Heiz, U.; Del Vitto, A.; Giordano, L.; Pacchioni, G. *Chem. Phys. Lett.* **2004**, *399*, 266−270.

(88) Giordano, L.; Del Vitto, A.; Pacchioni, G.; Ferrari, A. M. *Surf. Sci*. **2003**, *540*, 63−75.

(89) Abbet, S.; Riedo, E.; Brune, H.; Heiz, U.; Ferrari, A. M.; Giordano, L.; Pacchioni, G. *J. Am. Chem. Soc*. **2001**, *123*, 6172−6178.

(90) Abbet, S.; Heiz, U.; Ferrari, A. M.; Giordano, L.; Di Valentin, C.; Pacchioni, G. *Thin Solid Films* **2001**, *400*, 37−42.

(91) Neyman, K.; Rosch, N.; Pacchioni, G. *Appl. Catal., A* **2000**, *191*, 3−13.

(92) Abbet, S.; Sanchez, A.; Heiz, U.; Schneider, W. D.; Ferrari, A. M.; Pacchioni, G.; Rosch, N. *J. Am. Chem. Soc*. **2000**, *122*, 3453−3457.

(93) Pacchioni, G.; Rosch, N. *J. Chem. Phys.* **1996**, *104*, 7329−7337.

(94) Neyman, K. M.; Vent, S.; Rosch, N.; Pacchioni, G. *Top. Catal*. **1999**, *9*, 153−161.

(95) Matveev, A. V.; Neyman, K. M.; Pacchioni, G.; Rosch, N. *Chem. Phys. Lett*. **1999**, *299*, 603−612.

(96) Ferrari, A. M.; Xiao, C.; Neyman, K. M.; Pacchioni, G.; Rosch, N.; *Phys. Chem. Chem. Phys*. **1999**, *1*, 4655−4661.

(97) Giordano, L.; Pacchioni, G.; Ferrari, A. M.; Illas, F.; Rosch, N. *Surf. Sci*. **2001**, *473*, 213−226.

(98) Giordano, L.; Pacchioni, G.; Illas, F.; Rosch, N. *Surf. Sci*. **2002**, *499*, 73−84.

(99) Di Valentin, C.; Giordano, L.; Pacchioni, G.; Rosch, N. *Surf. Sci.* **2003**, *522*, 175−184.

(100) Giordano, L.; Di Valentin, C.; Goniakowski, J.; Pacchioni, G. *Phys. Rev. Lett*. **2004**, *92*, 096105.

(101) Oviedo, J.; Sanz, J. F.; Lopez, N.; Illas, F. *J. Phys. Chem. B* **2000**, *104*, 4342−4348.

(102) Goniakowski, J. *Phys. Rev. B* **1998**, *58*, 1189−1192.

(103) Zhukovskii, Y. F.; Kotomin, E. A.; Fucks, D.; Dorfman, S. *Superlattices Microstruct.* **2004**, *36*, 63−72.

(104) Zhukovskii, Y. F.; Kotomin, E. A.; Jacobs, P. W. M.; Stoneham, A. M. *Phys. Rev. Lett*. **2000**, *84*, 1256−1259.

(105) Zhukovskii, Y. F.; Kotomin, E. A.; Fucks, D.; Dorfman, S.; Stoneham, A. M.; Sychev, O.; Borstel, G. *Appl. Surf. Sci*. **2004**, *226*, 298−305.

(106) Zhukovskii, Y. F.; Kotomin, E. A.; Fucks, D.; Dorfman, S. *Surf. Sci*. **2004**, *566−568*, 122−129.

(107) Zhukovskii, Y. F.; Kotomin, E. A.; Jacobs, P. W. M.; Stoneham, A. M.; Harding, J. H. *J. Phys: Condens. Matter* **2000**, *12*, 55−66.

(108) Zhukovskii, Y. F.; Kotomin, E. A. *Phys. Status Solidi* **2005**, *2*, 347−350.

(109) Herschend, B.; Hermansson, K.; Alfredsson, M.; Zhukovskii, Y. F.; Kotomin, E. A.; Jacobs, P. W. M. *J. Phys. Chem. B* **2003**, *107*, 11893−11899.

(110) Fucks, D.; Dorfman, S.; Zhukovskii, Y. F.; Kotomin, E. A.; Stoneham, A. H. *Surf. Sci.* **2002**, *499*, 24−40.

(111) Vervisch, W.; Mottet, C.; Goniakovskii, J. *Phys. Rev. B* **2002**, *65*, 245411.

(112) Vervisch, W.; Mottet, C.; Goniakovskii, J. *Eur. Phys. J. D* **2004**, *24*, 311−314.

(113) Mottet, C.; Goniakovskii, J. *Surf. Sci*. **2004**, *566*−*568*, 443−450.

(114) Baroni, S.; Del Corso, A.; de Gironcoli, S.; Giannozzi, P. *PWscf*; Scuola Internazionale Superiore di Studi Avanzati and DEMOCRITOS National Simulation Center: Trieste, Italy. http://www.pwscf.org.

(115) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671−6687.

(116) Fortunelli, A. *J. Phys. Chem.* **1995**, *99*, 9056−9061.

(117) Del Vitto, A.; Pacchioni, G.; Delbecq, F.; Sautet, P. *J. Phys. Chem. B* **2005**, 109, 8040−8048.

CT050073E

JCTC Journal of Chemical Theory and Computation

Intrinsic Relative Stabilities of the Neutral Tautomers of Arginine Side-Chain Models

Jan Norberg,*,† Nicolas Foloppe,‡ and Lennart Nilsson†

Center for Structural Biochemistry, Department of Biosciences at Novum, Karolinska Institutet, SE-141 57 Huddinge, Sweden

Received December 18, 2004

Abstract: The specific protonation state of amino acids is crucial for the physicochemical properties of proteins and their biological functions. These protonation states influence, for instance, properties related to hydrogen bonding, solubility, and folding. pK_a calculations for proteins are, therefore, important and require, in principle, a specification of the most stable protonated and deprotonated forms of each titratable group. This is complicated by the existence of multiple tautomers, like the five neutral tautomers of the guanidine moiety in arginine. In this study, the compounds *N*-methyl-guanidine and *N*-ethyl-guanidine were used to model the charged and all neutral protonation states of the arginine side chain. The relative stabilities of all five neutral tautomers were investigated systematically for the first time, using quantum-mechanical calculations. These relative stabilities were obtained in vacuo, water and chloroform, by combining the quantum-mechanical calculations with a continuum solvation model. The water model was used to represent arginines exposed to an aqueous solution, whereas the chloroform model has a polarity representative of a protein core or a membrane. This allowed determining the relative pK_a's associated with each neutral tautomer in these environments. A key result is that significant differences in stability are found between the neutral tautomers, in both water and chloroform. Some tautomers are consistently found to be the most stable. These findings will be helpful to refine pK_a calculations in proteins.

Introduction

The guanidine is found at the extremity of arginine side chains and is, therefore, an important biochemical building block. This guanidine functionality is usually assumed to be in the protonated guanidinium form because of its high intrinsic basicity, with an effective pK_a of about 12.0 in aqueous solution at 298 K.[1] There are several examples where arginine is found to play a key role in relation to the mechanism of action of a protein.[2−5] For instance, the amino acid compositions of the interfaces of a large number of protein−protein[6−8] and protein-nucleic acid[9] complexes have been characterized, showing that arginine occurs frequently at the interface of the complexes.[8,9] Other specific examples include the activation of G protein coupled receptors,[10] the voltage gating of ion channels,[11] a possible role for neutral arginine in porin channel selectivity,[12] and the substrate binding of the human FLAP endonuclease-1 enzyme.[13] Peptides containing arginine have an important role in proteomics projects, which use mass spectrometry for rapid and reliable identification of proteins.[14−16] The fragmentation efficiency depends strongly on the protonation state of arginine.

Despite being very basic, the guanidine functionality is included as a titratable site in methods that aim at determining the protonated state of proteins.[17−20] These methods typically rely on the solution of the Poisson−Boltzmann equation in the framework of continuum electrostatics[18,20] and calculate the pK_a shift of a titratable group when it is transferred from

* Corresponding author tel.: +46 8 608 9263, fax: +46 8 608 9290, e-mail: jan.norberg@biosci.ki.se.

† Karolinska Institutet.

‡ Present address: Vernalis (R&D) Ltd., Granta Park, Abington, Cambridge CB1 6GB, U. K.

10.1021/ct049849m CCC: $30.25
Published on Web 07/13/2005

aqueous solution to the protein. The calculations involve both the protonated and the nonprotonated states of each titratable group in the protein environment.

In these pK_a calculations, one proton has to be removed from the guanidinium functionality of each arginine side chain. Details of the positioning of the polar hydrogens can have a significant impact on the pK_a's calculated in a protein environment,[21−23] and there are, a priori, five acidic protons on the guanidinium group, corresponding to five different neutral guanidine tautomers. Because none of these five protons are rigorously chemically equivalent, it would be very useful to have a ranking of their relative intrinsic acidities. This information is, to our knowledge, not available, probably because it would be difficult to determine the corresponding microscopic pK_a values experimentally. This, indeed, would require one to distinguish experimentally between each of these five neutral tautomers in aqueous solution. Alternatively, the guanidinium and each of the five neutral tautomers can be studied independently using a computational approach.

Recent studies have shown that the relative pK_a's of a series of analogous compounds in aqueous solution can be calculated with reasonable accuracy, using high level ab initio quantum-chemistry calculations, in combination with a continuum model to represent the condensed phase dielectric properties.[24−26] One study that tested several continuum models[24] found a good agreement between experimental and calculated pK_a's when the aqueous environment was represented using the isodensity surface polarized continuum model (IPCM).[27] The present report applies a similar protocol to rank the relative intrinsic acidities associated with the arginine side chain, modeled with two compounds, *N*-methylguanidine (Me-Gua) and *N*-ethyl-guanidine (Et-Gua), to test the effect of the length of the hydrocarbon side chain on the results. The relative acidities have been investigated in water, as well as in a medium with a low dielectric constant, to represent a protein core or a membrane environment. The results do indicate that there are significant stability differences between the neutral tautomers of an arginine side chain.

Computational Methodology

The compounds used to model the arginine side chain were the protonated and neutral forms of Me-Gua (Figure 1) and Et-Gua (Figure 2). The protonated forms (guanidinium) of Me-Gua (Figure 1) and Et-Gua (Figure 2) are denoted Me-GuaP and Et-GuaP, respectively. The five neutral tautomers are referred to as Me-GuaA (Et-GuaA) to Me-GuaE (Et-GuaE) (Figure 1), with Me-GuaX/Y (Et-GuaX/Y) denoting any of the neutral forms. Although larger molecules, even arginine itself, would be tractable on modern computers, it is on purpose that the model compounds were kept as simple as possible. This is indeed needed to assess the true intrinsic properties of the guanidine group, without the complications of conformational issues that inevitably arise with larger compounds. Larger compounds also open the possibility of intramolecular interactions involving the guanidine moiety (e.g., folded structures), which would only obscure the results regarding the intrinsic preferences for the protonation states.

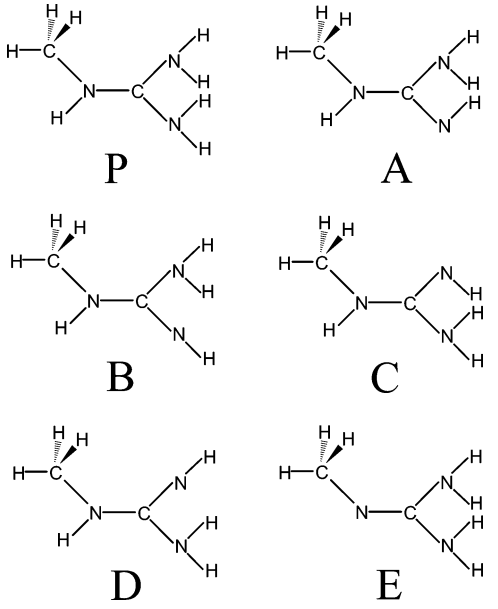


Figure 1. Structures of the protonated (Me-GuaP) and neutral (Me-GuaA to Me-GuaE) forms of Me-Gua.

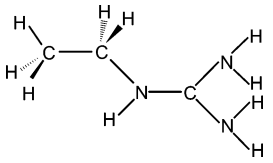


Figure 2. Structure of the protonated (Et-GuaP) form of Et-Gua. The same naming scheme was used for the Et-Gua series as for the Me-Gua series; therefore, the neutral tautomers of Et-Gua are not shown explicitly.

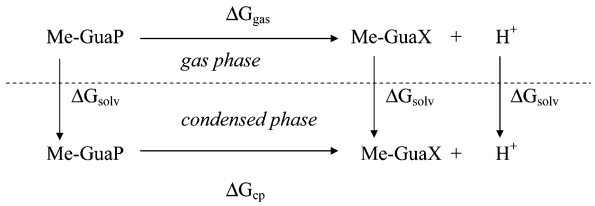


Figure 3. Thermodynamics cycle used for deriving the relative pK_a's, using Me-Gua as an example.

The dissociation constant K_a of an acid HA into its conjugate base A^- and a proton H^+ is related to the Gibbs free energy change (ΔG) of this dissociation reaction by

$$\Delta G = -RT \ln K_a = G(H^+) + G(A^-) - G(HA) \quad (1)$$

where R is the molar gas constant and T is the temperature; our calculations used $T = 298$ K. The absolute pK_a for the dissociation reaction is then

$$pK_a = \frac{1}{2.303RT} \Delta G \quad (2)$$

The absolute pK_a is difficult to calculate, in particular, because of problems associated with the determination of $G(H^+)$.[28−30] This study, therefore, focuses on the relative stabilities of the neutral tautomers of Me-Gua and Et-Gua. This can be investigated using a thermodynamic cycle (Figure 3), from which the pK_a difference, ΔpK_a, corre-

sponding to the difference in acidity associated with two different neutral tautomers can be obtained. The cycle relates the acid dissociation in the condensed phase (ΔG_{cp}) to its dissociation in the gas phase (ΔG_{gas}) and to the solvation free energies (ΔG_{solv}) of all the chemical species involved. Accordingly, if the acid Me-GuaP dissociates to produce one of its conjugate bases Me-GuaX, the corresponding free energy difference and the associated microscopic absolute pK_a in the condensed phase are given by

$$\Delta G_{cp}(\text{Me-GuaX}) = 2.303RT\,\text{p}K_a(\text{Me-GuaX}) = \Delta G_{solv}(\text{Me-GuaX}) + \Delta G_{solv}(H^+) - \Delta G_{solv}(\text{Me-GuaP}) + \Delta G_{gas}(\text{Me-GuaX}) \quad (3)$$

where ΔG_{solv}(Me-GuaX) and ΔG_{solv}(Me-GuaP) are the free energies of solvation of a neutral guanidine tautomer and the guanidinium, respectively. $\Delta G_{solv}(H^+)$ is the free energy of solvation for the proton, and ΔG_{solv}(Me-GuaX) is the deprotonation free energy of the guanidinium in the gas phase. Subtracting eq 3, obtained with a neutral tautomer, Me-GuaY, from its counterpart for another neutral tautomer, Me-GuaX, yields

$$\Delta\Delta G_{cp}(\text{Me-GuaX} - \text{Me-GuaY}) = \Delta G_{solv}(\text{Me-GuaX}) - \Delta G_{solv}(\text{Me-GuaY}) + \Delta G_{gas}(\text{Me-GuaX}) - \Delta G_{gas}(\text{Me-GuaY}) \quad (4)$$

A key point is that the free energies of solvation of the proton and the protonated guanidinium from eq 3 cancel when eq 4 is formed. Therefore, the relative pK_a difference between two distinct neutral guanidine tautomers depends only on (i) the difference in free energy of solvation between these neutral tautomers [ΔG_{solv}(Me-GuaX) − ΔG_{solv}(Me-GuaY)] and (ii) the difference in the gas-phase protonation free energies associated with these neutral tautomers [ΔG_{gas}(Me-GuaX) − ΔG_{gas}(Me-GuaY)]. The gas-phase protonation free energies were computed using ab initio quantum mechanics, and the solvation free energies have been calculated using a continuum solvation model.[27]

All calculations were performed using the Gaussian 98 suite of programs.[31] The 6-31G(d,p) basis set was used for all compounds. Further, the 6-31+G(d,p) and 6-311+G(d,p) basis sets were also applied to test the influence of the basis set on selected calculations in vacuo. First, the geometries of all the compounds were optimized by energy minimization in the gas phase to the default tolerances of the program. The gas-phase energy-minimized geometries were then used directly for all the calculations in the condensed phase. This is expected to be a reasonable approximation, given that there is no torsional freedom within the model compounds that could be drastically affected by the presence of a solvent. Also, a recent study on similar compounds tested the effect of a solvent on the internal geometry, and this influence was found to be minor, as expected.[32]

The solvation free energies were obtained using the reaction-field IPCM,[27] which previously led to satisfactory agreement with experimental results for the relative pK_a's in a series of pyridine derivatives.[24] The solvation free energies were obtained at 298 K by calculating single-point energies with the IPCM solvation model, for both water and chloroform. The dielectric constants of water and chloroform were set to 78.56 and 4.806, respectively.[33] In the IPCM model, the isodensity cutoff value was 0.001 atomic units.[24]

All calculations were performed at three levels of theory: (i) Hartree−Fock (HF), (ii) second-order Møller−Plesset (MP2),[34] and (iii) with the density functional theory using the B3LYP method.[35] Gas-phase free energy differences were calculated by adding the vibrational, rotational, and translational corrections at 298 K to the differences in heat of formation obtained directly from Gaussian.[31] This thermal free energy correction was determined from a normal-mode analysis of the previously geometry-optimized structures in vacuo at the HF and B3LYP levels of theory. The gas-phase proton affinity (PA) was computed as the change in the heat of formation and in the zero-point vibrational energy:

$$\text{PA} = H_f(\text{Me-GuaX}) - H_f(\text{Me-GuaP}) + \text{ZPVE(Me-GuaX)} - \text{ZPVE(Me-GuaP)} \quad (5)$$

where H_f(...) and ZPVE(...) are the heat of formation and the zero-point vibrational energy, respectively.

Results

The properties of interest were determined for all compounds at the HF/6-31G(d,p), B3LYP/6-31G(d,p), and MP2/6-31G(d,p) levels in the gas phase and in water and at the MP2/6-31G(d,p) level in chloroform. For the Me-GuaD, Me-GuaE, and Et-GuaE model compounds, a tendency toward a more pyramidal conformation of the amino groups was observed, as has also been reported for the amino groups of guanidine and guanidinium cations.[36] The HF and B3LYP results are presented for the sake of completeness, but our analysis focuses on the MP2 results. Overall, for a given model compound, the results for the different levels of theory were quite consistent and do reveal differences across the neutral tautomers.

Basis Set Dependence and Bond Order Analysis. Given that the amino groups were found to be planar or pyramidal in different tautomers of methyl-guanidine with the 6-31G(d,p) basis set, the influence of larger basis sets on these geometries was tested. These calculations were done with the 6-31+G(d,p) and 6-311+G(d,p) basis sets and the B3LYP and MP2 levels of theory for Me-GuaB, Me-GuaD, and Me-GuaE. These were chosen because the Me-GuaB was planar at the MP2/6−31(d,p) level, whereas Me-GuaD and Me-GuaE were slightly pyramidal. Me-GuaB remained planar for both the larger basis sets. Also, Me-GuaD and Me-GuaE showed the same pyramidal tendency for the larger basis sets as they did for the 6-31G(d,p) basis set.

Heats of formation computed for the three compounds using B3LYP and MP2 with the two larger basis sets were ranked in the same order as with the 6-31G(d,p) basis set. These differences in the heats of formation, for a given compound, computed with the three different basis sets were <0.5 kcal/mol with the MP2 method and <1.4 kcal/mol with B3LYP.

A bond order (BO) analysis was made at the MP2 level with the 6-31G(d,p) basis set for the protonation sites that showed a tendency for a pyramidal conformation. The BO

Table 1. Relative Stabilities (kcal/mol) ($\Delta\Delta G$ from eq 4) between the Different Protonation States of Me-Gua and Et-Gua from the MP2/6-31G(d,p) Calculations

compound	water	chloroform
Me-GuaB−Me-GuaA	−1.05	−0.58
Me-GuaC−Me-GuaA	−1.92	−1.49
Me-GuaD−Me-GuaA	−4.70	−5.21
Me-GuaE−Me-GuaA	−3.35	−3.73
Me-GuaC−Me-GuaB	−0.87	−0.91
Me-GuaD−Me-GuaB	−3.65	−4.63
Me-GuaE−Me-GuaB	−2.30	−3.15
Me-GuaD−Me-GuaC	−2.78	−3.72
Me-GuaE−Me-GuaC	−1.43	−2.24
Me-GuaE−Me-GuaD	1.35	1.48
Et-GuaB−Et-GuaA	0.79	0.55
Et-GuaC−Et-GuaA	−0.69	−0.71
Et-GuaD−Et-GuaA	−1.25	−1.10
Et-GuaE−Et-GuaA	−3.21	−4.33
Et-GuaC−Et-GuaB	−1.48	−1.26
Et-GuaD−Et-GuaB	−2.04	−1.65
Et-GuaE−Et-GuaB	−4.00	−4.88
Et-GuaD−Et-GuaC	−0.56	−0.39
Et-GuaE−Et-GuaC	−2.52	−3.62
Et-GuaE−Et-GuaD	−1.96	−3.23

was about 0.99 and 1.22 for the C−N bond in the CNH_2 and CNH groups, respectively, and about 0.67−0.69 for the N−H bonds. These bond orders were observed for both Me-GuaD and Me-GuaE.

Relative Stabilities of Neutral Tautomers in Water and Chloroform. The relative stabilities of the five neutral tautomers for the two model compounds of the arginine side chain, Me-Gua and Et-Gua (Figures 1 and 2), were determined in water and in chloroform (Table 1 and Figures 4 and 5), using the thermodynamic cycle in Figure 3. The components of this cycle are presented in the sections below. For Me-Gua in water and chloroform, the A state was consistently the least stable and the D and E states were the most stable. The E state was also the most stable at the MP2 level with Et-Gua in both water and chloroform. The B rather than the A state, however, was found to be the least stable with Et-Gua. The B tautomer is also the second least stable state with Me-Gua at the MP2 level of theory. Arguably, the main differences between Me-Gua and Et-Gua are the

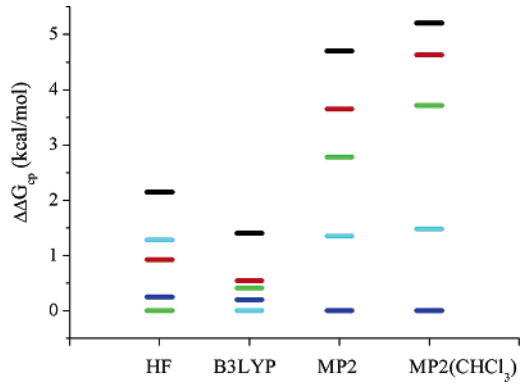


Figure 4. $\Delta\Delta G_{cp}$ in kcal/mol for the different protonation states of Me-Gua (A, black; B, red; C, green; D, blue; E, cyan).

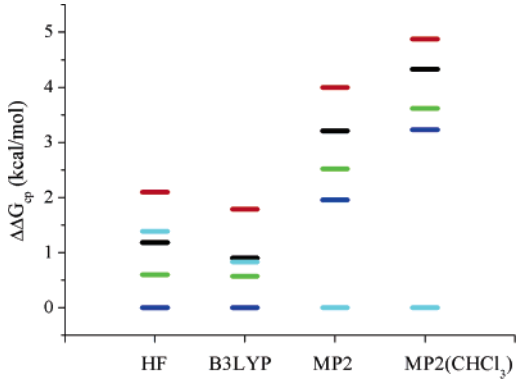


Figure 5. $\Delta\Delta G_{cp}$ in kcal/mol for the different protonation states of Et-Gua (A, black; B, red; C, green; D, blue; E, cyan).

Table 2. Gas-Phase Proton Affinity (kcal/mol) of Neutral Forms of Me-Gua and Et-Gua

compound	HF/ 6−31G(d,p)	B3LYP/ 6−31G(d,p)	MP2/ 6−31G(d,p)
Me-GuaA	252.35	249.25	247.65
Me-GuaB	252.62	249.62	247.94
Me-GuaC	251.19	248.85	246.68
Me-GuaD	251.57	249.00	241.88
Me-GuaE	250.90	246.76	243.54
Et-GuaA	253.93	251.00	249.14
Et-GuaB	254.19	251.34	249.42
Et-GuaC	252.79	250.56	248.31
Et-GuaD	253.20	250.76	248.64
Et-GuaE	253.50	250.36	242.60

relative rankings of states A and B and states D and E. The longer hydrocarbon side chain of Et-Gua, thus, has an influence on the relative stability of the neutral tautomers; however, the pattern of relative stabilities observed with the two model compounds is largely consistent.

Not only do the present results reveal differences in stability for the neutral tautomers, but these differences spread on a significant free energy range. In water, the difference between the most and least stable tautomers is at least 4.0 kcal/mol, well-above the thermal energy at 298 K. This difference is even more pronounced in chloroform. Interestingly, however, the relative ranking remains the same in chloroform as in water, for a given compound (Figures 4 and 5). Therefore, a largely consistent picture emerged regarding the relative intrinsic stabilities of the neutral tautomers of the compounds representing the arginine side chain.

Gas-Phase Proton Affinities and Free Energies. The gas-phase proton affinities at the MP2 level ranged from 241.88 to 247.94 kcal/mol for Me-Gua and from 242.60 to 249.42 kcal/mol for Et-Gua (Table 2). The B3LYP method led to slightly larger gas-phase proton affinities than with MP2; however, the differences in proton affinity across the tautomers are less pronounced when obtained with B3LYP than with MP2. For all states except E, the MP2 gas-phase proton affinities were slightly larger for the compounds with the longer hydrocarbon side chain, by about 1% on average. This increase of the proton affinity with an increase in the side chain length is compatible with the slightly smaller

Table 3. Change in Heat of Formation (ΔH_f) between the Neutral and Protonated Forms and Thermal Correction (kcal/mol) of the Neutral Forms of Me-Gua and Et-Gua Used in the MP2/6-31G(d,p) Calculations

compound	ΔH_f	thermal correction[a]
Me-GuaA	258.05	−8.58
Me-GuaB	258.30	−8.74
Me-GuaC	257.09	−9.27
Me-GuaD	252.24	−9.00
Me-GuaE	252.46	−7.83
Et-GuaA	259.52	−8.52
Et-GuaB	259.75	−8.66
Et-GuaC	258.75	−9.30
Et-GuaD	258.98	−8.97
Et-GuaE	253.39	−9.41

[a] The thermal correction to get the gas-phase Gibbs free energy difference at 298 K from ΔH_f.

proton affinities (about 233−238 kcal/mol) obtained quantum-mechanically in the absence of any side chain for guanidine.[36−38] These differences with previous theoretical results may also be, in part, ascribed to the differences in the levels of theory used. These results are quite similar to the experimental value of 235.7 kcal/mol obtained for guanidine,[39] suggesting that the protonated and neutral forms of these two compounds are valid models of the protonation states of the arginine side chain.

The variation in the MP2 proton affinities in a range of about 6.0 kcal/mol clearly shows that there are intrinsic differences associated with the protonation sites and their chemistry in the guanidine moiety. For both Me-Gua and Et-Gua, the largest gas-phase proton affinities were obtained for the B state (Table 2); however, the proton affinities of states A and C were close, within 1.3 kcal/mol. The behavior of tautomer E is also consistent across the two model compounds, with its relatively low proton affinity. This peculiarity of state E is not surprising, given that it is associated with the deprotonation of the nitrogen that differs chemically most from the others. The main surprise came from state D, for which the relative proton affinity strongly depends on the model compound; that is, it is the lowest in Me-Gua but much higher in Et-Gua.

Gas-phase free energies were obtained from the heats of formation and the thermal corrections (Table 3), which, as expected,[40] were quite similar for our compounds. The correction contributed less than 0.9 kcal/mol to the $\Delta\Delta G_{cp}$ in all cases except two involving Me-GuaE, for which it was up to 1.4 kcal/mol because of the smaller correction obtained for Me-GuaE.

Solvation Free Energy in Water and Chloroform. The solvation free energy in water of the protonated species ranged from about −64 to −62 kcal/mol for Me-GuaP and from −62 to −60 kcal/mol for Et-GuaP, depending on the level of theory (Table 4). As expected, the corresponding free energies for the neutral tautomers are much less favorable, by about 50 kcal/mol. The free energies of solvation of the neutral tautomers in water vary in a range of 2.7 kcal/mol in Me-Gua and 4.1 kcal/mol in Et-Gua. In that respect, states A, B, and C are rather similar, differing by 1.1 kcal/mol at most in each model compound. Neutral tautomer E stands apart in both model compounds, with a relatively unfavorable free energy of solvation in water. Surprisingly, state D in water goes from being the least well solvated in Me-GuaD to being the best solvated in Et-Gua.

As expected, the chloroform solvation free energies were found to be significantly less favorable compared to the aqueous solvation free energies for both charged and neutral species. Et-Gua was more soluble than Me-Gua in chloroform. The ranking of the MP2 solvation free energies of the neutral states in chloroform and water was the same for Et-Gua and only slightly different for Me-Gua.

Dipole Moments. The electric dipole moments of the neutral compounds are reported in the gas phase, water, and chloroform (Table 5) to examine their relationship with solvation properties. In that context, the present work is only concerned with the magnitude of these dipoles and not their orientation. In the gas phase, the dipole varies from ∼2.5 to ∼4.5 D with both model compounds, showing that the neutral tautomers vary in overall polarity. For these dipoles, the same pattern of similarities across the tautomers is observed as those for the proton affinities and the solvation energies. Indeed, states A, B, and C of Me-Gua and states A, B, C, and D of Et-Gua have similar dipole moments. State E is an

Table 4. Solvation Free Energy in Water and Chloroform (kcal/mol) of the Protonated and Neutral Forms of Me-Gua and Et-Gua

	HF/6−31G(d,p) water	B3LYP/6−31G(d,p) water	MP2/6−31G(d,p) water	MP2/6−31G(d,p) chloroform
Me-GuaP	−64.09	−62.45	−62.92	−49.48
Et-GuaP	−62.28	−60.12	−60.52	−47.32
Me-GuaA	−12.23	−10.79	−10.43	−7.29
Me-GuaB	−13.53	−11.83	−11.58	−7.97
Me-GuaC	−12.53	−10.98	−10.71	−7.14
Me-GuaD	−12.88	−11.24	−8.90	−6.27
Me-GuaE	−10.93	−9.30	−8.95	−6.19
Et-GuaA	−15.81	−13.50	−12.78	−8.80
Et-GuaB	−14.96	−12.78	−12.08	−8.34
Et-GuaC	−14.54	−12.44	−11.92	−7.96
Et-GuaD	−15.76	−13.64	−13.04	−8.91
Et-GuaE	−14.69	−12.46	−8.97	−6.11

Table 5. Dipole Moment (D) of the Neutral Forms of Me-Gua and Et-Gua in the Gas Phase, Chloroform, and Aqueous Solution

	HF/6−31G(d,p)		MP2/6−31G(d,p)			B3LYP/6−31G(d,p)	
compound	gas phase	aqueous	gas phase	chloroform	aqueous	gas phase	aqueous
Me-GuaA	3.98	5.04	4.08	4.46	4.78	3.63	4.76
Me-GuaB	4.19	5.48	4.30	4.84	5.23	3.84	5.22
Me-GuaC	3.93	5.02	4.03	4.46	4.82	3.73	4.87
Me-GuaD	4.11	5.13	3.04	3.22	3.42	3.86	4.94
Me-GuaE	2.57	3.51	2.47	2.77	3.22	2.30	3.06
Et-GuaA	4.02	5.40	4.12	4.69	5.10	3.68	5.12
Et-GuaB	4.34	5.60	4.45	4.96	5.32	4.00	5.34
Et-GuaC	3.82	5.20	3.92	4.49	4.96	3.62	5.02
Et-GuaD	4.17	5.60	4.27	4.96	5.41	3.92	5.46
Et-GuaE	3.80	4.86	2.45	2.71	3.04	3.62	4.66

outlier, with a significantly smaller dipole in both model compounds.

As expected, the dipole moments become systematically larger when going from the gas phase to a solvent, reflecting the influence of the solvent reaction field. This polarization is significantly more pronounced in water than in chloroform, as anticipated. When compared to the gas phase, the dipoles in water are on average 20.3% (Me-Gua series) and 24.1% (Et-Gua series) larger. When going from the gas phase to chloroform, the dipoles increase by 12.1% (Me-Gua) and 13.3% (Et-Gua) on average. These averages mask significant differences among the tautomers; for example, the dipole of Me-GuaD increases by only 12.5% in water, although the corresponding increase is 30.4% for Me-GuaE. It is interesting to note that Me-GuaE is both the least polar and yet the most polarizable tautomer in the Me-Gua series. Therefore, the neutral tautomers differ not only in polarity but also in polarizability.

Concluding Discussion

Although it is well-known that the guanidine group of arginine is strongly basic, the relative intrinsic stabilities of the corresponding neutral tautomers are unknown. In particular, this means that one of the neutral tautomers is chosen arbitrarily when performing pK_a calculations on proteins, using the increasingly popular continuum electrostatic techniques. This adds practical concerns to the broader and more fundamental question of the relative basicities of the guanidine tautomers. Indeed, it is now well-documented that pK_a calculations on full proteins are influenced by the detailed positioning of the polar hydrogens.[21−23] This is why the selection of a given guanidine tautomer will not only impact the calculated pK_a of this guanidine but also influence the electrostatic environment of other titratable groups in the vicinity of this guanidine.

Here, a variety of quantum-mechanical methods and model compounds were used to investigate the relative stabilities of the five neutral guanidine tautomers. This analysis was carried out in aqueous solution and in chloroform, which was used to mimic a protein core or a membrane environment. Like the presentation of the results, this discussion concentrates on the results obtained at the MP2 level. When examining each component of the free energy cycle, it is clear that the dominant contribution to $\Delta\Delta G_{cp}$ is from the heat of formation, with the thermal correction and the free energy of solvation providing a smaller modulation.

The model compounds Me-Gua and Et-Gua were selected because their size allows high level ab initio calculations to be carried out. In addition, they contain as few torsional degrees of freedom as possible, which allows derivation of properties that are truly intrinsic to the guanidinium moiety, without being confounded by conformational issues. For instance, using these compounds avoids formation of intramolecular hydrogen bonds that could be formed between the guanidine moiety and other polar groups in a full arginine,[41] which would cloud the interpretation of the results. We tested the effect of this simplification by using two model compounds, differing by the length of the aliphatic side chain. The pattern of behavior for the five tautomers is mostly similar between the Me-Gua and Et-Gua series, with the exception of state D. Me-GuaD presents surprising differences as compared to Me-GuaB, with respect to proton affinity and free energy of solvation in water and chloroform. These differences are surprising because the local topological arrangement of the protons seems quasi-equivalent in neutral states B and D and would be expected to yield similar results. This could be due to the tendency of the amino groups of the D state to form a pyramidal conformation instead of a planar conformation. States B and D, however, are much more similar in the Et-Gua series, as anticipated intuitively. This, combined with the fact that Et-Gua is, a priori, a closer analogue of the arginine side chain than Me-Gua, suggests that the interpretation of the results should put more weight on the Et-Gua series. It is, however, the similarities between both series, rather than their differences, that dominate the results.

Tautomer E is consistently an outlier, with relatively favorable proton affinities but relatively unfavorable solvation free energies in water or chloroform. In differential terms, however, the proton affinity of Et-GuaE more than compensates for its solvation free energy. That leads to tautomer E in water or chloroform being clearly the most stable in the Et-Gua series, by a distinctive margin (Figure 5). That tautomer E should differ from the other neutral tautomers is not surprising, given that it involves the deprotonation of the nitrogen that differs chemically most from the others.

The local symmetry of the guanidinium also suggests that tautomers A and C may behave similarly, which is indeed observed in the Et-Gua series. The same reasoning would also imply a similar behavior for tautomers B and D; however, that is not observed. Instead, tautomer D is systematically more stable than B, illustrating the limits of chemical intuition and the need for detailed calculations. In sum, these calculations show that the relative intrinsic stabilities of the neutral tautomers differ, be it in a medium of high or low dielectric value.

With both media and model compounds, tautomers E and D appear to be the most stable and should, therefore, be deprotonated in priority, for example, during pK_a calculations. Relying on the Et-Gua series, tautomer E may be selected as the most likely neutral form, given that the free energy difference between Et-GuaE and its congeners in water is nearly 2.0 kcal/mol or more. The preference for Et-GuaE is even more marked in a medium of low dielectric value, which also represents the conditions in which the neutral guanidine is most likely to exist. In both media, however, the intrinsic energy differences between the neutral tautomers (Table 1) are comparable to that associated with a single hydrogen bond, and therefore, the actual tautomeric state of the guanidine will be strongly influenced by the details of its environment. The present results should provide an energetic basis for a more rational positioning of the protons around a neutral guanidine moiety, a situation encountered, at least, during most pK_a calculations with proteins and continuum electrostatic methods.

References

(1) Nozaki, Y.; Tanford, C. Examination of titration behavior. *Methods Enzymol.* **1967**, *11*, 715−734.

(2) Ballesteros, J.; Kitanovic, S.; Guarnieri, F.; Davies, P.; Fromme, B. J.; Konvicka, K.; Chi, L.; Millar, R. P.; Davidson, J. S.; Weinstein, H.; Sealfon, S. C. Functional microdomains in G-protein-coupled receptors. The conserved arginine-cage motif in the gonadotropin-releasing hormone receptor. *J. Biol. Chem.* **1998**, *273*, 10445−10453.

(3) Barth, P.; Guillouard, I.; Setif, P.; Lagoutte, B. Essential role of a single arginine of photosystem I in stabilizing the electron-transfer complex with ferredoxin. *J. Biol. Chem.* **2000**, *275*, 7030−7036.

(4) Borgstahl, G. E.; Williams, D. R.; Getzoff, E. D. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* **1995**, *34*, 6278−6287.

(5) Demchuk, E.; Genick, U. K.; Woo, T. T.; Getzoff, E. D.; Bashford, D. Protonation states and pH titration in the photocycle of photoactive yellow protein. *Biochemistry* **2000**, *39*, 1100−1113.

(6) Jones, S.; Daley, D. T.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M. Protein-RNA interactions: a structural analysis. *Nucleic Acids Res.* **2001**, *29*, 943−954.

(7) Nadassy, K.; Wodak, S. J.; Janin, J. Structural features of protein-nucleic acid recognition sites. *Biochemistry* **1999**, *38*, 1999−2017.

(8) Lo Conte, L.; Chothia, C.; Janin, J. The atomic structure of protein−protein recognition sites. *J. Mol. Biol.* **1999**, *285*, 2177−2198.

(9) Jones, S.; van Heyningen, P.; Berman, H. M.; Thornton, J. M. Protein-DNA interactions: A structural analysis. *J. Mol. Biol.* **1999**, *287*, 877−896.

(10) Oliveira, L.; Paiva, A. C. M.; Vriend, G. A low resolution model for the interaction of G proteins with G protein-coupled receptors. *Protein Eng.* **1999**, *12*, 1087−1095.

(11) Jager, H.; Grissmer, S. Regulation of a mammalian Shaker-related potassium channel, hKv1.5, by extracellular potassium and pH. *FEBS Lett.* **2001**, *488*, 45−50.

(12) Karshikoff, A.; Spassov, V.; Cowan, S. W.; Ladenstein, R.; Schirmer, T. Electrostatic Properties of 2 Porin Channels from Escherichia-Coli. *J. Mol. Biol.* **1994**, *240*, 372−384.

(13) Qiu, J.; Bimston, D. N.; Partikian, A.; Shen, B. Arginine residues 47 and 70 of human flap endonuclease-1 are involved in DNA substrate interactions and cleavage site determination. *J. Biol. Chem.* **2002**, *277*, 24659−24666.

(14) Csonka, I. P.; Paizs, B.; Suhai, S. Modeling of the gas-phase ion chemistry of protonated arginine. *J. Mass Spectrom.* **2004**, *39*, 1025−1035.

(15) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J. Am. Chem. Soc.* **1996**, *118*, 8365−8374.

(16) Paizs, B.; Suhai, S.; Hargittai, B.; Hruby, V.; Somogyi, A. Ab initio and MS/MS studies on protonated peptides containing basic and acidic amino acid residues I. Solvated proton vs salt-bridged structures and the cleavage of the terminal amide bond of protonated RD-NH2. *Int. J. Mass Spectrom.* **2002**, *219*, 203−232.

(17) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* **1994**, *238*, 415−436.

(18) Bashford, D.; Karplus, M. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **1990**, *29*, 10219−10225.

(19) Beroza, P.; Case, D. A. Including side chain flexibility in continuum electrostatic calculations of protein titration. *J. Phys. Chem.* **1996**, *100*, 20156−20163.

(20) Demchuk, E.; Wade, R. C. Improving the continuum dielectric approach to calculating p*K*(a)s of ionizable groups in proteins. *J. Phys. Chem.* **1996**, *100*, 17373−17387.

(21) Alexov, E. G.; Gunner, M. R. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* **1997**, *72*, 2075−2093.

(22) You, T. J.; Bashford, D. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. *Biophys. J.* **1995**, *69*, 1721−1733.

(23) Nielsen, J. E.; Andersen, K. V.; Honig, B.; Hooft, R. W. W.; Klebe, G.; Vriend, G.; Wade, R. C. Improving macromolecular electrostatics calculations. *Protein Eng.* **1999**, *12*, 657−662.

(24) Chen, I. J.; MacKerell, A. D. Computation of the influence of chemical substitution on the p*K*(a) of pyridine using semiempirical and ab initio methods. *Theor. Chem. Acc.* **2000**, *103*, 483−494.

(25) Shapley, W. A.; Bacskay, G. B.; Warr, G. G. Ab initio quantum chemical studies of the p*K*(a)'s of hydroxybenzoic acids in aqueous solution with special reference to the hydrophobicity of hydroxybenzoates and their binding to surfactants. *J. Phys. Chem. B.* **1998**, *102*, 1938−1944.

(26) Tran, N. L.; Colvin, M. E. The prediction of biochemical acid dissociation constants using first principles quantum chemical simulations. *THEOCHEM* **2000**, *532*, 127−137.

(27) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. Solvent effects. 5. Influence of cavity shape, truncation of electrostatics, and electron correlation ab initio reaction field calculations. *J. Phys. Chem.* **1996**, *100*, 16098−16104.

(28) Mejias, J. A.; Lago, S. Calculation of the absolute hydration enthalpy and free energy of H^+ and OH^-. *J. Chem. Phys.* **2000**, *113*, 7306−7316.

(29) Schaefer, M.; Karplus, M. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578−1599.

(30) Lim, C.; Bashford, D.; Karplus, M. Absolute pK_a calculations with continuum dielectric methods. *J. Phys. Chem.* **1991**, *95*, 5610−5620.

(31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A. J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T. A.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, rev. A.7; Gaussian Inc: Pittsburgh, PA, 1998.

(32) Alagona, G.; Ghio, C.; Nagy, P. I.; Durant, G. J. Theoretical studies on the continuum solvation of some *N*,*N*′-dimethyl- and *N*-methyl, *N*′-acetyl-guanidine and guanidinium conformers. *J. Phys. Chem. A* **1999**, *103*, 1857−1867.

(33) Lide, D. R. E. *Handbook of Chemistry and Physics*, 72nd ed.; CRC Press: Boca Raton, FL, 1991.

(34) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618−622.

(35) Becke, A. D. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(36) Gobbi, A.; Frenking, G. Y-Conjugated Compounds−The Equilibrium Geometries and Electronic-Structures of Guanidine, Guanidinium Cation, Urea, and 1,1-Diaminoethylene. *J. Am. Chem. Soc.* **1993**, *115*, 2362−2372.

(37) Koskinen, J. T. Experimental and computational studies on aminoguanidine free base, monocation and dication. Part III: Proton affinities of guanidine, aminoguanidine and glyoxal bis(amidinohydrazone). *Z. Naturforsch., B: Chem. Sci.* **1998**, *53*, 386−392.

(38) Maksic, Z. B.; Kovacevic, B. Absolute proton affinity of some polyguanides. *J. Org. Chem.* **2000**, *65*, 3303−3309.

(39) Julian, R. R.; Beauchamp, J. L. Site specific sequestering and stabilization of charge in peptides by supramolecular adduct formation with 18-crown-6 ether by way of electrospray ionization. *Int. J. Mass Spectrom.* **2001**, *210*, 613−623.

(40) Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. Ab initio calculation of absolute pK_a values in aqueous solution. II. aliphatic, alcohols, thiols, and halogenated carboxylic acids. *J. Phys. Chem. A* **2000**, *104*, 2402−2409.

(41) Rak, J.; Skurski, P.; Simons, J.; Gutowski, M. Low-energy tautomers and conformers of neutral and protonated arginine. *J. Am. Chem. Soc.* **2001**, *123*, 11695−11707.

CT049849M

JCTC Journal of Chemical Theory and Computation

Intrinsic Carbon−Carbon Bond Reactivity at the Manganese Center of Oxalate Decarboxylase from Density Functional Theory

Christopher H. Chang† and Nigel G. J. Richards*

Department of Chemistry, University of Florida, Gainesville, Florida 32611-7200

Received March 10, 2005

Abstract: The detailed manganese-dependent chemistry employed by oxalate decarboxylase (OxDC) to catalyze the nonoxidative decarboxylation of oxalic acid remains poorly understood. For example, enzyme activity requires the presence of dioxygen even though this compound is not a formal substrate in the reaction. We now report density functional theory (DFT) calculations upon a series of hypothetical OxDC active site model structures. Our results suggest that the function of the metal ion may be to position dioxygen and oxalate such that electrons can be shuttled directly between these species, thereby removing the need for the existence of Mn(III) as an intermediate in the mechanism. These calculations also indicate that the intrinsic, gas-phase reactivity of the *Bacillus subtilis* oxalate *decarboxylase* active center is to *oxidize* oxalate. Since this reactivity is not observed for OxDC, our DFT results suggest that protein environment modulates the intrinsic metallocenter reactivity, presumably by affecting the electronic distribution at the manganese center during catalysis.

Introduction

Oxalate decarboxylase (E.C. 4.1.1.2) (OxDC) catalyzes the nonoxidative decarboxylation of oxalic acid to formic acid and carbon dioxide.[1] This transformation is interesting because the enzyme employs manganese to cleave the relatively inert C−C bond in the substrate,[2] although the role of the metal ion in catalysis is poorly understood.[3] A further puzzling observation is that OxDC is inactive under anaerobic conditions even though the decarboxylation is a redox-neutral reaction.[2−4] Chemical precedent for direct homolytic C−C bond cleavage in the breakdown of carboxylic acids is provided by the Kolbe[5] and Hunsdiecker[6] reactions, and similar, radical-dependent mechanisms likely mediate oxalate decomposition in the presence of light and flavins[7] or high-valent Mn species.[8] On the other hand, heavy atom (^{13}C and ^{18}O) isotope effect measurements provide support for a mechanism in which heterolytic bond C−C cleavage takes place in a Mn-bound oxalate radical anion.[9] Studies of the pH dependence of OxDC activity have also established that monoprotonated oxalate is the substrate that initially binds within the OxDC active site.[9] These observations are consistent with a mechanistic model in which the radical anion intermediate and Mn(II) are generated by reversible, proton-coupled electron transfer[10] from Mn(III)-bound, monoprotonated oxalate prior to decarboxylation (Scheme 1). Alternate proposals involving the involvement of higher-valency Mn(IV)-containing intermediates[11] could not, however, be ruled out on the basis of these heavy-atom isotope effect experiments.

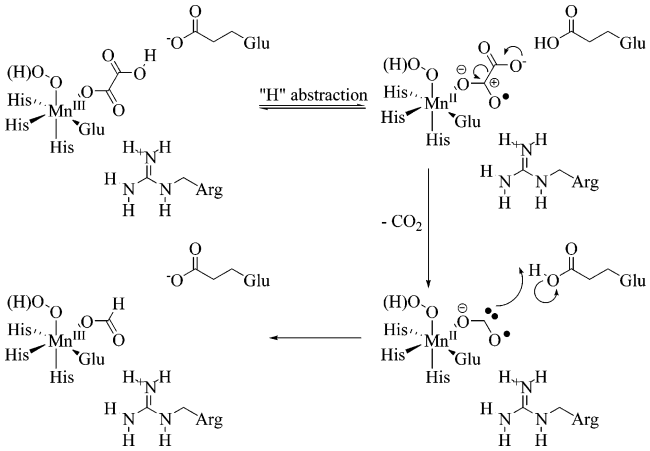
Modern density functional theory (DFT) calculations represent an alternative strategy for investigating (i) the likely involvement of high-valent Mn(III) or Mn(IV) species in OxDC-catalyzed oxalate breakdown and (ii) the effects of protonation on the electronic properties of hypothetical OxDC active site model structures corresponding to potential Michaelis complexes and intermediates that might be formed prior to C−C bond cleavage (Scheme 2).[12] In combination with bonding analyses of the Kohn−Sham wave function

* Corresponding author phone: (352)392-3601; e-mail: richards@qtp.ufl.edu.

† Present address: National Renewable Energy Laboratory, 1617 Cole Boulevard, Mail Stop 1608, Golden, CO 80401.

10.1021/ct050063d CCC: $30.25
Published on Web 07/06/2005

Scheme 1: Hypothetical Catalytic Mechanism for Oxalate Decarboxylase Based on Heavy-Atom Isotope Effect Studies[9]



employing the Natural Bond Orbital (NBO) framework,[13] we have employed DFT methods (i) to elucidate the resting valence of the OxDC Mn center in the presence and absence of dioxygen, (ii) to investigate whether formation of superoxide or peroxide can take place in Mn-bound substrate complexes, (iii) to determine the effects of protonating Mn-bound oxalate and dioxygen on metal charge and spin redistribution, and (iv) to identify structures in which the oxalate C−C bond might be weakened significantly. Prior computational studies of the metal centers in Fe- and Mn-dependent enzymes have yielded considerable insights into mechanistic details that are difficult to obtain using current experimental methods.[14,15] The results discussed herein have interesting implications for the general understanding of the mechanistic basis of the decarboxylation and oxidation reactions catalyzed by OxDC and oxalate oxidase (OxOx), respectively, and the role of dioxygen in generating Mn(III) and Mn(IV) complexes for OxDC-catalyzed C−C bond cleavage.

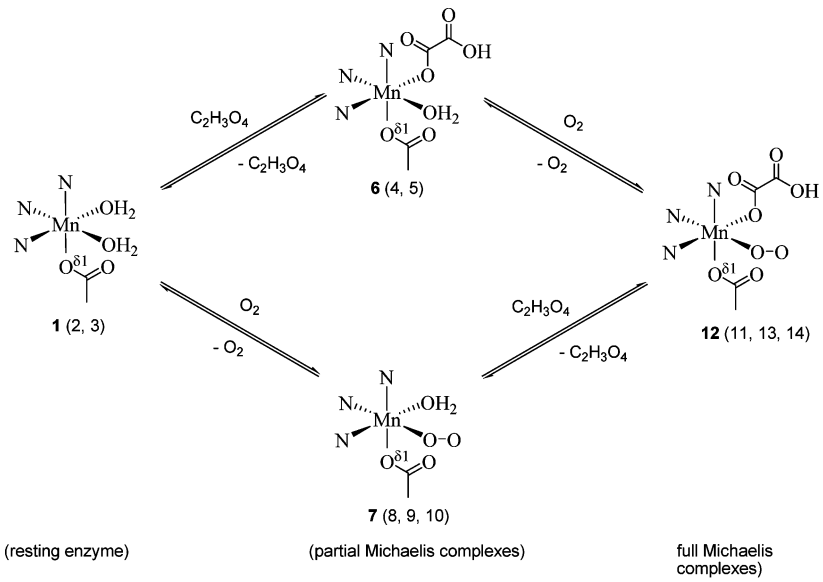
Computational Methods

Construction of Active Site Models 1−14. The initial coordinates for each of the active site model structures were obtained by excising the Mn-bound residues corresponding to His-95, His-97, Glu-101, and His-140 (*Bacillus* subtilis residue numbering) from the crystal structure of OxDC (code: 1J58)[11] deposited in the Protein Data Bank,[16] together with the metal ion and its ligands (water and formate).

The N-terminal site was chosen for the template because (1) formate was bound to the metal despite the absence of this ligand in preparation or crystallization buffers, and (2) oxalate could not be modeled into the C-terminal site of the 1J58 crystal structure without unrealistic steric clashes. This choice is also consistent with recent literature evidence supporting the N-terminal Mn-binding site as the primary site of oxalate breakdown in this enzyme.[17] Backbone atoms of the protein residues, with the exception of the C_α carbon, were then deleted, and hydrogens added to give ethylimidazole and butyrate ligands coordinating the metal (Figure 1). The six C−C distances involving the carbon atoms of the capping methyl groups were fixed at their crystallographic values so as to prevent significant distortions of the active site model structures during geometry optimization.[18]

Geometry Optimization of Model Complexes. Spin-unrestricted geometry optimizations of the active site model structures were performed with the TURBOMOLE software package.[19] Calculations upon neutral and cationic complexes used the BLYP exchange-correlation functional[20] with a split-valence + polarization (SVP) double-ζ basis set[21] and employed the "resolution-of-the-identity" approximation for Coulomb integrals (RI).[22] In the case of the anionic complexes (**4**, **5**, and **11**), for which the inclusion of diffuse functions was critical, the lack of a suitable fitting basis for the RI algorithm required standard methods for evaluating Coulomb integrals, and so these calculations employed an SVP basis on H, C, and Mn and a 6-31+G* basis on the O and N centers.[23] DFT calculations on model complexes

Scheme 2: Schematic Overview of the Relationship of the Model Complexes to Steps in the OxDC-Catalyzed Reaction[a]



[a] Models that are listed *but not shown explicitly* differ from their parent complex (bold numbers) in the scheme by the loss or addition of protons from/to the water, metal-bound dioxygen or oxalate ligands

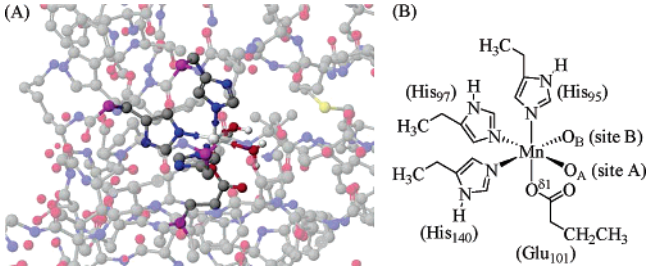


Figure 1. (A) Visualization of the putative, N-terminal active site in the crystal structure of *Bacillus subtilis* OxDC,[11] with the observed Mn-bound formate replaced by a water molecule to give the H_2O-H_2O complex **1**. Highlighted atoms represent those used to construct active site models **1**−**14**. The purple atoms correspond to the α-carbons that are transformed into methyl groups in the model complexes. The relative coordinates of the purple atoms were fixed during geometry optimization. The more solvent accessible coordination site (site B), to which oxalate is assumed to bind in these studies, faces away from the reader and toward a selenomethionine residue. With the exception of the two Mn-bound water molecules, only non-hydrogen atoms are shown for clarity. Atoms are colored using the following scheme: C, gray; H, white; N, blue; O, red; Se, yellow; Mn, white. (B) Schematic representation of model complexes **1**−**14** showing the location of sites A and B relative to the ethylimidazole and butyrate moieties corresponding to the protein residues that coordinate manganese.

containing Mn-bound dioxygen (**7**−**14**) were performed at both $M_S = 3/2$ and $M_S = 5/2$ spin configurations, while all other calculations assumed that Mn was present in its high-spin $M_S = 5/2$ electronic configuration for reasons that are described below. Geometry optimizations were converged to 10^{-6} Ha in total energy and 10^{-3} Ha/Bohr in the Cartesian gradients, unless bond cleavage took place in which case the calculation was terminated since one or more dimensions on the potential surface became essentially flat. Unrestricted B3LYP[24] optimizations on active site models **11**−**14** were also undertaken to examine the effects of including exact exchange and employed an SVP(+) basis.

NBO Analysis of the Kohn−Sham Wave Functions. The Kohn−Sham wave functions determined for the optimized complexes were analyzed within the Natural Bond Orbital formalism,[13] using the NBO 5.0 package. Input files for this analysis were obtained by performing single point energy calculations for each of the active site models at its optimized geometry using a PBE/LACVP(O/N+)** model chemistry, as implemented in Jaguar V5.0 (Schrödinger LLC, Portland OR). Thus, an effective core potential was used on Mn[27] together with a 6-31G** basis on H, C, and Mn and a 6-31+G** basis on the O and N centers. The extent of covalency in metal−ligand bonds was evaluated using a procedure developed previously for open-shell transition metal complexes,[28] with the modification that 3-centered, 4-electron hyperbonding was accounted for implicitly by analysis of the Natural Localized Molecular Orbitals[29] and Natural Bond Orbitals. Atomic charges and populations were derived from the Natural Population Analysis.[30]

Results and Discussion

Structural Properties and NBO Analysis of Complexes 1−3 (Resting Enzyme). The polypeptide framework of *Bacillus subtilis* OxDC provides residues (His-95, His-97, His-140, and Glu-101) that occupy four of the six Mn-coordination sites in the N-terminal, Mn-binding site.[11] In the original crystal structure of the enzyme, the two remaining coordination sites were occupied by a water molecule (site A) and formate (site B).[11] The latter site appears to be more exposed to bulk solvent in this high-resolution structure, consistent with the hypothesis that Mn-bound oxalate is decarboxylated to yield formate. We therefore assumed in constructing our active site model complexes that (i) the less solvent-accessible and more hydrophobic site A may be occupied by water, hydroxide, or dioxygen derivatives, and (ii) the more solvent-accessible site B may be occupied by water, hydroxide, formate, or mono- or diprotonated oxalate (Figure 1). In our initial calculations, we examined the BLYP-optimized structure and chemical bonding of complexes that model the OxDC Mn-center in the resting enzyme, which could be derived by replacing formate by a water molecule (Figure 1). There were four possible active site complexes due to the two unique binding sites, which we represented as H_2O-H_2O **1**, H_2O-HO^- **2**, HO^--H_2O **3**, and HO^--HO^-. In this notation (Figure 1B), the initial ligand is assumed to occupy coordination site A and the second is therefore located in the more solvent accessible site B. Calculations on the complex containing two Mn-bound hydroxide ligands were not performed, however, given that this species is unlikely to be relevant to any catalytic mechanism given that the enzyme exhibits optimum activity under acidic conditions.[4,9] In addition, only the $M_S = 5/2$ spin configuration was considered for these resting-state model complexes because (i) experimental EPR measurements indicate that the d^5 high-spin state of Mn(II) is the ground state of the resting enzyme,[2a,3b] and (ii) histidine, glutamate, water and hydroxide anion are all weak field ligands.

The geometries of model complexes **1**−**3** were optimized using the BLYP exchange-correlation functional and a double-ζ SVP basis (Table 1), in part to obtain reference structures that could be used to assess the effects of oxalate and/or dioxygen binding on metal−ligand bonding, molecular geometry, and the spin preferences of the Mn center. These calculations showed that replacing a water by an anionic hydroxide ligand (i) shortened the Mn−O bond length for this ligand, (ii) systematically lengthened the metal−imidazole bonds by approximately 0.1 Å while leaving the Mn−carboxylate distance little changed, and (iii) decreased the N−Mn−N bond angles in almost all of the active site model structures (Table 1). In addition, the angle between the hydroxyl and carboxylate ligands was increased with a concomitant decrease in the water/carboxylate angle. These changes are consistent with those expected due to the increased negative charge of the metal-bound ligand after deprotonation.

To obtain a picture of the chemical bonding in active site model structures **1**−**3**, we undertook a detailed analysis of the PBE wave function, as represented by a single determi-

Table 1. Selected Structural Properties of BLYP-Optimized Geometries for OxDC Active Site Models **1**−**3**[a]

internal coordinate[b,c]	H_2O-H_2O **1**	H_2O-HO^- **2**	HO^--H_2O **3**
Mn−H95N$^\epsilon$	2.279	2.345	2.356
Mn−H97N$^\epsilon$	2.245	2.342	2.326
Mn−E101O$^{\delta1}$	2.161	2.158	2.174
Mn−H140N$^\epsilon$	2.258	2.369	2.329
Mn−O_A	2.307	2.400	2.051
Mn−O_B	2.345	2.053	2.452
H95N$^\epsilon$−Mn−H97N$^\epsilon$	92.4	87.14	92.15
H95N$^\epsilon$−Mn−E101O$^{\delta1}$	176.4	169.1	176.8
H95N$^\epsilon$−Mn−H140N$^\epsilon$	90.84	84.03	85.05
H95N$^\epsilon$−Mn−O_A	98.84	95.12	85.31
H95N$^\epsilon$−Mn−O_B	85.07	83.57	100.4
H97N$^\epsilon$−Mn−E101O$^{\delta1}$	85.21	84.02	85.12
H97N$^\epsilon$−Mn−H140N$^\epsilon$	102.1	93.08	95.98
H97N$^\epsilon$−Mn−O_A	161.8	175.2	165.1
H97N$^\epsilon$−Mn−O_B	100.3	113.1	96.50
E101O$^{\delta1}$−Mn−H140N$^\epsilon$	92.25	90.06	93.47
E101O$^{\delta1}$−Mn−O_A	82.88	93.24	97.77
E101O$^{\delta1}$−Mn−O_B	92.76	105.8	81.70
H140N$^\epsilon$−Mn−O_A	91.99	82.96	98.45
H140N$^\epsilon$−Mn−O_B	157.3	150.3	166.2
O_A-Mn−O_B	66.73	71.46	69.59

[a] All calculations are for the $S = 5/2$ spin state of Mn(II) and are spin-unrestricted. [b] All bond distances and angles are reported in units of Å and deg, respectively. [c] O_A and O_B correspond to the ligating oxygen atoms in the buried (site A) and solvent-exposed (site B) coordination sites (Figure 1), respectively, and $O^{\delta1}$ is the oxygen of the butyrate moiety corresponding to the Glu-101 residue that is bonded directly to the metal.

nant of canonical Kohn−Sham orbitals, within the NBO formalism for each structure at its ground-state, BLYP-optimized geometry. More specifically, we sought to obtain (i) the distribution of electrons in these complexes, as determined by spin magnetization density (SMD) and partial charges obtained using Natural Population Analysis (NPA),[13,28] (ii) the extent to which our results were subject to spin contamination (Table 2),[14c] and (iii) the extent of metal−ligand bond covalency (Table 3). For all three models of the OxDC center in the resting enzyme (**1**−**3**), the bonding appeared characteristic of a typical, high-spin Mn(II) complex in that the difference in α- and β-spin density on the metal center was 4.7, and the total *d*-orbital population on Mn was approximately 5.3 electrons, consistent with the metal ion having a d^5 configuration in the $M_S = 5/2$ spin state (Table 2). We also observed that the NPA charges on Mn were unaffected by the presence or absence of hydroxide ion, having a value of approximately +1.4, and that there was some delocalization of ligand β-spin density onto the metal with a corresponding, slight spin polarization of the ligands. More variation was seen in the total metal−ligand covalencies computed for the three active site models **1**−**3** (Table 3). For example, when hydroxide was present in site A (model **3**), the covalency was significantly higher than that computed for the other two complexes. This increase did not appear correlated with Mn−hydroxide binding, however, but with the covalency of the bonds involving Mn and the ligands modeling the side chains of residues His-95, His-97, and Glu-101. In addition, a decreased covalency in the metal−ligand bond involving the ethylimidazole corresponding to the His-140 side chain was observed in both of the models containing a hydroxide ligand (**2** and **3**), consistent with our finding that this ligand dissociates from the metal when the geometry of the anionic complex containing two hydroxide ions is optimized.

Structural Properties and NBO Analysis of Complexes 4−6 (Oxalate-Bound Enzyme). Having analyzed the structural models corresponding to the Mn-center in the resting enzyme, we next examined three models (**4**−**6**) in which oxalate occupies a coordination site (site B) on the metal in the absence of dioxygen (Figure 1B). Geometry optimization, within the RI approximation, of an active site model (**4**) containing the monoprotonated form of oxalate in which the sixth ligand was a hydroxide anion, merely led to dissociation of the ethylimidazole ligands corresponding to His-95 and His-140 in the putative OxDC active site. We therefore performed a similar calculation using a BLYP/SVP(O/N 6-31+G*) model chemistry to account for potentially diffuse electron density in this anionic complex. Even with this expanded basis, geometry optimization still resulted in dissociation of an ethylimidazole ligand (His-140 in OxDC) (Table 4). Similar observations were made for the anionic, oxalate-containing, H_2O-Ox model **5** in that cleavage of the Mn−N bond involving the ethylimidazole corresponding to His-140 took place. The physical and/or mechanistic significance of these results is unclear, however, since the tendency of the Mn−ligand bond to undergo cleavage in these active site models may (i) reflect a change in metal coordination that actually takes place on substrate binding or (ii) be a consequence of neglecting residues defining the protein environment about the metal center in OxDC in our calculations.

On the other hand, it seems unlikely that hydroxide would occupy the sixth coordination site in preference to water in any OxDC/oxalate complex over the pH range of 3.8−5.0 at which OxDC exhibits significant activity.[9] Additional support for this assumption is provided by the measured pK_1 and pK_2 values of free $Mn(H_2O)_6^{2+}$, which are estimated to be 10.4 and 22, respectively.[31,32] Thus, the Mn-bound water molecule seems likely to have a pK_a of greater than 20 when oxalate monoanion is also coordinated to the metal. This pK_a may be increased further by the glutamate ligand that coordinates manganese. Efforts to obtain an optimized structure for a model structure in which hydroxide and oxalate dianion occupied coordination sites A and B, respectively, also gave results that were dependent on the model chemistry used in the calculations.[33] Although oxalate dianion might conceivably adopt bidentate coordination at the mononuclear manganese center similar to that recently seen in a putative oxalate decarboxylase from *Thermatoga maritima*,[34] we did not observe such a binding mode in this series of calculations. This may, however, reflect the lack of a suitable proton donor in our active site model containing dianionic oxalate and/or the energy of a free vs bound hydroxide ion in the absence of a solvation potential, and, therefore, participation of bidentate binding in the OxDC active site is not definitively ruled out by this computational investigation.

Table 2. NPA-Derived Partial Charges and Spin Magnetization Densities (SMD) of Selected Atoms in Complexes **1**−**14** at Their DFT-Optimized Geometries[a]

		net charge	$\langle S^2\rangle_{KS}$[b]	partial charge (Mn)	partial charge (O_A)[c]	partial charge (O_C)[c]	SMD (Mn)	SMD (O_A)	SMD (O_C)	SMD (oxalate)[d]	C−C bond order
H_2O-H_2O **1**	$M_S = 5/2$	+1	8.75	+1.44	*f*	*f*	4.70	*f*	*f*	*g*	*g*
H_2O-HO^- **2**	$M_S = 5/2$	0	8.75	+1.38	*f*	*f*	4.67	*f*	*f*	*g*	*g*
HO^--H_2O **3**	$M_S = 5/2$	0	8.75	+1.39	*f*		4.66	*f*	*f*	*g*	*g*
HO^-−OxH **4**	$M_S = 5/2$	−1	8.76	+1.48	*f*	*f*	4.65	*f*	*f*	0.06[e]	0.826
H_2O−Ox **5**	$M_S = 5/2$	−1	8.76	+1.46	*f*	*f*	4.62	*f*	*f*	0.16[e]	0.874
H_2O−OxH **6**	$M_S = 5/2$	0	8.75	+1.45	*f*	*f*	4.68	*f*	*f*	0.08[e]	0.928
O_2-H_2O **7**	$M_S = 5/2$	+1	8.92	+1.39	−0.22	−0.16	3.95	0.33	0.46	*g*	*g*
O_2-HO^- **8**	$M_S = 5/2$	0	8.84	+1.35	−0.28	−0.22	3.77	0.42	0.52	*g*	*g*
HO_2-H_2O **9**	$M_S = 3/2$	+2	3.84	+1.13	−0.25	−0.41	2.76	0.14	−0.01	*g*	*g*
HO_2-HO^- **10**	$M_S = 3/2$	+1	3.82	+1.12	−0.28	−0.43	2.68	0.10	0.00	*g*	*g*
O_2−Ox **11**	$M_S = 5/2$	−1	9.21	+1.48	−0.29	−0.19	4.26	0.31	0.44	0.25	0.824
O_2−OxH **12**	$M_S = 5/2$	0	9.07	+1.41	−0.21	−0.07	4.10	0.20	0.36	0.12	0.861
HO_2−Ox **13**	$M_S = 5/2$	0	8.76	+1.34	−0.61	−0.58	4.56	0.18	0.06	0.00	0.001
HO_2−OxH **14a**	$M_S = 3/2$	+1	3.86	+1.15	−0.28	−0.54	2.80	0.11	−0.02	0.12	0.912
HO_2−OxH **14b**	$M_S = 5/2$	+1	8.95	+1.40	−0.21	−0.54	3.99	0.29	0.43	0.10	0.908

[a] Complexes are named as discussed in the text and shown in Figure 1B. [b] Noninteracting values of $\langle S^2\rangle_{KS}$ were computed from a single determinant of Kohn−Sham orbitals, using the algorithm implemented in Gaussian98. [c] O_A and O_C correspond to the atoms in the dioxygen or hydroperoxy ligand in the buried coordination site (site A). O_A is bonded to the Mn center. [d] This value was computed using $\Sigma\rho_{\alpha-\beta}$ where the difference in α and β spin densities is summed over all the atoms in the oxalate ligand, including hydrogens when present. [e] Spin polarization on the oxalate moiety in complexes **4**−**6** is associated with the oxygen atom covalently bonded to the Mn center. [f] Not applicable since dioxygen or hydroperoxide is not present in the complex. [g] Not applicable since oxalate is not present in the complex.

Table 3. Calculated Covalent Character of Mn−Ligand Bonds in Complexes **1**−**3** at Their BLYP-Optimized Ground-State Geometries[a]

bond	bond covalency (|e$^-$|)	total covalency (|e$^-$|)
	H_2O-H_2O **1**	
Mn−H95N$^\epsilon$	0.100	0.536
Mn−H97N$^\epsilon$	0.106	
Mn−E101O$^{\delta1}$	0.101	
Mn−H140N$^\epsilon$	0.097	
Mn−O_A	0.066	
Mn−O_B	0.066	
	H_2O-HO^- **2**	
Mn−H95N$^\epsilon$	0.060	0.603
Mn−H97N$^\epsilon$	0.073	
Mn−E101O$^{\delta1}$	0.121	
Mn−H140N$^\epsilon$	0.065	
Mn−O_A	0.080	
Mn−O_B	0.204	
	HO^--H_2O **3**	
Mn−H95N$^\epsilon$	0.196	0.868
Mn−H97N$^\epsilon$	0.214	
Mn−E101O$^{\delta1}$	0.137	
Mn−H140N$^\epsilon$	0.053	
Mn−O_A	0.197	
Mn−O_B	0.071	

[a] Atom labels correspond to those used in Table 1.

Examination of the chemical bonding in complexes **4**−**6** showed that the NPA charges, spin magnetization densities, and *d*-orbital populations on the Mn center were very similar to those computed for the active site structures modeling the resting state of the enzyme (Table 2). Unlike complexes **1**−**3**, however, substantially increased covalent bonding to Mn in the 5-coordinate complexes resulting from dissociation of the His-140 ligand during geometry optimization was not observed (Table 5). Instead, the oxalate C−C NPA/NLMO bond order was reduced in models **4** and **5** relative to that in the 6-coordinate active site model **6** (Table 2). This reduction in C−C bond order was mitigated in **5**, however, by hydrogen bonding of the oxalate moiety to the Mn-bound water molecule, which also formed a hydrogen bond to the distal oxygen on the butyrate moiety corresponding to Glu-101 in the enzyme. We therefore interpret the reduced oxalate C−C bond order in this structure to arise from an inductive effect that is enhanced by the loss of His-140 imidazole during geometry optimization. A possible interpretation of these results is that electrostatic polarization alone can weaken the oxalate C−C bond when this substrate becomes bound to the Mn center in OxDC. In addition, hydrogen bonding effects and/or protonation of the distal carboxylate of Mn-bound, monodentate oxalate may operate to modulate the effects of charge polarization so as to create an exquisitely controlled switch,[9] as discussed below.

Structural Properties and NBO Analysis of Complexes 7−10 (Dioxygen-Bound Enzyme). Although OxDC activity is observed only under aerobic conditions, the mechanistic role of dioxygen in catalyzing the decarboxylation of oxalate has yet to be defined unambiguously. We,[9] and others,[2,11] have speculated that dioxygen is employed in the oxidation of Mn(II) to Mn(III) or Mn(IV), which are species that can participate in proton-coupled electron transfer to yield a metal-bound oxalate radical anion (Scheme 1). Recent EPR studies aimed at observing any interaction between Mn(II) and dioxygen under steady-state conditions, however, yielded little direct evidence to support this mechanistic proposal,[3b] although recent work has suggested that dioxygen can bind to Mn(II) if there is sufficient anionic character present in the ligand field.[35] Such behavior can be rationalized in terms of a requirement for formal electron transfer from Mn(II) to dioxygen, thereby yielding a complex that is best described as involving a superoxide anion bound to Mn(III), at least

Table 4. Selected Structural Properties of BLYP-Optimized Geometries for OxDC Active Site Models **4**−**6**[a]

internal coordinate[b,c]	HO^-−OxH **4**	H_2O−Ox **5**	H_2O−OxH **6**
Mn−H95N$^\epsilon$	2.430	2.289	2.372
Mn−H97N$^\epsilon$	2.303	2.346	2.338
Mn−E101O$^{\delta1}$	2.176	2.119	2.205
Mn−H140N$^\epsilon$	*d*	*d*	2.358
Mn−O_A	2.013	2.226	2.207
Mn−O_B	2.142	1.992	2.108
C_A−C_B	1.553	1.566	1.550
H95N$^\epsilon$−Mn−H97N$^\epsilon$	87.94	85.22	86.35
H95N$^\epsilon$−Mn−E101O$^{\delta1}$	164.0	142.3	165.3
H95N$^\epsilon$−Mn−H140N$^\epsilon$	*d*	*d*	85.06
H95N$^\epsilon$−Mn−O_A	81.24	85.71	101.5
H95N$^\epsilon$−Mn−O_B	90.42	96.79	88.21
H97N$^\epsilon$−Mn−E101O$^{\delta1}$	82.66	80.09	82.18
H97N$^\epsilon$−Mn−H140N$^\epsilon$	*d*	*d*	92.26
H97N$^\epsilon$−Mn−O_A	130.8	144.0	170.4
H97N$^\epsilon$−Mn−O_B	101.2	107.0	95.20
E101O$^{\delta1}$−Mn−H140N$^\epsilon$	*d*	*d*	86.17
E101O$^{\delta1}$−Mn−O_A	95.17	86.15	89.13
E101O$^{\delta1}$−Mn−O_B	104.0	120.7	102.0
H140N$^\epsilon$−Mn−O_A	*d*	*d*	82.99
H140N$^\epsilon$−Mn−O_B	*d*	*d*	169.6
O_A−Mn−O_B	126.5	108.7	90.64
O_B−C_A−O_D	130.4	123.6	128.9
O_E−C_B−O_F	121.5	127.2	121.7

[a] All calculations are for the $S = 5/2$ spin state of Mn(II) and are spin-unrestricted. [b] All bond distances and angles are reported in units of Å and deg, respectively. [c] O_A and O_B correspond to the ligating oxygen atoms in the buried (site A) and solvent-exposed (site B) coordination sites (Figure 1), respectively, and $O^{\delta1}$ is the oxygen of the butyrate moiety corresponding to the Glu-101 residue that is bonded directly to the metal. C_A and O_D are the remaining atoms in the carboxylate coordinated directly to Mn, and atoms in the other carboxyl moiety in oxalate are defined as C_B, O_E, and O_F. [d] Internal coordinates involving for the ethylimidazole moiety corresponding to His-140 are not presented because this ligand dissociated from the metal during geometry optimization.

Table 5. Calculated Covalent Character of Mn−Ligand Bonds in Complexes **4**−**6** at Their BLYP-Optimized Ground-State Geometries[a]

bond	bond covalency (|e$^-$|)	total covalency (|e$^-$|)
	HO^-−OxH **4**	
Mn−H95N$^\epsilon$	0.056	0.526
Mn−H97N$^\epsilon$	0.076	
Mn−E101O$^{\delta1}$	0.102	
Mn−H140N$^\epsilon$	0.001[b]	
Mn−O_A	0.209	
Mn−O_B	0.082	
	H_2O−Ox **5**	
Mn−H95N$^\epsilon$	0.086	0.495
Mn−H97N$^\epsilon$	0.068	
Mn−E101O$^{\delta1}$	0.107	
Mn−H140N$^\epsilon$	0.000[b]	
Mn−O_A	0.090	
Mn−O_B	0.144	
	H_2O−OxH **6**	
Mn−H95N$^\epsilon$	0.075	0.541
Mn−H97N$^\epsilon$	0.071	
Mn−E101O$^{\delta1}$	0.094	
Mn−H140N$^\epsilon$	0.070	
Mn−O_A	0.110	
Mn−O_B	0.121	

[a] Atom labels correspond to those used in Table 4. [b] The ethylimidazole ligand corresponding to His-140 dissociated from the metal during geometry optimization of this complex.

within a standard ligand field model. Support for such an idea is provided by studies on copper-dependent amine oxidases.[36] The interaction of dioxygen with the Mn(II) center in our active site models was therefore investigated, although these DFT calculations were complicated by the need to consider the effects of spin coupling between triplet O_2 and the metal. Three possible systems can be envisaged in which $S = 3/2$, 5/2, or 7/2. For the case in which $S = 7/2$, the Mn(II)/dioxygen interaction might take place by (i) mixing a formally empty 4*s* orbital on the metal with a filled π MO on the ligand, (ii) mixing of one (or more) half-filled Mn 3*d* orbital(s) with the low-lying half-filled π* MOs on dioxygen, or (iii) one-electron oxidation of histidyl or glutamyl ligand by dioxygen to give the corresponding radical and superoxide radical anion, resulting in a system containing seven unpaired electrons. Of these possibilities, the first is unlikely, and the second is disfavored because it would require a spin flip (within the standard single determinantal model in which excess unpaired electrons are α-spin) to satisfy the Pauli exclusion principle and permit electron transfer from the Mn *d* manifold to the O_2 π* MOs.[37] In addition, while ligand oxidation is observed for thiolates[38] and high-valent ferryl heme intermediates,[39] such a process seems unlikely for the histidine and glutamate ligands coordinating Mn(II) in the OxDC active site. We therefore focused our computational studies on dioxygen-containing systems for which the total spin corresponded to $S = 3/2$ and $S = 5/2$ and obtained optimized structures for each of the four possible dioxygen-containing active site models **7**−**10** at both of these spin states (Table 6). The use of this single determinantal method to describe open-shell transition metal complexes, which are best described by multiple configurations, is fraught with difficulties.[14c,40] For example, such calculations may yield wave functions exhibiting significant spin contamination, complicating any interpretation of their electronic structure and/or energetics.

In the case of DFT calculations, spin contamination manifests itself as unpaired β-spin density resulting in a value of $\langle S^2\rangle_{KS}$ (at least to the degree that this two-electron quantity can be computed from a determinant of Kohn−Sham orbitals[41]) that is larger than $M_S(M_S + 1)$.[15a] We note that methods have been devised to extract reasonable spin-state energies from such broken-symmetry solutions[42] and applied in studies of Fe−S clusters for which spin coupling and valence delocalization patterns have been well characterized by experimental methods, including Mössbauer spectroscopy.[43] Although such methodology might have been employed to examine active site models **7**−**10**, a conceptual parsing of valence and spin subsystems, together with quantitation of the resonance and exchange interactions between them, remains to be established for Mn−O_2 systems. We therefore chose to examine the properties of the lowest energy, DFT-optimized structure for each *only* if it exhibited

Table 6. Selected Structural Properties of BLYP-Optimized Geometries for OxDC Active Site Models **7**−**10**[a]

internal coordinate[b,c]	O_2-H_2O **7** $M_S = 5/2$	O_2-HO^- **8** $M_S = 5/2$	HO_2-H_2O **9** $M_S = 3/2$	HO_2-HO^- **10** $M_S = 3/2$
Mn−H95N$^\epsilon$	2.141	2.406	2.073	2.112
Mn−H97N$^\epsilon$	2.146	2.555	2.106	2.148
Mn−E101O$^{\delta1}$	1.918	2.079	1.941	1.960
Mn−H140N$^\epsilon$	2.256	2.132	2.005	2.125
Mn−O_A	**2.068**[d]	**2.122**	**1.859**	**1.851**
Mn−O_B	2.330	1.818	2.033	1.804
O_A-O_C	**1.301**	**1.313**	**1.415**	**1.424**
H95N$^\epsilon$−Mn−H97N$^\epsilon$	90.18	74.82	91.22	87.55
H95N$^\epsilon$−Mn−E101O$^{\delta1}$	170.67	147.6	175.5	170.7
H95N$^\epsilon$−Mn−H140N$^\epsilon$	89.65	85.68	88.14	87.92
H95N$^\epsilon$−Mn−O_A	86.82	77.25	87.16	90.31
H95N$^\epsilon$−Mn−O_B	85.89	90.92	90.50	90.60
H97N$^\epsilon$−Mn−E101O$^{\delta1}$	88.23	73.95	85.58	85.89
H97N$^\epsilon$−Mn−H140N$^\epsilon$	98.19	95.39	99.25	93.72
H97N$^\epsilon$−Mn−O_A	173.7	151.9	169.8	177.8
H97N$^\epsilon$−Mn−O_B	93.40	91.04	87.35	89.44
E101O$^{\delta1}$−Mn−H140N$^\epsilon$	99.68	88.79	89.25	85.99
E101O$^{\delta1}$−Mn−O_A	93.83	134.2	96.55	96.27
E101O$^{\delta1}$−Mn−O_B	85.02	98.09	92.51	95.87
H140N$^\epsilon$−Mn−O_A	87.39	85.90	90.77	86.45
H140N$^\epsilon$−Mn−O_B	167.6	171.7	173.3	176.4
O_A−Mn−O_B	80.82	85.94	82.60	90.33
Mn−O_B−O_C	119.1	117.5	117.5	117.2

[a] All calculations are spin-unrestricted. M_S is the sum of the Mn and dioxygen spin states. [b] All bond distances and angles are reported in units of Å and deg, respectively. [c] O_A and O_B correspond to the ligating oxygen atoms in the buried (site A) and solvent-exposed (site B) coordination sites (Figure 1), respectively, and O$^{\delta1}$ is the oxygen of the butyrate moiety corresponding to the Glu-101 residue that is bonded directly to the metal. O_C is the atom in the dioxygen ligand that is not bound to Mn. [d] Highlighted internal coordinates are discussed explicitly in the text.

insignificant amounts of spin contamination as computed from $\langle S^2\rangle_{KS}$.[15a] For active site structures containing protonated (**9** and **10**) and unprotonated dioxygen (**7** and **8**), these criteria were met by systems for which $S = 3/2$ and $S = 5/2$, respectively.

The structural properties of these DFT-optimized active site models were consistent with qualitative expectations based on ligand field theory (Table 6). For example, hydroxide was bound more tightly to the metal center than water in these four active site models (**7**−**10**), and the O−O bond length (O_A-O_C) in the dioxygen moiety was correlated with its protonation state. Thus, there was a lengthening of the O_A-O_C bond with a concomitant decrease in the metal−ligand bond (Mn−O_A) when a proton was attached to the distal oxygen (Table 6), suggesting that protonation enhances the extent of metal to ligand charge transfer in these complexes. We also note that binding dioxygen and hydroxide in the nonproteinaceous coordination sites A and B, respectively, causes an increase in the bond lengths between the metal and the ligands corresponding to OxDC residues His95, His97, and Glu101 relative to the cognate geometric parameters in the other three complexes (Figure 2). This effect may be associated with the uncompensated negative charge in the hydroxide-bound model **8**. Similar, albeit

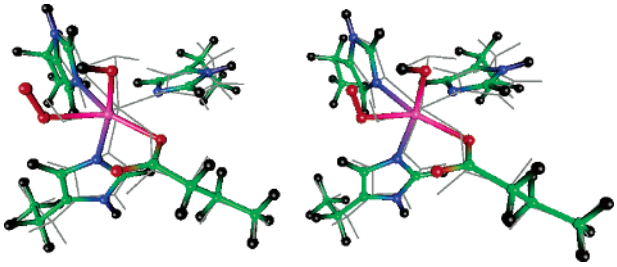


Figure 2. Stereoscopic image (cross-eyed) showing the superposition of the BLYP-optimized structures of the O_2-H_2O (**7**) and the O_2-HO^- (**8**) active site model complexes. Structures are rendered according to the following scheme: (**8**) C, green; H, black; N, blue; O, red; Mn, magenta; (**7**) gray.

smaller, changes in metal−dioxygen bond length are observed for complexes **9** and **10** in which this ligand is protonated. Thus, protonation of dioxygen appears to drive electron density from the metal center, and the ensuing partial reduction of the HO_2 ligand is therefore facilitated by the presence of the negatively charged hydroxide anion. Spin coupling to give complexes with net $S = 3/2$ or $S = 5/2$ spin states permits either metal-to-ligand charge donation or outright electron transfer between Mn(II) and dioxygen to yield an active site model that can be formally described as a superoxide-Mn(III) species, as proposed in previous mechanistic hypotheses.[2,9]

For the two complexes **7** and **8**, in which hydroxide or water was present in the "oxalate-coordination" site, the final DFT wave functions for these optimized structures exhibited little spin contamination, based on the $\langle S^2\rangle_{KS}$ values of 8.8 and 8.9 (Table 2). Their ground-state electronic configuration ($S = 5/2$) corresponded to four α-spins in the Mn *d*-manifold with approximately one α-spin delocalized over the dioxygen moiety (Table 2). Despite this unpaired electron distribution, the NPA charges on Mn of 1.4 computed for models **7** and **8** do not deviate substantially from that observed in complexes **9** and **10**. The NPA charges on the dioxygen moiety sum to −0.38 and −0.50 in the water- and hydroxide-containing models **7** and **8**, respectively. Examination of the α- and β-spin *d* NBOs for these complexes also shows that approximately 5 electrons are associated with the *d*-manifold, implying that the formal valence of the metal/dioxygen complex when hydroxide or water are bound in the remaining coordination site corresponds to Mn(II). These results are therefore consistent with a scenario in which dioxygen can bind to Mn(II), with a *net* partial transfer of charge to dioxygen from the ligand sphere. Hence, assuming that such an intermediate $S = 5/2$ spin coupling is feasible at the OxDC active site, dioxygen can associate with the metal to give a complex in which manganese retains divalent d^5 character.[44]

Given that the wave functions computed for the BLYP-optimized active site models were free of spin contamination, we analyzed the *d*-orbital SMD so as to assign a formal valence to the metal center (Table 2). We generally prefer to employ this measure of valence assignment rather than partial charge values because the latter are not uniquely defined from the wave function and often have limited similarity to the formal charges of transition metal-containing systems.[15e,45]

Table 7. Calculated Covalent Character of Mn−Ligand Bonds in Complexes **7**−**10** at Their BLYP-Optimized Ground-State Geometries[a,b]

bond	bond covalency (|e⁻|)	total covalency (|e⁻|)
	O_2−H_2O **7** $M_S = 5/2$	
Mn−H95N$^{\epsilon}$	0.163	0.994
Mn−H97N$^{\epsilon}$	0.065	
Mn-E101O$^{\delta 1}$	0.322	
Mn−H140N$^{\epsilon}$	0.125	
Mn−O_A	0.245	
Mn−O_B	0.074	
	O_2−HO^- **8** $M_S = 5/2$	
Mn−H95N$^{\epsilon}$	0.078	1.288
Mn−H97N$^{\epsilon}$	0.059	
Mn-E101O$^{\delta 1}$	0.192	
Mn−H140N$^{\epsilon}$	0.125	
Mn−O_A	0.260	
Mn−O_B	0.574	
	HO_2−H_2O **9** $M_S = 3/2$	
Mn−H95N$^{\epsilon}$	0.306	2.193
Mn−H97N$^{\epsilon}$	0.226	
Mn-E101O$^{\delta 1}$	0.357	
Mn−H140N$^{\epsilon}$	0.380	
Mn−O_A	0.720	
Mn−O_B	0.204	
	HO_2−HO^- **10** $M_S = 3/2$	
Mn−H95N$^{\epsilon}$	0.262	2.308
Mn−H97N$^{\epsilon}$	0.178	
Mn−E101O$^{\delta 1}$	0.340	
Mn−H140N$^{\epsilon}$	0.179	
Mn−O_A	0.656	
Mn−O_B	0.693	

[a] Atom labels correspond to those used in Table 4. [b] M_S is the sum of the Mn and dioxygen spin states.

When applied to the ground-state $S = 3/2$ active site models **9** and **10**, however, SMD analysis suggested that electron transfer had taken place to give Mn(IV) and bound hydroperoxide anion. This was an unexpected finding, especially given that the NPA charges on Mn decreased in complexes **9** and **10** by almost 0.3 electrons (Table 2). We therefore analyzed the extent of metal−ligand covalency in these active site models (Table 7). This analysis showed that the extent of covalency was significantly increased relative to the modest levels seen in the active site models for the resting enzyme (**1**−**3**) and the E-oxalate complex (**4**−**6**). Thus, a total of approximately 1 and 1.3 electrons are shared between the metal and its ligands in complexes **7** and **8**, respectively. The extent of covalency increases further to 2.2 and 2.3 electrons in active site models **9** and **10** (Table 7). As a result, we conclude that SMD quantitation of the formal charge on the metal in the latter complexes is flawed due to a breakdown of the ligand-field approximation associated with high metal−ligand covalency resulting from substantial mixing of the empty Mn *d*-orbital(s) with filled ligand orbitals and thus effective partial electron transfer from the Mn *d*-manifold to the superoxide or peroxide moiety. Thus, rather than Mn(III) or Mn(IV) ligand field complexes, these active site models might be chemically better described in terms of resonance forms in which oxidized ligands and Mn(II) also participate.

Structural Properties and NBO Analysis of Complexes 11−14 (Dioxygen-Oxalate Reactive Enzyme). We found the effect of binding both oxalate and dioxygen to the manganese center in the OxDC active site models **11**−**14** was critically dependent on the protonation states of these two ligands (Table 8). Thus, when both nonproteinaceous ligands were protonated in model **14**, the length of the BLYP-optimized oxalate C−C bond was 1.56 Å, identical to its value in model **6**, and only 0.01 Å longer than the cognate bond in model **5**, in which water is present in the putative dioxygen binding site. Similarly, the O−O bond length in the hydroperoxy ligand was 1.40 Å, a value slightly shorter than in either of the $S = 3/2$ model complexes **9** and **10**, in which water or hydroxide are bound in the putative oxalate-binding site, respectively. We therefore conclude that when both ligands are protonated, there is no significant perturbation in electronic structure relative to the models in which each is bound separately.

Binding unprotonated dioxygen in oxalate-containing active site models **11** and **12** gave systems for which the lowest energy, ground-state spin multiplicity was 5/2 rather than 3/2, similar to our findings for complexes **7** and **8**, in which water or hydroxide occupy site A, respectively. The O−O distance, however, shortens to 1.28 Å in models **11** and **12**, consistent with a reduction in superoxide or peroxide character (Table 8). The same decrease of approximately 0.1 Å is also observed in complexes **9** and **10** containing the HO_2 ligand (Table 4), suggesting that this behavior arises primarily from local bonding considerations within the dioxygen moiety. Likewise, binding oxalate dianion and dioxygen to give active site model **11** did not have dramatic structural consequences. The bond lengths in the optimized complex **11** were generally longer than those observed in either O_2−OxH **12** or HO_2−OxH **14**, consistent with its overall anionic character (Table 8). It is therefore interesting to note that (i) the Mn−oxalate bond (Mn−O_B) is shorter than the cognate bond in **12**, (ii) the oxalate C−C bond is slightly longer than complexes in which this ligand is protonated, including models **4** and **6**, and (iii) the bond between the ethylimidazole ligand, corresponding to His-95, and the metal is 2.4 Å. The latter distance generally correlated with the overall charge on each model, with protonation of oxalate shortening the bond more than protonation of dioxygen. Simultaneous binding of dioxygen and either oxalate monoanion (**12**) or dianion (**11**) gave BLYP-optimized structures having an $S = 5/2$ ground state that again exhibited only small amounts of spin contamination (Table 2). The NPA charges and SMD on Mn and bound O_2 are consistent with a formal Mn(III) bound to superoxide, as also observed for the dioxygen-containing models **7** and **8**. Some spin magnetization density finds its way onto the oxalate molecule, however, in both the O_2−Ox **11** and O_2−OxH **12** complexes (Table 2). Tellingly, the NPA/NLMO bond order for the oxalate C−C bond is comparable to that computed for the ground-state sextet HO^-−OxH **4** and H_2O−Ox **5** models, although the correlation between oxalate protonation state and C−C bond order appears reversed

Table 8. Selected Structural Properties of BLYP-Optimized Geometries for OxDC Active Site Models **11**−**14**[a]

internal coordinate[b,c]	O_2−Ox **11** $M_S = 5/2$	O_2−OxH **12** $M_S = 5/2$	HO_2−Ox **13** $M_S = 5/2$	HO_2−OxH **14a** $M_S = 3/2$	HO_2−OxH **14b** $M_S = 5/2$
Mn−H95N$^{\epsilon}$	2.425	2.276	2.326	2.096	2.264
Mn−H97N$^{\epsilon}$	2.282	2.169	2.327	2.155	2.137
Mn−E101O$^{\delta 1}$	2.123	1.988	2.182	1.952	2.276
Mn−H140N$^{\epsilon}$	2.236	2.352	2.298	2.092	2.132
Mn−O_A	2.096	2.041	1.979	1.860	2.067
Mn−O_B	1.986	2.093	2.419	1.894	1.929
C_A−C_B	1.605	1.561	**3.880**[d]	1.558	1.558
O_A−O_C	1.322	1.281	1.509	1.395	1.296
H95N$^{\epsilon}$−Mn−H97N$^{\epsilon}$	84.48	89.51	85.85	91.23	91.94
H95N$^{\epsilon}$−Mn−E101O$^{\delta 1}$	166.6	175.5	163.4	172.4	174.4
H95N$^{\epsilon}$−Mn−H140N$^{\epsilon}$	82.98	82.97	86.56	86.41	90.15
H95N$^{\epsilon}$−Mn−O_A	82.76	86.79	89.28	86.33	86.98
H95N$^{\epsilon}$−Mn−O_B	84.58	83.40	103.6	95.03	91.91
H97N$^{\epsilon}$−Mn−E101O$^{\delta 1}$	85.02	152.3	84.42	85.62	87.45
H97N$^{\epsilon}$−Mn−H140N$^{\epsilon}$	88.52	92.29	112.6	93.08	94.30
H97N$^{\epsilon}$−Mn−O_A	166.8	175.2	132.2	177.6	177.9
H97N$^{\epsilon}$−Mn−O_B	92.64	92.61	97.08	88.88	95.70
E101O$^{\delta 1}$−Mn−H140N$^{\epsilon}$	88.52	93.43	84.77	86.83	84.35
E101O$^{\delta 1}$−Mn−O_A	107.1	95.56	107.2	96.83	93.45
E101O$^{\delta 1}$−Mn−O_B	104.1	100.4	91.01	91.85	93.69
H140N$^{\epsilon}$−Mn−O_A	86.48	84.20	114.5	87.01	83.94
H140N$^{\epsilon}$−Mn−O_B	167.3	165.5	149.4	177.5	169.7
O_A−Mn−O_B	89.61	90.03	38.51	91.10	86.11
Mn−O_A−O_C	115.8	117.1	86.72	120.1	120.4
O_B−C_A−O_D	127.8	130.0	173.8	128.0	128.2
O_E−C_B−O_F	131.6	122.1	179.1	125.3	124.3

[a] All calculations are spin-unrestricted. M_S is the sum of the Mn and dioxygen spin states. [b] All bond distances and angles are reported in units of Å and deg, respectively. [c] O_A and O_B correspond to the ligating oxygen atoms in the buried (site A) and solvent-exposed (site B) coordination sites (Figure 1), respectively, and O$^{\delta 1}$ is the oxygen of the butyrate moiety corresponding to the Glu-101 residue that is bonded directly to the metal. O_C is the atom in dioxygen that is not bound to Mn. C_A and O_D are the remaining atoms in the carboxylate coordinated directly to Mn, and atoms in the other carboxyl moiety in oxalate are defined as C_B, O_E, and O_F. [d] Highlighted internal coordinate is discussed explicitly in the text.

(Table 2). The relationship that oxalate protonation appears to maintain C−C bond order in this ligand, presumably by preventing electron transfer to the Mn center, is still evident when the HO^-−OxH **4** and H_2O−Ox **5** models are compared with O_2−Ox **11** and O_2−OxH **12** (Table 2). Finally, the overall metal−ligand covalency in the O_2−OxH complex **12** is slightly enhanced relative to that in O_2−Ox **11** (Table 9), the majority of the difference arising from enhanced covalency in the bond between the imidazole moiety corresponding to Glu-101 and the metal in **12** being only partially counterbalanced by enhanced metal-O_2 covalency in **11**. The addition of a proton to the dioxygen ligand in **12** to give the deprotonated model **14** results in a ground spin multiplicity of $S = 3/2$ based on the lowest energy structure exhibiting modest spin contamination (Table 2). The electronic structure of HO_2−OxH **14**, based on the NPA charges and SMD computed for the Mn center and hydroperoxy ligand, resembles the hydroperoxy complexes **9** and **10**. Since the total Mn *d*-orbital occupancy is 5.5 e^-, we consider the metal to be Mn(II) in a quartet state spin state (Table 2). Protonation also increases the oxalate C−C bond order in **14** relative to **12**. Hence, in a result that has important mechanistic implications, merely protonating dioxygen in the OxDC does not appear to be sufficient to weaken the oxalate C−C bond, either by inductive effects or direct electron transfer.

In sharp contrast to the results described above, no stable, minimum energy structure was found for the active site model HO_2−Ox **13** in which intact oxalate dianion and protonated dioxygen are both bound to the Mn center. Thus, efforts to optimize **13** using the BLYP functional gave a structure in which the oxalate C−C bond had clearly broken to give two O−C−O fragments with bond angles similar to that in CO_2 (Figure 3).[46] Furthermore, the dioxygen ligand becomes bidentate so as to fill the open coordination site produced as a result of C−C bond cleavage, giving Mn−O_C and O_A−O_C bond lengths of 1.98 and 1.51 Å, respectively (Table 8). Although it is hard to make any firm mechanistic conclusions in the absence of more detailed reaction path studies,[46] C−C bond cleavage in **13** therefore appears to be correlated with the formation of a Mn-bound peroxide. Notably, the ground spin multiplicity of this complex is 5/2, suggesting that the metal can be formally regarded as Mn(II).

The remarkable finding that C−C bond cleavage occurred in active site model **13** on BLYP optimization led us to investigate whether our choice of model chemistry had been appropriate. The geometries of the model complexes O_2−Ox **11**, O_2−OxH **12**, HO_2−Ox **13**, and HO_2−OxH **14** were therefore reoptimized using the B3LYP functional in combination with an SVP(+) basis set. This functional was chosen given (i) its inclusion of orbital-dependent exchange, which tends to stabilize high-spin rather than low-spin

Table 9. Calculated Covalent Character of Mn−Ligand Bonds in Complexes **11**−**14** at Their BLYP-Optimized Ground-State Geometries[a,b]

bond	bond covalency (|e⁻|)	total covalency (|e⁻|)
	O_2−Ox **11** $M_S = 5/2$	
Mn−H95N$^{\epsilon}$	0.067	0.872
Mn−H97N$^{\epsilon}$	0.085	
Mn−E101O$^{\delta 1}$	0.132	
Mn−H140N$^{\epsilon}$	0.096	
Mn−O_A	0.319	
Mn−O_B	0.173	
	O_2−OxH **12** $M_S = 5/2$	
Mn−H95N$^{\epsilon}$	0.103	0.927
Mn−H97N$^{\epsilon}$	0.125	
Mn−E101O$^{\delta 1}$	0.244	
Mn−H140N$^{\epsilon}$	0.083	
Mn−O_A	0.234	
Mn−O_B	0.138	
	HO_2−Ox **13** $M_S = 5/2$	
Mn−H95N$^{\epsilon}$	0.072	0.682
Mn−H97N$^{\epsilon}$	0.078	
Mn−E101O$^{\delta 1}$	0.108	
Mn−H140N$^{\epsilon}$	0.083	
Mn−O_A	0.277	
Mn−O_C	0.064[c]	
	HO_2−OxH **14a** $M_S = 3/2$	
Mn−H95N$^{\epsilon}$	0.256	2.058
Mn−H97N$^{\epsilon}$	0.176	
Mn−E101O$^{\delta 1}$	0.472	
Mn−H140N$^{\epsilon}$	0.122	
Mn−O_A	0.664	
Mn−O_B	0.368	
	HO_2−OxH **14b** $M_S = 5/2$	
Mn−H95N$^{\epsilon}$	0.128	1.163
Mn−H97N$^{\epsilon}$	0.154	
Mn−E101O$^{\delta 1}$	0.080[d]	
Mn−H140N$^{\epsilon}$	0.168	
Mn−O_A	0.374[d]	
Mn−O_B	0.259	

[a] Atom labels correspond to those used in Table 4. [b] M_S is the sum of the Mn and dioxygen spin states. [c] Optimization resulted in C−C bond cleavage to generate two dissociated molecules of carbon dioxide. The coordinating ligand in site B was therefore the protonated oxygen atom (O_C) of the hydroperoxide anion formed by electron transfer (Figure 3). [d] Optimization resulted in transfer of a proton from the hydroperoxide ligand to the O$^{\delta 1}$ oxygen of the butyrate moiety corresponding to Glu-101 (*Bacillus subtilis* numbering). This likely reflects the greater charge on the carboxylate oxygens versus O_2 in the $M_S = 5/2$ complex. Notably, all complexes with protonated dioxygen in $M_S = 3/2$ complexes did not show such a transfer, suggesting that the low-spin coupling (physically between $S = 5/2$ high-spin Mn and $S = 1$ dioxygen) promotes electron transfer to the O_2 moiety. This is consistent with Pauli exclusion and implies that dioxygen protonation will not occur prior to oxalate deprotonation.

multiplicities due to stabilization of parallel spins by Fermi exchange,[47] and (ii) its excellent performance when used with basis sets of modest size.[48] In addition, all optimizations were carried out for $M_S = 5/2$, in part because this configuration is expected to maximize Pauli repulsion and hence disfavor electron redistribution from oxalate to the HO_2 moiety, thereby biasing the calculation against oxidation. Even when

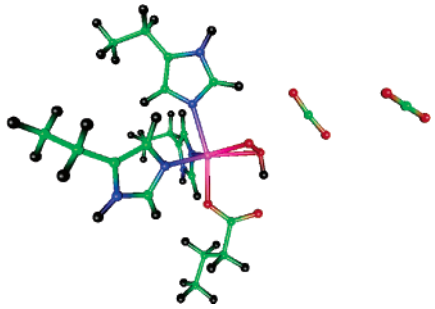


Figure 3. Low-energy structure of the HO_2−Ox active site model complex at $M_S = 5/2$ spin multiplicity, showing C−C bond scission to yield two CO_2 molecules and a hydroperoxide ligand. Atoms are colored using the following scheme: C, green; H, black; N, blue; O, red; Mn, magenta.

the B3LYP functional was used to optimize model **13**, the C−C bond in the oxalate ligand elongated to over 1.88 Å, an observation clearly consistent with significant weakening even though complete dissociation did not occur. Presumably, this difference arises from the description of exchange in BLYP (a pure functional) and B3LYP (a hybrid functional with some explicit orbital-dependent exchange) involving orbitals centered on each oxalate-CO_2^{δ}- fragment. Similarly, the dioxygen O_A−O_C bond elongated to 1.43 Å, rather than 1.51 Å as seen in the BLYP-optimization. Model complexes **11**, **12**, and **14**, however, showed no significant differences in the optimized geometries obtained in calculations employing either BLYP or B3LYP functionals.

In addition to the geometric changes that take place during BLYP optimization of the HO_2−Ox complex **13**, substantial reorganization of the electronic structure occurs when dioxygen is protonated and oxalate deprotonated. The ground state of this complex is a spin sextet, which deviates from the behavior of the analogous HO_2−H_2O and HO_2−HO complexes, both of which were ground-state quartets. Although the NPA charge on Mn is unremarkable compared to that in complexes, **11**, **12**, and **14**, the charges on both atoms of dioxygen in **13** are noticeably lower than in any other dioxygen-containing model examined in these studies (Table 2). Perhaps more significantly, the SMD of the Mn ion in the final structure of **13** suggests that the manganese is best considered as divalent, which is consistent with an overall metal−ligand covalency of 0.68 |e⁻|, lower than that observed for dioxygen-containing models **4**−**6**, **12**, and **14**, and even the HO−H_2O resting complex **3**. Overall, simultaneous O_2 protonation and deprotonation of oxalate therefore results in an apparent ligand oxidation to give a complex in which Mn(II) is bound to hydroperoxide (Figure 3). Thus, even the truncated models considered here exhibit chemistry that is essentially that of the manganese center in oxalate oxidase.

Mechanistic Relevance. Several mechanistically interesting points can be inferred from these model calculations. First, C−C bond cleavage in Mn-bound oxalate with concomitant oxidation of the substrate and dioxygen reduction requires both the formation of an oxalate anion *and* polarization of the dioxygen-derived ligand. Although we model the latter requirement in these calculations by the

addition of a proton to form a Mn-bound hydroperoxy species in complexes **13** and **14**, examination of the crystal structure of *Bacillus subtilis* OxDC does not show any residues that might function as general acids to accomplish such a protonation. It is therefore likely that if dioxygen indeed coordinates the active site Mn, electrostatic polarization of the O−O bond would have to result from the protein environment surrounding the metal center. Modulation of the extent of such polarization might be a potential means to discriminate between oxalate decarboxylation and oxidation, which we have suggested differ by the transfer of one or two electrons from oxalate to metal-bound dioxygen, respectively. Second, the need to deprotonate the distal carboxylic acid of the oxalate substrate is consistent with experimental proposals that rationalize observed kinetic isotope effects,[9] and the functional importance of Glu-162, which is an excellent candidate for the general base on the basis of site-directed mutagenesis studies.[17] On this point, we note that mutation of this residue to alanine results in an E162A OxDC mutant that exhibits oxalate oxidase rather than decarboxylase activity. This interesting observation can be rationalized on the basis of these calculations by postulating direct proton transfer from Mn-bound oxalate to the dioxygen-derived ligand in an adjacent coordination site as a consequence of greater conformational freedom and the absence of the general base to yield an intermediate similar to **13** with subsequent C−C bond cleavage and electron transfer to form two molecules of CO_2. In the presence of Glu-162, however, hydrogen bonding between the distal end of bound oxalate monoanion and the Glu-162 protein side chain, together with the absence of a proton donor in the hydrophobic pocket to which we suggest dioxygen binds, would prevent such a proton transfer thereby preventing substrate oxidation and favoring decarboxylation to yield a formal metal-bound formyl radical anion (Scheme 1). Verification or refutation of this idea awaits further experimental and computational study.

Conclusions

All mechanistic proposals for OxDC to date have postulated the involvement of formal oxidation states for manganese that are higher than that observed (Mn(II)) in the resting enzyme by electron paramagnetic resonance spectroscopy.[2a,9,11] In part, these hypotheses reflect the need to rationalize (i) the generation of an oxalate radical anion as an intermediate and (ii) the requirement of the enzyme for oxygen during catalytic turnover.[2a] This conflicts, however, with the total Mn *d*-orbital occupancy of slightly more than five electrons that is a common element of all the ground-state structures computed in this study. In the case of active site models **1**−**6**, this electronic configuration can be explained from simple ligand field theory, i.e., that Mn(II) is coordinated by relatively inert ligands that transfer a small amount of charge into the empty Mn 4*s* orbital to form stable complexes. In the other cases, such as **7**−**14**, in which some form of dioxygen is bound to the metal, divalency arises instead from the presence of bonds with significant covalency that are formed by extensive mixing of atomic orbitals on Mn (*d*) and ligand atoms (*s*,*p*) to give molecular orbitals with significant Mn *d* character. One might think of this in terms of ligand redox noninnocence,[49] i.e., there is effectively partial electron transfer to the metal ion. Thus, although any assignment of formal or spectroscopic oxidation state in the latter models is complicated by the nature of the metal−ligand bonding,[49a,50] if such a state is demanded our calculations suggest that Mn(II) persists even when the metal is bound to oxidizing species, such as dioxygen or hydroperoxide radical. Our results therefore raise the possibility that the Mn(II) center may play an alternative role in the OxOx and OxDC-mediated cleavage of the oxalate C−C bond. Such a proposal is also consistent with recent failures to observe Mn(III)-containing species when OxDC is undergoing steady-state turnover.[3b]

Thus, the function of the metal ion may be to juxtapose dioxygen and oxalate, or formate radical anion (its breakdown product), and establish an electronic structure such that electrons can be shuttled directly between these species, thereby removing the need for the existence of Mn(III) as a discrete intermediate oxidant that would be formed by metal-to-dioxygen charge transfer. In a mechanism in which divalent manganese serves as a mediator, active site models **12**−**14** are particularly intriguing. In models **12** and **14** the oxalate C−C bond order is approximately 0.9 (Table 2), and there is a small amount of electron transfer from Mn(II) to the bound form of dioxygen. Removal of a proton from oxalate, however, yields a model (**13**) for which no stable minimum containing an intact oxalate C−C bond is obtained upon geometry optimization, at least in the absence of solvation, perhaps implying that the intrinsic reactivity of the active center in *Bacillus subtilis* oxalate *decarboxylase* is to *oxidize* oxalate. Although it might be argued that this result reflects a poor choice of functional, the behavior of this active site model when optimization was performed using B3LYP, which includes a significant amount of orbital-dependent exchange, was qualitatively consistent. While the change of exchange functional yielded a structure containing an oxalate C−C bond, once again the presence of oxalate anion and the hydroperoxy ligand in model **13** resulted in a significant weakening of this bond. Thus, we tentatively conclude that the intrinsic reactivity of the Mn(II) center is to promote proton-coupled electron transfer from oxalate to protonated dioxygen to form a metal-bound peroxide anion.

Since this intrinsic reactivity is not observed in OxDC, we believe that these calculations provide evidence to support the hypothesis that OxDC evolved from OxOx by gene duplication leading to a bicupin structure in which the protein environment modulates the intrinsic metallocenter reactivity.[51] The overall cupin fold around the metallocenter is largely unchanged by this hypothetical duplication, so our finding that a common intrinsic reactivity for the manganese centers in the two enzymes in the absence of a detailed description of their polypeptide surroundings is not too surprising. The different transformations of oxalate that are catalyzed by OxOx and OxDC may therefore reflect subtle changes in electrostatics, hydrogen bonding, and differential molecular dynamics associated with the protein structure in which the metal center is embedded. One plausible scenario is that mutations close to the manganese center in OxDC

have accumulated to ensure that (i) manganese-bound dioxygen cannot be protonated and/or (ii) the intermediate formate radical anion is protonated prior to transfer of a second electron to the Mn-superoxyl system, which would presumably be coupled to flattening of the O−C−O bond angle in the formyl radical.[52] In this regard, it is interesting to note that the removal of one metal binding site in OxDC by site-directed mutagenesis largely eliminates the catalytic activity of the other site,[53] raising the intriguing possibility that the two sites in the enzyme may be redox-coupled. Regardless of whether such intersite communication holds up to experimental scrutiny,[17] the molecular basis for the apparent evolutionary honing of OxOx to OxDC reactivity remains to be elucidated.

Acknowledgment. We thank the National Science Foundation (CHE-0079008) and the National Institutes of Health (DK61666) for support of this work. C.H.C. also gratefully acknowledges support from a Ruth L. Kirchstein National Research Service Award (DK61193).

Supporting Information Available: Complete citation for ref 34 and Cartesian coordinates for the B3LYP-optimized structures of models **11**−**13** and **14b**. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Dutton, M. V.; Kathiara, M.; Gallagher, I. M.; Evans, C. S. *FEMS Microbiol. Lett.* **1994**, *116*, 321−326. (b) Magro, P.; Marciano, P.; Di Lenna, P. *FEMS Microbiol. Lett.* **1988**, *49*, 49−52. (c) Lillehoj, E. B.; Smith, F. G. *Arch. Biochem. Biophys.* **1965**, *109*, 216−220. (d) Shimazono, H.; Hayaishi, O. *J. Biol. Chem.* **1957**, *227*, 151−159. (e) Shimazono, H. *J. Biochem. (Tokyo)* **1955**, *42*, 321−340.

(2) (a) Tanner, A.; Bowater, L.; Fairhurst, S. A.; Bornemann, S. *J. Biol. Chem.* **2001**, *276*, 43627−43634. (b) Tanner, A.; Bornemann, S. *J. Bacteriol.* **2000**, *182*, 5271−5273.

(3) (a) Svedružić, D.; Jónsson, S.; Toyota, C. G.; Reinhardt, L. A.; Ricagno, S.; Lindqvist, Y.; Richards, N. G. J. *Arch. Biochem. Biophys.* **2005**, *433*, 176−192. (b) Chang, C. H.; Svedružić, D.; Ozarowski, A.; Yeagle, G.; Walker, L.; Britt, R. D.; Angerhofer, A.; Richards, N. G. J. *J. Biol. Chem.* **2004**, *279*, 52840−52849.

(4) (a) Emiliani, E.; Riera, B. *Biochim. Biophys. Acta* **1968**, *167*, 414−421. (b) Emiliani, E.; Bekes, P. *Arch. Biochem. Biophys.* **1964**, *105*, 488−493.

(5) (a) Borodine, A. *Annalen* **1861**, *119*, 121−123. (b) Kolbe, H. *Annalen* **1849**, *69*, 257−294.

(6) Hunsdiecker, H.; Hunsdiecker, C. *Chem. Ber.* **1942**, *75*, 291−297.

(7) Halliwell, B. *Biochem. J.* **1972**, *129*, 497−498.

(8) Drummond, A. Y.; Waters, W. A. *J. Chem. Soc.* **1953**, 435−443.

(9) Reinhardt, L. A.; Svedruzic, D.; Chang, C. H.; Cleland, W. W.; Richards, N. G. J. *J. Am. Chem. Soc.* **2003**, *125*, 1244−1252.

(10) (a) Huynh, M. H. V.; Meyer, T. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13138−13141. (b) Qu, S. J.; Klinman, J. P. *Biochemistry* **1998**, *37*, 12513−12525.

(11) Anand, R.; Dorrestein, P. C.; Kinsland, C.; Begley, T. P.; Ealick, S. E. *Biochemistry* **2002**, *41*, 7659−7669.

(12) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH: Weinheim, 2001. (b) Siegbahn, P. E. M.; Blomberg, M. R. A. *Annu. Rev. Phys. Chem.* **1999**, *50*, 221−249. (c) Siegbahn, P. E. M. *Adv. Chem. Phys.* **1996**, *93*, 333−387.

(13) (a) Weinhold, F. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; pp 1792−1811. (b) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899−926.

(14) Fe-dependent enzymes: (a) Solomon, E. I.; Szilagyi, R. K.; DeBeer George, S.; Basumallick, L. *Chem. Rev.* **2004**, *104*, 419−458. (b) Bassan, A.; Borowski, T.; Siegbahn, P. E. M. *J. Phys. Chem. B* **2004**, *108*, 13031−13041. (c) Boone, A. J.; Chang, C. H.; Greene, S. N.; Herz, T.; Richards, N. G. J. *Coord. Chem. Rev.* **2003**, *238*−*239*, 291−314. (d) Friesner, R. A.; Baik, M. H.; Gherman, B. F.; Guallar, V.; Wirstam, M.; Murphy, R. B.; Lippard, S. J. *Coord. Chem. Rev.* **2003**, *238*−*239*, 267−290. (e) Ghosh, A.; Steene, E. *J. Biol. Inorg. Chem.* **2001**, *6*, 739−752. (f) Harris, D. L. *Curr. Opin. Chem. Biol.* **2001**, *5*, 724−735. (g) Spiro, T. G.; Zgierski, M. Z.; Kozlowski, P. M. *Coord. Chem. Rev.* **2001**, *219*−*221*, 923−936. (h) Li, S.; Hall, M. B. *Inorg. Chem.* **2001**, *40*, 18−24. (i) Siegbahn, P. E. M.; Blomberg, M. R. A. *Chem. Rev.* **2000**, *100*, 421−437.

(15) Mn-dependent enzymes: (a) Noodleman, L.; Lovell, T.; Han, W. G.; Himo, F. *Chem. Rev.* **2004**, *104*, 459−508. (b) Sinnecker, S.; Neese, F.; Noodleman, L.; Lubitz, W. *J. Am. Chem. Soc.* **2004**, *126*, 2613−2622. (c) Ivanov, I.; Klein, M. L. *Proteins: Struct. Funct. Genet.* **2003**, *54*, 1−7. (d) Himo, F.; Siegbahn, P. E. M. *Chem. Rev.* **2003**, *103*, 2421−2456. (e) Siegbahn, P. E. M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 227−235. (f) Siegbahn, P. E. M. *Theor. Chem. Acc.* **2001**, *105*, 197−206.

(16) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. B. *Nucl. Acids Res.* **2000**, *28*, 235−242.

(17) Just, V. J.; Stevenson, C. E. M.; Bowater, L.; Tanner, A.; Bornemann, S. *J. Biol. Chem.* **2004**, *279*, 19867−19874.

(18) In structural comparisons, active site models were superimposed by overlapping the constrained methyl carbons using algorithms implemented in a custom Python script, which employed the following freely available supplementary modules: Numerical Python (Dubois, P. F. available on the World Wide Web at http://sourceforge.net/projects/numpy as of January 2005) and Scientific Python (Hinson K. module library available on the World Wide Web at http://starship.python.net/∼hinsen/ScientificPython/ as of January 2005). This script and/or the Cartesian coordinates for the initial and optimized structures of active site models **1**−**14** can be obtained from the authors on request.

(19) (a) Von Arnim, M.; Ahlrichs, R. *J. Comput. Chem.* **1998**, *19*, 1746−1757. (b) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165−169.

(20) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(21) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571−2577.

(22) (a) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119−124. (b) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283−290. (c) Neese, F.; Olbrich, G. *Chem. Phys. Lett.* **2002**, *362*, 170−178.

(23) (a) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294−301. (b) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222. (c) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257−2261.

(24) We note that the implementation of B3LYP within TURBOMOLE differs from the standard form implemented in Gaussian,[25] in that the fifth (VWN-V) rather than the third (VWN-III) form of the VWN local correlation is employed.[26]

(25) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(26) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Chem.* **1980**, *58*, 1200−1211.

(27) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299−310.

(28) (a) Chang, C. H.; Boone, A. J.; Bartlett, R. J.; Richards, N. G. J. *Inorg. Chem.* **2004**, *43*, 458−472. (b) Greene, S. N.; Richards, N. G. J. *Inorg. Chem.* **2004**, *43*, 7030−7041.

(29) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736−1740.

(30) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735−746.

(31) Reed, G. H.; Markham, G. D. In *Biological Magnetic Resonance*; Berliner, L. J., Reuben, J., Eds.; Plenum Press: New York, 1984; Vol. 6, pp 73−142.

(32) Baes, C. F., Jr.; Mesmer, R. E. *The Hydrolysis of Cations*; Wiley-Interscience: New York, 1976.

(33) For example, the oxalate moiety cleaved to give two CO_2 molecules when a dianionic HO^-−ox active site model structure was optimized within the RI approximation using a basis that lacked diffuse functions. When the same complex was optimized without this Coulomb integral approximation, its behavior was similar to that observed for the HO^-−OxH complex **4** in that the ethylimidazole ligands corresponding to His-95 and His-140 merely dissociated from the metal center. Given these radically different results, we attribute the apparent oxalate oxidation in the RI-BLYP/SVP calculation to spurious "electron transfer" from dianionic oxalate to the metal complex. Such observations are clearly the result of an inadequate description of these strongly charged species in vacuo, an interpretation supported by occupation of positive-energy molecular orbitals even in the calculations including diffuse functions. The latter is a hallmark of either a lack of sufficiently diffuse basis functions which would allow proper description of a (marginally) stable anion in vacuo or a lack of continuum basis functions which would permit a description of the ionization that would actually occur physically in vacuo.

(34) Schwarzenbacher, R. et al. *Proteins: Struct. Funct. Genet.* **2004**, *56*, 392−395.

(35) (a) Coleman, W. M.; Taylor, L. T. *Coord. Chem. Rev.* **1980**, *32*, 1−31. (b) Que, L.; Reynolds, M. F. *Met. Ions Biol. Syst.* **2000**, *37*, 505−525. (c) Vetting, M. W.; Wackett, L. P.; Que, L.; Lipscomb, J. D.; Ohlendorf, D. H. *J. Bacteriol.* **2004**, *186*, 1945−1958.

(36) (a) Mills, S. A.; Klinman, J. P. *J. Am. Chem. Soc.* **2000**, *122*, 9897−9904. (b) Qu, S. J.; Klinman, J. P. *Biochemistry* **1998**, *37*, 8572−8581.

(37) (a) Metz, M.; Solomon, E. I. *J. Am. Chem. Soc.* **2001**, *123*, 4938−4950. (b) Brunold, T. C.; Solomon, E. I. *J. Am. Chem. Soc.* **1999**, *121*, 8288−8295.

(38) (a) van Gastel, M.; Lubitz, W.; Lassmann, G.; Neese, F. *J. Am. Chem. Soc.* **2004**, *126*, 2237−2246. (b) Witting, P. K.; Mauk, A. G. *J. Biol. Chem.* **2001**, *276*, 16540−16547. (c) Licht, S.; Gerfen, G. J.; Stubbe, J. *Science* **1996**, *271*, 477−481. (d) Balagopalakrishna, C.; Abugo, O. O.; Horsky, J.; Manoharan, P. T.; Nagababu, E.; Rifkind, J. M. *Biochemistry* **1998**, *37*, 13194−13202. (e) Cadenas, E. *Methods Enzymol.* **1995**, *251*, 106−116. (f) McEldoon, J. P.; Dordick, J. S. *J. Biol. Chem.* **1989**, *266*, 14288−14293.

(39) (a) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Mayes, S. A.; Benson, D. E.; Sweet, B. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615−1622. (b) Gouet, P.; Jouve, H. M.; Williams, P. A.; Anderson, I.; Andreoletti, P.; Nussaume, L.; Hadju, J. *Nat. Struct. Biol.* **1996**, *3*, 951−956. (c) Ghosh, A. K.; Steene, E. *J. Biol. Inorg. Chem.* **2001**, *6*, 739−752. (d) Fabian, M.; Palmer, G. *Biochemistry* **2001**, *40*, 1867−1874.

(40) (a) Fouqueau, A.; Mer, S.; Casida, M. E.; Daku, L. M. L.; Hauser, A.; Mineva, T.; Neese, F. *J. Chem. Phys.* **2004**, *120*, 9473−9486. (b) Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. *J. Phys. Chem. A* **2004**, *108*, 5479−5483. (c) Ghosh, P.; Bill, E.; Weyhermuller, T.; Neese, F.; Wieghardt, K. *J. Am. Chem. Soc.* **2003**, *125*, 1293−1308. (d) Staroverov, V. N.; Davidson, E. R. *Chem. Phys. Lett.* **2001**, *340*, 142−150.

(41) (a) Gräfenstein, J.; Cremer, D. *Mol. Phys.* **2001**, *99*, 981−989. (b) Wang, J.; Becke, A. D.; Smith, V. H., Jr. *J. Chem. Phys.* **1995**, *102*, 3477−3480. (c) Baker, J.; Scheiner, A.; Andzelm, J. *Chem. Phys. Lett.* **1993**, *216*, 380−388.

(42) (a) Noodleman, L.; Peng, C. Y.; Case, D. A.; Mouesca, J.-M. *Coord. Chem. Rev.* **1995**, *144*, 199−244. (b) Noodleman, L.; Case, D. A. *Adv. Inorg. Chem.* **1992**, *38*, 423−470. (c) Noodleman, L.; Case, D. A.; Aizman, A. *J. Am. Chem. Soc.* **1988**, *110*, 1001−1005.

(43) (a) Yoo, S. J.; Angove, H. C.; Burgess, B. K.; Hendrich, M. P.; Münck, E. *J. Am. Chem. Soc.* **1999**, *121*, 2534−2545. (b) Papaefthymiou, V.; Millar, M. M.; Münck, E. *Inorg. Chem.* **1986**, *25*, 3010−3014.

(44) Such a model is by definition oversimplified given the covalency of bonding and consequent breakdown of the ligand field approximation. It is nevertheless useful to frame the discussion of the formal valence of the manganese ion within that given in the literature.

(45) The discrepancy between formal and partial charge in first-row transition ions often arises as a result of charge transfer from the ligand into the metal 4*s* orbital associated with coordinate-covalent bonding.

(46) The geometry optimization of complexes for which oxalate C−C bond cleavage occurs is partial, given the difficulty of locating a well-defined minimum energy structure due to flattening of the potential surface. In any event, oxalate C−C bond clearly breaks, and the linear O−C−O angles strongly suggest that the observed "cleavage" is oxidative. While recognizing that it is technically fallacious to discuss changes during a geometry optimization in terms of a chemical reaction, these changes nevertheless imply that a stable,

minimum-energy structure in which the oxalate C−C bond is intact was not be found with this model chemistry, starting geometry, and gradient and Hessian update algorithms. We have therefore framed our discussion of the behavior of model complex **13** relative to that of other *minimized* structures obtained in these studies.

(47) (a) Salomon, O.; Reiher, M.; Hess, B. A. *J. Chem. Phys.* **2002**, *117*, 4729−4737. (b) Reiher, M.; Salomon, O.; Hess, B. A. *Theor. Chem. Acc.* **2001**, *107*, 48−55.

(48) This performance enhancement has been suggested to arise from a fortuitous cancellation of errors between truncated bases and the erroneous long-range behavior of the exchange-correlation potential (see ref 12).

(49) (a) Bill, E.; Bothe, E.; Chaudhuri, P.; Chlopek, K.; Herebian, D.; Kokatam, S.; Ray, K.; Weyhermüller, T.; Neese, F.; Wieghardt, K. *Chem. Eur. J.* **2005**, *11*, 204−224. (b) Chaudhuri, P.; Verani, C. N.; Bill, E.; Bothe, E.; Weyhermüller, T.; Neese, F.; Wieghardt, K. *J. Am. Chem. Soc.* **2001**, *123*, 2213−2223.

(50) Jörgensen, C. K. *Oxidation Numbers and Oxidation States*; Springer: Berlin, 1969.

(51) (a) Dunwell, J. M.; Khuri, S.; Gane, P. J. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 153−179. (b) Dunwell, J. M.; Culham, A.; Carter, C. E.; Sosa-Aguirre, C. R.; Goodenough, P. W. *Trends Biochem. Sci.* **2001**, *26*, 740−745.

(52) (a) Krauss, M.; Neumann, D. *Chem. Phys. Lett.* **1972**, *14*, 26−27. (b) Pacansky, J.; Wahlgren, U.; Bagus, P. S. *J. Chem. Phys.* **1975**, *62*, 2740−2744.

(53) Chakraborty, S.; Chakraborty, N.; Jain, D.; Salunke, D. M.; Datta, A. *Protein Sci.* **2002**, *11*, 2138−2147.

CT050063D

JCTC Journal of Chemical Theory and Computation

Improving the QM/MM Description of Chemical Processes: A Dual Level Strategy To Explore the Potential Energy Surface in Very Large Systems

Sergio Martí and Vicente Moliner*

Departament de Ciències Experimentals, Universitat Jaume I, Box 224, 12080 Castellón, Spain

Iñaki Tuñón*

Departament de Química Física/IcMol, Universidad de Valencia, 46100 Burjasot, Valencia, Spain

Received May 31, 2005

Abstract: Potential energy surfaces are fundamental tools for the analysis of reaction mechanisms. The accuracy of these surfaces for reactions in very large systems is often limited by the size of the system even if hybrid quantum mechanics/molecular mechanics (QM/MM) strategies are employed. The large number of degrees of freedom of the system requires hundreds or even thousands of optimization steps to reach convergence. Reactions in condensed media (such as enzymes or solutions) are thus usually restricted to be analyzed using low level quantum mechanical methods, thus introducing a source of error in the description of the QM region. In this paper, an alternative method is proposed, coupled to the use of a micro/macroiteration algorithm during the optimization. In these algorithms, the number of microsteps involved in the QM region optimization is usually much smaller than the number of macrosteps required to optimize the MM region. Thus, we define a new potential energy surface in which the gas-phase energy of the QM subsystem and the interaction energy with the MM subsystem are calculated at different computational levels. The high computational level is restricted to the gas-phase energy, which is only requested during the microsteps. The dual level strategy is tested for two reactions in solution (the Menshutkin and the oxy-Cope reactions) and an enzymatic one (the nucleophilic substitution of 1,2-dichloroethane in DhlA). The performance of the proposed computational scheme seems to be quite promising for future applications in other systems.

1. Introduction

The theoretical study of chemical processes taking place in condensed media (such as enzymes or solution) has become very promising in the last years because of the development of specific computational techniques. The strategy designed to afford such a study requires the consideration of at least three key steps. The first one is the selection and/or construction of the adequate potential energy function to describe the desired process. The quality of this potential energy function is crucial as far as it determines the possible molecular mechanisms of a given chemical reaction. The second step is the location and characterization of the relevant stationary structures (reactants, products, intermediates, and transition structures). This step is essential to discriminate among several mechanisms and to obtain all the information which the transition state theory can provide about the reaction (as, for example, transition vectors, rate constants,

* Corresponding authors phone: +34-6-354-48-80 ext 46100; fax: +34-6-354-45-64; e-mail: Ignacio.Tunon@uv.es (I.T.) and phone +34964728084; fax: +34964728066; e-mail: moliner@exp.uji.es (V.M.).

10.1021/ct0501396 CCC: $30.25
Published on Web 08/06/2005

kinetic isotopic effects, etc.). In any case, a third step involving molecular simulations is required in order to obtain well suited averaged macroscopic properties for the set of thermodynamic states involved in the studied chemical process. So far, one of the most important properties which can be obtained is the free energy profile, from which the activation and reaction free energies are computed. Nonetheless except in the simplest cases, the information obtained during the exploration of the potential energy surface is crucial to perform a proper sampling along adequate reaction coordinates.

The main specific problem related with the study of processes in enzymes or in solution is associated with the size of the systems being considered: the huge number of electrons and degrees of freedom. For the first problem a reasonable solution is provided by hybrid quantum mechanics/molecular mechanics (QM/MM) strategies.[1−3] Usually, electronic reordering due to breaking and forming bonds is located in a reduced portion of the system. One then needs to describe just the atoms of this region using quantum mechanics, while the rest of the system is described using MM potentials. In this way the number of electrons to be considered is drastically reduced. The successful application of this methodology is nowadays well reported in the scientific literature.[1−5] One could then wonder if it would be possible to perform such a reduction of the number of degrees of freedom to be considered, including only those atoms directly involved in the reaction. The answer is generally negative, consisting of a lot of examples of enzymatic and in solution processes where important changes in the environment appear as long as the reaction proceeds. For example, in nucleophilic substitution reactions in solution, solvent molecules reorganize following the reaction charge flow.[6−8] Another example is provided by the enzyme catechol O-methyl transferase, where the activation of the reactants require an important change in the coordination shell of the magnesium cation with respect to the X-ray structure.[9] So, meanwhile in most cases the number of electrons which need to be explicitly treated can be clearly reduced, whereas the same is not generally true for the degrees of freedom of the system. Thus, an adequate study of such kind of processes must consider the very high dimensionality of the potential energy surfaces involved. Fortunately there are efficient optimization algorithms capable of working with a very large number of degrees of freedom.[10]

This paper is organized as follows. In the following section we first summarize the characteristics of the micro/macroiterations optimization algorithm, which is the most successful, in its different variants, to study large systems described with QM/MM methods. Afterward, and taking advantage of the use of this algorithm, we propose a dual-level strategy that can be used to obtain more accurate descriptions of a chemical reaction in very large systems at a still reasonable cost. Finally, the method is tested for three different reactions, two in aqueous solution and one in an enzymatic active site.

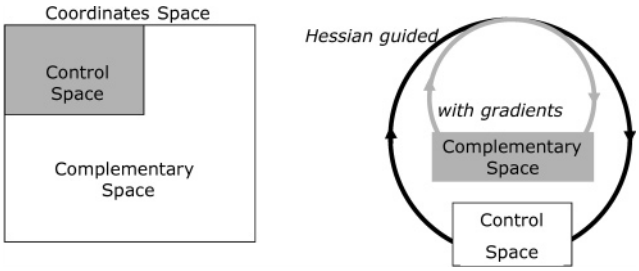


Figure 1. Schematic representation of the micro/macroiteration algorithm. The coordinate space is divided into two subsets: the control space and the complementary space. Optimization steps on the control space (microiterations) make use of a Hessian-based algorithm. At each step of this algorithm the complementary space is minimized using only the gradient vector (macroiterations).

2. Theory

2.1. The Micro/Macroiteration Optimization Algorithm. Most of the standard transition state localization and characterization techniques make use of the Hessian matrix.[11−17] While it is usually straightforward to obtain it in gas-phase calculations, this is not feasible when using a QM/MM based methodology: the large amount of degrees of freedom present in system models turns impracticable its evaluation. To overcome this problem, a partition of the full coordinates space of the system into a control space and complementary space subsets can be done: those atoms or molecules directly involved in the reaction process (plus may be some rounding molecules or residues) are included in the control space, while the rest defines the complementary space. Then, optimization of structures can be efficiently carried out in coupling iterations over these two subspaces: at each step of the control space Hessian guided optimization, the rest of the system is kept fully relaxed merely using gradient vectors. This strategy is also known as the micro/macroiteration method[18−22] (see Figure 1) and leads to stationary structures with the adequate number of negative eigenvalues for a reduced Hessian matrix (the one defined for the control space). In a typical application of the micro/macroiteration approach, the control space usually contains the coordinates of up to ∼100 atoms, while the complementary space can include a number of atoms 2 orders of magnitude larger. When searching stationary structures this is usually translated into about 10^1 Hessian guided optimization steps in the control space and up to 10^2-10^3 gradient based optimization steps in the complementary space at each control space movement. This last number of cycles is of course highly dependent on the size of the system and of the gradient of the considered structure. Anyway, a typical application in enzymes or condensed media can amount up to 10^3 or even more energy and gradients evaluations to fully relax a stationary structure, which means that only low cost computational methods can be employed to describe the QM region (except in those cases where the QM subsystem contains very few atoms). Whereas upcoming computational power would allow using higher electronic Hamiltonian approaches such as ab initio or based in density functional theory (DFT) to describe the quantum region, the usually large amount of gradient vector evaluations needed during

the macrosystem minimization procedure turns this way almost impracticable, still making semiempirical methods the most suitable one.

2.2. The Dual-Level Scheme. We here present a computational scheme to improve the description of QM/MM systems taking advantage of the characteristics of the micro/macroiteration algorithms. We propose a dual level method based on the combination of a high level (HL) and a low level (LL) description of the quantum region. A convenient decomposition of the total energy allows describing the QM gas-phase energy using a HL method, while the interaction with the environment is obtained at a LL one. In this way, the HL term must be evaluated only during the optimization of the control space, while the LL terms are devoted to avoid the more time consuming complementary space optimization. This method has shown to provide considerably better descriptions than standard QM/MM calculations at the semiempirical level, while the computational cost is still reasonable.[23]

The energy for a qm/mm system can be split into different terms

$$E = E_{\mathrm{qm}} + E_{\mathrm{qm/mm}} + E_{\mathrm{mm}} = \langle \Psi | \hat{H}_0 | \Psi \rangle + \langle \Psi | \hat{V} | \Psi \rangle + E_{\mathrm{mm}} \quad (1)$$

where $\hat{H}_0$ is the in vacuo Hamiltonian for the selected qm method, Ψ is the polarized wave function due to the presence of a nonrigid distribution of charges q_{mm} (representing the mm atoms), $\bar{V}$ is the coupling (electrostatic and van der Waals) operator between the qm and mm subsystems, and E_{mm} represents the force field energy. This equation must be solved at each optimization step.

However, having in mind the characteristics of the micro/macrooptimization scheme, there are several considerations that can be applied into the general qm/mm formulation. For each optimization step in the control space (microiteration), the energy function depending on the complementary space is minimized (gradient based macroiterations). Some strategies can be introduced to improve the treatment of the quantum subsystem taking advantage of the characteristics of this optimization procedure. Effectively, eq 1 can be also written as

$$E = \langle \Psi^0 | \hat{H}_0 | \Psi^0 \rangle + [\langle \Psi | \hat{H}_0 | \Psi \rangle - \langle \Psi^0 | \hat{H}_0 | \Psi^0 \rangle] + \langle \Psi | \hat{V} | \Psi \rangle + E_{\mathrm{mm}} = E^{\circ} + E_{\mathrm{pol}} + E_{\mathrm{int}} + E_{\mathrm{mm}} \quad (2)$$

where Ψ^0 stands for the unpolarized (gas phase) wave function of the qm subsystem, E_{pol} for the energy spent in polarizing this wave function, and E_{int} for the interaction between the (polarized) qm subsystem and the mm subsystem. During a micro/macroiteration algorithm only a part of the total energy function is needed at each type of iteration step as either the qm subsystem (in the macrosteps) or the mm subsystem (in the micro steps) is kept frozen. Then, the energy functions explored are

$$E_{\mathrm{micro}} = E_0 + E_{\mathrm{pol}} + E_{\mathrm{int}} \quad (3)$$

$$E_{\mathrm{macro}} = E_{\mathrm{pol}} + E_{\mathrm{int}} + E_{\mathrm{mm}} \quad (4)$$

Equations 3 and 4 allows for improving the quantum treatment of the qm subsystem just describing it in a different way during the micro- and the macroiterations. As far as the flexible mm region is large enough, the number of macroiterations largely exceeds the number of microiterations, a high level (HL) description can be selected only during the microiterations, precisely when the qm subsystem is being optimized, while a low level (LL) description can be used during the macroiterations. One of the first attempts to reduce the more time-consuming step, i.e. the environment relaxation, was accomplished by describing the qm subsystem as an ensemble of classical charges during the macroiterations:[15−17,24]

$$E_{\mathrm{micro}} = E_0^{\mathrm{HL}} + E_{\mathrm{pol}}^{\mathrm{HL}} + E_{\mathrm{int}}^{\mathrm{HL}} \quad (5)$$

$$E_{\mathrm{macro}} = \left[\sum \frac{q_{\mathrm{mm}} q_{\mathrm{qm}}^{\mathrm{fitted}}}{r_{\mathrm{mm,qm}}} + E_{\mathrm{LJ,qm/mm}} \right] + E_{\mathrm{mm}} \quad (6)$$

Albeit nuclear geometry of the core region and fitted charges ($q_{\mathrm{qm}}^{\mathrm{fitted}}$) are kept frozen during the environment optimization, skipping in this way the effect of the core electronic polarization as long as the classical mm electric field changes, this approach can lead to slower convergences (the global number of micro/macroiterations is increased) as far as the energy function contains an intrinsic discontinuity between the micro- and the macroiterations, due to the different treatment of the interaction and polarization terms.

In this point, the dual level method that we propose is based on two different quantum treatments for the qm region, one LL method (usually a semiempirical Hamiltonian) and one HL method (usually, DFT or ab initio correlated methods). The main difference is that we combine these two descriptions into a unique dual level energy surface:

$$E = E_0^{\mathrm{HL}} + E_{\mathrm{pol}}^{\mathrm{LL}} + E_{\mathrm{int}}^{\mathrm{LL}} + E_{\mathrm{mm}} \quad (7)$$

According to this definition of the dual level energy surface, the part of the global energy that is explored during the micro- and macrosteps are

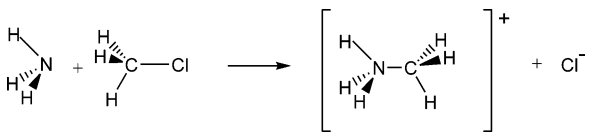
$$E_{\mathrm{micro}} = E_0^{\mathrm{HL}} + E_{\mathrm{pol}}^{\mathrm{LL}} + E_{\mathrm{int}}^{\mathrm{LL}} \quad (8)$$

$$E_{\mathrm{macro}} = E_{\mathrm{pol}}^{\mathrm{LL}} + E_{\mathrm{int}}^{\mathrm{LL}} + E_{\mathrm{mm}} \quad (9)$$

Then, this computational scheme allows a high level description of the qm subsystem during its optimization while avoiding the discontinuities between micro- and macrosteps. In other words, during the environment optimization, the core region is described by a computationally low cost Hamiltonian (such as semiempirical ones); meanwhile a high level correction in the pure gas-phase term is included during the microiterations, keeping the interaction and polarization terms at a low description level. Of course, this computational scheme is not only applicable to the energy but also to the gradient vector and the Hessian matrix:

$$H_{ij} = \frac{\partial^2 E}{\partial x_i \partial x_j} = \frac{\partial^2 E_0^{\mathrm{HL}}}{\partial x_i \partial x_j} + \frac{\partial^2 E_{\mathrm{pol}}^{\mathrm{LL}}}{\partial x_i \partial x_j} + \frac{\partial^2 E_{\mathrm{int}}^{\mathrm{HL}}}{\partial x_i \partial x_j} + \frac{\partial^2 E_{\mathrm{mm}}}{\partial x_i \partial x_j} \quad (10)$$

Scheme 1



It is important to stress here the differences with respect to the popular ONIOM methods.[25,26] Briefly, the strategy here proposed is an electronic embedding method, where the qm region is described using two different computational levels, being the low-level employed to obtain the interaction between the two subsystems. A more detailed discussion on the differences is presented in ref 23.

This dual level strategy, that can be generally denoted as HL:LL/MM calculations, has been coded into the Fortran90 DYNAMO library.[27,28] In the first choice, eqs 5 and 6, the microiteration is obtained by successive calls to Gaussian03[29] for both the energy and the fitted ESP charges[30,31] (considering the nuclear geometry of the core rounded by a distribution of point charges). For the second method, eqs 7−9, Gaussian03 is used just for the high level gas-phase calculation (given the geometry of the core region), meanwhile low level qm/mm and gas-phase calculations are carried out by DYNAMO. In the case of the macroiterations, the L-BFGS-B minimization routines[10] have been incorporated. On the other hand, microiterations are guided by the BAKER algorithm:[32,33] the positive eigenvectors of the Hessian matrix are minimized, while those with a negative eigenvalue are maximized. Control over sequential optimizations steps has been introduced through eigenvector overlapping, allowing the system to undo wrong steps. The BAKER algorithm has been slightly modified from the one already implemented in DYNAMO, to be able to work in a micro/macroscheme.

3. Results and Discussion

Preliminary results of our dual level scheme (eqs 7−9) were presented in ref 23, where the transition state of the chorismate to prephenate rearrangement catalyzed by BsCM was localized, and the kinetic isotopic effects were computed and compared to experiment.[34,35] In this work we present a more systematic test of the dual level QM/MM scheme, exploring the potential energy surface for different reactions in water solution and in an enzymatic environment.

Menshutkin Reaction in Aqueous Solution. The Menshutkin reaction[36] is a S_N2 process in which neutral reactants yield charged products, and then it is a prototypic example of a chemical process where the solvent effects play a decisive role on the process stabilizing the charge separation. We have selected the simplest example of this kind of reaction, as shown in Scheme 1.

This reaction has been thoroughly studied by means of different theoretical strategies involving continuum and QM/MM models,[37−41] and it provides a very good example to test the reliability of the proposed approach. In our calculations the quantum subsystem, composed by methyl chloride plus ammonia, was placed in a water box of 31.4 Å of side containing a total of 1031 TIP3P water molecules.[42] To explore the potential energy surface and locate the transition structure we used a control space, for which the Hessian matrix is obtained, including only the coordinates of the QM subsystem. Structures were converged to a gradient norm lower than 0.1 kJ·mol^{-1}·Å^{-1}. We used different descriptions for our system: AM1/MM, where the quantum subsystem is described by means of the AM1 Hamiltonian; MP2/6-31+G**/MM, where the quantum subsystem is described at the MP2 level of theory; MP2/6-31+G**:CHARGE/MM, where the quantum subsystem is described at the MP2 level of theory and the coupling with the environment is made through the use of MK-ESP charges[30,31] (see eqs 5 and 6), and, finally, MP2/6-31+G**:AM1/MM, where the quantum subsystem is described at the MP2 level of theory and the coupling with the environment (polarization and interaction) is obtained through the AM1 Hamiltonian. The Lennard-Jones parameters for the qm/mm interaction are taken from the OPLS-AA potential,[43,44] except for the chlorine atom, for which we used those of ref 38. We employed a switched cutoff of 14.5 Å for all interactions. The most important characteristics of the transition states located with these methodologies are presented in Table 1.

Table 1. Transition Structures for the Menshutkin Reaction in Water Obtained at Different Computational Levels[a]

	AM1/MM	MP2:AM1/MM	MP2:CHARGE/MM	MP2/MM
d_{NC} (Å)	1.960	2.121	2.080	2.245
d_{CCl} (Å)	2.105	2.187	2.230	2.094
NCCl (deg)	177.7	177.8	177.0	176.4
ν (cm^{-1})	659i	536i	657i	467i
t_{rel}[b]	1.0	22.4	22.4	954

[a] See text. The most significant geometrical parameters (the distances associated with the broken and formed bonds and the angle between these two bonds) are provided together with the frequency of the transition vector and the relative CPU time of the different computational strategies. [b] Relative CPU time per optimization step. A standard optimization step is defined as 1 microiteration followed by 50 macroiterations.

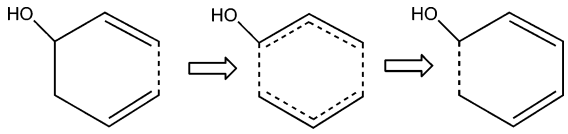
The AM1/MM calculation provides a transition structure with a too short NC distance when compared to the MP2/MM result. This result is clearly improved when the description is made using dual level schemes, either using MP2:AM1/MM or MP2:CHARGE/MM. In general, when compared to the MP2/MM description both dual levels provide a more advanced transition structure, where the NC distance is shorter and the CCl distance is larger, probably due to the underestimation of the interaction between the solute and the solvent molecules. This effect is more important in the MP2:CHARGE/MM calculations than in the MP2:AM1/MM one, which is reflected also in the value of the transition frequency. With respect to the computational efficiency, both dual schemes provide very similar CPU times, and the cost is much more modest than for a pure MP2 optimization.

Once we demonstrated the ability of the MP2:AM1/MM dual scheme to improve the description of the transition state of the model Menshutkin reaction, we traced down steepest descent reaction paths up to the reactant and product valleys. These calculations were followed by complete optimization to obtain stationary reactant and product structures. Table 2

Table 2. Relative Potential Energies (in kcal/mol) for the Reactants, Products, and Transition Structures of the Model Menshutkin Reaction in Aqueous Solution Obtained at the AM1/MM and MP2:AM1/MM Levels

	AM1/MM	MP2:AM1/MM
reactants	0.0	0.0
transition structure	23.1	16.0
products	−18.9	−43.9

Scheme 2



compares the relative potential energies of these stationary structures obtained using MP2:AM1/MM and AM1/MM levels.

The experimental activation and reaction free energies are 23.5 and −34 kcal/mol.[45,46] The potential energy differences presented in Table 2 cannot be directly compared to these experimental free energies as far as entropic effects are expected to make an important contribution. Effectively, during the reaction a charge separation process is taking place, and then the solvent disorder is considerably diminished. However, when the AM1/MM potential energy surface is used to trace the potential of the mean force corresponding to this reaction, the activation free energies are overestimated and the reaction free energies severely underestimated (obtaining 29.3 and −10.4 kcal/mol, respectively).[47] Taking into account these results and from the comparison of the energetic values presented in Table 2, the improvement obtained using the MP2:AM1/MM potential energy surface, at a still moderate computational cost, is quite evident. Assuming that the contribution of the entropy to the activation and reaction free energies are the same at both computational levels we can estimate these quantities in the MP2:AM1/MM approach as

$$\Delta G(\text{MP2:AM1/MM}) = \Delta G(\text{AM1/MM}) + [\Delta E(\text{MP2:AM1/MM}) - \Delta E(\text{AM1/MM})] \quad (11)$$

The results for the activation and reaction free energies are 22.1 and −35.4 kcal/mol, respectively, in very good agreement with the experimental estimations.

Oxy-Cope Reaction in Aqueous Solution. The rearrangement of 1,5-hexadiene-3-ol to 2,4-hexadiene-1-ol (see Scheme 2) is an oxy-Cope reaction, a pericyclic type reaction that has recently received much attention because its implication in antibody catalyzed processes.[48−50]

Solvent molecules, acting as proton donors with the alcohol oxygen atom, increase the energy barrier, while proton acceptor solvent molecules decrease the energy barrier. Bulk electrostatic effects (as obtained from continuum models) slightly increase the energy barrier. These solvent effects can be rationalized in terms of electron withdrawing or electron donor substituent effects.[51]

We selected this reaction as a model of an intramolecular process to test our proposed computational schemes. Location and characterization of the transition structure in aqueous solution has been carried out at the AM1/MM, MPW1K:AM1/MM, MPW1K:CHARGE/MM, and MPW1K/MM levels. Density functional calculations employed the MPW1K hybrid functional,[52] that has recently shown to provide reasonable energy barriers,[53] with the 6-31G* basis set. The quantum system, composed by the substrate (see Scheme 2), was placed in the center of a water box of 31.4 Å of side containing a total of 1026 TIP3P water molecules. The Lennard-Jones parameters of the qm subsystem were taken from the OPLS-AA force field. In all cases we employed a switched cutoff of 14 Å for all interactions. Also in this example we used a control space that includes only the coordinates of the QM subsystem. As before, the convergence criteria was a gradient norm lower than 0.1 $kJ \cdot mol^{-1} \cdot Å^{-1}$. The main results are collected in Table 3.

Table 3. Transition Structures for the Oxy-Cope Reaction in Water Obtained at Different Computational Levels[a]

	AM1/MM	MPW1K:AM1/MM	MPW1K:CHARGE/MM	MPW1K/MM
d_{CC} (Å)	1.646	1.857	1.969	1.948
$d_{CC(OH)}$ (Å)	1.716	1.967	1.871	1.876
ν (cm^{-1})	533i	456i	501i	471i
n_{micro}	42	31	54	40
t_{rel}	1.0	17.9	16.3	610

[a] See text. The distances associated with the broken CC(OH) bond and the new CC bond are provided together with the frequency of the transition vector, the number of microiterations needed to reach convergence, and the relative CPU time per optimization step of the different computational strategies.

The AM1 Hamiltonian gives too short distances for the broken and formed bonds, a feature already observed in other pericyclic reactions, for example, in the Claisen rearrangement.[54,55] Both dual schemes, MPW1K:AM1/MM and MPW1K:CHARGE/MM, give results in good agreement with the pure DFT/MM calculation. The geometrical agreement is somewhat better when using fitted charges to estimate the interaction with the environment. In the case of MPW1K:AM1/MM the transition structure presents a more advanced character, being the distance of the new C−C bond shorter than the distance of the broken bond (the C−C(OH) one). This is a feature also observed in the AM1/MM structure. However, the curvature of the PES, as given by the imaginary frequency associated with the transition vector, is better reproduced by the MPW1K:AM1/MM dual scheme. With respect to the computational efficiency, both dual schemes employ a similar CPU time per optimization step. As in the previous example the cost is about 20 times larger than at the AM1/MM level but much more modest than at the ab initio/MM or DFT/MM levels. It is important to note that, as explained in the preceding section, the convergence of the dual levels based on fitted charges is poorer, because the potential energy surface is not unique and then there is a discontinuity between the micro- and macroiterations. This discontinuity is avoided defining a HL:LL/MM surface with a semiempirical Hamiltonian as low level, which allows for the polarization of the wave function during the macrosteps. This is clearly reflected in the lower number of microiterations needed to reach convergence.

Table 4. Potential Energy Barriers (in kcal/mol) for the Oxy-Cope Reaction in Gas Phase and in Aqueous Solution at Different Computational Levels

	AM1 (gas)	AM1/MM	MPW1K (gas)	MPW1K:AM1/MM	MPW1K:CHARGE/MM
$\Delta E^{\ddagger}$	35.2	40.1	24.1	30.6	29.8

From the transition structures located at the AM1/MM and dual levels we traced down a steepest descent reaction path to the reactant valleys. The energy barriers obtained as the difference between the transition structure and reactants energy are given in Table 4, together with the gas-phase values. The AM1 Hamiltonian overestimates the energy barrier both in gas phase and in solution, a feature also found in other pericyclic reactions.[55] The effect of the solvent observed on the PES exploration, excluding enthalpic and entropic temperature effects, is to increase the energy barrier significantly: between 4.9 and 6.5 kcal/mol. The differences in the estimation of the solvent effect on the energy barrier among the different approaches are not larger than 1.6 kcal/mol. Both dual schemes, using the AM1 Hamiltonian or fitted charges to obtain the interaction between the qm and mm subsystems, provide a very similar energy barrier, although, as said before, the convergence of the MPW1K:AM1/MM scheme is better than using the MPW1K:CHARGE/MM one.

Haloalkane Dehalogenase. Haloalkane dehalogenase from *Xanthobacter Autotrophicus* GJ10 (DhlA) catalyzes the conversion of dichloroethane (DCE) to 2-chloroethanol and chloride.[56] The complete reaction takes place in two steps.[57,58] In the first one, dichloroethane undergoes a S_N2 displacement of a chloride anion by means of the carboxylate group of Asp124, resulting in an ester covalently bound to the enzyme. In a second process, a crystal water molecule hydrolyzes the ester (see Scheme 3).

The first step of the reaction has been the subject of different theoretical studies, in which the activation free energy has been estimated from QM/MM calculations based on semiempirical or empirical valence bond Hamiltonians.[59−61] Semiempirical descriptions of this S_N2 process leads to an important overestimation of the activation free energy, which has been usually corrected with ab initio estimations of the error introduced in the potential energy barrier. We have now explored the PES for this process at the PM3/MM and B3LYP:PM3/MM levels, using in the later the 6-31+G* basis set. With this purpose the 1,2-dichloroethane and part of the Asp124 residue were chosen to be the qm subsystem, while the rest of the enzyme and the water molecules were in the mm subsystem (see ref 60 for details). A link atom was added to the qm subsystem in order to complete its valence. This link atom was placed between CA and CB atoms of Asp124. The mm subsystem was described using the OPLS-AA potential for the enzyme and the flexible TIP3P potential for the water molecules. The Lennard-Jones parameters of the qm subsystem were also taken from the OPLS potential, except for the chlorine atoms, for which we used those of ref 38. A switched cutoff radius of 12 Å was used for all kind of interactions. To explore the potential energy surface and locate the transition structure, the control space, for which the Hessian matrix is obtained, includes only the coordinates of the qm subsystem. As before, the convergence criteria was a gradient norm lower than 0.1 $kJ \cdot mol^{-1} \cdot Å^{-1}$. We first located the transition structure (see Figure 2) using the PM3/MM and B3LYP:PM3/MM descriptions, and afterward we traced the corresponding reaction paths up to the reactant valleys. The main results concerning both structures are collected in Table 5.

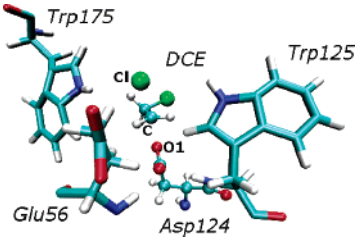


Figure 2. Transition structure for the nucleophilic reaction between DCE and Asp124 in the active site of DhlA. Only some relevant residues are shown for the sake of clarity.

In this case we provide not only intramolecular geometrical parameters but also some important hydrogen bond distances established between the qm and the mm subsystems. In particular the leaving chlorine anion can form hydrogen bonds with the H^{ϵ} of Trp125 and Trp175, while the nucleophilic oxygens of Asp124 (O1 is the nucleophilic oxygen and O2 is the other one) can establish H-bonds with H^{N} atoms of Glu56 and Trp125. The two computational levels give a very similar description of the reactant structure, both regarding intra- and intermolecular parameters. However, important differences appear in the transition structure. Using the dual level scheme, the transition structure presents significantly larger distances for the broken and formed bonds (the CCl and the CO1 distances). The difference is especially

Scheme 3

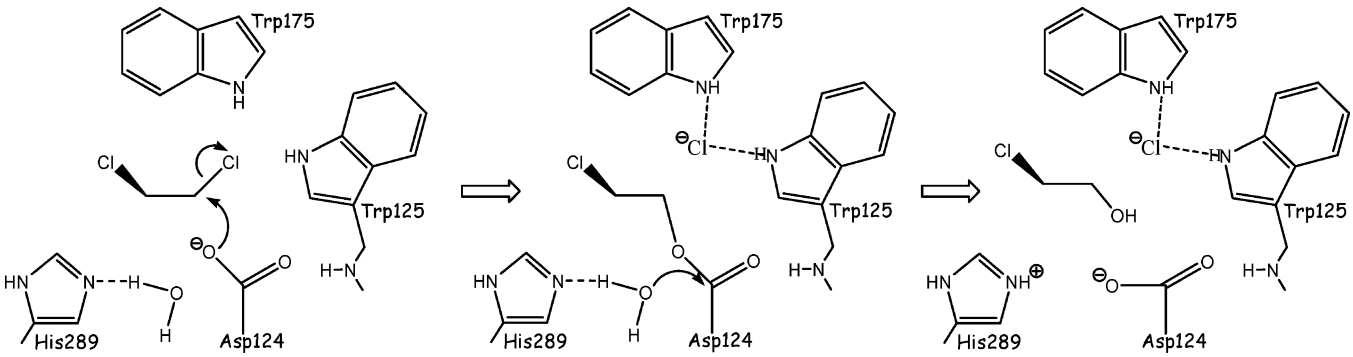


Table 5. Relevant Distances (in Å) and Relative Potential Energies (in kcal/mol^{-1}) for the Reactants and Transition Structure of the Nucleophilic Displacement of Chlorine by Aspartate in the Active Site of DhlA as Obtained from PM3/MM and B3LYP:PM3/MM Calculations

	PM3/MM		B3LYP:PM3/MM	
	R	TS	R	TS
d_{O1C}	2.883	1.905	2.905	2.017
d_{CCl}	1.784	2.256	1.835	2.491
$d_{Cl-Trp175}$	2.875	2.498	2.735	2.444
$d_{Cl-Trp125}$	3.701	2.623	3.777	2.569
$d_{O1-Trp125}$	1.810	2.861	1.888	2.866
$d_{O2-Glu56}$	1.842	1.853	1.829	1.853
ΔE	0	32.9	0	18.2

important in the CCl distance, which is about 0.23 Å larger in the B3LYP:PM3 transition structure than in the PM3 one. This intramolecular change has also consequences on the interaction pattern established with the residues of the active site. In particular, the H-bond distances established by the chlorine leaving group with Trp125 and Trp175 are shorter in the dual level description than in the PM3/MM exploration. This result can be important in the analysis of enzyme effects as it indicates that the use of the semiempirical description can lead to an underestimation of the enzymatic stabilization of the transition structure. We must keep in mind that much of the recent debate about the origin of an enzyme's ability to speed up chemical reactions is based on the comparison between the magnitude of transition state stabilization and reactant state destabilization.[9,62]

The energetic description of the reaction seems to be clearly improved with our dual treatment. Effectively, potential of mean force estimations of the activation free energy using PM3/MM energy functions lead to a severe overestimation. For example, in ref 60 we recently obtained a PM3/MM activation free energy of 28.4 kcal/mol to be compared to the experimentally derived value of 15.3 kcal/mol. If we look now at the potential energy barriers given in Table 5 we can see that the potential energy barrier provided by the B3LYP:PM3/MM method is about 14.7 kcal/mol lower than the PM3/MM value. Assuming that other contributions to the activation free energy remain essentially constant we could estimate (see eq 11) a dual level activation free energy of about 13.7 kcal/mol in much better agreement with the experimental value.

As a final test about the goodness of our dual level PES, heavy atom kinetic isotopic effects on this latter enzyme catalyzed reaction have been computed. We will compare our results with the available experimental data of intrinsic chlorine KIEs, as these effects provide information about the single chemical step of the full dehalogenation catalyzed process. The required structures of reactants (Michaelis complex) and TS were taken from the previously located and characterized by means of both methods: PM3/MM and B3LYP:AM1/MM. Once the stationary structures on the PES were refined, the rigid-rotor/harmonic-oscillator approximations were used with the CAMVIB/CAMISO programs[63,64] to calculate semiclassical KIEs without the scaling of vibrational frequencies, as explained and applied in previous papers.[23,65] A core of 15 atoms was equivalent to the QM region, which corresponds to the substrate. This subset of atoms was used to define the Hessian for these KIE calculations, consistent with the "cut-off rule" and the local nature of isotope effects.[66] The calculated values were 1.0047 and 1.0071 at the PM3/MM and the B3LYP:AM1/MM levels, respectively. Both methods predict small, normal heavy-atom effects for ^{37}Cl substitutions, close to the experimentally determined[67] k_{35}/k_{37} equal to 1.0066. Nevertheless, as we expected, the use of a high level theory to describe the QM region renders a value much closer to the experimental results, as was previously observed for the chorismate mutase KIE calculations.[66] A better agreement with experiments has been obtained not only because the structures obtained at a higher level are more reliable but also because of the good quality of vibrational frequencies. This limitation of the semiempirical methods to predict good chlorine kinetic isotope effects was already observed by Paneth and co-workers[68] who obtained the best estimation by means of the ONIOM QM/MM method on a reduced active site model. Our final results give much confidence in our developed dual level method.

4. Conclusions

We have here presented a new dual level strategy to be used in qm/mm methods combined with a micro/macroiteration algorithm for searching stationary structures. A new potential energy surface is defined in which the gas-phase and the interaction energies of the qm subsystem are obtained using different computational levels: a high level and a low level, respectively.

We have analyzed the performance of dual level methods in two chemical reactions in solution (mensutkin and oxy-Cope processes in aqueous solution) and one enzymatic reaction (a S_N2 reaction catalyzed by DhlA haloalkane dehalogenase). When compared to typical semiempirical/MM computations, the dual level approach (HL:LL/MM) allows for a considerable improvement of the description of bond breaking and bond forming processes. We have tested two different implementations, one in which the low level method is a semiempirical description and another one in which the low level method is a set of potential-derived charges. While the computational efficiency of both implementations is very similar, the use of a qm method as low level ensures the continuity of the potential energy surface between micro- and macroiterations. This characteristic leads to a reduction in the number of steps needed to reach full convergence, and then it seems to be the most promising strategy of the two proposed here. Obviously, the computational scheme is not restricted to a particular choice of the high level and low level methods, and then extension to other qm methods will be available in future applications.

Acknowledgment. We thank DGI (Spain) for project BQU2003-04168-C03, BANCAIXA for project P1A99-03 and Generalitat Valenciana for projects GRUPOS04/28, GV04B-21, and GV04B-131.

References

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(2) Gao, J. In *Methods and applications of combined quantum mechanical and molecular mechanical potentials*; Lipkowitz, Boyd, Eds.; VCH Inc.: New York, 1995.

(3) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

(4) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Chem. Eur. J.* **2003**, *9*, 984.

(5) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar D. G. *Science* **2004**, *303*, 186.

(6) Gertner, B. J.; Wilson, K. R.; Hynes, *J. Chem. Phys.* **1989**, *90*, 3537.

(7) Gertner, B. J.; Whitnell, R. M.; Wilson, K. R.; Hynes, J. T. *J. Am. Chem. Soc.* **1991**, *113*, 74.

(8) Kim, H. J.; Hynes, J. T. *J. Am. Chem. Soc.* **1992**, *114*, 10528.

(9) Roca, M.; Martí, S.; Andrés, J.; Moliner, V.; Tuñón, I.; Bertrán, J.; Williams, I. H. *J. Am. Chem. Soc.* **2003**, *125*, 7726.

(10) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. *J. Sci. Comput.* **1995**, *16*, 1190.

(11) Farkas, Ö.; Schlegel, H. B. *J. Chem. Phys.* **1999**, *111*, 10806.

(12) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514.

(13) Pu, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 54.

(14) Nudged Elastic Band (NEB) methods make use of a discrete representation of the reaction path from the reactant structure to the product one, along a string of replicas which are usually built by linear interpolation. Then optimization algorithms (mostly gradient based) are used to relax each of the structures of the chain towards the minimum energy path (MEP), providing this way an idea of the energy profile and of the transition structure. The key of the method relies on a proper definition of the constraints between consecutive points of the chain. For further information see: (a) Henkelman, G.; Jóhannesson, G.; Jónsson, H. *Methods for Finding Saddle Points and Minimum Energy Paths*; Schwartz, S. D., Ed.; Kluwer Academic Publishers: 2000. (b) Olsen, R. A.; Kroes, G. J.; Henkelman, G.; Arnaldsson, A.; Jónsson, H. *J. Chem. Phys.* **2004**, *121*, 9776.

(15) Galván, I. F.; Sánchez, M. L.; Martín, M. E.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Chem. Phys.* **2003**, *118*, 255.

(16) Galván, I. F.; Sánchez, M. L.; Martín, M. E.; Olivares del Valle, F. J.; Aguilar, M. A. *Comput. Phys. Commun.* **2003**, *115*, 244.

(17) Galván, I. F.; Martín, M. E.; Aguilar, M. A. *J. Comput. Chem.* **2004**, *25*, 1227.

(18) (a) Moliner, V.; Turner, A. J.; Williams, I. H. *Chem. Com.* **1997**, 1271. (b) Turner, A. J.; Moliner, V.; Williams, I. H. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323.

(19) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H.; Frisch, M. *J. Comput. Chem.* **2003**, *24*, 760.

(20) Prat-Resina, X.; González-Lafont, A.; Lluch, J. M. *J. Mol. Struct. Theochem* **2003**, *632*, 297.

(21) Monard, G.; Prat-Resina, X.; González-Lafont, A.; Lluch, J. M. *Int. J. Quantum. Chem.* **2003**, *93*, 229.

(22) Prat-Resina, X.; Bofill, J. M.; González-Lafont, A.; Lluch, J. M. *Int. J. Quantum. Chem.* **2004**, *98*, 367.

(23) Martí, S.; Moliner, V.; Tuñón, I.; Williams, I. H. *J. Phys. Chem. B* **2005**, *109*, 3707.

(24) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483.

(25) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. *J. Mol. Struct. Theochem* **1999**, *21*, 461.

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

(27) Field, M. J.; Albe, M.; Bret, C.; Proust-de Martin, F.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088.

(28) Field, M. J. *A practical introduction to the simulation of molecular systems*; Cambridge University Press: 1999.

(29) Gaussian 03, by 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.

(30) Singh U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(31) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(32) Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192.

(33) Baker, J. *J. Comput. Chem.* **1997**, *18*, 1079.

(34) Lia Addadi, E. K. J. *Biochemistry* **1983**, *22*, 4494.

(35) Gustin, D. J.; Mattei, P.; Kast, P.; Wiest, O.; Lee, L.; Cleland, W. W.; Hilvert, D. *J. Am. Chem. Soc.* **1999**, *121*, 1756.

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

(37) Solà, M.; Lledos, A.; Duran, M.; Bertran, J.; Abboud, J. M. *J. Am. Chem. Soc.* **1991**, *113*, 2873.

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

(39) Dillet, V.; Rinaldi, D.; Bertran, J.; Rivail, J. L. *J. Chem. Phys.* **1996**, *104*, 9437.

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

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

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

(43) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(44) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(45) Okamoto, K.; Fukui, S.; Shingu, H. *Bull. Chem. Soc. Jpn.* **1967**, *40*, 1920.

(46) Okamoto, K.; Fukui, S.; Nitta, I.; Shingu, H. *Bull. Chem. Soc. Jpn.* **1967**, *40*, 2354.

(47) Ruiz-Pernia, J. J.; Silla, E.; Tunon, I.; Marti, S.; Moliner, V. *J. Phys. Chem. B* **2004**, *108*, 8427.

(48) Braisted, A. C.; Schultz, P. G. *J. Am. Chem. Soc.* **1994**, *11*, 2211.

(49) Black, K. A.; Leach, A. G.; Kalani, M. Y. S.; Houk, K. N. *J. Am. Chem. Soc.* **2004**, *126*, 9695.

(50) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Angew. Chem., Int. Ed.* **2005**, *44*, 904.

(51) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. unpublished results.

(52) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715.

(53) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012.

(54) Ganem, B. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 936.

(55) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Theor. Chem. Acc.* **2001**, *105*, 207.

(56) Pries, F.; Kingma, J.; Krooshof, G. H.; Jeronimus-Stratingh, C. M.; Bruins, A. P.; Janssen, D. B. *J. Biol. Chem.* **1995**, *270*, 10405.

(57) Kennes, C.; Pries, F.; Krooshof, G. H.; Bokma, E.; Kingma, J.; Janssen, D. J. *Eur. J. Biochem.* **1995**, *228*, 403.

(58) Schanstra, J. P.; Kingma, J.; Janssen, D. B. *J. Biol. Chem.* **1996**, *271*, 14747.

(59) Nam, K.; Prat-Resina, X.; Garcia-Viloca, M.; Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2004**, *126*, 1369.

(60) Soriano, A.; Silla, E.; Tunon, I.; Ruiz-Lopez, M. F. *J. Am. Chem. Soc.* **2005**, *127*, 1946.

(61) Olsson, M. H. M.; Warshel, A. *J. Am. Chem. Soc.* **2004**, 126, 15167.

(62) Martí, S.; Roca, M.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Chem. Soc. Rev.* **2004**, *33*, 98.

(63) Williams, I. *Chem. Phys. Lett.* **1982**, *88*, 462.

(64) Williams, I. *J. Mol. Struct. Theochem* **1983**, *11*, 275.

(65) Martí, S.; Moliner, V.; Tuñón, I.; Williams, I. H. *Org. Biomol. Chem.* **2003**, *1*, 483.

(66) Ruggiero, G. D.; Guy, S. J.; Martí, S.; Moliner, V.; Williams, I. H. *J. Phys. Org. Chem.* **2004**, *17*, 592.

(67) Grimsrud, E. P.; Taylor, J. W. *J. Am. Chem. Soc.* **1970**, *92*, 739.

(68) Lewandowicz, A.; Rudzinski, J.; Tronstad, L.; Widersten, M.; Ryberg, P.; Mattson, O.; Paneth, P. *J. Am. Chem. Soc.* **2001**, *123*, 4550.

CT0501396

JCTC Journal of Chemical Theory and Computation

Evaluating the Accuracy of the Quasiharmonic Approximation

Chia-En Chang,[†] Wei Chen,[‡] and Michael K. Gilson*[,‡]

Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California, and Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received April 6, 2005

Abstract: The quasiharmonic approximation (QH) allows the configurational entropy of a molecule to be estimated on the basis of a molecular dynamics simulation, through construction of a Gaussian probability distribution of conformations with variances equal to those provided by the simulation. At its introduction in 1981, the QH method was successfully applied to simple molecular systems with only one highly occupied energy well, and fluctuations were analyzed in a system of internal bond-angle-torsion coordinates. However, more recent studies have applied the QH method to complex biomolecular systems and have relied upon Cartesian coordinates. The present study evaluates the accuracy of the QH method through comparisons with more detailed methods. The chief findings are that the QH method can markedly overestimate the configurational entropy for systems with multiple occupied energy wells and that such errors tend to be magnified by the use of Cartesian coordinates instead of bond-angle-torsion coordinates.

1. Introduction

Methods of calculating free energies are of central interest in molecular modeling and computational chemistry as a basis for interpretation, prediction, and design. Statistical thermodynamics dictates that free energy is essentially an integral over all configurations of the system that are accessible at ambient temperature. This can be a very difficult quantity to compute, especially for large, flexible molecules. One of the more rigorous approaches to this problem is to use free energy integration (see refs 1−11 and citations in ref 12), which uses molecular dynamics (MD) or Monte Carlo simulations to compute the reversible work, or free energy change, for a molecular process such as a change in conformation. The chief drawback of this approach is that it tends to be computationally demanding[13] and, hence, subject to convergence problems.[14,15] Less rigorous but more efficient approaches include linear interaction energy models[16−18] and the MMPB/SA method (e.g., refs 19−21).

Recently, we have developed a formally rigorous method of computing free energies called the second-generation mining minima (M2) algorithm. This comprises two major steps: identifying the low-energy conformations of the molecular system and calculating local configuration integrals—effectively free energies or standard chemical potentials—within each energy well. These local integrals employ an enhanced form of the harmonic approximation that accounts for anharmonicity via local scans of the energy surface. The contributions from the various energy wells are then combined to provide the full free energy of the system. Thus, a binding free energy, for example, is computed as the difference in the standard chemical potential of the product complex versus the free reactants. This approach requires the use of an implicit solvation model in order to keep the number of energy minima to a manageable number. The M2 method has proven to be both accurate and tractable for a range of host−guest systems, including synthetic receptors and cyclodextrins.[22,23] However, in its present form, at least, the M2 algorithm is not directly applicable to proteins

* Corresponding author phone: (240) 314-6217, fax: (240) 314-6255, e-mail: gilson@umbi.umd.edu.

[†] University of California—San Diego.

[‡] Center for Advanced Research in Biotechnology.

10.1021/ct0500904 CCC: $30.25
Published on Web 08/16/2005

because of the vast number of local energy minima. We are, therefore, exploring adaptations of the M2 method, as well as other approaches to computing molecular free energies.

One such approach is quasiharmonic (QH) analysis, which was introduced by Karplus and Kushick in 1981 as a means of capturing the anharmonicity of the energy surface when computing configurational entropies.[55] Briefly, the QH method involves carrying out MD simulations, computing the covariance matrix of atomic coordinates, interpreting the covariances as resulting from a harmonic energy surface rather than the actual anharmonic energy surface, and computing effective, or "quasiharmonic", force constants on this basis. These effective force constants can then be used to compute approximations to thermodynamic properties such as entropy, mean energy, and free energy (see, e.g., ref 24). A number of variations in technique can be accommodated within this general description; two key issues are the choice of coordinate system and the range of conformations sampled in the MD simulation, as now briefly reviewed.

In Karplus and Kushick's seminal paper, Cartesian coordinates from the MD trajectory were separated into 6 external (translational and rotational) coordinates and $3N - 6$ internal coordinates comprising bond lengths, bond angles, and bond torsions [i.e., bond-angle-torsion (BAT) coordinates], prior to calculation of the covariance matrix, where N is the number of atoms. (See also refs 25−28.) This separation of coordinates is well-suited to the calculation of thermodynamic properties via classical statistical mechanics because external rotations do not affect the molecular conformation. This is in contrast to the separation of coordinates via the Eckart−Sayvetz conditions,[29−31] which results in external rotations that distort the molecular conformation if they are more than infinitesimal.[27,32] Karplus and Kushick focused on changes in configurational entropy associated with the transition of a molecule from one well-defined energy well to another; that is, butane from trans to gauche and decaglycine from an extended to an α-helical conformation. A subsequent paper by Rojas et al,[33] which derived a cubic correction to the QH approximation, similarly focused on individual energy wells and mapped trajectory snapshots to BAT coordinates before computing the covariance matrix.

More recently, a number of groups have used other forms of the QH method to address important problems in molecular and biomolecular function. Typically, in these applications, Cartesian coordinates from the MD trajectory are not explicitly separated into external and internal coordinates before calculation of the covariance matrix. In addition, the MD simulation is not constrained to remain in a single energy well, although proteins at room temperature are believed to access many different energy wells.[34−37] At least two methods of handling the external coordinates have been described, see refs 38−40. In one study, Schafer et al. examine the entropy of a molten globule state of the enzyme α-lactalbumin. In another, Hsu et al. address changes in configurational entropy upon association of two proteins involved in human immunodeficiency virus infection,[41] also evaluating the convergence of the entropy as a function of simulation length. In a yet a third study, Jusuf et al. use the QH method to argue that changes in vibrational entropy are responsible for the cooperativity of ligand binding by glycopeptide antibiotics,[42] and Rinaldo and Field have used the QH method to examine entropy changes associated with a conformational change of the protein transferrin.[43] Several groups have also used a variant of the QH method in which some components of the covariance matrix are neglected. Thus, Luo and Sharp's application of the QH method to compute binding affinities neglects couplings (covariances) among the translational, rotational, and internal coordinates of each molecule, as well as couplings between the internal coordinates of the two molecules,[44] and thereby permits the equilibrium constant to be factorized into translational, rotational, and internal contributions. Swanson et al.'s examination of binding affinities also factorizes the equilibrium constant and applies the QH method only to the translational/rotational component of the covariance matrix.[45]

Thus, the QH method is finding application to molecular systems of increasing complexity. However, the literature to date provides remarkably little information on the numerical accuracy of the approach, that is, on how well it reproduces the entropy or free energy associated with the actual anharmonic energy surface. In particular, we are not aware of any paper comparing the QH approach against accurate reference results for a flexible molecule that accesses multiple energy minima during the MD simulation. Interestingly, as pointed out by Schlitter,[38] the Gaussian probability distribution assumed in the QH method yields the upper limit of the entropy associated with a given coordinate variance.[46] It is, thus, important to ask whether the upper limit is an accurate approximation for the actual entropy.[47] Interestingly, one can readily construct an energy surface for which it is not a good approximation. Consider, for example, a one-dimensional energy function with two equivalent, deep energy wells separated by a distance $2R$ that is much larger than the width of the wells. The variance of the solitary coordinate is clearly R^2, but one can readily show that the entropy of such a system is much less than that of a single harmonic energy corresponding to a variance of R^2. In fact, the entropy of the dual-well system is independent of R, in contrast to the harmonic well for which the variance cannot be increased without increasing the entropy. On the other hand, this dual-energy model is artificial and, therefore, does not bear directly on the real molecular systems to which the QH approximation is usually applied.

The present study addresses this issue by evaluating the accuracy and convergence of the QH method for a range of molecular systems, from a simplified alkane molecule to a host−guest system for which experimental binding data are available. Implementations of the QH method in both Cartesian and BAT coordinates are examined here, because the choice of coordinate system is known to influence accuracy and convergence in normal mode and related calculations.[28] In particular, Cartesian coordinates tend to provide lower accuracy because modes constructed from a Cartesian basis set necessarily consist of linear displacements of atoms, rather than more natural bond rotations. For the simplest system examined here, QH results are compared to analytic results, while for the more complex systems, reference results are obtained from the M2 algorithm (see

above), which has been numerically validated[22,23,28] and has provided good agreement with experimental binding affinities for a range of host−guest systems.

2. Theory

2.1. Chemical Potentials, Configuration Integrals, and Coordinate Systems. The binding constant K_b of a receptor R and ligand L, which form a noncovalent complex RL, is given by[48,49]

$$-RT \ln K_b = \Delta G^\circ \equiv \mu^\circ_{RL} - \mu^\circ_R - \mu^\circ_L \tag{1}$$

where RT is the gas constant multiplied by the absolute temperature and μ° is the standard chemical potential of the species indicated in the subscript. The standard chemical potential of a molecule in solution may be written as[48,50]

$$\mu^\circ = -RT \ln\left(\frac{1}{VC^\circ} Z'\right) \tag{2}$$

$$Z' = \int \mathrm{e}^{-\beta[U(\mathbf{r}')+W(\mathbf{r}')]}\, \mathrm{d}\mathbf{r}' \tag{3}$$

where $\beta = 1/RT$, C° is the standard concentration, V is the volume of the system, Z' is a configuration integral, and U and W are the potential and solvation energies of the molecule, respectively, as a function of the molecule's $3N$ Cartesian coordinates $\mathbf{r}'$. It is convenient to separate the coordinates into 6 external coordinates and $3N - 6$ internal coordinates $\mathbf{r}$ using the criterion that changing the external coordinates does not affect U or W. This allows the integrals over the external coordinates to be carried out trivially, yielding a factor of $8\pi^2 V$, so

$$\mu^\circ = -RT \ln\left(\frac{8\pi^2}{C^\circ} Z\right) \tag{4}$$

$$Z = \int J(\mathbf{r})\, \mathrm{e}^{-\beta[U(\mathbf{r})+W(\mathbf{r})]}\, \mathrm{d}\mathbf{r} \tag{5}$$

where Z is a configuration integral over only the internal coordinates $\mathbf{r}$ and $J(\mathbf{r})$ is the Jacobian determinant for the coordinate transformation.[27,51,52]

It is worth noting that the separation of coordinates afforded by the harmonic oscillator−rigid rotor approximation does not meet the criterion that U and W are independent of the external coordinates because it yields external rotations that correspond to linear displacements of atoms. These alter the conformation and, hence, the energy of the molecule.[27,53] This problem can be avoided by a different approach in which three atoms bonded in sequence are used to establish a molecular frame of reference. The three coordinates of atom 1 then define the translational coordinates of the molecule, the direction of the bond joining atoms 1 and 2 defines two rotational coordinates, and the angle of atom 3 defines the third (body) rotation of the molecule.[27] Thus, atom 1 is the origin of coordinates; the bond joining atoms 1 and 2 represents the z axis; the plane defined by atoms 1, 2, and 3 is the $x-z$ plane; and the y axis is normal to this plane. All of the $3N - 6$ Cartesian coordinates of the molecule can then be referenced to this coordinate system to generate "anchored Cartesian" coordinates.[27] The Jacobian determinant for this coordinate transformation is $b_2^2 b_3 \sin\theta_3$, where b_2 is the bond length of atoms 1 and 2, b_3 is the bond length of atoms 2 and 3, and θ_3 is the bond angle between b_2 and b_3.[26,27] The $3N - 6$ coordinates can be, furthermore, converted from Cartesian to BAT coordinates, in which the position of each atom is defined by one bond length, one bond angle, and one torsion angle, defined relative to three other atoms.[28,51,54] The Jacobian determinant when this transformation is included becomes $b_2^2 \Pi_{i=3}^N b_i \sin\theta_i$, where b_i and θ_i are the bond lengths and bond angles defined by the BAT coordinates.

2.2. Harmonic Approximation. The harmonic approximation is introduced here as a basis for subsequent discussion of the quasiharmonic and M2 methods, below. It is most applicable when a single deep energy well i at location $\mathbf{r}_i$ dominates the energy landscape. The harmonic approximation is established by expressing the energy $E = U + W$ as a Taylor series expansion around the energy minimum at $\mathbf{r}_i$ and retaining terms only up to the second order:

$$E(\mathbf{r}) \approx E(\mathbf{r}_i) + \frac{1}{2}(\Delta\mathbf{r})^T \mathbf{H}(\mathbf{r}_i)\, \Delta\mathbf{r} \tag{6}$$

$$\Delta\mathbf{r} \equiv \mathbf{r} - \mathbf{r}_i \tag{7}$$

where $\mathbf{H}(\mathbf{r}_i)$ is the second derivative, or Hessian, matrix of E at $\mathbf{r}_i$ and the first derivative terms are absent because all first derivatives equal zero at a local minimum. With this approximation, the chemical potential can be written as

$$\mu^\circ_{harm} = -RT \ln\left(\frac{8\pi^2}{C^\circ} J_i \int \exp\{-\beta[E_i + {}^1/_2(\Delta\mathbf{r})^T \mathbf{H}(\mathbf{r}_i)\, \Delta\mathbf{r}]\}\, \mathrm{d}(\Delta\mathbf{r})\right) \tag{8}$$

$$= E_i - RT \ln\left(\frac{8\pi^2}{C^\circ} J_i\right) - \frac{RT}{2} \ln\left\{\left(\frac{2\pi}{\beta}\right)^{3N-6} [\det \mathbf{H}(\mathbf{r}_i)]^{-1}\right\} \tag{9}$$

where $E_i \equiv E(\mathbf{r}_i)$ and $J_i \equiv J(\mathbf{r}_i)$ and we have made the approximation that the Jacobian determinant does not vary significantly within the low-energy part of the energy well. This is reasonable because the Jacobian does not depend on torsion angles, and bond lengths and bond angles are relatively rigid.

2.3. Quasiharmonic Approximation. The harmonic energy expression in eq 6 implies a Gaussian probability distribution of conformations around the energy minimum at $\mathbf{r}_i$, with the covariance matrix $\mathbf{C}$:

$$\mathbf{C} = (\beta\mathbf{H})^{-1} \tag{10}$$

In the QH method, the covariance matrix $\mathbf{C}$ is computed from snapshots of a MD simulation, and eq 10 is inverted to provide an estimate of $\mathbf{H}$,[33,44,45,55−57] which is then substituted into eq 9 and related equations for other thermodynamic quantities, such as the entropy. When the molecular dynamics run used to compute $\mathbf{C}$ is restricted to a single energy well (e.g., refs 55−57), the value of E_i in eq 9 can be obtained by energy minimization within the energy well. When the MD simulation samples multiple energy wells, it is not clear what value to use for E_i. However, the MD simulation provides the Boltzmann average of the energy, $\langle E \rangle$, and this

may be set equal to $E_i + (3N - 6)/2\beta$ by applying the equipartition theorem to the positional degrees of freedom, as is appropriate, given the assumption of harmonicity. From eqs 9 and 10, then

$$\mu^\circ_{\mathrm{QH}} = \langle E \rangle - \frac{3N-6}{2\beta} - RT \ln\left(\frac{8\pi^2}{C^\circ} J_i\right) - \frac{RT}{2} \ln[(2\pi)^{3N-6} \det \mathbf{C}] \quad (11)$$

The sum of the second, third, and fourth terms on the right-hand side is identifiable as $-TS^\circ$, while the sum of the first and second terms can be viewed as an effective energy minimum associated with the QH calculation, although it is not expected to correspond to the actual energy of any particular conformation.

2.4. Predominant States Approximation: Summation over Energy Wells. For a molecular system with multiple energy wells that are significantly occupied at ambient temperature, the chemical potential may be estimated via a sum of contributions z_i from the M lowest energy wells:

$$\mu^\circ \approx RT \ln\left(\frac{8\pi^2}{C^\circ} \sum_{i=1}^{M} z_i\right) \quad (12)$$

$$z_i \equiv J_i \int_i \mathrm{e}^{-\beta E(\mathbf{r})}\, \mathrm{d}\mathbf{r} \quad (13)$$

where the subscript on the integral symbol implies an integration domain limited to energy well i. The M2 algorithm (Section 3.1) uses this approach, computing z_i with an extension of the harmonic approximation that accounts for anharmonicity in the shapes of the energy wells.

3. Methods

3.1. Second-Generation Mining Minima Method. The M2 method is based on the predominant states approximation and has two main steps: finding the low energy conformations $i = 1, ..., M$ and evaluating the configuration integral z_i within each energy well i. The first step is carried out with the Tork algorithm.[58] In Tork, the molecule is energy-minimized and the Hessian matrix of the energy with respect to BAT coordinates is computed and diagonalized. The molecule is iteratively distorted along the eigenvectors with lower force constants (eigenvalues) and energy-minimized again to arrive at new energy minima. The method is similar to low mode search;[59−61] perhaps the chief difference is the use of BAT coordinates, which lead to more physically reasonable molecular distortions. Repeat conformations are removed to prevent double-counting; as part of this process, a symmetry-detection algorithm[62] is used to detect topological symmetries[62,63] of the molecule and eliminate conformations that are the same after rotational interchange of chemically equivalent atoms. For example, two conformations that differ by a 120° rotation of a methyl group are considered identical.

The second step, evaluation of the configuration integral z_i in each energy well i, is carried out with the harmonic approximation/mode scanning (HA/MS) method, which uses the harmonic approximation along with mode scanning, a fast correction for local anharmonicity. The Hessian matrix of the energy with respect to BAT coordinates is computed at the energy minimum, and the Hessian is diagonalized. The harmonic approximation of the configuration integral is then computed along each eigenvector with an integration range of ±3 standard deviations of the corresponding Gaussian. For each eigenvector with a force constant less than 10 kcal/mol-A^2, the integral is also evaluated by mode scanning,[28] in which the molecule is distorted stepwise along the eigenvector, the energy is computed at each step, and the integral of the Boltzmann factor is computed numerically. The numerical result for a given eigenvector is substituted for the harmonic result when the numerical integral deviates from the harmonic by >1 kcal/mol, indicating substantial local anharmonicity along the mode. The configuration integral for energy well i becomes

$$z_i \approx J_i \prod_j^{N_{\mathrm{scan}}} S_j \prod_k^{N_{\mathrm{harm}}} \left[\sqrt{\frac{2\pi RT}{K_k}}\, \mathrm{ERF}\left(\frac{w_k}{\sqrt{\frac{2RT}{K_k}}}\right)\right] \quad (14)$$

where j ranges over the N_{scan} numerically integrated modes, S_j is the numerical integral for mode j, k ranges over the N_{harm} modes treated as harmonic, K_k is the eigenvalue of mode k, w_k is the integration range of mode k, and ERF is the error function. The chemical potential is then calculated via a sum of M local integrals z_i, according to eq 12. The mean energy $\langle E \rangle$ can also be computed with the M2 method as

$$\langle E \rangle = \frac{\sum_i^M J_i \int_i E(\mathbf{r})\, \mathrm{e}^{-E(\mathbf{r})/RT}\, \mathrm{d}\mathbf{r}}{\sum_i^M z_i} \quad (15)$$

and the entropy can be computed from the difference between μ° and $\langle E \rangle$.

3.2. Molecular Systems and Computational Methods. This study uses test systems of increasing complexity to evaluate and characterize the QH approximation. Section 4.1.1 considers a united-atom representation of butane in which hydrogens are merged with the carbons to which they are bonded and all nonbonded interactions are turned off, leaving only terms related to bond lengths, bond angles, and the solitary dihedral angle. This model is useful because the configurational integral can be computed to high accuracy by analytic or semianalytic methods. Section 4.1.2 considers linear alkanes in the all-atom representation with the full CHARMM22[64] vacuum potential from the program QUANTA,[65] including nonbonded interactions, but with no solvation model. Finally, Section 4.2 considers the binding affinity of a cyclic urea with a synthetic receptor in chloroform, a system for which experimental binding data are available. The all-atom CHARMM22 representation is used for this system, along with an implicit solvation model, which is implemented as follows. Initial coordinates for all molecules were generated with the program QUANTA[65] and then energy-minimized by the conjugate gradient method and, then, the Newton−Raphson method until the energy gradient was $<10^{-3}$ kcal/mol/Å.

As previously discussed,[22,23] the solvation energy is estimated with a generalized Born (GB) model[66] during the Tork, HA/MS, and MD calculations. In addition, the M2 calculations adjust the free energy of each energy well i toward the Poisson−Boltzmann/surface area (PB/SA) model by using the program UHBD[67] to solve the linearized PB equation and compute the surface area for the energy minimum conformation and replacing the GB contribution with these more detailed solvation terms. Thus, if the initial value of the chemical potential for energy well i is $\mu^{\circ}_{i,\mathrm{init}}$, then the corrected value is

$$\mu^{\circ}_{i,\mathrm{corr}} = \mu^{\circ}_{i,\mathrm{init}} - W_{i,\mathrm{GB}} + W_{i,\mathrm{PB}} + W_{i,\mathrm{SA}} \qquad (16)$$

where the W terms are solvation contributions computed at $\boldsymbol{r}_i$. The PB calculations use a chloroform dielectric constant of 4.8 and a chloroform radius of 2.4 Å.[68] The surface area term, $W_{i,\mathrm{SA}}$, is the work of forming a nonpolar cavity in the solvent with the shape of the solute in conformation i and also includes the van der Waals interactions between the solvent and solute. It is approximated by a linear function of the molecular surface area A_i: $W_{i,\mathrm{SA}} = aA_i + b$. For the chloroform solvent considered here, a and b are set to −0.04 kcal/(mol Å^2) and 1.41 kcal/mol, respectively.[69,70] As noted in Section 2.3, the MD calculation used in implementing the QH method often is not associated with a single, well-defined energy minimum. Therefore, to enable comparison with the M2 calculations, the solvation correction just described is applied on the basis of the energy-minimized conformation used to initiate the MD run.

The QH method is implemented here with the stochastic MD method at 300 K,[71] as implemented in the program UHBD.[67] This approach ensures that the systems are fully thermalized despite their small size. All simulations used a 1 fs time step, and the duration of each simulation is stated in the Results section. In every case, 2000 evenly spaced snapshots were saved and used for the QH analysis, except that 4000 snapshots were used in the case of the 6 ns convergence tests for the urea receptor. The conformations in the snapshots were aligned on the basis of the three arbitrarily selected atoms that define the molecular coordinate system. Both anchored Cartesian coordinates and BAT coordinates were computed, along with the corresponding Jacobian determinants, so that QH results in both coordinate systems could be computed and compared. The resulting coordinates were then used to obtain the covariance matrices for use in eq 11. Because methyl rotations were not considered to generate new conformations when counting energy minima, a ±60° constraint was applied to all methyl groups in the alkanes and the urea receptor during the MD simulations.

For some QH simulations, we wished to restrict the molecular dynamics to conformations in the neighborhood of a single energy well. This was accomplished by restraining all rotatable bonds with added energy terms of the form $E_{\mathrm{rest}} = 0.005[(\phi - \phi_0)/\Delta\phi]$,[14] where ϕ is the rotation to be restrained, ϕ_0 is the center of the restraint range, and $\Delta\phi$, the range of the restraint, was set to 60° for all rotatable bonds, except that a range of 90° was used for rotatable bonds attached to phenyl groups.

All calculations were carried out on Linux computers with Athlon 1.67 GHz processors. The parameters for Tork conformational search are as described previously.[58]

4. Results

This section compares various methods of computing chemical potentials, including the QH approximation in both Cartesian and BAT coordinates and the M2 method. All results are based upon integrals over $3N - 6$ internal degrees of freedom, defined with respect to the molecular frame of reference, and include the appropriate Jacobian determinants and the factor of $8\pi^2/C^{\circ}$, which results from integrating over external coordinates. The values of $\langle E \rangle$ reported for the QH calculations are mean potential energies obtained from the MD calculations.

4.1. Linear Alkanes. This section compares QH results for a simplified representation of butane for which analytical results are available and then compares QH results for more fully represented linear alkanes with the results of the M2 method, whose numerical accuracy has previously been validated.[28] We begin by establishing that the QH calculations have converged adequately for this series of systems. This is done by examining convergence for the most complicated one, heptane in an all-atom representation with no conformational restraints. Figure 1 graphs the entropic contribution to the chemical potential computed with the QH method as a function of the number of sequential snapshots included from simulations of 2 ns total duration, to a maximum of 2000 snapshots. Four of these convergence graphs are shown. Two use a BAT basis set and are based upon simulations starting from two different conformations of heptane, while the other two use an anchored Cartesian basis set and are based upon the same simulations used for the BAT results. The two BAT graphs plateau at values agreeing to within a few tenths of kilocalories per mole about $^3/_4$ of the way through the 2 ns simulation, while the anchored Cartesian result plateaus about halfway through the simulation. These results document adequate convergence for the most complex alkane studied here using 2000 snapshots from a 2 ns MD trajectory.

4.1.1. Simplified Butane. The simplest system considered here is united-atom butane with all nonbonded interactions artificially turned off. This system possesses only four particles and six energy terms (three bond stretches, two angle bends, and one torsion), and its configuration integral can be factorized and computed semianalytically to high precision. It possesses three equivalent energy wells because of the 3-fold periodicity of the dihedral energy component, and Table 1 presents results for a single energy well and for all three energy wells taken together. The QH results for a single well were obtained by adding a restraining function during the MD calculation, as described in the Methods section.

For a single energy well, chemical potentials computed by QH in BAT coordinates and by the M2 method agree with the semianalytic reference results to within 0.05 kcal/mol. However, QH in Cartesian coordinates yields a chemical potential that is about 0.5 kcal/mol lower than the reference result. For integration over all three energy wells, the M2 method still agrees with the reference results to within 0.03

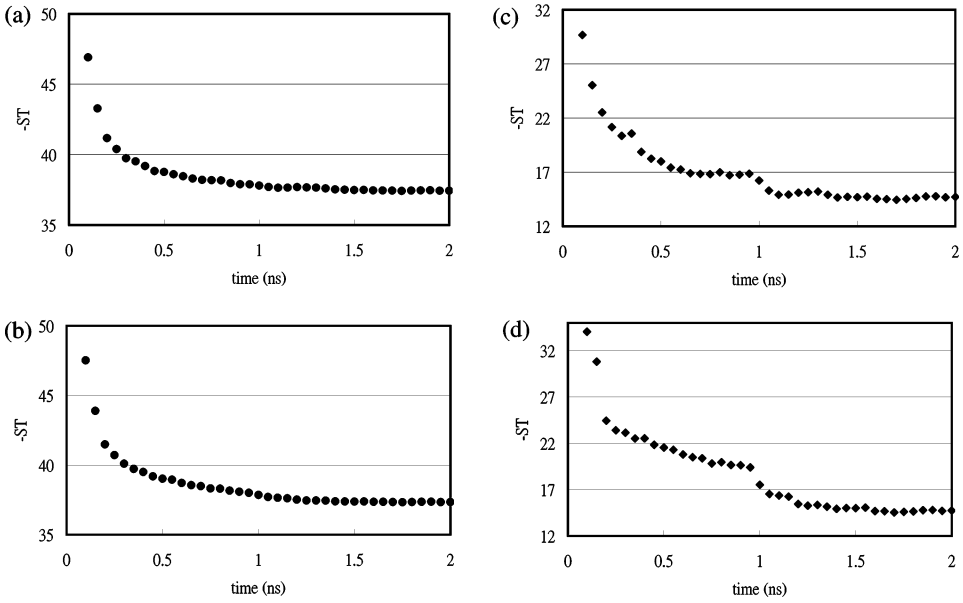


Figure 1. Convergence analysis of QH calculations, showing the entropy contribution to the chemical potential (kcal/mol) as a function of the duration of MD simulations of heptane, which sample all energy wells. Top (a, c) and bottom (b, d) graphs show results for simulations starting from the QUANTA-generated conformation and the least-stable local energy minimum found by a Tork conformational search. Left graphs (a, b): BAT coordinates. Right graphs (c, d): Cartesian coordinates.

Table 1. Computed Chemical Potential μ° and Average Potential Energy $\langle E \rangle$ (kcal/mol) of Simplified Butane Obtained Analytically, from the M2 Method, and from QH Calculations Based upon Simulations of 100 ps and 2 ns for One and Three Energy Wells, Respectively[a]

	analytic	M2		QH		
	μ°	$\langle E \rangle$	μ°	$\langle E \rangle$	μ°_{BAT}	μ°_{Cart}
single energy well	−1.71	1.74	−1.69	1.84	−1.76	−2.19
three energy wells	−2.37	1.74	−2.34	1.85	−2.94	−4.53

[a] μ°_{BAT} and μ°_{Cart}: QH results based upon BAT and Cartesian coordinates.

kcal/mol, but the QH calculations yield significantly larger errors, −0.57 kcal/mol and −2.16 kcal/mol, respectively, for BAT and Cartesian coordinates. Another way to look at these data is to recognize that the chemical potential associated with three equivalent energy wells must be $-RT \ln 3 = -0.66$ kcal/mol lower than that associated with only one of the wells. The semianalytic and M2 results yield this difference, but the QH calculations with BAT and Cartesian coordinates yield differences of −1.18 and −2.34 kcal/mol, respectively.

The cause of the error can be elucidated by additional calculations focusing exclusively on the variation of dihedral energy with butane's solitary dihedral degree of freedom. First, Figure 2 shows the exact dihedral potential for one of the three dihedral-energy energy wells (dots), along with the harmonic (bold solid) and the QH approximations (light solid) to the shape of this energy well. The approximations match the exact curve well, and the chemical potentials associated with the harmonic and QH approximations, 0.40 and 0.37 kcal/mol, respectively, agree well with the analytic result, 0.36 kcal/mol. Thus, the QH treatment accounts well for the small amount of anharmonicity in the single energy well. For comparison, Figure 3 shows the exact potential

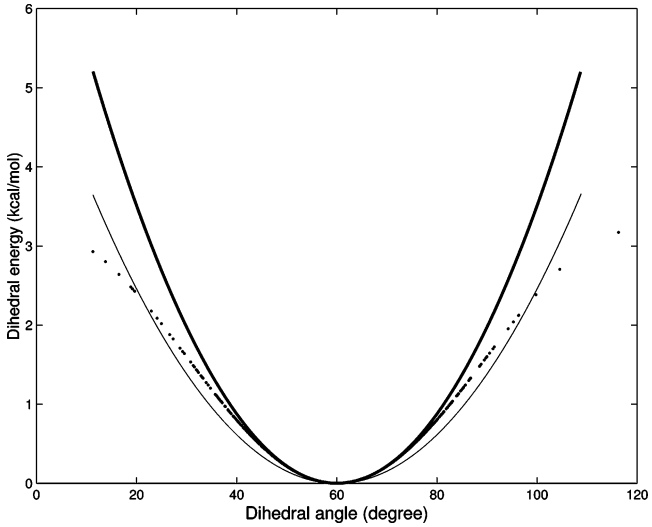


Figure 2. Dihedral energy of simplified butane (see Methods) as a function of the central dihedral angle, in a single energy well. Dots are results from MD trajectory snapshots during a 2 ns MD simulation. Bold line: harmonic approximation. Thin line: QH approximation based on the MD results (dots).

for all three dihedral energy wells (dots), along with the QH approximation (solid). It is evident that the QH approximation effectively merges the three relatively narrow wells into a single broad well. Not surprisingly, then, it yields a significantly lower chemical potential, −0.8 kcal/mol, than the analytic result, −0.29 kcal/mol. On the other hand, using the harmonic approximation to approximate the integral in each well and summing the three resulting integrals, as in the M2 method, yields an accurate chemical potential of −0.26 kcal/mol. Thus, the QH method generates significant errors because it merges multiple narrow energy wells into one broad energy well.

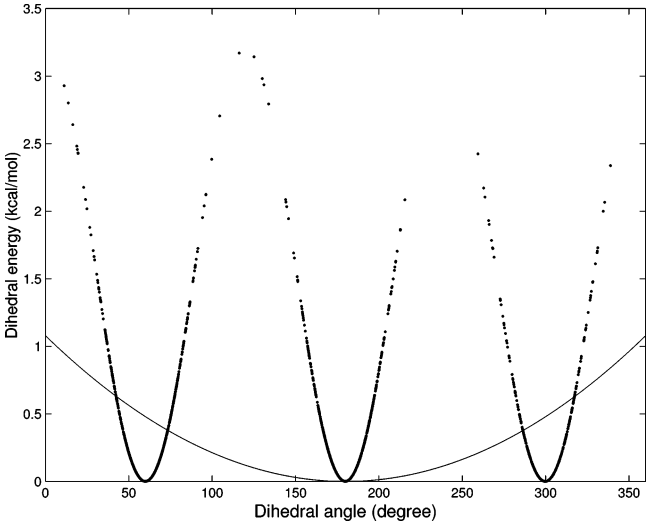


Figure 3. Dihedral energy of simplified butane across all three energy wells. Dots and thin line have the same meaning as in Figure *2*.

It is also of interest that the QH results are particularly inaccurate when they are based upon Cartesian coordinates. We conjecture that this is because the Cartesian basis set effectively magnifies the fluctuations associated with bond rotations. Rotating one bond causes an atom to move along a curved path, producing changes in at least two Cartesian degrees of freedom and, thus, generating excessive contributions to the covariance matrix and, hence, to the entropy obtained by the QH approximation. This error is especially large when large bond rotations occur; thus, going from BAT to Cartesian coordinates produces a much larger error when all three energy wells of simplified butane are considered than when only one is considered, as shown in Table 1.

4.1.2. All-Atom Alkanes. The results for simplified butane suggest that the QH approximation is particularly inaccurate for a system with multiple distinct energy wells. If so, then its accuracy should decrease for linear alkanes of increasing length, because they possess more energy wells. This expectation is evaluated here by applying the QH approximation to linear alkanes from butane through heptane. All hydrogens are treated explicitly and nonbonded interactions are included, to provide a more physically realistic test than simplified butane. Analytic reference results are not available for these cases, so reference results are obtained by the M2 method, on the basis of validation of its accuracy for simplified butane (above) and other systems.[28] Data are presented for unconstrained sampling over all energy wells and also for sampling only in the all-trans global energy minimum of each alkane.

The results for sampling over all energy wells, in Table 2a, confirm that the QH results become more inaccurate with increasing chain length. The errors result primarily from an excessively large entropy, consistent with overestimation of motional freedom by the QH method. This is also consistent with the fact that the Gaussian distribution yields the maximum entropy for a given variance.[46] As for simplified butane, the errors are most severe when Cartesian coordinates are used as the basis set. Thus, for heptane, the QH chemical potentials with BAT and Cartesian basis sets differ from the reference M2 results by −1.9 and −24.7 kcal/mol, respectively.

The QH method yields considerably more accurate results when it is used to sample over only the single all-trans energy well of each alkane, as shown in Table 2b. QH in BAT coordinates now agrees closely with the reference M2 calculations: chemical potentials computed with BAT and Cartesian coordinates are accurate to within 1.15 and 4.85 kcal/mol. It is worth remarking on the fact that going from multiple energy wells (Table 2a) to a single energy well (Table 2b) does not raise the entropy by $-RT \ln N$, where N is the number of energy wells. This is because the all-trans conformation studied here is the global energy minimum, so the other energy wells are not as highly populated and,

Table 2. Standard Chemical Potential μ°, Average Potential Energy $\langle E \rangle$, and Configurational Entropy $-TS^\circ_{\text{config}}$ (kcal/mol, 1 mol/L Standard Concentration) of All-Atom Linear Alkanes, Computed with the M2 Method and the QH Method Using 2 ns MD Simulations[a]

(a) All Energy Wells

		M2			QH				
						bond-angle-torsion		Cartesian	
	number of wells	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	$-TS^\circ_{\text{config}}$	μ°
butane	3	10.95	20.81	31.55	11.33	19.86	31.19	15.13	26.46
pentane	6	13.42	27.48	40.90	14.09	25.82	39.82	14.38	27.47
hexane	15	15.92	34.04	49.96	16.90	31.53	48.43	14.51	31.41
heptane	52	18.53	40.72	58.95	19.58	37.43	57.01	14.71	34.27

(b) Single Energy Well

						bond-angle-torsion		Cartesian	
	number of wells	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	$-TS^\circ_{\text{config}}$	μ°
butane	1	10.71	21.06	31.77	10.40	21.26	31.64	19.50	29.88
pentane	1	13.18	27.86	41.04	13.59	27.99	41.58	23.93	37.52
hexane	1	15.54	34.65	50.19	16.07	34.40	50.47	29.45	45.72
heptane	1	17.87	41.48	59.35	18.59	41.90	60.50	35.95	54.54

[a] (a) Results computed for all energy wells. (b) Results for the all-trans energy well only.

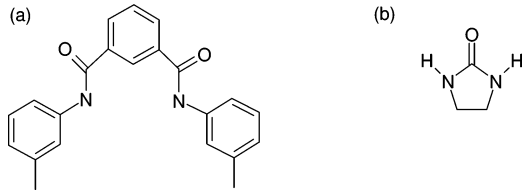


Figure 4. Host−guest system. (a) Receptor and (b) ethylenurea ligand.

Table 3. Standard Chemical Potential, Average Potential Energy, and Configurational Entropy for Binding of the Urea Receptor with Ethylenurea from the M2 Method[a]

	initial energy	number of wells	$\langle E \rangle$	$-TS^{\circ}_{\text{config}}$	μ°	$\mu^{\circ}_{\text{corr}}$
ethylenurea	−12.1	1	−12.4	19.3	6.9	4.1
receptor	−38.5	15	−2.9	104.2	101.3	83.0
complex	−74.6	87	−28.0	134.3	106.3	83.7
change on binding			**−12.7**	**10.8**	**−1.9**	**−3.4**

[a] Last row shows changes in these quantities upon binding. Results are based upon the GB solvation model except that $\mu^{\circ}_{\text{corr}}$ has been corrected toward the PB/SA solvent model, as detailed in the Methods section. See Table 1 for other symbols. Experimental binding free energy is −4.1 kcal/mol.

therefore, make smaller contributions to the configuration integral.

It is also of interest that the values of average energy $\langle E \rangle$ obtained by the M2 method agree with the QH values to within 1 kcal/mol. This agreement is striking because the methods are so different. The M2 method computes $\langle E \rangle$ with the harmonic approximation in multiple energy wells, correcting anharmonicity via mode scanning if needed, while the QH values are obtained directly from MD simulations. The agreement found here supports the validity of viewing the molecular configuration integral as a sum of contributions across distinct energy wells.

Note that the BAT results are insensitive to the choice of the three root atoms that define the molecular frame of reference and to the four-atom sets that define torsion angles in the BAT coordinate system. For example, changing to a different set of root atoms in heptane, and consequently also changing the torsion definitions, yields the same QH and M2 entropies to within 0.2 and 0.01 kcal/mol, respectively.

4.2. Urea Receptor. Because the QH method can provide standard chemical potentials, it can be used to compute the binding free energy of two molecules via the expression $\Delta G_{\text{bind}} = \mu^{\circ}_{\text{RL}} - \mu^{\circ}_{\text{R}} - \mu^{\circ}_{\text{L}}$, where R and L indicate a receptor and a its ligand, respectively. This section tests such an application for a simple host−guest system, a synthetic receptor and its urea guest (Figure 4) that interact primarily via hydrogen bonding and bind in chloroform with a standard free energy change of −4.1 kcal/mol.[72] As shown in Table 3, the M2 method yields a binding free energy $\Delta G^{\circ}_{\text{corr}}$ of −3.4 kcal/mol, in good agreement with the experimental result, the error being less than 1 kcal/mol. However, the QH method yields binding free energies $\Delta G^{\circ}_{\text{corr}}$ that are about 8 kcal/mol too strong, with either BAT or Cartesian coordinates, as shown in Table 4. For QH in Cartesian coordinates, this error would be even greater if not for partial cancellation of errors, since the chemical potentials from the QH calculations in Cartesian coordinates deviate from the M2 results by on the order of 30 kcal/mol. QH in BAT coordinates yields more accurate chemical potentials, with maximal errors of only about 9 kcal/mol

It is worth asking whether the inaccuracy of the QH method results from the somewhat arbitrary method by which the solvation correction is applied. As noted in the Methods section, there is no definite prescription for selecting the conformation(s) for which it should be computed, unlike for the M2 method; here, the QH correction is obtained for the first conformation of the MD calculation. It is, thus, of interest to consider the accuracy of the M2 and QH methods when the solvation correction is neglected; the only solvation model then is the GB model (see Methods). As shown in Tables 3 and 4, the uncorrected M2 results are still much closer to those of the experiment than the uncorrected QH results: ΔG° is −1.9 kcal/mol from M2 and −9 kcal/mol from QH with both BAT and Cartesian coordinates. These results suggest that the errors in the QH calculations result from inaccuracy in the configuration integral rather than the solvation model.

It is also important to assess the convergence of the QH results, which are based upon 2000 snapshots from a 2 ns MD simulation. Convergence was assessed by recomputing the QH results with a new MD calculation starting from the global energy minima of the receptor and the complex obtained from the M2 calculations. (The free ligand is rigid and, therefore, has only one highly occupied conformation.) The new QH calculation also overestimates the binding affinity, as shown in Table 5, but by a different amount. This difference indicates that the MD calculations are not adequately converged, even though they are rather lengthy and the molecular systems are simple. Examination of the two MD simulations indicates that both sample the same energy minima but spend different amounts of time in them.

Table 4. QH Calculations of Standard Chemical Potential, Average Potential Energy, and Configurational Entropy for Binding of the Urea Receptor with Ethylenurea, Based upon 2 ns MD Simulations Started from QUANTA-Generated Conformations[a]

	initial energy	$\langle E \rangle$	bond-angle-torsion			Cartesian		
			$-TS^{\circ}_{\text{config}}$	μ°	$\mu^{\circ}_{\text{corr}}$	$-TS^{\circ}_{\text{config}}$	μ°	$\mu^{\circ}_{\text{corr}}$
ethylenurea	−21.1	−12.1	19.4	7.3	4.3	18.4	6.3	3.5
receptor	−38.5	−2.1	101.1	99.0	81.1	78.4	76.4	58.4
complex	−74.6	−26.5	123.9	97.4	74.3	99.9	73.4	50.3
change on binding		**−12.3**	**3.4**	**−8.9**	**−11.1**	**3.1**	**−9.3**	**−11.6**

[a] See Table 1 for symbols.

Table 5. QH Calculations of Standard Chemical Potential, Average Potential Energy, and Configurational Entropy for Binding of the Urea Receptor with Ethylenurea Based upon 2 ns Simulations Started from Conformations of Lowest Potential Energy as Found by a Tork Conformational Search

			bond-angle-torsion			Cartesian		
	initial energy	$\langle E \rangle$	$-TS^\circ_{\text{config}}$	μ°	μ°_{corr}	$-TS^\circ_{\text{config}}$	μ°	μ°_{corr}
ethylenurea	−21.1	−12.1	19.4	7.3	4.3	18.4	6.3	3.5
receptor	−39.4	−2.2	103.0	100.8	82.2	82.3	80.1	61.5
complex	−75.6	−26.5	122.5	96.0	73.5	97.1	70.7	48.1
change on binding		**−12.2**	**0.1**	**−12.1**	**−13.0**	**−3.6**	**−15.7**	**−16.9**

Table 6. QH Calculations for the Free Urea Receptor and the Bound Complex Using BAT Coordinates and Based upon 6 ns MD Simulations Started from QUANTA-Generated Conformations (Simulation 1) and from the Global Energy Minima (Simulation 2)[a]

	initial energy	$\langle E \rangle$	μ°
	Simulation 1		
receptor	−38.5	−2.0	97.3
complex	−74.6	−26.5	95.5
	Simulation 2		
receptor	−39.4	−2.1	97.5
complex	−75.6	−26.5	93.9

[a] Results are based on the GB solvation model only. See Table 4 for symbols.

We sought to improve convergence by extending the 2 ns simulations of the complex and the free receptor to 6 ns. QH analysis in BAT coordinates of these longer simulations still leaves a 1.6 kcal/mol discrepancy between the two MD runs for the complex, though the results for the free receptor now agree to within 0.2 kcal/mol. (See Table 6.) The binding free energies computed from these QH calculations are still inaccurate, at −6.1 and −7.9 kcal/mol for Simulations 1 and 2, respectively. Presumably, the free receptor results converge more readily because it is simpler than the complex. The 0.2 kcal/mol difference between the two 6 ns results for the free receptor is traceable to slightly different probabilities of the two main conformations of the receptor. In conformation A, the dihedral angle marked in Figure 5 is near 0, and in conformation B, this angle is near π radians. As illustrated in Figures 6 and 7, both MD runs find conformations A and B. However, simulation 1 assigns a 50% probability to conformation A, while simulation 2 assigns 35.7% probability to conformation A. The implications of this conformational transition are highlighted by the graphs of computed free energy versus time in Figure 6 and 7, which show sudden drops in free energy (top graph) when the receptor shifts from conformation A to conformation B (bottom graph).

It is also of interest to test a prior suggestion for achieving well-defined results from the QH method, that is, using short MD simulations in order to limit the number of transitions among different energy wells.[73] We tested this approach by repeating the QH calculations with 100-fold-shorter MD runs, 20 ps in length. As shown in Table 7, this method does not improve the agreement with the experiment results or with the M2 calculations. Similar results were obtained when

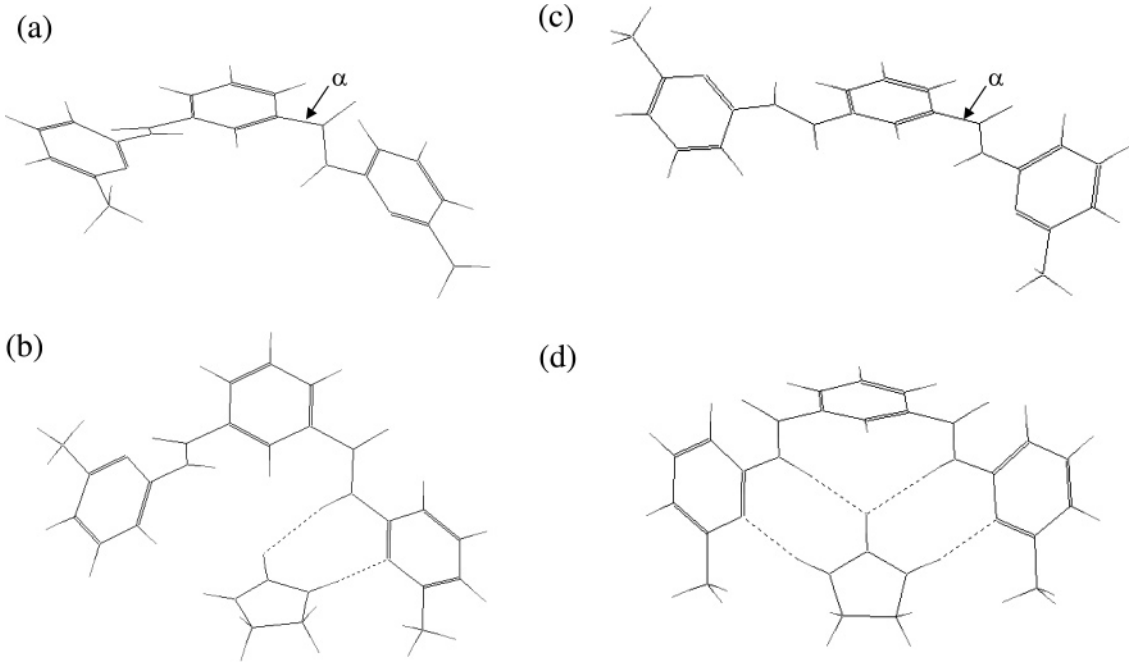


Figure 5. Conformations of free host and host−guest complex. Top (a, c): free receptor. Bottom (b, d): receptor−ligand complex. Left (a, b): QUANTA-generated conformations. Right (c, d): global energy minima obtained from a Tork conformational search. A key dihedral angle, α, is noted. (See Results and Figures 6 and 7.)

Table 7. QH Calculations of Standard Chemical Potential, Average Potential Energy, and Configurational Entropy for Binding of the Urea Receptor with Ethylenurea Based upon Only the First 20 ps of the MD Simulations

			bond-angle-torsion			Cartesian		
	initial energy	$\langle E \rangle$	$-TS^\circ_{config}$	μ°	μ°_{corr}	$-TS^\circ_{config}$	μ°	μ°_{corr}
ethylenurea	−21.1	11.9	19.5	7.6	4.7	17.8	6.9	4.1
receptor	−38.5	−1.4	102.8	101.5	83.5	82.6	81.2	63.3
complex	−74.6	−26.4	126.2	99.8	76.7	102.4	76.0	52.9
change on binding		**−13.1**	**3.9**	**−9.3**	**−11.5**	**2.0**	**−12.1**	**−14.5**

another set of 20 ps simulations was started from the global energy minimum conformation of each species (results not shown). Inspection of the first trajectory shows that, although the free receptor does stay in one energy well during the short simulation, the more flexible complex still samples multiple energy minima during the 20 ps simulation. Thus, shortening the simulation to 20 ps does not accomplish the goal of eliminating conformational transitions. More importantly, failure to sample over multiple energy wells that are occupied at the ambient temperature is expected to produce its own errors. Thus, using artificially short simulations does not appear to be a helpful approach to improving the QH method.

It is encouraging that the change in mean energy $\langle E \rangle$ upon binding computed with the M2 method (−12.7 kcal/mol) agrees well with that from the 2 ns MD simulations used in the QH calculations (−12.3 and −12.2 kcal/mol, as shown in Tables 3−5). This result is consistent with the alkane test cases and further supports the validity of the equations and approximations used in the M2 method, which does not involve any MD calculations. It also indicates that the difference in the binding free energies from the QH and M2 methods results primarily from differences in the entropies they yield. In particular, the QH approximation overestimates entropies, especially for relatively flexible systems. Thus, QH with BAT coordinates and M2 yield very similar values of $-TS^\circ$ for the rigid ethylenurea ligand (19.3 and 19.4 kcal/mol, respectively), but the deviation is larger for the more flexible receptor (104.2 versus 101.1 or 103.0 kcal/mol) and larger still for the complex (134.3 versus 123.9 or 122.5 kcal/mol). As a consequence, the QH calculations markedly overestimate the binding affinity.

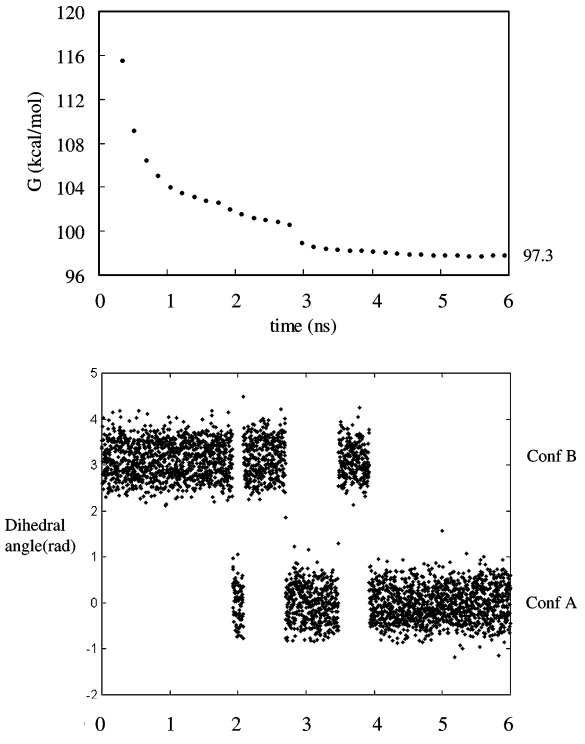


Figure 6. Analysis of QH calculations for the urea receptor, based upon simulation 1 (Table 6), which starts from the QUANTA-generated conformation. Top: convergence of μ°. Bottom: values of dihedral angle α (Figure 5) defining conformations A ($\alpha = 0$) and B ($\alpha = \pi$).

5. Discussion and Conclusions

This study has characterized the accuracy of the QH approximation by studying molecular systems that are small enough to allow comparison with accurate reference calculations, including a host−guest system for which experimental binding data are available. We find that the QH approximation can be accurate for a system with a single, highly occupied energy well, when the BAT coordinate system is

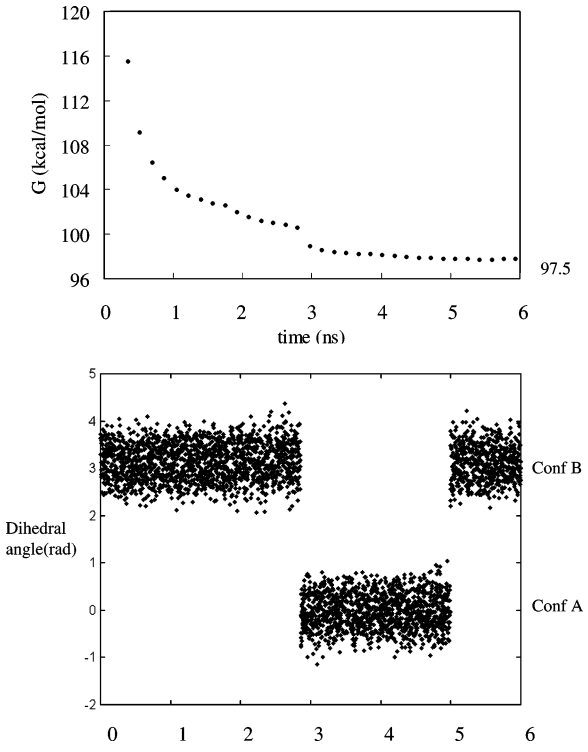


Figure 7. Same as Figure 6, but for simulation 2 (Table 6), which starts from the global energy minimum found by a Tork conformational search.

used as a basis set. Somewhat surprisingly, however, the use of a Cartesian basis set can introduce errors of several kilocalories per mole, even if the system is as simple as a linear alkane and attention is limited to a single energy well. Furthermore, for systems that explore multiple conformations (energy wells), the QH method can readily generate errors in the tens of kilocalories per mole. As in the case of a single energy well, the errors are much greater when a Cartesian basis set is used, rather than a BAT basis set. Although some cancellation of error occurs when a binding free energy is computed, the residual errors for the simple host−guest system studied here are still substantial. In contrast, the M2 method, which treats each energy well separately and uses BAT coordinates, agrees with experimental results to within 0.7 kcal/mol. It is expected that the QH method will also produce large errors for proteins, because proteins also access multiple distinct energy wells at room temperature.[34−36]

The chief reason for the inaccuracy of the QH method appears to be that it effectively merges multiple narrow energy wells and represents them, instead, as one wide energy well whose entropy is much greater than that associated with the individual energy wells (see Figure 3). This explanation is consistent with the fact that the entropy from the QH method is an upper limit to the actual entropy associated with the observed conformational variance.[46] That Cartesian coordinates yield particularly large errors, even for fluctuations within a single energy well (Table 2b), seems to result from the fact that fluctuation of a single torsional degree of freedom produces fluctuations of multiple Cartesian degrees of freedom and, hence, leads to higher computer entropies. This problem is least severe for small fluctuations, and the errors tend to cancel when taking free energy differences. However, there is no guarantee of adequate cancellation in any given calculation.

The quasiharmonic method also poses significant convergence challenges. Thus, two separate 2 ns MD calculations of the free and bound host−guest systems were found to yield binding free energies that differed by about 3 kcal/mol, and even 6 ns simulations are not fully converged for the receptor ligand complex. These are rather simple systems, so much longer calculations are presumably needed for a system as complicated as a protein. This expectation is consistent with Rinaldo and Field's QH results for the protein transferrin; Figure 15 in their paper shows that the computed entropy rises monotonically by 7 kcal/K mol during the course of a 1.5 ns simulation and is clearly still not converged at the end.[43] One reason convergence is difficult to achieve is that merely visiting all relevant energy wells is not enough; it is also necessary for the relative populations of the various wells to have converged fairly well. In contrast, converging the M2 method only requires visiting and integrating each occupied energy well a single time. Indeed, for the systems studied here, the M2 method can be run to convergence in about the time required to carry out a 2 ns MD calculation, and it yields more accurate and better converged results.

Overall, the present study indicates that the QH method can be accurate for a molecular system that remains in a single energy well but that it substantially overestimates the configurational entropy of systems that access multiple energy wells. The errors are particularly severe when the covariance matrix is computed in Cartesian coordinates. It is also found that the cancellation of errors does not suffice to provide accurate entropies for changes of state and that achieving good convergence can be surprisingly difficult even for a simple system, if it accesses multiple energy wells at ambient temperature.

Acknowledgment. This publication was made possible by Grant GM61300 from the National Institute of General Medical Sciences (NIGMS) of the NIH. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS. The authors thank Dr. Ioan Andricioaei for a helpful conversation.

References

(1) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281−283.

(2) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

(3) Miyamoto, S.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 226−245.

(4) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830−3832.

(5) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564−568.

(6) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−92.

(7) Lybrand, T. *Rev. Comput. Chem.* **1990**, *1*, 295−320.

(8) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.

(9) Warshel, A.; Tao, H.; Fothergill, M.; Chu, Z.-T. *Isr. J. Chem.* **1994**, *34*, 253−256.

(10) Marrone, T. J.; Briggs, J. M.; McCammon, J. A. *Annu. Rev. Pharmacol. Toxicol.* **1997**, *37*, 71−90.

(11) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430−437.

(12) Jorgensen, W. L. *Science* **2004**, *303*, 1813−1818.

(13) Reddy, M. R.; Erion, M. D. *J. Enzyme Inhib.* **1998**, *14*, 1−14.

(14) Mitchell, M. J.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 271−275.

(15) Balbes, L. M.; Mascarella, S. W.; Boyd, D. B. *Rev. Comput. Chem.* **1994**, *5*, 337−379.

(16) Aqvist, J.; Medina, C.; Samuelsson, E. J. *Protein Eng.* **1994**, *7*, 385−391.

(17) Wang, W.; Wang, J.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 395−402.

(18) Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. *J. Phys. Chem. B* **2001**, *105*, 10388−10397.

(19) Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.

(20) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. *Biopolymers* **2003**, *68*, 16−34.

(21) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891−913.

(22) Chang, C.-E.; Gilson, M. K. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.

(23) Chen, W.; Chang, C.-E.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 3035−3049.

(24) Brooks, B. R.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.

(25) Go, N.; Scheraga, H. *J. Chem. Phys.* **1969**, *51*, 4751−4767.

(26) Go, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535−542.

(27) Potter, M. J.; Gilson, M. K. *J. Phys. Chem. A* **2002**, *126*, 563−566.

(28) Chang, C.-E.; Potter, M. J.; Gilson, M. K. *J. Phys. Chem. B* **2003**, *107*, 1048−1055.

(29) Eckart, C. *Phys. Rev.* **1935**, *47*, 552−558.

(30) Wilson, E. B., Jr.; Howard, J. B. *J. Chem. Phys.* **1936**, *4*, 260−268.

(31) Sayvetz, A. *J. Chem. Phys.* **1939**, *7*, 383−389.

(32) Kolossvary, I. *J. Phys. Chem. A* **1997**, *101*, 9900−9905.

(33) Rojas, O. L.; Levy, R. M.; Szabo, A. *J. Chem. Phys.* **1986**, *85*, 1037−1043.

(34) Ansari, A.; Berendsen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E. T.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 5000−5004.

(35) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318−321.

(36) Kitao, A.; Hayward, S.; Go, N. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 496−517.

(37) Darian, E.; Hnizdo, V.; Fedorowicz, A.; Singh, H.; Demchuk, E. *J. Comput. Chem.* **2005**, *26*, 651−660.

(38) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617−621.

(39) Schafer, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **2000**, *113*, 7809−7817.

(40) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289−6292.

(41) Hsu, S.-T. D.; Peter, C.; van Gunsteren, W. F.; Bonvin, A. M. J. J. *Biophys. J.* **2005**, *88*, 15−24.

(42) Jusuf, S.; Loll, P. J.; Axelsen, P. H. *J. Am. Chem. Soc.* **2002**, *124*, 3490−3491.

(43) Rinaldo, D.; Field, M. J. *Biophys. J.* **2003**, *85*, 3485−3501.

(44) Luo, H.; Sharp, K. *PNAS* **2002**, *99*, 10399−10404.

(45) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67−74.

(46) Shannon, C. E.; Weaver, W. *University of Illinois Press* 1949.

(47) Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620−630.

(48) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.

(49) Mihailescu, M.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 23−26.

(50) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover: New York, 1986; Chapter 15.

(51) Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652−1661.

(52) Boresch, S.; Karplus, M. *J. Chem. Phys.* **2001**, *105*, 5145−5154.

(53) Kolossvary, I.; McMartin, C. *J. Math. Chem.* **1992**, *9*, 359−367.

(54) Pitzer, K. S. *J. Chem. Phys.* **1946**, *14*, 239−243.

(55) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325−332.

(56) Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. *Macromolecules* **1984**, *17*, 1370−1374.

(57) Levy, R. M.; Rojas, O. L.; Friesner, R. A. *J. Phys. Chem.* **1984**, *88*, 4233−4238.

(58) Chang, C.-E.; Gilson, M. K. *J. Comput. Chem.* **2003**, *24*, 1987−1998.

(59) Kolossvary, I.; Guida, W. *J. Am. Chem. Soc.* **1996**, *118*, 5011−5019.

(60) Kolossvary, I.; Guida, W. *J. Comput. Chem.* **1999**, *20*, 1671−1684.

(61) Keserû, G. M.; Kolossvary, I. *J. Am. Chem. Soc.* **2001**, *123*, 12708−12709.

(62) Chen, W.; Huang, J.; Gilson, M. K. *J. Chem. Inf. Comput. Sci.* **2004**, *44*(4), 1301−1313.

(63) Ivanov, J.; Schûûrmann, G. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 728−737.

(64) *CHARMm*, version 22; Molecular Simulations Inc.: Waltham, MA, 1992.

(65) *QUANTA*; Accelrys, Inc.: San Diego, CA.

(66) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(67) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187−197.

(68) Luo, R.; Head, M. S.; Given, J. A.; Gilson, M. K. *Biophys. Chem.* **1999**, *78*, 183−193.

(69) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.

(70) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101−1109.

(71) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simul.* **1988**, *1*, 173−185.

(72) Goswami, S.; Mukherjee, R. *Tetrahedron Lett.* **1997**, *38*, 1619−1622.

(73) Nola, A. D.; Berendsen, H. J. C.; Edholm, O. *Macromolecules* **1984**, *17*, 2044−2050.

CT0500904

JCTC Journal of Chemical Theory and Computation

New General Tools for Constrained Geometry Optimizations

Luca De Vico

Department of Theoretical Chemistry, Lund University, Chemical Centre, P.O. Box 124, S-221 00 Lund, Sweden

Massimo Olivucci*

Dipartimento di Chimica, Università di Siena, via A. De Gasperi 2, I-53100 Siena, Italy, and Centro per lo Studio dei Sistemi Complessi, Via Tommaso Pendola 37, I-53100 Siena, Italy

Roland Lindh*

Department of Chemical Physics, Lund University, Chemical Centre, P.O. Box 124, S-221 00 Lund, Sweden

Received April 8, 2005

Abstract: A modification of the constrained geometry optimization method by Anglada and Bofill (Anglada, J. M.; Bofill, J. M. *J. Comput. Chem.* **1997**, *18*, 992−1003) is designed and implemented. The changes include the choice of projection, quasi-line-search, and the use of a Rational Function optimization approach rather than a reduced-restricted-quasi-Newton−Raphson method in the optimization step. Furthermore, we show how geometrical constrains can be implemented in an approach based on nonredundant curvilinear coordinates avoiding the inclusion of the constraints in the set of redundant coordinates used to define the internal coordinates. The behavior of the new implementation is demonstrated in geometry optimizations featuring single or multiple geometrical constraints (bond lengths, angles, etc.), optimizations on hyperspherical cross sections (as in the computation of steepest descent paths), and location of energy minima on the intersection subspace of two potential energy surfaces (i.e. minimum energy crossing points). In addition, a novel scheme to determine the crossing point geometrically nearest to a given molecular structure is proposed.

1. Introduction

Whereas the standard (i.e. unconstrained) optimization of equilibrium structures on the potential energy surface (PES) is considered a relatively easy task, the optimization of transition structures and constrained geometry optimizations represent a more complex problem. For a survey of currently established methods for optimizing minima and transition state structures see refs 1 and 2 and references therein. For certain applications as, for example, the location of a minimum on a PES hyperspherical cross section (e.g. used for mapping steepest descent paths)[3] or of a stationary point located along the intersection subspace of two or more PESs[4] the implementation of novel efficient constrained optimization methods is highly desirable.

In addition to projection methods[5] and penalty functions[6] based methods, constrained geometry optimizations are mostly based on variations of the Lagrange multiplier method.[4,7,8] Here the space is extended from $3n$-6(5) to $3n$-6(5)+m, where n is the number of atoms and m is the number

* Corresponding authors e-mail: olivucci@unisi.it (M.O.) and Roland.Lindh@chemphys.lu.se (R.L.).

10.1021/ct0500949 CCC: $30.25
Published on Web 07/09/2005

of constraints. For each constraint the Hessian of the Lagrangian carries a negative eigenvalue. However, when locating a minimum energy crossing point (MECP),[4,7] the Lagrange multiplier method can sometime, in our and others[5] experience, show poor convergence, and the reason for this is associated with the Hessian. In fact (1) the separation of the space into the constraint subspace and the disjoint subspace, in which the optimization is carried out, is, in standard quasi-Newton max−min optimizations,[9] defined implicitly by the eigenvectors of the Hessian of the Lagrangian, and (2) the so-called Broyden−Fletcher−Goldfarb−Shanno (BFGS)[10] Hessian update method cannot be exploited because the Lagrangian Hessian is not positive definite.

The first matter is under control in standard optimization with geometrical constraint(s), since here the second-order derivative of the constraint(s) with respect to the internal coordinates can be computed without much effort. However, in the location of a MECP, the computation of second-order derivatives would be an impractical and expensive procedure. Hence, constrained optimizations to locate MECPs are performed with approximate Hessian in combination with Hessian update methods.

An alternative to the Lagrange multiplier method was proposed by Bearpark et al.[5] In this approach, the so-called "direct" approach, the subspace of the constraint(s), is identified to first order and treated separately from the rest of the optimization. However, the minimization is still performed in the full space in combination with projection operators. The major advantage of the "direct" approach is that the definition of the subspace of the constraint(s) is not implicitly defined by the Hessian of the Lagrangian but rather from the gradient of the constraint(s) with respect to the nuclear displacements.

Recently Anglada and Bofill[11] described the connection between the "direct" and the Lagrange multiplier approach. They demonstrated that the Lagrange multiplier method can, by a simple linear transformation, be recast into *one subspace of the constraint(s) (of dimension m) and one geometrical coordinate subspace (of* dimension 3*n*-6(5)-*m*) in which a minimization is performed. *The latter is guaranteed to have a positive definite* Hessian. Hence, it combines the advantage of space subdivision, as suggested by the so-called direct approach, with a conventional minimization in combination with a BFGS Hessian update procedure. The method was implemented in the semiempirical program package AMPAC[12] and tested in combination with the AM1 Hamiltonian,[13] using a reduced-restricted-quasi-Newton−Raphson method and explicit line-search techniques.

The combination of general geometrical constraints with optimizations in Cartesian internal coordinates is trivial.[14] While a scheme for constrained optimizations in conjunction with curvilinear internal coordinates has been suggested by Baker et al.,[15] this requires that the geometrical constraint(s) are included in the coordinate space. It would be of advantage if the constraint(s) could be excluded from the coordinate space since they do not usually qualify as a natural internal coordinate.

In this paper we will explore the method of Anglada and Bofill, in combination with the Rational Function approach[16] and no line-search techniques. The method will be assessed at the (state average) CASSCF level of theory (even if any kind of level of theory could have been used, given that it was suitable for the studied system) and in combination with force-constant weighted redundant internal coordinates[17] and Hessian model function[18] implemented in the MOLCAS 6.0 quantum chemistry program package.[19,20] Further modifications of the approach will be suggested, and it will be demonstrated how constraints can be treated in combination with curvilinear internal coordinates and without their inclusion in the subspace of the redundant internal coordinates. In particular, we will present *an original procedure to optimize the crossing point geometrically nearest to a given* structure, that in the future will be made automatic.

2. Method

The methods section is subdivided into four parts. In the first part we will review and modify the projected constrained optimization (PCO) method of Anglada and Bofill. This is followed by an explicit description and motivation of our implementation in combination with a Rational Function optimization (RFO) and no line-search. Afterward, in the third subsection we describe the tools needed for combining arbitrary geometrical constraints with curvilinear internal coordinates. Finally, the last subsection will report some technical computational details.

2.1. The PCO Approach by Anglada and Bofill. The constrained geometry optimization using the Lagrange multiplier method is formulated as a minimization of the expression

$$L(\mathbf{q},\lambda) = E(\mathbf{q}) - \lambda^T \mathbf{r}(\mathbf{q}) \tag{1}$$

where E is the energy to be minimized with respect to the constraints $\mathbf{r}(\mathbf{q})$, and λ is the vector with the so-called Lagrange multipliers. A Taylor expansion to second order of $L(\mathbf{q}, \lambda)$ around $\mathbf{q}_0$ and λ_0 gives

$$L(\mathbf{q}_0 + \Delta\mathbf{q}, \lambda_0 + \Delta\lambda) = E(\mathbf{q}_0) + \Delta\mathbf{q}^T \frac{\partial E(\mathbf{q}_0)}{\partial \mathbf{q}} + (1/2)\Delta\mathbf{q}^T \mathbf{W} \Delta\mathbf{q} - \lambda^T \left(\mathbf{r}(\mathbf{q}_0) + \frac{\partial \mathbf{r}(\mathbf{q}_0)}{\partial \mathbf{q}} \Delta\mathbf{q} \right) \tag{2}$$

where $\mathbf{W}$ is defined as

$$\mathbf{W}(\mathbf{q}_0,\lambda_0) = \frac{\partial^2 E(\mathbf{q}_0)}{\partial \mathbf{q}^2} - \sum_{i=1,m} (\lambda_0)_i \frac{\partial^2 (\mathbf{r}(\mathbf{q}_0))_i}{\partial \mathbf{q}^2} \tag{3}$$

This sets up the equation for the generalized elimination method.[1,21] We note that the last term of the right-hand side of eq 2 controls to first order the constraint. In the subspace which fulfills the constraint any displacement $\Delta\mathbf{q}$ must be such that

$$\mathbf{r}(\mathbf{q}_0) + \frac{\partial \mathbf{r}(\mathbf{q}_0)}{\partial \mathbf{q}} \Delta\mathbf{q} = \mathbf{0} \tag{4}$$

This defines a linear transformation which will to first order subdivide the original 3*n*-6(5) space into a m-dimensional space in which the constraints are fulfilled and

a 3*n*-6(5)-*m* subspace in which a normal optimization is made. The unitary transformation matrix **T** contains two parts and transforms as

$$\Delta\mathbf{q}_0 = [\mathbf{T}_c\mathbf{T}_m]\begin{pmatrix}\Delta\mathbf{y}\\ \Delta\mathbf{x}\end{pmatrix} = \mathbf{T}_c\Delta\mathbf{y} + \mathbf{T}_m\Delta\mathbf{x} \tag{5}$$

where **y** and **x** are the new parameters. **y** is of *m* dimensions and **x** is of 3*n*-6(5)-*m* dimension. In particular we note that at $\mathbf{q}_0$

$$\frac{\partial\mathbf{r}(\mathbf{q}_0)}{\partial\mathbf{q}}\mathbf{T}_c \neq \mathbf{0} \tag{6}$$

and

$$\frac{\partial\mathbf{r}(\mathbf{q}_0)}{\partial\mathbf{q}}\mathbf{T}_m = \mathbf{0} \tag{7}$$

These two equations are sufficient for the definition of **T** via a Gram-Schmidt procedure.[22]

We now proceed by introducing the transformation matrix into the Lagrangian expression (eq 2). The equation now splits into two parts (one of *m* and a second of 3*n*-6(5)*m* dimensions), one of which depends only on **y**

$$\Delta\mathbf{y} = -\left(\frac{\partial\mathbf{r}(\mathbf{q}_0)}{\partial\mathbf{q}}\mathbf{T}_c\right)^{-1}\mathbf{r}(\mathbf{q}_0) \tag{8}$$

and a second part that depends on both **x** and **y**

$$Q(\mathbf{q}_0+\Delta\mathbf{q}) = E(\mathbf{q}_0) + \Delta\mathbf{y}^T\mathbf{T}_c{}^T\frac{\partial E(\mathbf{q}_0)}{\partial\mathbf{q}} + (1/2)\Delta\mathbf{y}^T\mathbf{T}_c{}^T\mathbf{W}\mathbf{T}_c\Delta\mathbf{y} + \Delta\mathbf{x}^T\mathbf{T}_m{}^T\left(\frac{\partial E(\mathbf{q}_0)}{\partial\mathbf{q}} + \mathbf{W}\mathbf{T}_c\Delta\mathbf{y}\right) + (1/2)\Delta\mathbf{x}^T\mathbf{T}_m{}^T\mathbf{W}\mathbf{T}_m\Delta\mathbf{x} \tag{9}$$

The latter is the so-called projected energy expression, with $\mathbf{T}_m{}^T\mathbf{W}\mathbf{T}_m$ being the reduced Hessian and $\mathbf{T}_m{}^T([\partial E(\mathbf{q}_0)]/[\partial\mathbf{q}] + \mathbf{W}\mathbf{T}_c\Delta\mathbf{y})$ is the reduced gradient. Solving eq 8 and substituting the result into eq 9 we minimize *Q* with respect to $\Delta\mathbf{x}$. The minimization of eq 9 can be done with any standard optimization technique. At convergence we have found the solution to the Lagrangian in eq 1.

Using the quasi-Newton condition applied to eq 1 we find that the effective gradient to be used in an Hessian update procedure applied only to the molecular part of the Lagrangian Hessian is

$$\mathbf{h}(\mathbf{q},\lambda) = \frac{\partial E(\mathbf{q})}{\partial\mathbf{q}} - \frac{\partial\mathbf{r}(\mathbf{q})}{\partial\mathbf{q}}\lambda \tag{10}$$

The update procedure is commenced by evaluating a series of $\mathbf{h}(\mathbf{q},\lambda)$ for different values of **q** and a fixed value of λ. A suitable value of λ is the first order estimate of λ at convergence as given by

$$\mathbf{h}(\mathbf{q}_0,\lambda_0) = \mathbf{0} \tag{11}$$

To conclude the presentation of the PCO approach of Anglada and Bofill let us here summarize the major advantages of the presented method, compared to an optimization/maximization procedure applied to eq 1: (1) the PCO approch has an explicit separation of the two subspaces, while an optimization/maximization indirectly separates the two subspaces by identifying the positive and negative eigenvectors of the Hessian of the Lagrangian, (2) the presence of negative eigenvalues in the Hessian of the Lagrangian restricts the selection of Hessian update (variable metric) methods, while the PCO approach allows the use of the BFGS update method to be applied to the reduced Hessian, and (3) the technique by Bofill and Anglada does not explicitly require λ to be determined.

In particular the first point is of importance since it allows us to perform constrained optimization in cases where we cannot or would not like to compute explicit second-order derivatives of the constraints with respect to nuclear displacement. One such important case is in the search of a minimum energy cross point (MECP) on the search of a (conical) intersection.

2.2. An RFO Implementation without Line-Search. In this section we will sketch on an implementation of the PCO approach in combination with Rational Function optimization and a back-feed mechanism rather than a line-search.

For a given $\mathbf{q}_1$ and a maximum step length of β, set $k = 1$, $y_\beta = 1$, $g_\beta = 1$, $x_\beta = 1$, $dy = \beta$, $g = \beta$, and $dx = \beta$.

(1) Compute $E(\mathbf{q}_k)$, $\frac{\partial E(\mathbf{q}_k)}{\partial\mathbf{q}}$, an empirical estimate of $\frac{\partial^2 E(\mathbf{q}_k)}{\partial\mathbf{q}^2}$, $\mathbf{r}(\mathbf{q}_k)$, $\frac{\partial\mathbf{r}(\mathbf{q}_k)}{\partial\mathbf{q}}$, and $\frac{\partial^2\mathbf{r}(\mathbf{q}_k)}{\partial\mathbf{q}^2}$

(2) For $i = 1,k$

Evaluate $\mathbf{T}(\mathbf{q}_i)$, $\Delta\mathbf{y}(\mathbf{q}_i)$, $\lambda(\mathbf{q}_i)$, $\mathbf{x}(\mathbf{q}_i)$, $\Delta\mathbf{x}(\mathbf{q}_i)$ and the reduced gradient, $\mathbf{g}_r(\mathbf{q}_i)$, as defined in previous equations.

If $dy < 0.75|\Delta\mathbf{y}|$ or $|\Delta\mathbf{y}| < 1.0\cdot10^{-2}$
set $y_\beta = \min(2, y_\beta \times \sqrt{2})$
Else if $(dy > 1.25|\Delta\mathbf{y}|)$ and $(dy \geq 1.0\cdot10^{-5})$
set $y_\beta = \max(1/10, y_\beta/\sqrt{2})$
Set $dy = |\Delta\mathbf{y}|$

and

If $i \neq k$
If $dx < |\Delta\mathbf{x}|$ and $|\Delta\mathbf{x}| < \beta$
set $x_\beta = \min(1, x_\beta \times \sqrt{2})$
Else if $dx > 1.25|\Delta\mathbf{x}|$ or $dx \geq \beta$
set $x_\beta = \max(1/5, x_\beta/\sqrt{2})$
Set $dx = |\Delta\mathbf{x}|$

and

If $g < 0.75|\mathbf{g}_r|$ and $g < \beta$
set $g_\beta = \min(1, g_\beta \times \sqrt{2})$
Else if $g > 1.25|\mathbf{g}_r|$ or $g \geq \beta$
set $g_\beta = \max(1/5, g_\beta/\sqrt{2})$
Set $g = |\mathbf{g}_r|$.

(3) For $i = 1,k$

Evaluate $\mathbf{h}(\mathbf{q}_i,\lambda_k)$ as expressed by eq 10.

(4) Compute $\mathbf{W}$ and update with the BFGS procedure.

(5) Assemble the reduced Hessian according to eq 9.

(6) Minimize Q (see eq 9) with respect to $\Delta\mathbf{x}(\mathbf{q}_k)$ with RFO in association with a step restriction of $\beta \times \max(y_\beta \times \min(x_\beta, g_\beta), (1/10))$.

(7) Evaluate $\Delta\mathbf{q}(\mathbf{q}_k)$ from $\Delta\mathbf{y}(\mathbf{q}_k)$ and $\Delta\mathbf{x}(\mathbf{q}_k)$.

(8) If not converged, set $k = k + 1$ and go to (1).

The design of this scheme has a major objective to give priority to the constraint(s) condition over the overall minimization procedure. Once the constraints are fulfilled, the algorithm will protect this condition. The implemented method was tailored in order to achieve a better convergence on some test cases. Mainly, this refinement consisted in tuning the parameters relative to the components of the step length of the two subspaces.

2.3. Derivatives of Constraints. Constrained geometry optimization requires first- and second-order derivatives of the constraint(s) with respect to the internal coordinates. This is trivial for Cartesian coordinates. For curvilinear coordinates Baker et al.[15] suggested to include the constraint(s) in the definition of the redundant internal coordinate space, followed by a linear transformation after the generation of the nonredundant coordinates, such that the coefficient(s) of the constraint(s) equals(equal) one(some) of the nonredundant coordinates. In this paper we will present the required derivatives for the case when the constraints not necessarily are explicitly expressed in the redundant space of internal coordinates.

Given

$$\frac{\partial}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{q}}\frac{\partial \mathbf{q}}{\partial \mathbf{x}} \tag{12}$$

The first-order derivative is expressed as

$$\frac{\partial \mathbf{r}}{\partial \mathbf{q}} = \left(\frac{\partial \mathbf{q}}{\partial \mathbf{x}}\right)^{-1}\frac{\partial \mathbf{r}}{\partial \mathbf{x}} \tag{13}$$

And the second derivative is expressed as

$$\frac{\partial^2 \mathbf{r}}{\partial \mathbf{q}^2} = \left(\frac{\partial \mathbf{q}}{\partial \mathbf{x}}\right)^{-1}\left[\frac{\partial^2 \mathbf{r}}{\partial \mathbf{x}^2} - \frac{\partial \mathbf{r}}{\partial \mathbf{q}}\frac{\partial^2 \mathbf{q}}{\partial \mathbf{x}^2}\right]\left(\frac{\partial \mathbf{q}}{\partial \mathbf{x}}\right)^{-1} \tag{14}$$

2.4. Computational Details. The method described above was used to perform constrained geometry optimizations, testing the response with constraints of various kinds. The standard thresholds of the MOLCAS SlapAf module were applied to check convergence. Namely, a RMS value of 0.12×10^{-2} au and a largest component smaller than 0.18×10^{-2} au was considered as a threshold for the Cartesian displacements; a RMS value of 0.30×10^{-3} au and a largest component smaller than 0.45×10^{-3} au was considered as a threshold for the gradient vector. All optimizations reported in the Results section were calculated at the CASSCF level of theory. Nonetheless, the method could have been applied using any level of theory, as long as it is appropriate to the studied problem. Computations in subsections 3.1 and 3.2 employed a ANO-RCC basis set[23] and a minimum active space comprising 2 electrons in 2 orbitals. The 6-31G* basis set[24] was chosen for the optimizations presented in subsections 3.3−3.6. For *N*-methylthioacetamide (NMTAA, see subsection 3.3) an active space comprising 10 electrons in 8 orbitals (see ref 25 for details on orbitals' choice) was chosen, while a minimum (6 electrons in 5 orbitals) active space comprising the π orbitals and one of the oxygen lone pairs was preferred for acrolein (subsections 3.4−3.6), respectively.

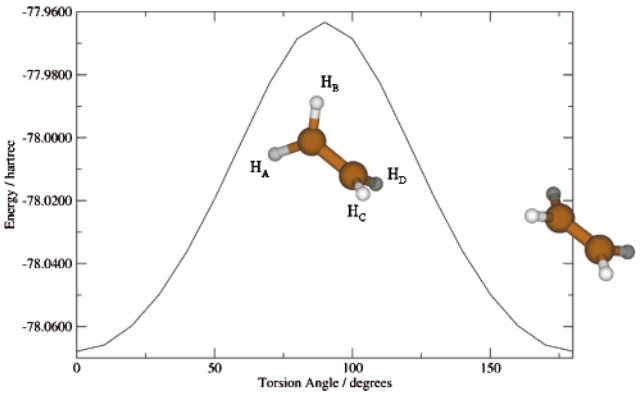


Figure 1. Ethene hindered internal rotation.

3. Results and Discussion

In this section we test the implemented method with respect to (i) geometry optimizations with single and multiple constraints, (ii) optimizations on a PES hyperspherical cross section applied to the mapping of steepest descent paths, and (iii) the optimization of MECPs.

3.1. Single and Multiple Geometrical Constraint Optimization. To evaluate the efficiency of the geometry optimization with a single geometrical constraint, we compute the energy profile for ethene double bond twisting (see Figure 1). This is done running a set of geometry optimizations where the H_ACCH_C dihedral angle is required to assume a predefined final value. Thus, the constraint is

$$r_1 = \phi_{H_ACCH_C} - \phi^0_{H_ACCH_C} \tag{15}$$

On average, we find that the optimizations converge after 6 iterations for each selected angle.

To evaluate if this efficiency is maintained in the case of an increased number of geometrical constrained, we have mapped a two-dimensional cross section of the triplet potential energy surface of the 1,2-dioxoethane biradical. This has been carried out running a grid of geometry optimizations where the O−O bond distance and the O−C−C−O dihedral angle are required to assume predefined final values (see surface in Figure 2). In this case the constraints are

$$r_1 = r_{OO} - r_{OO}{}^0 \tag{16}$$

and

$$r_2 = \phi_{OCCO} - \phi_{OCCO}{}^0 \tag{17}$$

As for the case of a single constraint we have found that each optimization typically converge after 5 iterations.

3.2. Steepest Descent Path Optimizations. Steepest descent paths can be constructed by using a variety of algorithms, see ref 2 and references therein. We chose to

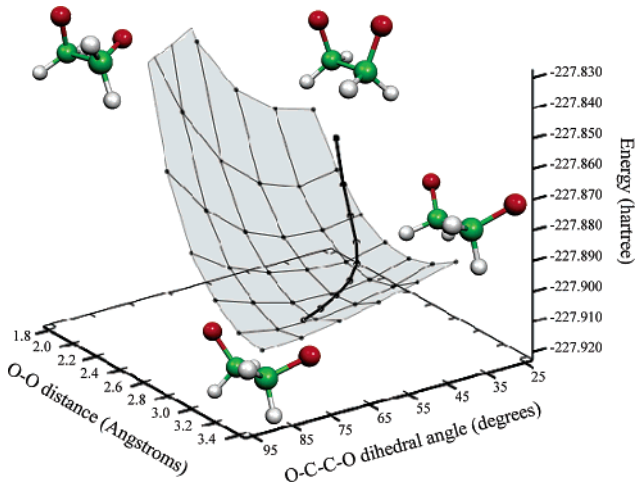


Figure 2. The potential energy surface of triplet 1,2-dioxetane as a function of the O−O bond distance (angstrom) and the O−C−C−O dihedral angle (degrees), shaded surface. Each point of the grid was obtained through the procedure described in section 3.1. The minimum energy path of 1,2-dioxoethane on the same state surface starting at the geometry of the singlet transition state is also reported in black line. The method used to calculate the path is described in section 3.2. The first part of the path (higher energy) is mainly characterized by the O−O stretching coordinate, while the second part (lower energy) is mostly described by the O−C−C−O dihedral torsional coordinate.

follow a Müller−Brown approach[26] and construct them through a series of geometry optimizations, each requiring the minimization of the potential energy on a hyperspherical cross section of the PES centered on a given reference geometry and characterized by a predefined radius. One usually starts the calculation from a high energy reference geometry, which may correspond to the Franck−Condon structure on an excited-state PES or to a transition structure (TS). Once the first lower energy optimized structure is converged, this is taken as the new hypersphere center, and the procedure is iterated until the bottom of the energy surface is reached. Notice that in the TS case a pair of steepest descent paths, connecting the TS to the reactant and product structures (i.e. following the forward and reverse orientation of the direction defined by the transition vector) provides the minimum energy path (MEP) for the reaction. If mass-weighted coordinates $\mathbf{R}(\mathbf{q})$ are used, the MEP coordinate corresponds to the so-called Intrinsic Reaction Coordinates (IRC).

$$\mathbf{R}(\mathbf{q}) = (\sqrt{m_1}\mathbf{r}_1(\mathbf{q}),\sqrt{m_2}\mathbf{r}_2(\mathbf{q}),...,\sqrt{m_n}\mathbf{r}_n(\mathbf{q})) \qquad (18)$$

Choosing this type of coordinates our methods requires that the following constrain is satisfied at the end of the optimization

$$r_1 = \frac{\sqrt{(\mathbf{R}(\mathbf{q}) - \mathbf{R}(\mathbf{q}_{\mathrm{ref}}))^2} - R}{\sqrt{M_{\mathrm{tot}}}} \qquad (19)$$

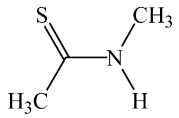
where $\mathbf{R}$ is the radius of the hypersphere and M_{tot} is the total mass of the system.

The performance of our implementation is, again, assessed on the triplet 1,2-dioxoethane biradical. As a starting geometry we take that of the syn singlet biradical structure that defines the transition state for the ring-opening of dioxethane. A hypersphere radius of 0.1 au was chosen. A small value for the radius implies not only less iterations before convergence is reached but also a larger number of points to be calculated to describe the MEP. We found that 0.1 au represents a balanced value to calculate the entire MEP. As we show in Figure 2 the steepest descent path in mass-weighted coordinates is easily constructed even in the presence of a turn in the path coordinate. Each individual optimization typically takes 9 iterations to converge.

3.3. Optimization of Minimum Energy Crossing Points: Nongeometrical Constraints. A MECP is computed by optimizing the structure corresponding to a local excited state (e.g. E_1) energy minimum requiring that a degeneracy constraint with a lower potential energy surface (e.g. E_0) is satisfied at the end of the optimization. Thus the constraint is the energy difference between the two electronic states:

$$r_1 = E_1 - E_0 \qquad (20)$$

Notice that in this case the required constraint is not geometrical. To test the methodology we compute the intersection between the ^{3}A and ^{1}A state of *N*-methylthioacetamide (NMTAA).



N-methylthioacetamide

NMTAA was chosen as a test molecule for calculating MECP because, as reported in a recent paper,[25] we failed to obtain fully optimized singlet/triplet crossings of this molecule by using other methods. The changes of the excited-state energy (E_1) and of the energy difference (r_1) as a function of the number of iteration (Figure 3) display a fast convergence (in about five iterations, both energy difference and excited state energy are nearly converged).

3.4. Multiple Heterogeneous Constraints: MECP Search on an Hyperspherical Cross Section. *The ultimate target of the test presented in this section is to compute the crossing point geometrically closest to a given molecular structure.* To achieve this target two qualitatively different (heterogeneous) types of constraints are imposed during the optimization. The first constraint determines the value of the E_1-E_0 energy difference as seen in section 3.3 above. The second constraint is of the type described in section 3.2 and corresponds to the distance between the target geometry and a given reference molecular structure. Recently, we presented a method[27] for searching PES crossing points over a hyperspherical cross section, not directly based on Lagrangian multipliers. Even if it has been used in various cases,[28] it needed some tailoring specific for each considered molecular system. The present method, on the other hand, should be more generally applicable.

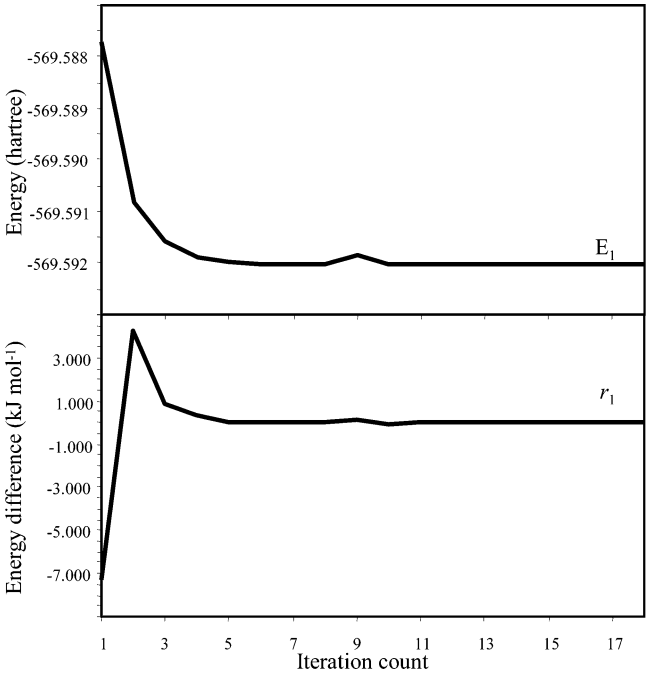


Figure 3. The excited triplet state energy of *N*-methylthioacetamide (indicated as E_1, see text) and the energy difference (indicated as *r*, see eq 20) as a function of the iteration count during the MECP optimization.

Due to its considerable experimental and theoretical interest,[29,30] the simple conjugated carbonyl compound acrolein was chosen for our test. Accordingly, the chosen reference structure is the previously reported[30] ^{3}n-π^* planar minimum of the fundamental triplet state (T_1). Notice that the target structure, i.e., the acrolein crossing point geometrically closest to the selected reference, will, in general, not correspond to a MECP.

As will become clear later, the present implementation of our optimization scheme for finding the T_1/S_0 crossing point nearest (NCP) to the reference T_1 minimum, is not automatic but is based on a series of sequential *singlet/triplet energy difference minimization coupled to a geometry minimization on a hyperspherical cross section of the T_1 energy surface. Each optimization is characterized by a different value of the hypersphere radius that is regularly decreased along the series.* As seen in the previous sections, the two constraints to be applied in each optimization are

$$r_1 = E_1 - E_0 \tag{21}$$

$$r_2 = \frac{\sqrt{(\mathbf{R}(\mathbf{q}) - \mathbf{R}(\mathbf{q}_{\text{ref}}))^2} - R}{\sqrt{M_{\text{tot}}}} \tag{22}$$

At the end of the optimization the r_1 and r_2 values must be zero. More specifically, given a hyperspherical cross section of the PES defined by the radius *R* and assuming that the *N*-1-dimensional T_1/S_0 intersection subspace (where *N* is the number of freedom degrees, equal to 3*n*-6) we look for a structure belonging to the T_1/S_0 intersection and having the lowest possible T_1 energy. As displayed in Figure 4, where one optimization step is analyzed, about 5 iterations were sufficient to achieve nearly convergence.

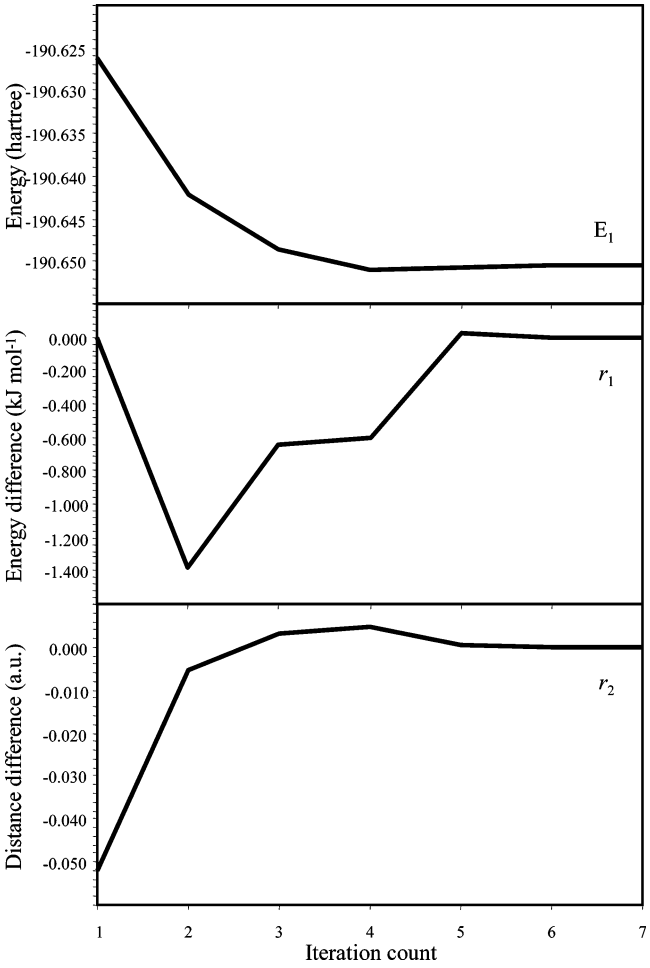


Figure 4. Energy of the ^{3}n-π^* state (indicated as E_1, see text), energy difference (indicated as r_1, see eq 21), and distance difference from the constrained value (indicated as r_2, see eq 22) as a function of iteration count during the optimization of the point of the ^{3}n-π^*/S_0 seam at 0.35 au from the T_1 planar minimum (see Figure 5).

As anticipated above, the minimization of *R* was performed "manually". To better follow such procedure and provide an initial guess we optimized, following the method described in section 3.3, the T_1/S_0 MECP. The optimization yielded a structure located 7 kcal mol^{-1} below the reference ^{3}n-π^* T_1 minimum that shows a twist of 93.3° degrees around the C═C bond, a lengthening of C═C and C−C bonds, and shortening of the C═O bond, with respect to the reference. Moreover, such MECP connects the $^3\pi-\pi^*$ potential energy surface to that of the singlet ground state and is not related to the ^{3}n-π^* state. The first optimization was performed taking a radius smaller than the 0.462 au distance between the reference and the MECP guess. Once an optimized geometry was obtained, its structure was used as starting geometry for the following optimization with a shorter *R*. The procedure was iterated until it was not possible to achieve convergence: at this stage the hyperspherical cross section was no more intersecting the T_1/S_0 intersection subspace. *The final NCP structure, corresponding to a ^{3}n-π^*/S_0 crossing, was located 0.228 au from and 108 kcal mol^{-1} higher in energy with respect to the reference T_1 planar minimum, Its geometrical parameters are reported*

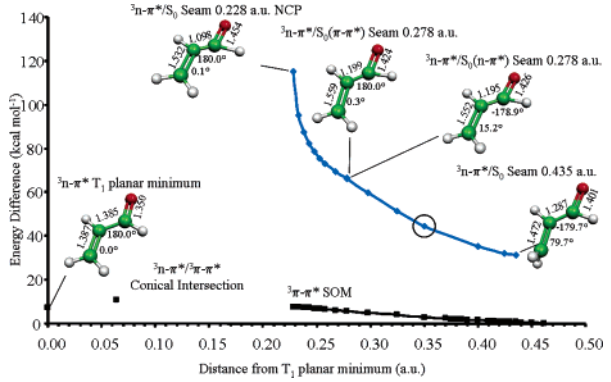


Figure 5. Acrolein, ^{3}n-π* T$_1$ planar minimum, ^{3}n-π*/S$_0$ NCP, ^{3}n-π*/$^3\pi$-π* conical intersection,[30] ^{3}n-π*/S$_0$ seam (full diamonds, blue line), $^3\pi$-π* SOM (series of minima, full squares, black line). All geometrical parameters are in angstrom and degrees. The optimization of the circled point of the ^{3}n-π*/S$_0$ seam at 0.35 au from the T$_1$ planar minimum is the one analyzed in Figure 4. The SOM points represent a path in the $^3\pi$-π* valley, even if they were not calculated as a rigorous IRC or steepest descent path. See section 3.6.

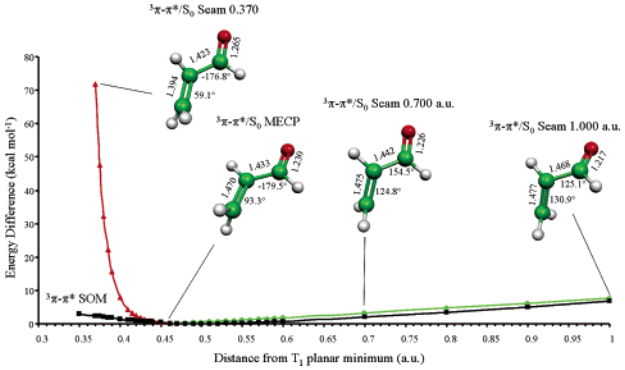


Figure 6. Acrolein, $^3\pi$-π*/S$_0$ seam (full triangles, red line plus full circles, green line), and $^3\pi$-π* SOM (full squares, black line). All geometrical parameters are in angstrom and degrees.

in Figure 5. Given the energy difference with the reference planar minimum and the unusual structure (see the C−C distance of 1.098 Å), it is very unlikely that the NCP structure could be physically reached.

The location of the NCP to a given structure may be important when investigating a photochemical process. In fact, it can give an idea of the accessibility of a certain crossing seam from a previously optimized excited-state intermediate (an energy minimum) or excited-state reaction path (a steepest descent path). *In the future, an algorithm to perform a simultaneous minimization of r_1, r_2, and of the radius (R) will be implemented, to provide a tool to locate, in an unbiased way, the geometrically NCP to a given structure.*

3.5. Mapping the Intersection Subspace: the ^{3}n-π*/S$_0$ Seam. The methodology used in section 3.4 can also be used to construct a low-lying ^{3}n-π*/S$_0$ cross section of the *N*-1-dimensional ^{3}n-π*/S$_0$ intersection space that originate at NCP. Taking the T$_1$ planar minimum as the hypersphere center we run a series of two-constraint (r_1 and r_2 as seen above) optimizations setting a hypersphere radius longer than 0.228 au and pointing in the direction of the optimized NCP. The optimized structure was then used as the guess geometry for the next optimization at an even longer radius and maintaining the planar minimum as the hypersphere center. The ^{3}n-π*/S$_0$ seam reported in Figure 5 was built in this way (blue line).

Notice that at the beginning of the seam the optimized structures are planar, showing only stretching distortions: C═C and C═O lengthening and C−C shortening, with respect to the T$_1$ planar minimum. Around 0.278 au twisting around C═C begins, along with inversion of the stretching movements. Actually, two different structures with nearly degenerate energies where found at 0.278 au distance. A change in the S$_0$ wave function, from a partial $\pi-\pi$* character to a partial n-π* character is likely to be at the basis of such a behavior. More specifically, it could happen that a ^{1}n-π*/$^1\pi-\pi$* seam crosses the located singlet/triplet seam. The ^{3}n-π*/S$_0$ seam ends (approximately) at 0.435 au as at longer distances the optimization leads, as explained below, to a $^3\pi-\pi$*/S$_0$ seam, that is lower in energy.

3.6. Mapping the Intersection Subspace: the $^3\pi$-π*/S$_0$ Seam. The existence of crossing points with $^3\pi-\pi$*/S$_0$ character (e.g. the MECP documented in section 3.4) and located lower in energy than any point of the ^{3}n-π*/S$_0$ seam documented in section 3.5 suggests that the low-energy region of T$_1$ must be dominated by a $\pi-\pi$* character. This is confirmed by relaxation of the ^{3}n-π*/S$_0$ seam point on T$_1$. In fact, a set of low-lying $^3\pi-\pi$* point (SOM) were located performing a series of standard T$_1$ geometry optimizations on hyperspherical cross sections (as discussed in section 3.2) centered on the usual T$_1$ planar minimum and starting at each ^{3}n-π*/S$_0$ seam point (i.e. using the same radius and hypersphere center of the seam). The optimized points are shown in Figure 5 (black line). Clearly, the SOM represents a sort of bottom of the T$_1$ valley closest to the computed ^{3}n-π*/S$_0$ seam. In other words, SOM is a scan of the T$_1$ potential energy surface at various distances from the T$_1$ planar minimum. Since it is a scan, it may not have the same properties of an IRC or MEP. Remarkably, it was found (via orbital occupancy analysis) that the T$_1$ SOM has a $^3\pi-\pi$* nature. The low-lying regions of the T$_1$ potential energy surface showing a $^3\pi-\pi$* and ^{3}n-π* character (as the ^{3}n-π* planar minimum in Figure 5) must be somehow connected. In Figure 5 we report the energy of a previously located ^{3}n-π*/$^3\pi-\pi$* conical intersection[30] that may provide a link between these two T$_1$ regions.

The existence of a low-lying $^3\pi-\pi$* region suggests the existence of a low-lying $^3\pi-\pi$*/S$_0$ seam. To locate such a seam we employed the same procedure used to locate the NCP point (see section 3.4) but starting with a longer radius. In fact, the computations were started at the structure of MECP (at 0.462 au from the usual hypersphere center) and performing iteratively a series of two constraints (r_1 and r_2) minimizations of the T$_1$ energy with a decreased or increased radius. The seam, reported in Figure 6 along with the T$_1$ SOM, is divided into two branches, characterized by shorter or longer *R* values. The branch on the left of the figure (red line) is characterized by twisting of the C═C bond mixed with C═C and C−C lengthening and C═O shortening. The branch on the right side (green line) is substantially an

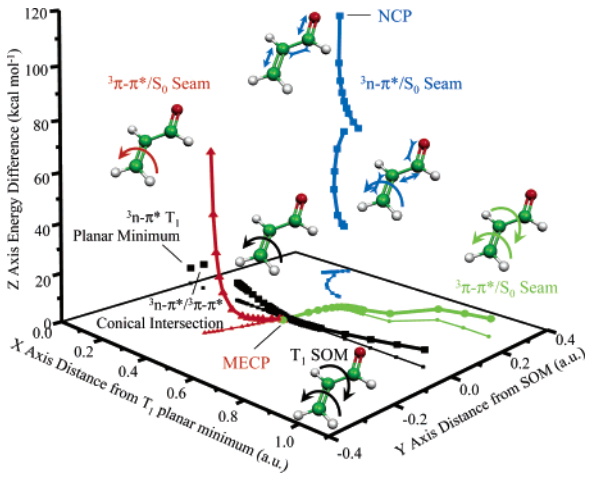


Figure 7. Acrolein, general view of the calculated seams, with relative distance from the corresponding points of the SOM. Arrows on the molecular structures represent the main movement characterizing each branch of the seams or the SOM. As stated in section 3.5, the ^{3}n-π^*/S_0 intersection seam is split in two parts, each characterized by different skeletal movements. The two degenerate structures at 0.278 au from the T_1 planar minimum do not have the same distance from SOM, originating two separate branches.

extension of the previous one both from the geometrical and electronic structure point of view. However, this branch features an additional twisting about the C−C bond, which leads to quite distorted structures. Notice the SOM curve and the right branch of the seam go along nearly in parallel and close to each other, being the energy difference between them never more than 1.5 kcal mol^{-1}. The point of the SOM at 0.490 au from the T_1 planar minimum has the lowest energy of all the computed structures. The energies of all structures reported in Figures 5−7 were calculated as difference with respect to that of this structure.

The difference in energy is not the only important parameter to be considered when comparing seams and valleys (even if SOM does not rigorously correspond to a valley). In fact, the difference in geometry provides information on the vibrational modes that most effectively could lead from the T_1 valley to the seam and, therefore, to intersystem crossing. The computed distances between the equivalent (i.e. featuring the same radius R from the reference T_1 planar minimum) points of the seams and SOM are summarized in Figure 7. The seams as well as the SOM follow the same coloring/symbol as in Figures 5 and 6 and the distance from the T_1 planar minimum is reported on the X axis. Similarly, the distances of the seam points from the corresponding SOM points are given on the Y axis. The projection on the XY plane is also reported, and it is indicated by smaller symbols. The energy difference is on the Z axis. As can be seen, being close in energy does not imply being close geometrically. It is also interesting to note that our results suggest that only the crossing points located in the vicinity of the MECP point could be vibrationally accessible at room temperature from the T_1 state. A more quantitative evaluation of such accessibility including quantum effects would require a more rigorous and expensive procedure and is out of the scope of this paper.

4. Conclusions

A modified algorithm based on the approach of Anglada and Bofill has been implemented and tested. The results indicate that such implementation can be used with success in different cases of constrained geometry optimizations. It has been shown that geometry optimizations subject to both geometrical and energy constraints are efficient. Most remarkably, we have provided evidence that geometry optimization in the presence of multiple and heterogeneous constraints (i.e. simultaneous geometrical and energy constraints) is also efficient. This last features allows for the mapping of very different and complex seams of intersection relating electronic states of different electronic nature with the ground state. In the present paper, only crossings between singlet and triplet state surfaces have been optimized. Nevertheless, the developed method is generally valid for any kind of constraint. Work on exploring the new optimization tools in connection with conical intersection optimization problems is in progress.

Acknowledgment. The authors wish to thank the Università di Siena (Progetto di Ateneo A.A. 02/04), the Swedish Research Council (VR), and the Swedish Foundation for Strategic Research (SSF).

References

(1) Gill, P. E.; Murray, W. In *Numerical Methods for Constrained Optimization*; Gill, P. E., Murray, W., Eds.; Academic Press: London, 1974.

(2) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514−1527.

(3) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363. Gonzales, C.; Schlegel, H. B. *J. Chem. Phys.* **1991**, *95*, 5853.

(4) Koga, N.; Morokuma, K. *Chem. Phys. Lett.* **1985**, *119*, 371.

(5) Bearpark, M. J.; Robb, M. A.; Schlegel, H. B. *Chem. Phys. Lett.* **1994**, *223*, 269−274.

(6) Fletcher, R. *Practical Methods of Optimization: Constrained Optimization*; Wiley & Sons: Chichester and New York, 1981; Vol. 2.

(7) Mana, M. R.; Yarkony, D. R. *J. Chem. Phys.* **1993**, *99*, 5251. Yarkony, D. R. *J. Chem. Phys.* **1993**, *97*, 4407.

(8) Farazdel, A.; Dupuis, M. *J. Comput. Chem.* **1991**, *12*, 276.

(9) Bofill, J. M.; Anglada, J. M. *Theor. Chem. Acc.* **2001**, *105*, 463−472. Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580−5601.

(10) Shanno, D. F. *Math. Comput.* **1970**, *24*, 647. Goldfarb, D. *Math. Comput.* **1970**, *24*, 23. Fletcher, R. *Comput. J.* **1970**, *13*, 317. Broyden, C. G. *J. Inst. Math. Appl.* **1970**, *6*, 222.

(11) Anglada, J. M.; Bofill, J. M. *J. Comput. Chem.* **1997**, *18*, 992.

(12) AMPAC program, local version, extended by D. Liotard, 1987.

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

(14) Baker, J.; Bergon, D. *J. Comput. Chem.* **1993**, *14*, 1339. Baker, J. *J. Comput. Chem.* **1992**, *13*, 240. Baker, J. *J. Comput. Chem.* **1993**, *14*, 1085.

(15) Baker, J. *J. Comput. Chem.* **1997**, *18*, 1079.

(16) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. *J. Phys. Chem.* **1985**, *89*, 52.

(17) Lindh, R.; Bernhardsson, A.; Schütz, M. *Chem. Phys. Lett.* **1999**, *303*, 567−575.

(18) Lindh, R.; Bernhardsson, A.; Karlström, G.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **1995**, *241*, 423−428.

(19) Andersson, K.; Barysz, M.; Bernhardsson, A.; Blomberg, M. R. A.; Carissan, Y.; Cooper, D. L.; Fülscher, M. P.; Gagliardi, L.; De Graaf, C.; Hess, B. A.; Hagberg, D.; 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; Lund University, Sweden, 2004.

(20) Veryazov, V.; Widmark, P.-O.; Serrano-Andrés, L.; Lindh, R.; Roos, B. O. *Int. J. Quantum Chem.* **2004**, *100*, 626−635. 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*, 222−239.

(21) Fletcher, R. *Practical Methods of Optimization*; Wiley: New York, 1981.

(22) We chose the transformation matrix **T** to be unitary, even if it is not strictly necessary, so not to have to include the computation of the inverse matrix, which is anyway needed. By the way, we note that Anglada and Bofill implicitly affirm that **T** should be unitary when stating that $\mathbf{T}_b^T\mathbf{T}_b = \mathbf{I}_b$. By consequence, their statement that $[\mathbf{R}(\mathbf{r})]^T\mathbf{T}_b = \mathbf{I}_b$ is not correct, given the fact that $\mathbf{R}(\mathbf{r})$ is not, in general, unitary.

(23) Widmark, P.-O.; Malmqvist, P.-Å.; Roos, B. O. *Theor. Chem. Acc.* **1990**, *77*, 291−306. Roos, B. O.; Lindh, R.; Malmqvist, P.-Å.; Veryazov, V.; Widmark, P.-O. *J. Phys. Chem. A* **2004**, *108*, 2851−2858.

(24) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(25) Helbing, J.; Bregy, H.; Bredenbeck, J.; Pfister, R.; Hamm, P.; Huber, R.; Wachtveil, J.; De Vico, L.; Olivucci, M. *J. Am. Chem. Soc.* **2004**, *126*, 8823−8834.

(26) Müller, K.; Brown, L. D. *Theor. Chem. Acc.* **1979**, *53*, 75−93.

(27) Garavelli, M.; Page, C. S.; Celani, P.; Olivucci, M.; Schmid, W. E.; Trushin, S. A.; Fuss, W. *J. Phys. Chem. A* **2001**, *105*, 4458−4469.

(28) Migani, A.; Sinicropi, A.; Ferré, N.; Cembran, A.; Garavelli, M.; Olivucci, M. *Faraday Discuss.* **2004**, *127*, 179−191. Migani, A.; Robb, M. A.; Olivucci, M. *J. Am. Chem. Soc.* **2003**, *125*, 2804.

(29) Schuster, D. I. In *The chemistry of enones*; Petai, S., Rappoport, Z., Eds.; John Wiley and Sons: Chichester, U.K., 1989; Vol. 2, pp 693−756. Schuster, D. I. In *Rearrangements in ground and exited states*; de Mayo, P., Ed.; Academic Press: London, 1980; Vol. 3, pp 161−279.

(30) Reguero, M.; Olivucci, M.; Bernardi, F.; Robb, M. A. *J. Am. Chem. Soc.* **1994**, *116*, 2103-2114.

CT0500949

JCTC Journal of Chemical Theory and Computation

Open Shell Dianions Likely To Be Stable in the Gas Phase with Respect to Autoionization

Carl Trindle*

Department of Chemistry, University of Virginia, Charlottesville Virginia 22904

Ayse Yumak

Department of Physics, Marmara University, Istanbul Turkey

Received December 21, 2004

Abstract: We address the challenge set by Dreuw and Cederbaum [Dreuw, A.; Cederbaum, L. S. *Chem. Rev.* **2002**, *102*, 181−200] to develop a set of "small" open shell stable dianions. We offer two families of such species, based on a central diradical center with attached anionic sites. Both families achieve dianion stabilization through alternating zones of positive and negative charge. First, quasi-linear systems such as $NB(C_2)_n-Q-(C_2)_nBN$ become diradical dianions stable to autoionization in two cases: (a) for Q a divalent (carbene) carbon when *n* is two or greater and (b) for Q a C_4 ring diradical when *n* is one or greater. Second, carbenes with certain anionic aromatic substituents can be stable with respect to autoionization. π-Acid substituted carbenes (A_2Q) have triplet ground states generally. If A is cyclopentadienyl anion stabilized by cyano substitution, the resulting triplet dianion is stable with respect to autoionization. In bisphenyl carbenes the triplet is relatively stabilized when ortho substituents force the system toward D_{2d} symmetry. The dianion of bis(orthochlorophenyl) carbene produced by para-substitution with BN retains the triplet preference and is stable with respect to autoionization. These results obtained first by density functional calculations in a small basis, B3LYP/6-31G(d), are confirmed and extended by ROMP2 and CCSD calculations in a more flexible basis, 6-31+G(d). DFT has a systematic bias against systems with excess electrons, which is paradoxically a virtue in the screening of candidate dianions since the DFT calculation does not make false predictions of stability.

Introduction

It seems quite implausible that small dianionic species would persist without stabilizing effects of a medium—one would think that the intense repulsion arising from the excess negative charge would preclude any extended lifetime. Nonetheless the study of small dianions long-lived in the gas phase has advanced strikingly both in experimental and theoretical directions since dianionic alkaline earth tetrahalides such as MgF_4^{2-} were first predicted to be stable to electron loss and dissociation in the gas phase.[1] Now the types of dianions known to be stable in these ways have come to include the linear clusters of alternating charges studied by Simons and Boldyrev[2] and also include a substantial number of variants on the center-ligand sphere in which concentric shells alternate in charge as established by Dreuw and Cederbaum.[3] Other means of stabilization of dianionic systems are known, the simplest of which are the separation of negatively charged sites[4] or the dispersal of the excess charge over large volumes. Substituent-stabilized aromatic dianionic systems have been characterized computationally, most notably derivatives of B_6H_6, e.g. the *closo* $B_6(CN)_6^{2-}$ devised by Zink, Dreuw, and Cederbaum.[5] All these dianionic sytems are closed shell singlet species. Are there also stable open shell dianionic species? It seems unlikely that the study of the triplet states of those dianions known to be stable in their closed shell ground states would

10.1021/ct049845h CCC: $30.25
Published on Web 07/15/2005

be rewarding, since the triplet would be an excitation from a ground state itself already marginal in stability. Stable dianionic species with open shell ground states might be expected to be even more elusive than stable closed shell dianionic systems, since few chemically stable neutral systems have open shell ground states. The task of devising a "small" open-shell dianionic molecule stable to autoionization over the transit time in a mass spectrometer has been set.[6] It is among those few systems which have open shell ground states, specifically carbenes and small rings, that we search. They can serve as cores of molecules bearing substituents capable of accommodating excess charges. Here we offer some triplet ground-state systems which according to our computational modeling should be stable with respect to autoionization.

Computational Methods

We employ B3LYP/6-31G(d) calculations throughout this work, as implemented in the Gaussian 03 suite.[7] This model chemistry makes feasible the investigation of the large systems at issue here but is open to objection on two grounds. The basis seems inadequate to represent loosely bound electrons, lacking diffuse functions. The density functionals now available are not capable *in principle* of representation of anions.[8] Their failure to cancel self-repulsion of electrons would be expected to produce a systematic underbinding of excess electrons. Careful investigations of electron affinities of neutral systems show that the case *in practice* is not bleak; accuracies of ca. 0.2 eV are routine for B3LYP in even modest basis sets.[9] It seems likely that a fortunate error cancellation is at work. The finite basis (lacking continuum or even extensive diffuse components) keeps weakly bound electrons localized at a molecule even when autoionization should be spontaneous. The finite-basis flaw produces a bias in opposition to the DFT underbinding and the errors may tend to cancel in the final estimate of stability of the anion. No such empirical reassurance is available for the electron affinity of monoanions (i.e., the ionization energy of dianions), which are less thoroughly studied. However the choice of B3LYP/6-31G(d) as a model chemistry for closed shell dianions has been tested by comparison to R(O)HF and MP2 calculations in more flexible basis sets and has proved to be a useful semiquantitative guide to the stability of singlet dianions[10] and also to the stability of neutral triplet carbenes.[11] In the following discussion we employ MP2/6-31+G(d) and in some cases CCSD/6-31+G(d) model chemistries, especially to provide an independent view of the dianions predicted in the simplest DFT calculations to be stable with respect to autoionization.

Design Stratagem

The principles guiding the design of stable dianions, mentioned above and summarized in Table 1, have produced a variety of closed shell singlet dianions stable both with respect to autoionization and unimolecular dissociation and are helpful in the search for an open shell stable dianion. Our design focuses on linear and quasi-linear systems, with the general three-fragment structure anion site−spacer−open-shell center−spacer−anion site. The spacers are always CC fragments, the anionic sites are C_2, BN, or BeO, and the open shell centers are carbene carbons C or cyclobutyl diyl rings CB, which have figured largely in the search for high spin building blocks for magnetic materials. We also investigate the properties of aryl-substituted carbenes, where the anionic site may be a substituent on an aryl ring or the cyclopentadienyl anion. In each type of systems we can find a candidate for the sought open shell dianion stable with respect to autoionization. In this report we define adiabatic ionization energies of the dianions but do not explore details of the dissociation pathways which produce product monoanions. Neither do we explore the repulsive coulomb barrier,[12] which can lend an extended (tunneling) lifetime to metastable dianions.

Table 1. Ways To Stabilize Molecular Dianions

mode of stabilization	example
separation of charges	long-chain dicarboxylates[4]
alternating charges (spherical shells)	MX_4, $O(BN)_4$[20]
alternating charges (linear)	XMXMX [2]
aromaticity or closed-shell stabilization and substituent effects	B_6H_6 and $B_6(CN)_6$[6]

Linear Systems I: Carbene Core R−(CC)$_n$-C−(CC)$_n$−R

The simplest X-core-X system is derived from the family of carbenes $NC(C_k)C(C_k)CN$ studied by Belbruno et al.[13] and Blanksby et al.[14] In analogy to these well-characterized systems we consider the isoelectronic dianions $NB-(C_2)_nC-(C_2)_n-BN$ and variants with anionic sites CC and BeO. Watts and Bartlett[15] studied carbon chain dianions by CCSD methods and suggest C_9^{2-} may be stable with respect to autoionization. Calabrese et al.[16] reported beam evidence for C_n^{2-}, $n = 7-13$, with the species with $n = 11$ most abundant but clear signals from dianions with $n = 7$, 9, and 13. Table 2 shows the computed crossover from autoionizing to stable dianion as n increases. For all n, BeO is the most effective in stabilizing excess negative charge and C_2 is least effective. As the size of the system increases the dianion takes on diradical character. Only for the smallest system is the closed shell singlet favored over the broken symmetry single-determinant unrestricted open shell "singlet." That wave function has an expectation value of S^2 near 1.0, typical of an equal admixture of singlet and triplet states. For systems with n greater than zero the triplet state is favored over a broken symmetry open shell "singlet" state by a small amount. In the limit characteristic of diradicals, i.e., identical geometry and orbitals in singlet and triplet states, we can take this energy difference to be twice the singlet−triplet gap. While no system we studied reached this limit, systems with the BeO anionic site are closest.

We expect enhanced stability of the dianion as n increases since the mean separation between anionic sites increases. When $n = 3$ all dianionic species are stable relative to the doublet monoanion produced by autoionization. The DFT calculation predicts triplet ground states for all dianions; the energy differences between symmetry breaking "singlet" and unrestricted triplets approach about 5 kcal/mol. This leads us to estimate the singlet−triplet gap as about 10 kcal/mol.

Table 2. Favored State of Dianions and Their Stability Relative to the Anion Produced by Autoionization, for $R-(C_2)_n-C-(C_2)_n-R$[a]

n-value	0	0	0	1	1	1	2	2	2	3	3	3
sink	CC	BN	BeO	CC	BN	BeO	CC	BN	BeO	CC	BN	BeO
B3LYP/6-31G(d)	−93	−66	−47	−35	−17	−6	−5	9	17	15	27	34
B3LYP/6-31+G(d)	−61	−38	−20	−16	1	13	11	24	34	28	40	
ROMP/6-31+G(d)	−65	−44	−14	−21	−9	15	8	40	37			
CCSD/6-31+G(d)	−66	−40	−17	−16	4	19	x	28	42			

[a] Adiabatic ionization energies for carbene-based systems, in kcal/mol. Positive values designate endothermic ionization energies and indicate that the triplet dianions are stable with respect to autoionization.

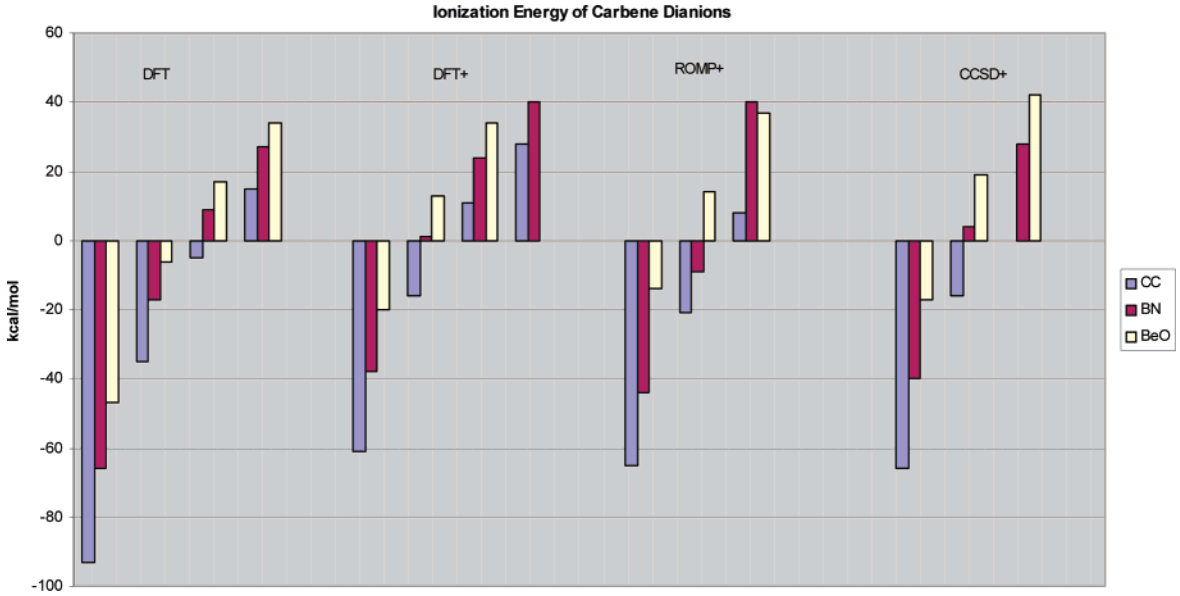


Figure 1. Adiabatic ionization energies of carbene-based dianions in kcal/mol. Labels: DFT = B3LYP/6-31G(d), DFT+ = B3LYP/6-31+G(d), ROMP+ = ROMP/6-31+G(d)//UMP2/6-31+G(d), CCSD+ = CCSD/6-31+G(d)//UMP2/6-31+G(d).

While it is the case that the terminal anionic sites each bear a local negative charge near the unit value, the interior is not uniformly neutral. Rather there is a marked charge alternation in the CC chain which might be considered a polarization in response to the marked negative charges at the termini. This would complicate, but not preclude, estimates of the coulomb barrier which rely on modeling the charges as well separated.[17]

The patterns discernible in the B3LYP/6-31G(d) model chemistry are also present in the alternative (and, in principle, superior) ROMP2/6-31+G(d)//UMP2/6-31+G(d) and CCSD/6-31+G(d)//UMP2/6-31+G(d) calculations. There is considerable spin contamination in the UMP2 results; ROMP2 (for which one has a pure spin state) gives predictions closely comparable with UCCSD. There are however significant (though not surprising) differences in the pictures provided by the density functional and wave function methods. In keeping with the expected underbinding of excess electrons expected for density functional methods, the stability of the dianions relative to monoanions is always underestimated by B3LYP compared to ROMP and CCSD, by roughly 20 kcal/mol. This is far from the accuracy claimed for electron affinities of medium-sized neutral molecules, 0.2 eV. One might kindly say that the small-basis DFT results are "conservative"; B3LYP/6-31G(d), systematically biased against excess charge, produces no "false positive" predictions of dianion stability. Every DFT prediction of dianion stability made by B3LYP/6-31G(d) is confirmed by ROMP/- and CCSD/6-31+G(d) estimates. However a small-basis DFT-based survey will overlook some dianions that may be stable with respect to autoionization. We note ROMP/6-31+G(d) and CCSD/6-31+G(d) calculations suggest that OBe(CC)C-(CC)BeO dianion might be stable relative to the monoanion, while B3LYP/6-31G(d) makes the opposite prediction. The obvious step of enhancing the basis set used in the DFT calculations has a strikingly good effect. The results of B3LYP/6-31+G(d) description of the carbene-based systems are in surprising agreement with the wave function based and correlation-corrected ROMP and CCSD calculations in the same basis. It does not appear that the systematic flaw of the functionals in use, which should produce systematic underbinding of excess electrons, has any practical impact here. Figure 1 shows the systematic error in B3LYP/6-31G-(d) relative to calculations in the basis augmented with diffuse functions, i.e., 6-31+G(d). A general consistency obtains among B3LYP, ROMP, and CCSD calculations in this basis.

Linear Systems II: Cyclobutadiyl Core, $R-(CC)_n$-CB-$(CC)_n-R$

Some small systems in the substituted cyclobutadiyl dianion family show imaginary frequencies in D_{2h}, but larger systems have a D_{2h} minimum energy structure. For the sink BeO the triplet and "singlet" dianion for $n = 1$ is predicted in B3LYP/

Table 3. Favored B3LYP/6-31G(d) State of Dianions and Their Stability Relative to the Anion Produced by Autoionization, for R−$(C_2)_n$−Cyclobutadiyl−$(C_2)_n$−R[a]

n-value	0	0	0	1	1	1	2	2	2
sink	CC	BN	BeO	CC	BN	BeO	CC	BN	BeO
B3LYP/6-31G(d)	−71	−132	−39	−27	−13	5	34	10	20
B3LYP/6-31+G(d)	−46	−7	−8	26	4	13		54	60
ROMP/6-31+G(d)	−47	−9	−10	25	34	46	19[b]	35[b]	72[b]
CCSD/6-31+G(d)	−50	−9	−20	x	37	44			

[a] Adiabatic ionization energies for cyclobutadiyl-based systems, in kcal/mol. Positive values designate endothermic ionization energies and indicate that the triplet dianions are stable with respect to autoionization. [b] The ROMP entries for $n = 2$ refer to ROMP/6-31+G(d)//UHF/6-31+G(d).

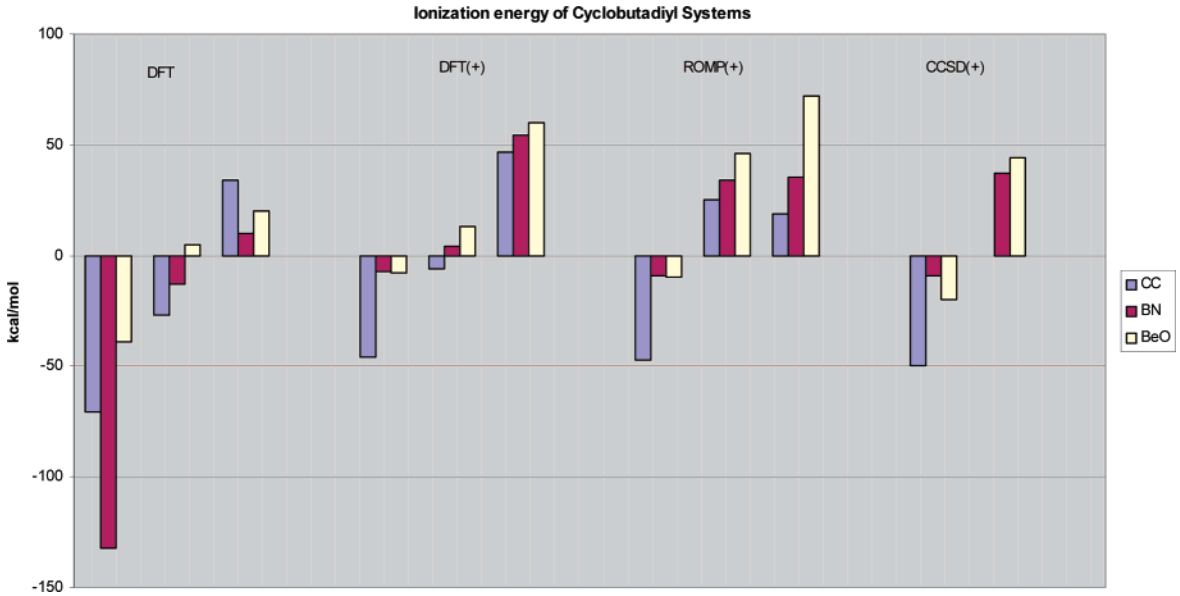


Figure 2. Adiabatic ionization energies of cyclobutadiyl-based dianions in kcal/mol. Labels: DFT = B3LYP/6-31G(d), DFT+ = B3LYP/6-31+G(d), ROMP+ = ROMP/6-31+G(d)//UMP2/6-31+G(d), CCSD+ = CCSD/6-31+G(d)//UMP2/6-31+G(d).

6-31G(d) to be stable with respect to autoionization, as is the case for all dianionic species with $n = 2$. (See Table 3.)

The cyclobutadiyl system is antiaromatic and thus should either prefer an open shell ground state or spontaneously distort from 4-fold symmetry to achieve a stable closed shell state. The cyclobutadiyl system could perhaps favor the open shell singlet, which is an intrinsically multiconfigurational state and hence not strictly representable by a UHF symmetry breaking single determinant wave function. Neither is this state strictly represented by the analogous broken symmetry DFT calculation reported here. The observed preference for this broken-symmetry singlet over the triplet in the larger systems is an interesting outcome, worth testing by independent methods. In these systems we are close to the diradical limit for which the energy gap between UHF "singlet" and triplet is about twice the value predicted by the simplest two-configuration representation of the open-shell singlet.

Ab initio MP2/6-31+G(d) (and where available, CCSD/6-31+G(d)) representations of the cyclobutadiyl-based dianions confirm that dianions with $n = 2$ are stable with respect to autoionization and also suggest that triplets with $n = 1$ are stable with respect to autoionization. These methods place the triplet state lower than any open or closed shell singlet. Unrestricted MP2 states are again strongly spin-contaminated—projection makes a large difference in energy and even increases the expectation value of S^2. Owing to the erratic behavior of the dianion−monoanion energy gaps predicted by unrestricted and projected-unrestricted MP2, we rely on ROMP results. These are assured to have pure spin states and produce energies consistent with available CCSD values. For the largest cyclobutadiyl-based systems (n=2) we used UHF-optimized geometries rather than the UMP2-optimized structures, the basis for all other estimates. Comparison for the $n = 2$ BN system was possible and showed that although there are appreciable differences in details of the geometry, the dianion−monoanion energy gaps are similar. The mean difference between wave function results in the augmented basis and B3LYP/.6-31G(d) results for the dianion-monoanion energy gap is systematic (DFT underbinds dianions) and large (roughly 40 kcal/mol). Enhancing the basis set employed in DFT to 6-31+G(d) does bring some improvement, but this is not so impressive as what we found for the carbene-based systems. We may observe that DFT estimates of stability of dianions in this structural class are conservative regardless of the basis, in the sense that all predictions of dianion stability made in DFT are confirmed by ab initio methods. As we found for the carbene systems, DFT does not produce "false positives". Figure 2 displays estimates of the adiabatic ionization energies and shows the systematic error in B3LYP/6-31G-(d) estimates relative to those obtained in an augmented

Table 4. Bisaryl Dianions in B3LYP/6-31G(d)[a]

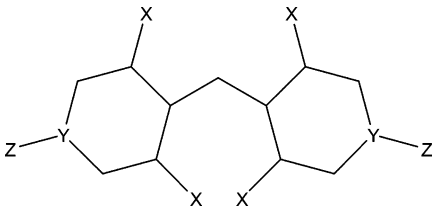
X	Y	Z	singlet angle	triplet angle	doublet angle	AEA kcal/mol
H	B	H	121	144	144	−68
H	B	F	121	143	143	−71
H	B	CN	121	146	143	−39
H	C	BN	120	142	139	−7
o-Cl	B	H	121	158	157	−42
o-Cl	B	F	129	154	155	−46
o-Cl	B	CN	132	178	159	−17
o-Cl	**C**	**BN**	**133**	**170**	**180**	**8**
perfluoro	B	F	124	144	145	−47
Candidate in ROMP2/6-31+G(d)//UHF/6-31+G(d)						
o-Cl	**C**	**BN**		**156**	**155**	**22**

[a] The entry in bold identifies an open shell dianion which is stable with respect to autoionization.

6-31+G(d) basis. For this system, in contrast to the perfectly linear carbene-based dianions, B3LYP underbinds dianions relative to ROMP and CCSD even in the augmented basis.

Bisaryl Carbenes

Bisphenyl carbenes are known to favor a triplet ground state,[18] so we constructed a series of dianions following the structural theme (aryl ring)−methylene−(aryl ring). We transformed the known neutral bisphenyl species to dianions by substituting B for C in the ring or by attaching a BN fragment at the para position. (See Table 4.)



The longest-lived bisaryl triplets are ortho-substituted with bulky CF_3, Br, or Cl groups.[19] This substitution has the effect of opening the central angle and orienting the ring systems orthogonal to one another. Attending this structural change is a shortening of the central CC bonds and considerable delocalization of the spin into the aryl rings. Fluoro-substitution (even perfluororination) is ineffective in inducing this major structural change, but ortho-chloro substitution accomplishes it. In the most extreme cases with near-linear central CCC angles, these systems approach the diradical limit, and an open shell singlet is close in energy and structure to the ground state triplet. The only system in this structural class we have found thus far to be autoionization stable according to B3LYP/6-31G(d) is such a linearized diradical, bis(2,6-dichloro-4-BN-phenyl) carbene. The DFT prediction is confirmed by ab initio estimates in ROMP2/6-31+G(d)//UHF/6-31G(d). The geometry differences between DFT and ab initio representations of these bisaryl systems have been discussed elsewhere.[11]

Biscyclopentadienyl Carbenes

The family of bisaromatic carbenes A−C−A could contain a dianion if the six-π-electron cyclopentadienyl anion is used as A. The unsubstituted species is liable to autoionization and has a closed shell singlet ground state. (See Table 5.) The perfluorinated dianion likewise has a closed shell ground state, but the monoanion is still more stable. Replacement of fluorine by the π-acid CN produces a triplet dianion stable with respect to autoionization. This is confirmed by an approximate ab initio calculation using the model chemistry MP2/6-31+G(d)//UHF/6-31+G(d). Spin contamination is minor in these systems. Related species bis(2,5-dicyanocyclopentadienyl)carbene dianion and bis(3,4-dicyanocyclopentadienyl)carbene dianion are not predicted to be autoionization-stable in this model chemistry.

The fully CN-substituted system has a broken symmetry UB3LYP/6-31G(d) "singlet" (expectation value of $S^2 = 1.0297$) at −1161.938014 Hartrees, about 4.1 kcal/mol below the triplet. Its geometry is similar to that of the triplet (1.3508 Å, 179.9 deg), so we may consider the stable open shell species to have a degree of diradical character, comparable with other long-lived triplet carbenes.

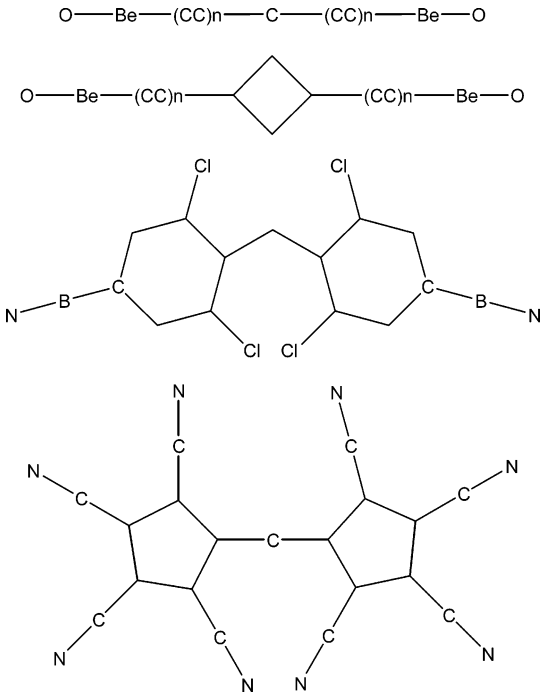
Conclusions

We present three kinds of open shell autoionization stable dianions, based on a simple structural theme. Chart 1 illustrates the structural types of the stable dianions: (1) systems built from diatomic anion centers connected by the methylene triplet linker; (2) systems built from diatomic anion centers connected by the cyclobutadiyl open shell singlet linker; and (3) systems built from diatomic anion centers connected by the bisphenyl carbene triplet linker.

Table 5. Variations on the Structure Cyclopentadienyl Anion−Carbene Carbon−Cyclopentadienyl Anion

substituent	(unsubstituted)	perfluoro	percyano	percyano
model chemistry	B3LYP/6-31G(d)	B3LYP/6-31G(d)	B3LYP/6-31G(d)	ROMP/6-31+G(d)//UHF/6-31+G(d)
		Dianions		
central bond length	1.4387	1.4255	1.4059	
central angle	120	124	124	
		Triplet		
central bond length	1.4118	1.4004	1.3592	1.4068
central angle	144	144	177	144
		Monoanion Doublet		
central bond length	1.3727	1.3690	1.3592	1.3694
central angle	143	145	177	147
gap (kcal/mol)	−102	−47	14	14

Chart 1. Structural Types of Triplet Dianions Stable with Respect to Autoionization



The triplet dianion biscyclopentadienyl methylene also appears to be stable with respect to autoionization, if the cyclopentadienyl ring is fully substituted by the stabilizing CN π-acid.

The challenge to devise open-shell dianionic systems potentially stable with respect to autoionization has been met insofar as calculations in a modest basis and in the B3LYP variant of DFT can be trusted. Comparison with MP2 and CCSD results obtained with a more flexible 6-31+G(d) basis show that the bias against dianions in these B3LYP/6-31G-(d) calculations is severe. This may be an advantage in the search for stable dianions however, since such a method will not yield false positives.

The systems described here include a number of dianionic species which may be stable with respect to autoionization. They are however not "small" as was specified in the challenge to identify small stable open-shell dianions. They should be considered only a step toward the study of much smaller systems. Smaller molecules can be more reliably characterized, and identification of stable smaller systems would be more responsive to the intention of the challenge.

Acknowledgment. This work was supported by the University of Marmara, Istanbul, The Turkpetrol Foundation, and the University of Virginia. We thank the Body Foundation for equipment, travel support, and general encouragement and wish to acknowledge the key role in making this collaboration possible of Professor Dr. Zikri Altun of the Physics Department, the University of Marmara, Istanbul.

Supporting Information Available: Detailed energies and structures of all species. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Weikert, H. G.; Cederbaum, L. S.; Tarantelli, F.; Boldyrev, A. I. *Z. Phys. D-Atoms, Molecules, Clusters* **1991**, *18*, 229.

(2) Boldyrev, A. I.; Simons, J. J. *Chem. Phys.* **1993**, *98*, 474.

(3) Dreuw, A.; Cederbaum, L. S. *Chem. Rev.* **2002**, *102*, 181−200.

(4) Maas, W. P. M.; Nibbering, N. M. M. *Int. J. Mass Spec. Ion Proc.* **1989**, *88*, 257

(5) Zint, N.; Dreuw, A.; Cederbaum, L. S. *J. Am. Chem. Soc.* **2002**, *124*, 4910.

(6) Reference 2 and also Dreuw, A.; Cederbaum, L. S. Free stable and long-lived multiply charged anions. In *Theoretical Prospect for Negative Ions*; Kalcher, J., Ed.; Research Signpost: Trivandrum, Kerala, India, 2002.

(7) Gaussian 03, Revision B.2, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Pittsburgh, PA, 2003.

(8) Koch, W.; Holthausen, M. C. *A Chemist's Guide to density Functional Theory*, 2nd ed.; Wiley-VCH: 2001; sections 6.7 and 9.5.

(9) Brown, S. T.; Rienstra-Kiracofe, J. C.; Schaefer, H. F. III. *J. Phys. Chem. A* **1999**, *103*, 4065.

(10) Trindle, C.; Yumak, A. *J. Chem. Theory Comput.* **2005**, *1*, 433.

(11) Trindle, C. *J. Org. Chem.* **2003**, *68*, 9669.

(12) Dreuw, A.; Cederbaum, L. S. *Phys. Rev. A* **2000**, *63*, 012501 and 049904(E) — C_n^{2-} for n = 2, 4, 6, 8. Dreuw, A.; Cederbaum, L. S. *J. Chem. Phys.* **2000**, *112*, 7400 — BeC_n^{2-}. Sommerfeld, T. *J. Phys. Chem. A* **2000**, *104*, 8806 — C_4^{2-}.

(13) BelBruno, J. J.; Tang, Z.-C.; Smith, R.; Hobday, S. *Mol. Phys.* **2001**, *99*, 957.

(14) Blanksby, S. J.; Dua, S.; Bowie, J. H.; Schröder, D.; Schwarz, H. *J. Phys. Chem. A* **2000**, *104*, 11248.

(15) Watts, J. D.; Bartlett, R. J. *J. Chem. Phys.* **1992**, *97*, 3445.

(16) Calabrese, D.; Covington, A. M.; Thompson, J. S. *J. Chem. Phys.* **1996**, *105*, 2936.

(17) Simons, J.; Skurski, P.; Barrios, R. *J. Am. Chem. Soc.* **2000**, *122*, 11893.

(18) Tomioka, H. *Acc. Chem. Res.* **1997**, *30*, 315.

(19) Tomioka's "almost bottleable carbene": Woodcock, H. L.; Moran, D.; Schleyer, P. v. R.; Schaefer, H. F. III *J. Am. Chem. Soc.* **2001**, *123*, 4331. Hirai, K.; Tomioka, H. *J. Am. Chem. Soc.* **1999**, *121*, 1 10213.

(20) Dreuw, A.; Schweinsberg, H.; Cederbaum, L. S. *J. Phys. Chem. A* **2002**, *106* 1406.

CT049845H

JCTC Journal of Chemical Theory and Computation

Microsolvation of Li^+ in Small He Clusters. Li^+He_n Species from Classical and Quantum Calculations

C. Di Paola, F. Sebastianelli, E. Bodo, I. Baccarelli, and F. A. Gianturco*

Department of Chemistry, University of Rome "La Sapienza" and INFM, Piazzale A. Moro 5, 00185 Rome, Italy

M. Yurtsever

Chemistry Department, Istanbul Technical University, 80626 Maslak, Istanbul, Turkey

Received March 22, 2005

Abstract: A structural study of the smaller Li^+He_n clusters with $n \leq 30$ has been carried out using different theoretical methods. The structures and the energetics of the clusters have been obtained using both classical energy minimization methods and quantum Diffusion Monte Carlo. The total interaction acting within the clusters has been obtained as a sum of pairwise potentials: Li^+−He and He−He. This approximation had been shown in our earlier study to give substantially correct results for energies and geometries once compared to full ab initio calculations. The general features of the spatial structures, and their energetics, are discussed in details for the clusters up to $n = 30$, and the first solvation shell is shown to be essentially completed by the first 8−10 helium atoms.

I. Introduction

The study of the nanoscopic forces which act within small aggregates of weakly interacting particles, especially within assemblies containing helium atoms, has received a great deal of attention in the past few years[1−4] because of the broad range of phenomena that can be probed under the very special conditions provided by the He nanodroplets as containers of atoms or molecules. They are, indeed, ultracold homogeneous matrices where the corresponding spectra often reach very high resolution due to the superfluid properties of the helium droplet.[5]

The possibility of causing electronic excitation and/or ionization of the dopant species can offer additional ways of probing the modified interaction between the new ionic species and the gentle matrix of the helium droplets[4] since the ensuing distribution surrounding the molecular impurity is usually markedly deformed as a consequence of the electrostriction effects on the solvent brought on by the induction forces between the charged dopant and the helium atoms in the droplet:[6−8] It simply means that the local density of the quantum solvent is going to markedly increase around the cationic dopant. The analysis of such effects in the case of lithium-containing impurities has been particularly intriguing because of the expected simplicity of the electronic structures involved and, at the same time, the unusual bonding behavior of such systems. Furthermore, experiments in which the droplet was ionized after capture of Li atoms[9] have revealed a wealth of newly formed species such as Li^+, Li_2^+, and $LiHe^+$ which are produced during droplet fragmentation and evaporation thereby triggering the corresponding analysis of structure and bonding behavior from theory and computations.

In a previous computational investigation[1] on the interaction of Li^+ with n helium atoms, with n varying from 1 to 6, we have shown, in fact, that the energy optimized structures of the clusters were largely determined by two-body (2B) forces, with the many-body effects being fairly negligible for determining the final geometry of those small aggregates. In the present study we have therefore decided to analyze in greater detail the quantum states and binding energies of the larger structures of Li^+ impurities within the

* Corresponding author fax: +39-05-49913305; e-mail: fa.gianturco@caspur.it.

10.1021/ct050072m CCC: $30.25
Published on Web 08/17/2005

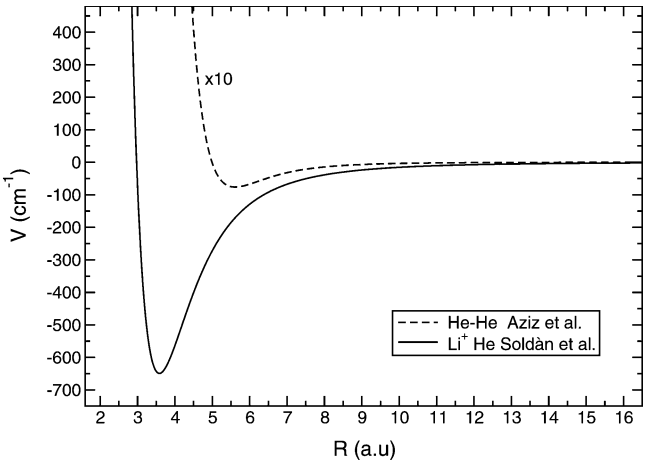


Figure 1. Comparison of the $LiHe^+$ potential from ref 17 and the He−He one from ref 18. The He−He potential has been multiplied by ten.

helium droplets in order to extend our computational knowledge on such species.

The next section briefly describes the $LiHe^+$ Potential Energy Curve (PEC) used here. The results for the trimer ground state are given and discussed in section III, where we report the features obtained using the distributed Gaussian functions (DGF) method[10−13] and the quantum Diffusion Monte Carlo (DMC) method,[14] in comparison with the values obtained via a classical minimization procedure. Section IV extends the discussion to the larger species, whose energetics and structural features, both obtained with DMC and classical methods, are analyzed.

II. The Pairwise Potentials

The actual potential energy curve (PEC) for Li^+He has been studied before because of its interest in modeling low-energy plasmas.[15] Alrichs and co-workers suggested a model potential as a variant of the Tang-Toennies model[16] which we shall here call the ATT potential. More recent calculations of Soldán et al.[17] employed a CCSD(T) treatment and used the aug-cc-pV5Z basis set. Recently, we have also carried out a new set of calculations on these clusters[1] using the MP4 method and the cc-pV5Z basis set. The three potentials provide very similar values of ZPE that are respectively 131.09, 129.18, and 128.75 cm^{-1}, values which roughly correspond to 20% of the well depth.

For the He_2 dimer we have employed the semiempirical potential (LM2M2) of Aziz and Slaman:[18] in Figure 1 we sketch the two pair potentials used in our calculations. The two potentials are completely different: the Li^+−He one is dominated by electrostatic and induction forces, while the He−He interaction is much weaker and purely van der Waals in nature.

As we mentioned above, the full interaction that acts in the M^+Rg_n clusters can be approximated by a sum of pairwise potentials. The main source of error in this model is due to the absence of the repulsive interactions between the induced dipoles on the helium atoms, especially those that are located near the ionic center. These interactions, however, can be taken to be very small for rare gas atoms and especially for helium partners. For example, in ref 19 for the analogous situation of the anionic dopant Cl^- in Ar clusters it has been shown that the inclusion of the leading terms of such three-body forces does not alter substantially the energetics and the structure of the clusters: in ref 1 we have discussed and confirmed this feature for the system analyzed here. Furthermore, since the charge delocalization in Li^+−He_n is very small[20] and the interaction is mostly determined by "physical" forces (charge/induced multipole) rather than by "chemical" effects, all the helium atoms that are attached directly to the ionic center can be considered to be equivalent because there is no detectable tendency for the latter to preferentially form chemically bound structures with only a subset of such adatoms, a fact which would therefore differentiate some of the binding effects with respect to those in the rest of the cluster.

III. An Outline of the Computational Tools

A. The Classical Optimization. The total potential in each cluster is described by the sum of pairwise potentials, and searching for the global minimum in this hypersurface V_{TOT} will give us the lowest energy structure for each aggregate from a classical picture viewpoint. All the classical minimizations were carried out with a modified version of the OPTIM code by Wales.[21] This code is based on the eigenvector following method. The basis of this method is the introduction of an additional Lagrange parameter into a "traditional" optimization framework,[22] which seeks to simultaneously minimize in all directions. All searches were conducted in Cartesian coordinates using projection operators to remove overall translation and rotation following Baker and Hehre.[23] Analytic first and second derivatives of the energy were employed at every step, and the resulting stationary point energies and geometries are essentially exact for the model potential in question. The details of the method as applied in our group have been given before and therefore will not be repeated here[21,24−26]

B. The Quantum Monte Carlo Calculations (DMC). The Diffusion Monte Carlo Method has been extensively discussed in a number of papers (refs 27−29 and references therein) and therefore will not be repeated here.

In our implementation a random walk technique is used to solve the diffusion equation where a large number of random walkers is propagated with time steps $\Delta\tau$ starting from an arbitrarily chosen initial distribution. The ground-state energy E_0 is obtained by averaging $E_L(\mathbf{r})$ over the final mixed distributions $f(\mathbf{r},t) = \psi_T(\mathbf{r},t)\psi(\mathbf{r},t)$:

$$\langle E_L \rangle = \frac{\int E_L(\mathbf{r}) f(\mathbf{r},\tau_f)\, d\mathbf{r}}{\int f(\mathbf{r},\tau_f)\, d\mathbf{r}} = \frac{\int \psi_0(\mathbf{r}) \hat{H} \psi_T(\mathbf{r})\, d\mathbf{r}}{\int \psi_0(\mathbf{r})\psi_T(\mathbf{r})\, d\mathbf{r}} = E_0 \quad (1)$$

The energy is therefore affected by a bias due to the use of a mixed distribution. The bias is however minimized by using very long propagation time and very short time steps. Expectation values of position operators $\hat{A}(\mathbf{r})$ are also given by averaging over $f(\mathbf{r})$, a procedure that leads to biased distributions. However, we believe that the bias does not modify the qualitative picture of the present calculations especially because whenever our spatial distributions are

compared with those from other quantum calculations (where possible) and with classically minimized structures, they are found to be in remarkable agreement with each other as we shall discuss further below.

The trial function used here for the He−He pairs is a product of trial wave function

$$\Psi_T = \prod_{i,j\in He} \exp\left(-\frac{p_5}{R_{ij}^5} - \frac{p_2}{R_{ij}^3} - p_0 \log R_{ij} - p_1 R_{ij}\right) \quad (2)$$

where the values of the coefficients have been taken by ref 30. The trial function for the Li^+−He pair has been chosen as a Gaussian function centered around the energy minimum of the relative interaction. However, to adjust each trial wave function to the size of the larger clusters, its parameters have been chosen so that they make it increasingly more delocalized as the number of adatoms is increased. Our trial functions are available on request from the correponding author.

C. The Distributed Gaussian Functions (DGF) Expansion. The DGF method[10−13] is a variational approach developed in our group which solves the trimer bound state problem written in terms of the atom−atom coordinates by employing a large ensemble of distributed Gaussian functions as a basis set.

First of all, the Hamiltonian for a triatomic system is expressed in terms of atom pair coordinates R_1, R_2, and R_3, i.e., in terms of the distances between each pair of atoms along which the Gaussian functions are distributed (see ref 12 for the expression of the Hamiltonian for a triatomic system with an impurity).

The total potential for the trimer is assumed to be the sum of the three two-body potentials as discussed previously, and the calculations are carried out for a zero total angular momentum. The total wave function for the υth vibrational states is then expanded in terms of symmetrized basis functions

$$\Phi_\upsilon(R_1,R_2,R_3) = \sum_j a_j^{(\upsilon)} \phi_j(R_1,R_2,R_3) \quad (3)$$

with

$$\phi_j(R_1,R_2,R_3) = N_{lmn}^{-1/2} \sum_{P\in S_2} P[\varphi_l(R_1)\varphi_m(R_2)]\varphi_n(R_3) \quad (4)$$

for the two-identical-particle system. Here, j denotes a collective index such as $j = (l \le m;n)$, N_{lmn} is a normalization constant expressed in terms of the overlap integrals $s_{pq} = \langle\varphi_p|\varphi_q\rangle$, and P is the permutation operator between the two helium atoms. Each one-dimensional function φ_p is chosen to be a Gaussian centered at the R_p position

$$\varphi_p(R_i) = \sqrt[4]{\mathrm{frac}2A_p}\tau e^{-A_p(R_i-R_p)^2} \quad (5)$$

With the DGF approach we can obtain several indicators on the spatial behavior of the bound states of the systems (the root mean value of the square area, the average of the cosine value—and the various moments—of any angle, etc.), along with several probability distribution functions such as the pair distribution function

$$D^{(\upsilon)}(R_1) = \int\int |\Phi_\upsilon(R_1,R_2,R_3)|^2 \, dR_2 \, dR_3 \quad (6)$$

The selection of a suitable set of Gaussian functions and their distribution within the physical space where the bound states are located is obviously of primary importance in order to finally obtain converged and stable results. We extensively experimented with different sets obtained by changing the number and location of the DGF depending on the features of the 2B potentials employed. The details of the basis set employed for the Li^+He_2 system are specifically given in the following section.

IV. The Li^+He_2 Trimer

Together with the classical optimization and the DMC calculations, we further carried out the analysis of the properties of the trimer Li^+He_2 using the DGF method. We employed an optimized DGF basis set in order to obtain well-behaving total wave functions at the triangle inequality boundaries (for further details see ref 31). The Gaussian functions are distributed equidistantly along the atom−atom coordinates starting from 2.55 a_0 and out to 9.18 a_0 with a step of 0.17 a_0, a choice which ensures converged results within about 0.1 cm^{-1} (0.01% of the total ground-state energy). In Table 1 we report the results obtained with DMC and DGF methods: they are seen to be in good agreement with each other. Due to the addition of the second light He atom, the Zero Point Energy (ZPE) of the trimer is about twice the ZPE value for the isolated dimer ion. The two ZPE values remain very similar as percentage values of the overall interaction showing that the Li^+ impurity strongly affects the features of the cluster, as expected from the involved potentials, while the additional helium atom has little effect on the bonding features. The ionic system, as expected, does not present the typical high degree of delocalization shown in pure He clusters (the ZPE for He_3 is more than 99% of its total well depth[32]) or by doped He aggregates with weakly bound impurities as, e.g., H^-, for which the impurity is clearly located outside the cluster (the ZPE for H^-He_2 is 90.66%[24]). Hence, we expect the Li^+−He_2 trimer (and its larger clusters) to behave in a different way with respect to the more weakly bound neutral He clusters and thus presume that the classical description of the ionic structures should give us realistic indications on the structure and energetics of such systems (as it was the case in, e.g., the H^-Ar_n clusters that we studied earlier[25]).

We therefore begin by looking at the average values of the radial distances and angles (see Table 1) which, together with the corresponding standard deviations, describe the ground state of the trimer and clearly confirm its having a rather floppy structure: on the other hand, the Li^+ species still is undoubtedly seen by our calculations to coordinate the two He atoms at a distance determined within 0.3 a_0. Information on the overall geometrical features of the trimer is further gained from the radial and angular distributions shown in Figure 2, where the values obtained with the classical optimization are also reported as vertical lines. DMC and DGF results are substantially coincident, and the small differences are mainly due to the bias contained in our DMC distributions. We notice that the floppiness of the system is

Table 1. Comparison of the Results for the Ground State of $(LiHe_2)^+$ Obtained with Both DMC and DGF Methods[a]

	$E_{v=0}$	ZPE (%)	r_{He-He}	r_{He-Li^+}	$\theta_{\widehat{He}}$	$\theta_{\widehat{Li^+}}$	$\sqrt{\langle S^2 \rangle}$
DMC	-1041.7 ± 0.68	20.24	6.59 ± 0.81	3.83 ± 0.32	28.43 ± 12.20	123.88 ± 24.01	5.54 ± 1.84
DGF	-1042.45	20.18	6.69 ± 0.73	3.75 ± 0.32	26.94 ± 17.69	127.36 ± 27.99	5.24

[a] The energies are in cm^{-1}, the distances in au, the angles in degrees, and the areas S in au.^{2} The ZPE is expressed as a percentage of the total well depth for the trimer $D_e = 1306.02$ cm^{-1} (from classical optimization).

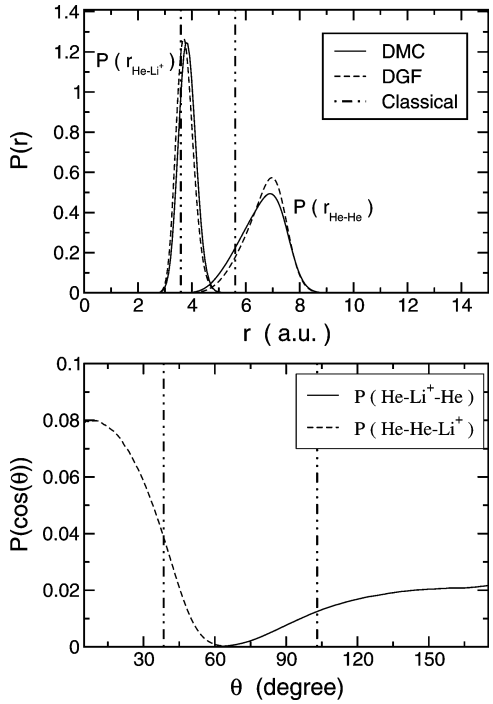


Figure 2. The radial distribution functions for the Li^+He_2 calculated via the DMC (solid line) and the DGF (dashed line) methods. In the upper panel we report the pair distribution functions of the relative atom−atom distances, while the lower panel shows the distribution functions for the internal angles. The (r,θ) values obtained classically are given as vertical lines in the two panels.

particularly evident when looking at the distribution function related to the He−He distance, whose standard deviation is more than twice as large as than the one for the Li^+−He bonds. The classical values obtained in the structure optimization are also closer to the quantum average values obtained for the Li^+−He distances.

In the classical description we do find that the ground state of the trimer is depicted by an isosceles triangle, with the two shorter sides associated with the two Li^+−He distances and a longer one corresponding to the He−He distance. On the other hand, the real quantum system cannot be described by one single structure only, and its distribution functions correctly show a delocalized triatom with a dominant contribution from the collinear arrangement. The image of a structure with the Li^+ coordinating the two He atoms at a rather well-defined distance (notice the compact distribution function related to Li^+−He distances) is not completely lost in the quantum description, meaning that the presence of the strong ionic forces from Li^+ is reducing the degree of delocalization which is always present in the pure He aggregates. This change determines the more rigid structure of the system in the sense that now the He atoms are more strongly coordinated directly to the Li^+ impurity (see the next section on the larger clusters).

We carried out an additional analysis of the structural features of the trimer by taking advantage of the DGF pseudoweights,[10] which allow us to pictorially describe the trimer in terms of types of triangular arrangements. In the upper panel of Figure 3 we thus report the total number of the employed basis set functions, grouped according to the triangular family to which they belong, and in the lower panel of the same figure we report the 'weight' of the dominant families when describing the ground state of the trimer. We can thus easily identify the most important arrangement to be given by the 'flat' isosceles, the collinear (with the impurity in the middle), and the scalene triangles, while all other possibilities do not contribute in a significant way. Again, we find that we cannot associate the system to one unique structure. while the marked delocalization features are now mainly related to the He−He binding and less to the Li^+−He ionic forces; hence we see that a conventional structure with the Li^+ coordinating the two He atoms can still be qualitatively identified. In the next section we shall further discuss how the situation evolves with the addition of more He atoms and to which extent the classical results can be still seen to qualitatively correspond to the quantum description of their structures surrounding the ionic dopant.

V. Energetics and Structures of Li^+−He_n ($3 \leq n \leq 30$) Clusters from Quantum and Classical Calculations

We start now to discuss the energetics and the geometrical features of the larger Li^+He_n clusters with $n \geq 3$. Up to $n = 20$, in fact, it was still possible to also carry out quantum DMC calculations which are not too demanding in terms of CPU-time. Hence, for clusters of such a size we can make a direct comparison between our quantum findings and the classical optimization results we obtained via the combination of the OPTIM procedure[21] with a random search for the minimum energy structures (see for details refs 25 and 26). In Table 2 we report the results for the energetics. The left part of that table shows the minimum potential energies obtained by means of the classical optimizations (column labeled 'classical') in comparison with the corresponding DMC ground-state energies (column labeled 'quantum'). The differences between the two sets of values are due to the ZPE effects of the nuclear motions: in the third column we also display the ZPE value for each cluster as a percentage of the well depth. The amount of the ZPE effects increases as the cluster grows, since the increasing addition of He atoms brings the ZPE percentage from about 20% for Li^+-He_2 to more than 40% for the last cluster studied here with the quantum DMC method (Li^+He_{20}). On the right part of Table 2 we report the total energies relative to the loss of

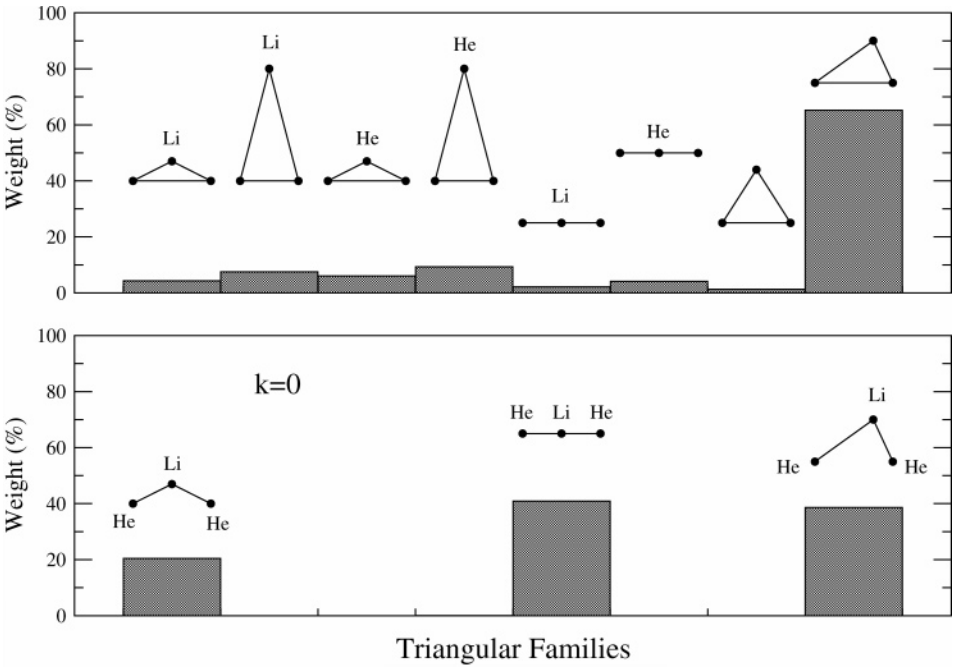


Figure 3. DGF description of the ground state of the Li^+He_2 in terms of triangular families. In the upper panel we report the relative weights of the triangular groups of basis functions (as percentage with respect to their total number) which correspond to the different types of triangular configurations sketched in that panel. In the lower panel we show the three configurations that contribute the most to the ground state of the trimer.

Table 2. Minimum Total Energies and Single-Particle Evaporation Energies Using Classical and Quantum Treatments for the Nuclear Motion[a]

	V^{TOT}			ΔE_{evap}	
n	classical	quantum	ZPE (%)	classical	quantum
2	−1306.0	−1041.7 ± 0.7	20.2	656.8	521.1 ± 0.7
3	−1970.5	−1556.3 ± 0.8	21.0	664.4	514.6 ± 1.5
4	−2640.1	−2038.1 ± 2.9	22.8	669.6	481.7 ± 3.7
5	−3287.0	−2474.1 ± 4.6	24.7	646.9	436.1 ± 7.5
6	−3940.2	−2895.3 ± 10.3	26.5	653.1	421.1 ± 14.9
7	−4281.4	−3000.7 ± 14.7	29.9	341.2	105.4 ± 25.0
8	−4627.6	−3161.5 ± 23.5	31.7	346.2	160.8 ± 38.2
9	−4789.3	−3234.3 ± 9.5	32.5	161.7	72.8 ± 33.0
10	−4936.0	−3271.2 ± 19.5	33.7	146.7	36.9 ± 29.0
14	−5284.2	−3292.4 ± 18.1	37.7	88.4	5.3 ± 37.6
20	−5785.6	−3432.5 ± 56.1	40.7	72.2	23.3 ± 74.2

[a] ZPE (%) means: $[(V^{TOT}_{classical} - V^{TOT}_{quantum})/V^{TOT}_{classical}]\cdot 100$. ΔE_{evap} means: $-[V^{TOT}(Li^+(He)_n) - V^{TOT}(Li^+(He)_{n-1})]$ the quantum evaporation energy for $n = 14$ has been calculated using the formula $E_{evap}(14) = (V^{TOT}(10) - V^{TOT}(14))/4$ and for $n = 20$ $E_{evap}(20) = (V^{TOT}(14) - V^{TOT}(20))/6$.

one He atom between the pairs of Li^+He_n and Li^+He_{n-1} clusters, calculated both with the classical and quantum methods. We notice that the evaporation energy is drastically reduced when passing from the cluster with $n = 6$ to that with $n = 7$ and correspondingly the ZPE percentage value increases most markedly (more than 3%). We can then surmise that the structure with $n = 6$ is a particularly stable cluster as we shall further discuss below.

The data presented in Table 2 are pictorially reported in the two panels of Figure 4 where the energies are plotted as functions of the number *n* of He atoms in each cluster. From the lower panel of Figure 4 we can see the similar behavior shown by the ΔE_{ev} values up to $n = 6$, while for $7 \leq n \leq 10$ both classical and quantum effects make the two curves show a marked drop in values. The single He atom evaporation energies, $\Delta E_{ev} = -[E(Li^+He_n) - E(Li^+He_{n-1})]$ (filled-in circles for the classical calculations and open square for those obtained with the DMC method), are plotted up to $n = 30$ (DMC results up to $n = 20$): again the two curves show very similar behavior, both presenting the same abrupt energy drop at $n = 7$ and $n = 9$. In contrast with what we found in our analysis of H^-He_n clusters,[24] the steplike structure shown by the classical treatment is now also present in the quantum calculations. Given such a correspondence between the classical and quantum description for these smaller clusters it then becomes reasonable to try to explain the sudden energy jumps of Figure 4 by looking at the lowest energy structures found with the classical minimizations (see Figure 5). In that figure we also report the corresponding symmetry groups, the total energy (in cm^{-1}), and the relevant distances between atoms (in au). We therefore see that the Li^+He_6 has the very symmetrical octahedral geometry (see the second panel in the upper part of the figure) where all the He atoms are equivalently coordinated to the central Li^+ at a distance very close to the R_{eq} (3.58 a_0) of the Li^+He PEC. On the other hand, when moving to the Li^+He_7, the repulsive forces acting between the rare gas species do not allow any more for such a symmetrical arrangement around the Li^+ ion, and the net effect is that of decreasing the energy contributions from the interactions between each He atom and the positive ion, i.e., the Li^+He distance becomes larger, as one can see in Figure 5 from the reported R_{LiHe} values. Similar reasoning can be applied to explain the second energy step between $n = 8$ and $n = 9$: the evaporation energy gives the mean value of the energy necessary to remove any of the He atoms, and the presence of a nonequivalent rare gas

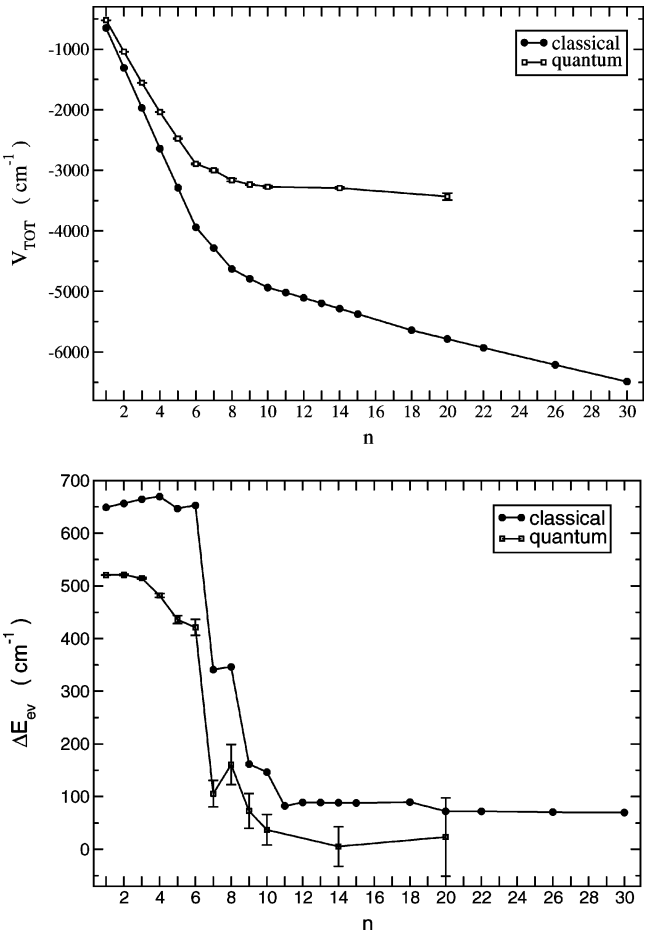


Figure 4. Upper panel: total energies (in cm^{-1}) as a function of the number n of helium atoms in Li^+H_n from classical optimizations (filled-in circles). Lower panel: single helium atom evaporation energies, $\Delta E = -[E(Li^+He_n) - E(Li^+He_{n-1})]$ in cm^{-1}, as a function of the number n of helium atoms in Li^+He_n from the classical optimizations (filled-in circles). The quantum results are given in both panels in open squares showing their corresponding numerical level of confidence.

atom in the apical position in Li^+He_9 (see second panel in the lower part of Figure 5) causes a significant decrease of the evaporation energy with respect to its value for the Li^+-He_8 cluster. Finally, for Li^+He_{10} the lowest energy structure we obtained with the classical minimization corresponds to the symmetrical *bicapped square antiprism* geometry. Hence, by using the classical geometries and energy minimization procedures, we found three relatively more stable structures for clusters with $n = 6$, 8, and 10, in correspondence with symmetrically compact structures. We therefore have here a relatively complex situation in which we clearly see the formation of a very stable "shell" with 6 He atoms, but, nevertheless, we also see the closing of two successive "shells" at $n = 8$ and $n = 10$. After the first ten He atoms have been placed around the ionic impurity in the bicapped antiprims geometry, the cluster growth is chiefly driven by weaker and essentially dispersive forces. While these results stem clearly from the classical calculations, the quantum evaporation energies are affected by large errors for $n > 6$, and therefore the quantum structures of the cluster for $n > 6$ are less easy to establish.

The correspondence between classical and quantum structural pictures is fairly well reproduced if we now compare the quantum distribution functions with the classical results for the relative distances and angles. In Figure 6 we report the DMC atom−atom distribution functions for the Li^+−He and He−He distances within each Li^+He_n cluster, normalized to the total number n of possible 'connections' between Li^+ and He atoms (solid lines) and to the total number $N = n(n - 1)/2$ possible 'connections' between the n He atoms (dashed lines). In this way we can make a direct comparison between the quantum calculations and the lowest energy geometrical structures obtained with the classical optimizations where the conventional picture of direct bonds existing between localized, pointlike, partners can be used. In that figure we also report the classically optimized distances: we have grouped together sets of close values and have given as horizontal bars their statistical standard deviations: each set has a height proportional to the number of distances which have the same value. Finally, the displayed numbers are the values of the integration along r for each broad peak in the quantum distribution functions and represent the number of connections within atoms which are in the given distance range under each integration. For some of the larger clusters the results of the integration depended slightly on the choice of the fitting functions (essentially we have used an asymmetric Gaussian for each peak), and therefore we have reported an interval of possible values. The numbers reported above each quantum distribution should be directly compared to the height of the classical vertical lines. In Figure 6 we report the results for four selected clusters ($n = 4$, 6, 7, 10) which we shall use in our discussion: we also obtained similar results for all the other clusters. When we look at the panel showing the Li^+He_4 clusters in the upper left of Figure 6, we see that the DMC distribution function for the Li^+−He distance peaks at 3.88 au, while the classical distances are represented by one 'stick' at 3.59 au whose height is 4, equal to the value of the area under the quantum distribution (solid line); we see also that the DMC distribution function for the He−He distance peaks at 6.13 au, while the classical distances are represented by one 'stick' at 5.85 au whose height is 6 (which is the value of $N = n(n - 1)/2$, with $n = 4$) a number which is indeed equal to the value of the area under the quantum distribution (dashed line). For the larger clusters (see the other panels of the same figure) the number of distinct sets of distances increases, but still the agreement between classical and quantum findings concerning the number of corresponding distances remains very good. Although the mean values are different as a consequence of the very diffuse behavior of the wave functions, we can clearly see that the classical results essentially provide the same qualitative structural picture. For example in the panel referring to Li^+He_{10} we see clearly also from the quantum calculation the formation of the first shell of 8 atoms and the second shell constituted by only two atoms. To compare even more in detail the classical and quantum results, we report in Figure 7 the DMC angular distribution functions $P(\theta)$ for the angles centered at the Li^+ ion and at any of the He atoms together with the results from classical optimization: we see again that the

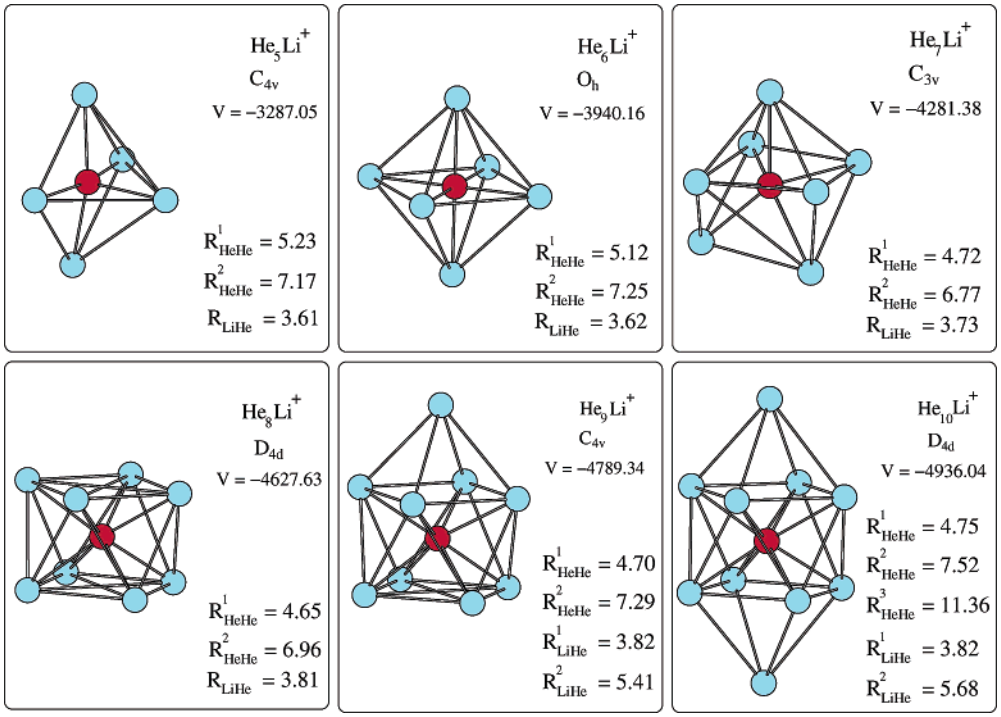


Figure 5. Optimized lowest energy structures obtained with the classical minimizations for Li^+He_n with n from 5 to 10.

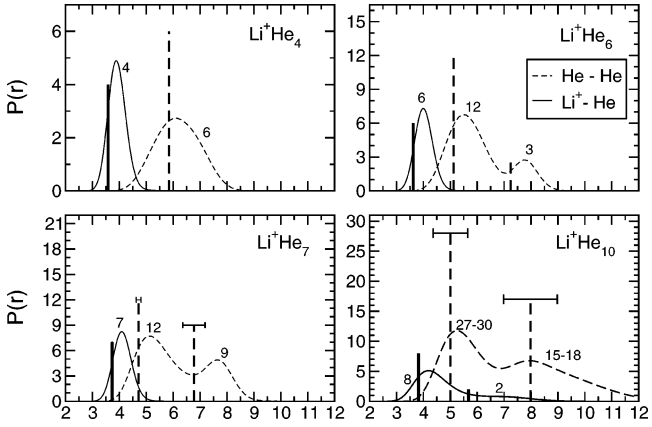


Figure 6. Radial distribution functions of the atom−atom relative distances $P(r)$, normalized to the number of 'connections' between the atoms in each cluster, as a function of the distance r (in au). DMC calculations for the Li^+He_n, $n = 4$, 6, 7, 10. In all panels the solid lines are for the Li^+−He distances, while the dashed ones refer to the He−He distances. The vertical lines indicate the computed mean values of the atom−atom distances (solid lines for Li^+−He and dashed lines for He−He) of the optimized structures from the classical calculations, with their height showing the number of bonds associated with that distance (marked also by its spread value on top of each "stick"). The presented numbers in each panel give the corresponding quantum density of bonds within each computed profile.

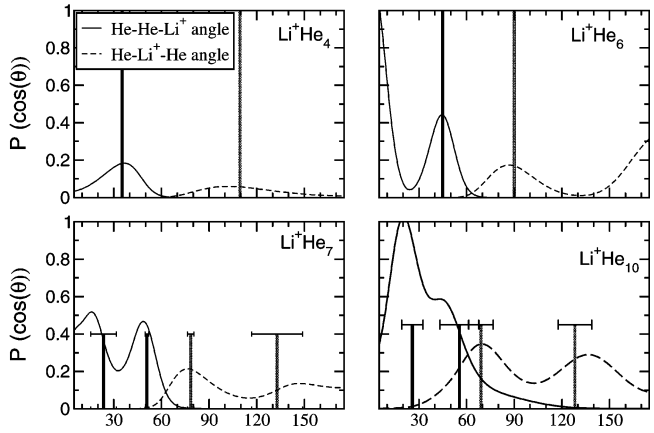


Figure 7. Angular distribution functions $P(\theta)$, obtained with the DMC calculations for the Li^+He_n, $n = 4$, 6, 7, 10. In all the panels the solid line is for the He−He−Li^+ angles, and the dashed line is for the He−Li^+−He angles. The vertical lines indicate the various mean values of the angles (solid lines for He−He−Li^+ and dashed lines for He−Li^+−He) of the optimized structures given by the classical calculations. The notation is the same as in Figure 7

agreement between quantum and classical values (vertical lines) indeed remains very close, at least at the qualitative level.

These findings allow us to make further comments on the microsolvation process that occurs when the Li^+ is inserted in a small He cluster. Both the classical and the quantum treatments concur in locating the Li^+ impurity inside the He_n moiety as one can clearly see from Figure 8 where we report the DMC distribution functions $P(r)$ of the Li^+ (solid lines) and of the He atoms (dashed lines) from the geometrical center of each Li^+He_n cluster. This quantity is defined as

$$\mathbf{r}_{gc} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i \qquad (7)$$

where N runs over the total number of atoms in the cluster. The Li^+ is always closer to the geometrical center with respect to the He atoms, and when the cluster size increases the corresponding distribution functions have reduced over-

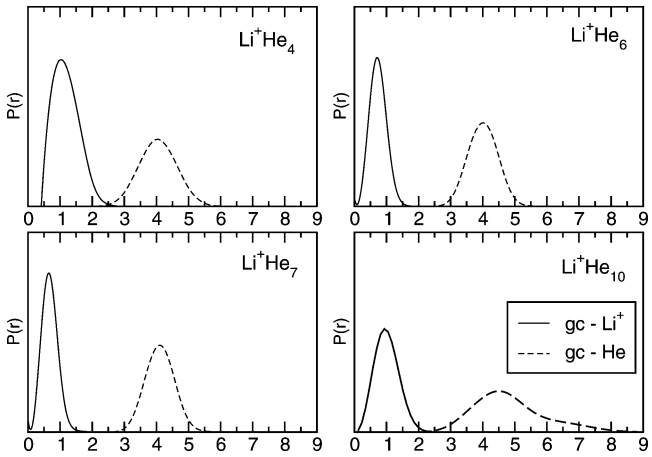


Figure 8. DMC radial distribution functions of the distances of the He atoms (dashed lines) and of the Li^+ (solid lines) from the geometrical center (gc) for selected clusters as functions of the distance r (in au). The density distributions are normalized so that the area under them is unitary.

lap, thereby showing that the cluster growth is accompanied by a slow "dropping" of the Li^+ toward the geometrical center of the latter, although the process is by no means monotonic with cluster size.

Finally, we carried out classical optimizations for larger clusters Li^+He_n ($n = 11-15$, 18, 20, 22, 26, 30) in order to better confirm what we have observed in the smaller ones; the lowest energy geometries for a selection of them are reported in Figure 9. For all the clusters under inspection the growth occurs around the highly symmetric Li^+He_{10} core represented by the bicapped square antiprism polyhedron enclosing the Li^+ impurity and drawn with thicker lines in the figure. The additional He atoms are now being placed further away from the impurity without perturbing the structure of the Li^+He_{10} moiety which therefore seems to constitute the last solvation shell of the atomic ion. We expect that the He atoms outside the shell will be characterized by a greater delocalization and weaker interactions with the central Li^+ that the less shielded inner core of 10 atoms. From now on we expect, therefore, that the cluster will grow by adding more solvent atoms in a nearly isotropic fashion driven mainly by He−He interaction, and we surmise that the binding energies of each atom will become increasingly similar to those of a pure He cluster. Correspondingly, the evaporation energy (see Figure 4) now shows a markedly different behavior with respect to what happens in the smaller clusters with $n \leq 10$: the steplike feature disappears, to be substituted by a plateau giving us the average energy needed to remove one of the nearly equivalent external He atoms.

This expected result is also shown by the quantum calculations (see Figure 4, lower panel) and is also borne out by the corresponding quantum distributions of the helium adatoms given by the data of Figure 10, where our DMC calculations for the $Li^+(He)_{20}$ clusters are reported: one clearly sees there that the radial distributions associated with the He distances from the Li^+ moiety show a set of more compact values related to the first shell of about 8 adatoms and a further distribution at larger distances (and broader than the first one) associated with the outer atoms that are chiefly bound by dispersion and by strongly screened induction interactions.

VI. Conclusions

In the present work we have analyzed the solvation process of a Li^+ ion in pure bosonic helium clusters in order to extend to a larger number of atoms the studies we had already

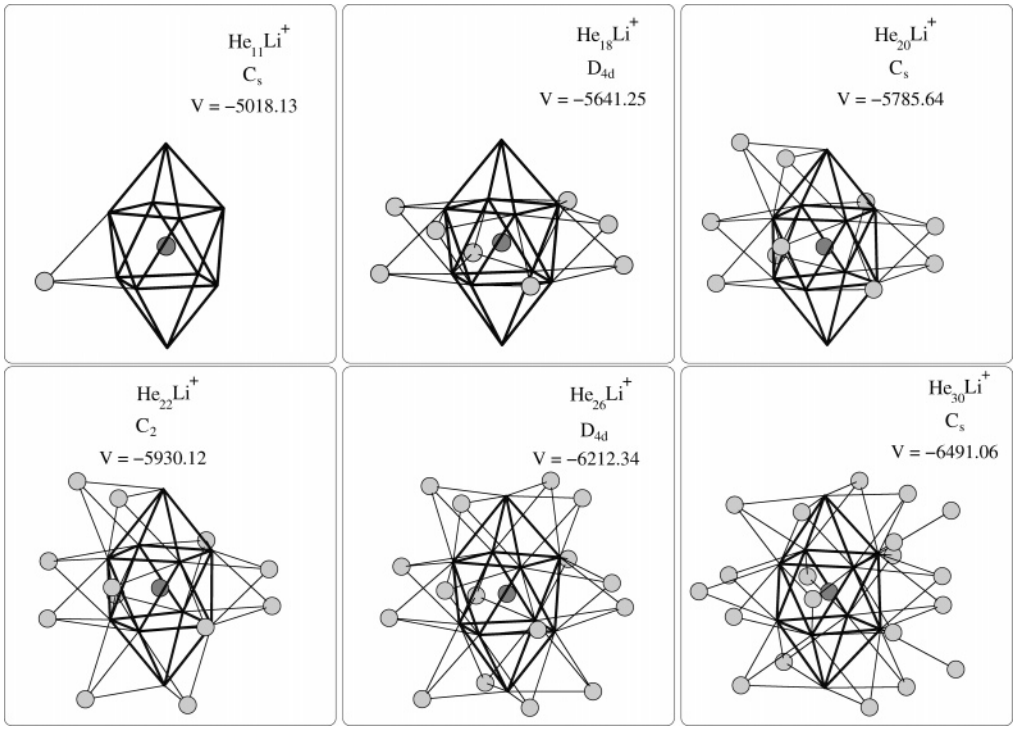


Figure 9. Optimized lowest energy structures obtained with the classical minimizations for Li^+He_n with $n = 11$, 18, 22, 26, 30. The central polyhedron including the Li^+ ion represents the Li^+He_{10} core.

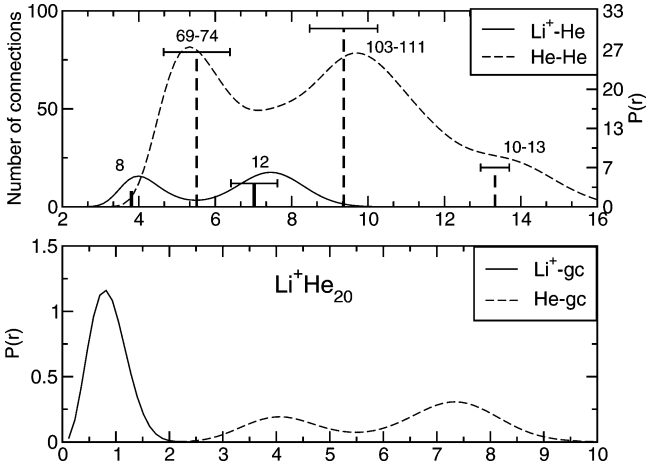


Figure 10. Quantum DMC radial distributions for the Li^+He_{20} moiety. Upper panel: atom−atom distance distributions compared with classical distance values as in Figure 7. Lower panel: distance distributions with respect to the geometric center of the cluster as in Figure 9.

carried out in previous work on the smaller aggregates.[1] Here a combination of classical energy minimization techniques and of "exact" quantum Monte Carlo methods have been employed in order to describe the structure and the energetics of the $Li^+(He)_n$ clusters with n up to 30 (20 for DMC calculations), employing always a description based on the sum-of-potential approximation.

The basic approximation which this study relies on is that of calculating the full cluster interaction as a sum of pairwise potentials. Although the error introduced by this assumption is in general found to be (in absolute terms) larger in ionic systems than in neutral ones, its relative weight remains rather small.[1] Thus, we believe that the resulting stuctures and energies would not be substantially altered by the correct use of a full many-body potential (see, for example, the discussion in ref 19 for the similar situations of H^- and Cl^- in Rg clusters and our calculations in ref 1).

We have therefore shown the following:

(1) The Li^+ is fully solvated inside larger ^{4}He clusters as already indicated by our earlier work on very small aggregates.[1]

(2) The ionic Li^+core does not form preferential "molecular cores" with surrounding He atoms but rather the helium adatoms remain equivalently bound within each solvation shell to the central, solvated lithium ion that persists in carrying the positive charge for more than 98%.[1]

(3) The ZPE corrections play a role in such systems, albeit strongly reduced with respect to the one found in neutral aggregates: this means that, at least in the initial solvation shell, the quantum adatoms are less delocalized and that the lithium−helium direct "bonds" are nearly rigid, classical structures.

(4) The classical optimization procedures can provide structural details which are reasonably close to those given by the quantum DMC calculations and can yield for ionic moieties the same structural picture as that given by the quantum treatment.

Furthermore, our comparison of single particle evaporation energies given by classical and quantum results suggests in both cases the formation of an initial shell of about 8 He atoms which are more strongly bound to the central ion. On the other hand, beyond that initial shell the cluster growth appears to be chiefly driven by He−He interactions, albeit at energies which are initially still kept larger than those of the neutral systems by the additional presence of the induction field due to the central ionic core. This behavior may be compared to that of Na^+ and K^+ doped helium clusters[6,7] where the first solvation shells were found to be of 9 and 12 atoms, respectively, from quantum calculations based on sum-of-potentials approximation.

Acknowledgment. The financial support of the FIRB project, of the University of Rome "La Sapienza" Scientific Committee and of the European Union "Cold Molecules" Collaborative Research Project no. HPRN-CT-2002-00290, is gratefully acknowledged. One of us (M.Y.) thanks the I.T.U. Research Fund for the financial support and the I.T.U. High Performance Computing Center for the computer time provided. We are grateful to Prof. G. Morosi and Dr. D. Bressanini for their help in improving our choice of trial functions. We also acknowledge the support of the INTAS Grant 03-51-6170.

References

(1) Sebastianelli, F.; Baccarelli, I.; Bodo, E.; Di Paola, C.; Gianturco, F. A.; Yurtsever, M. *J. Comput. Mater. Sci.* **2005**, in press.

(2) E.g., see: *Atomic and Molecular Beams, The state of the art 2000*; Camargue, R., Ed.; Springer: Berlin, 2000.

(3) Toennies, J. P.; Vilesov, A. F. *Angew. Chem., Int. Ed.* **2004**, *43*, 2622.

(4) Stienkemeyer, F.; Vilesov, A. F. *J. Chem. Phys.* **2001**, *115*, 10119.

(5) E.g., see: Brink, D. M.; Stringari, S. *Z. Phys. D* **1990**, *15*, 257.

(6) Galli, D. E.; Buzzacchi, M.; Reatto, L. *J. Chem. Phys.* **2001**, *115*, 10239.

(7) Rossi, M.; Verona, M.; Galli, D. E.; Reatto, L. *Phys. Rev. B* **2004**, *69*, 212510.

(8) Northby, J. *J. Chem. Phys.* **2001**, *115*, 10065.

(9) Scheidemann, A. A.; Kreisin, V. V.; Hess, H. *J. Chem. Phys.* **1997**, *107*, 2839.

(10) González-Lezana, T.; Rubayo-Soneira, J.; Miret-Artés, S.; Gianturco, F. A.; Delgado-Barrio, G.; Villarreal, P. *J. Chem. Phys.* **1999**, *110*, 9000.

(11) González-Lezana, T.; Rubayo-Soneira, J.; Miret-Artés, S.; Gianturco, F. A.; Delgado-Barrio, G.; Villarreal, P. *Phys. Rev. Lett.* **1999**, *82*, 1648.

(12) Baccarelli, I.; Delgado-Barrio, G.; Gianturco, F. A.; González-Lezana, T.; Miret-Artés, S.; Villarreal, P. *Phys. Chem. Chem. Phys.* **2000**, *2*, 4067.

(13) Gianturco, F. A.; Paesani, F.; Baccarelli, I.; Delgado-Barrio, G.; González-Lezana, T.; Miret-Artés, S.; Villarreal, P.; Bendazzoli, G. B.; Evangelisti, S. *J. Chem. Phys.* **2001**, *114*, 5520.

(14) Ceperley, D. M. *Rev. Mod. Phys.* **1995**, *67*, 279.

(15) Skullerud, H. R.; Løvaas, T. H.; Tsurugida, K. *J. Phys. B* **1999**, *32*, 4509.

(16) Alrichs, R.; Bohm, H. J.; Brode, S.; Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1988**, *88*, 6290.

(17) Soldan, P.; Lee, E. P. F.; Lozeille, J.; Murrel, J. N.; Wright, T. G. *Chem. Phys. Lett.* **2001**, *343*, 429.

(18) Aziz, R. A.; Slaman, M. J. *J. Chem. Phys.* **1991**, *94*, 8047.

(19) Lenzer, T.; Yourshaw, I.; Furlanetto, M. R.; Pivonka, N. L.; Neumark, D. M. *J. Chem. Phys.* **2001**, *115*, 3578.

(20) Bellert, D.; Breckenridge, W. H. *Chem. Rev.* **2002**, *102*, 1595.

(21) Wales, D. J. *J. Chem. Phys.* **1994**, *101*, 3750.

(22) Baker, J. *J. Comput. Chem.* **1986**, *7*, 385.

(23) Baker, J.; Hehre, W. H. *J. Comput. Chem.* **1991**, *12*, 606.

(24) Sebastianelli, F.; Di Paola, C.; Baccarelli, I.; Gianturco, F. A. *J. Chem. Phys.* **2003**, *119*, 8276.

(25) Sebastianelli, F.; Baccarelli, I.; Di Paola, C.; Gianturco, F. A. *J. Chem. Phys.* **2004**, *121*, 2094.

(26) Sebastianelli, F.; Baccarelli, I.; Di Paola, C.; Gianturco, F. A. *J. Chem. Phys.* **2003**, *119*, 5570.

(27) Hammond, B. L.; Lester, W. A., Jr.; Reynolds, P. J. *Monte Carlo Methods in Ab Initio Quantum Chemistry*; World Scientific: Singapore, 1994.

(28) Ceperley, D. M.; Alder, B. *Science* **1994**, *231*, 555.

(29) Suhm, M. A.; Watts, R. O. *Phys. Rev.* **1991**, *204*, 293.

(30) Lewerenz, M. J. *Chem. Phys.* **1997**, *106*, 4596.

(31) Baccarelli, I.; Gianturco, F. A.; González-Lezana, T.; Delgado-Barrio, G.; Miret-Artés, S.; Villarreal, P. *J. Chem. Phys.* **2005**, *122*, 084313.

(32) Rick, S. W.; Lynch, D. L.; Doll, J. D. *J. Chem. Phys.* **1991**, *95*, 3506.

CT050072M

JCTC Journal of Chemical Theory and Computation

Theoretical Study of the Interaction between the Guanidinium Cation and Chloride and Sulfate Anions

Isabel Rozas* and Paul E. Kruger

Centre for Synthesis and Chemical Biology, Department of Chemistry, Trinity College Dublin, Dublin 2, Ireland

Received January 14, 2005

Abstract: The interactions between the guanidinium cation and the chloride and sulfate anions have been explored theoretically. As a basis for this study, the structure of the guanidinium cation has been reexamined and confirmed to possess a nonplanar propeller structure. Thus, computations at the MP2/6-31+G* level yielded several guanidinium−chloride complexes in 1:1 and 2:1 stoichiometry, with the most stable being those in which the guanidinium cation(s) interact in pincerlike form with the spherical chloride anion. In the case of guanidinium−sulfate complexes, a large number of stable 1:1 dimers and 2:1 trimers were found at the same level of computation. The nature of all the interactions established has been analyzed by means of the AIMs and NBO approaches, and several correlations have been found between electron density at the bond critical points, hydrogen bond distances, and orbital interaction energies. The guanidinium moiety has been widely exploited as an anion receptor in supramolecular chemistry and bears relation to the arginine amino acid, which interacts with anionic substrates within proteins. Thus, the current study provides insight into the nature of the interactions involving the guanidinium cation within both abiotic and natural systems.

Introduction

There is considerable current interest in the development of molecular and supramolecular systems that have the ability to bind, identify, and signal the presence of negatively charged ions.[1] The motivation behind such studies is the recognition that anions enjoy an important role in biology, medicine, and the environment.[2] In the field of molecular recognition, the guanidinium cation has proven to be an extremely effective motif (host) for binding anionic guests.[3] Strong noncovalent interaction between this host and guest species is derived from ion pairing and H-bonding (HB) contributions, and these interactions have been shown to be particularly effective even in competitive solvent systems such as water.[4] Several features allow the guanidinium group to effectively bind anions. First, a very high pK_a value of around 12−13 allows it to remain protonated over a wide pH range, including physiological pH. Furthermore, its geometrical orientation allows it to align well with anionic groups, leading to a strong interaction. The incorporation of the guanidinium moiety into abiotic receptors is also driven in part by the recognition that the guanidinium group, in the form of the arginine amino acid, plays a crucial role in directing the formation of higher order structures in proteins through hydrogen-bonding as well as being a component found within enzymatic catalytic domains that bind anionic substrates.[5]

With the above concepts in mind, and in an effort to better understand those empirical contributions that lead to anion binding to the guanidinium cation, we undertook a theoretical study of the binding of the prototypical anions chloride (mononegative, spherical) and sulfate (dinegative, tetrahedral) to guanidinium. This empirical study provides a theoretical basis that might allow for better guanidinium receptor design and an impetus for our own formative structural studies into anion driven supramolecular chemistry.[6] Initial computational work focused on calculating the minimized structure of the guanidinium moiety itself. We then moved onto consider the interaction of chloride and sulfate anions with the guani-

* Corresponding author phone: (+353) 1 608 3731; fax: (+353)1 671 2826; e-mail: rozasi@tcd.ie.

10.1021/ct050009x CCC: $30.25
Published on Web 08/18/2005

dinium cation in 1:1 (dimer) and 2:1 (trimer) host-to-guest stoichiometry, as these are the stoichiometries most commonly found within abiotic host−guest systems. A systematic search of the Cambridge Structural Database was also conducted to corroborate the results found via computational methods. Results from this study are presented herein.

Methods

The geometry of the monomers (guanidinium cation and anion species) and host:guest complexes (guanidinium:anion) have been initially fully optimized with the Gaussian 98[7] program using the hybrid Hartree−Fock (HF) density functional theory (DFT) B3LYP/6-31G* computational level.[8] These structures have been used as the starting point for the optimization at the second-order Møller−Plesset perturbation method (MP2),[9] with the 6-31+G* basis set.[10] In the case of the guanidinium cation the MP2 method with the 6-31++G** basis set[11] was also used for comparison purposes. The combination of DFT and MP2 methods has proven to be sufficient for an accurate description of HBs and is not computationally expensive. In the case of the dimers, the minimum nature of the complexes has been confirmed by frequency calculations at both the B3LYP/6-31G* and MP2/6-31G* levels, whereas the trimers frequency calculations have been performed at the B3LYP/6-31G* level. Considering that the B3LYP and MP2 geometries were almost identical, B3LYP frequencies are enough to identify the trimers as minima.

The interaction energies have been calculated as the difference of the total energy of the complex and the sum of the isolated monomers. The interaction energies obtained in the majority of the cases were very high, and no basis set superposition error (BSSE) was calculated since, in those situations, such corrections should be irrelevant. The topological properties of the electron charge density have been studied using the atoms in molecules methodology (AIM) with the AIMPAC program package.[12] The natural bond orbital analysis was used to determine the nature of the interactions in the formation of the host:guest complexes. These calculations were performed with the NBO code implemented in Gaussian-98.[13] A search of the Cambridge Structural Database (CSD version 5.25 April 2004 Update)[14] for complexes of the guanidinium cation and chloride or sulfate anions was conducted employing the following restrictions: H-atoms were located, no disorder and a refinement (R-factor) ≤ 6.0%. All of the structures matching these criteria were examined manually and the following, deemed the most relevant to the current study, were fully analyzed and discussed in the text: guanidinium chloride (CSD refcode: GANIDC), guanidinium sulfate (CSD refcode: GUANSL01), and guanidinium bisulfate (CSD refcode: VETQIU).

Results and Discussion

Guanidinium Cation. For many years it was assumed that Y-conjugated molecules such as guanidinium were planar. Frenking[15] showed that guanidinium possesses a nonplanar minimum structure. In fact their calculations at the HF/6-31G* level predicted a planar form of the guanidinium cation, whereas at the MP2/6-31G* level such a planar form showed one imaginary frequency. The energy minimum structure at the MP2/6-31G* level had D_3 symmetry with the NH_2 groups rotated by 15° and a difference in energy from the planar form of 0.4 kcal mol^{-1}. To evaluate the interaction energy of the guanidinium−chloride and guanidinium−sulfate complexes and to analyze the behavior of DFT methods we have reinvestigated the nonplanarity of guanidinium at the B3LYP/6-31G*, MP2/6-31+G* and MP2/6-31++G** levels.

Table 1. Total Energy (E_T, au), Difference in Energy (ΔE, kcal mol^{-1}), ZPE and Frequencies for the Guanidinium Cation (G^+) Calculated at the B3LYP/6-31G*, MP2/6-31+G*, and MP2/6-31++G** Levels

B3LYP/6-31G*	E_T	ΔE	imaginary frequencies
G^+ nonplanar	−205.7593033		0
G^+ planar	−205.7588422	0.29	1
MP2/6-31+G*	**E_T**	**ΔE**	**imaginary frequencies**
G^+ nonplanar	−205.1307722		0
G^+ planar	−205.1297416	0.65	1
MP2/6-31++G**	**E_T**	**ΔE**	**imaginary frequencies**
G^+ nonplanar	−205.1861055		0
G^+ planar	−205.1852451	0.54	3

In Table 1 the results obtained for the planar and nonplanar forms are presented. In all cases the nonplanar form was the minimum energy structure, whereas the planar structure showed one to three imaginary frequencies. The difference in energy obtained with respect to the planar form was 0.29, 0.65, and 0.54 kcal mol^{-1} at the B3LYP/6-31G*, MP2/6-31+G*, and MP2/6-31++G** levels, respectively.

The good performance of the B3LYP/6-31G* method is noteworthy since the results obtained are very similar to those observed by Frenking at the MP2 level employing the same basis set ($\Delta E_T = 0.4$ kcal mol^{-1}, and one negative frequency for the planar form).[16] Frenking found that with inclusion of the ZPE corrections, the planar form was more stable than the nonplanar form by 0.2 kcal mol^{-1}. At the three levels of computation employed within the current study similar results were obtained; however, ZPE corrections cannot be applied to structures that show imaginary frequencies, as these data would be meaningless.

Regarding the equilibrium structures at the three levels of computation, guanidinium shows D_3 symmetry with the NH_2 groups rotated around the C−NH_2 bond by 14.5°, 16.9°, and 16.2°, respectively. The C−N bond lengths obtained in each case were the same for the planar and nonplanar forms and are very similar at the three levels of computation (B3LYP/6-31G*: 1.337 Å, MP2/631+G*: 1.336 Å, and MP2/6-31++G**: 1.336 Å). As in the case reported by Frenking these C−N distances are in excellent agreement with experimental values derived from vibrational spectra and X-ray diffraction studies (1.33−1.35 Å).[16]

The nonplanar form of guanidinium may be viewed as a left- or right-handed propeller. We have followed the

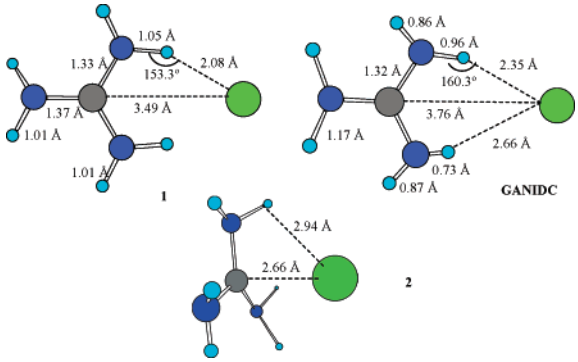


Figure 1. MP2/6-31+G* optimized structures of the two possible 1:1 guanidinium−chloride dimers, **1** and **2**, and the crystal structure of guanidinium chloride (GANIDC) as taken from the CSD.

transition between left- and right-handed propellers by a reaction coordinate in which the H−N−C−N dihedral angle was rotated from 15° (as observed in the minimum energy nonplanar structure) to −15° which would correspond to the mirror image structure. The structures with H−N−C−N = 15°, 10°, 5°, −5°, −10°, −15° are all minima, and the difference in energy between them is negligible. Only the planar structure (H−N−C−N = 0°) shows an imaginary frequency and a difference of energy of 0.24 kcal mol^{-1} with respect to the nonplanar forms. Therefore, the planar form of guanidinium can be considered as the transition state between these two 'propellers', and the fact that this transition requires such a small amount of energy could explain the perception of guanidinium being planar.

Guanidinium−Chloride Complexes. Two possible interactions between the guanidinium and chloride anions with 1:1 stoichiometry were considered, and their optimized structures at the MP2/6-31+G* level are shown in Figure 1. In one case the cation was looked as a pincerlike structure interacting with the spherical Cl^- anion, whereas in the other, the cation was considered as a positively charged delocalized plane interacting with the Cl^- anion in an orthogonal manner. Thus, complex **1** is similar to those observed in the crystal structure of guanidinium hydrochloride (GANIDC in Figure 1) and is at an energy minimum. The orthogonal dimer **2** (Figure 1) is a minimum as well.

Aspects of the geometry and nature of the interactions in dimer **1** are in agreement with those found in the crystal structure, GANIDC. The N−H···Cl distances and angles suggest the existence of hydrogen bond (HB) interactions. In complex **2**, the C···Cl distance suggests an ionic interaction, and the N−H···Cl distance is too long to correspond to a HB, even though there is a clear ionic attraction of the H atoms toward the Cl anion (the H atoms bend toward the chloride ion). Details of hydrogen bond interactions within GANIDC as derived from the CSD are brought together within Table 2 for comparison.

With regard to the existence of possible trimers, i.e. 2:1 host-to-guest complexes, they were constructed with consideration given to the computed structures of dimers **1** and **2** (see trimers **3**−**5** in Figure 2). Thus, the addition of a second guanidinium group to the dimer **1** yielded trimer **3**, whose MP2/6-31+G* optimized structure is shown in Figure 2. The N−H···Cl distances and angles to both guanidinium moieties are the same and indicative of a HB.

Table 2. Hydrogen Bond Parameters within Guanidinium Chloride (GANIDC), Guanidinium Sulfate (GUANSL01), and Guanidinium Bisulfate (VETQIU) as Derived from the CSD[14]

D−H···A[a]	D−H/Å	*d*(H···A)/Å	*d*(D···A)/Å	<(DH···A)/°
		GANIDC		
N1−H1···Cl^{i}	0.86	2.43	3.256	160.1
N1−H2···Cl^{ii}	0.96	2.35	3.272	160.3
N2−H3···Cl^{ii}	0.73	2.66	3.357	160.5
N2−H4···Cl^{iii}	0.87	2.41	3.256	163.6
N3−H5···Cl^{iii}	1.16	2.23	3.348	159.9
N3−H6···Cl^{i}	1.17	2.26	3.328	149.7
		GUANSL01		
N1−H1···O2	0.86	2.25	3.036	150.9
N1−H2···O2^{iv}	0.86	2.26	3.036	150.9
N2−H3···O3^{v}	0.86	2.23	2.975	145.0
N2−H4···O1^{vi}	0.86	2.11	2.914	154.8
N3−H5···O3^{vii}	0.86	2.45	3.157	140.1
N3−H6···O3^{viii}	0.86	2.45	3.157	140.1
N4−H7···O2^{ix}	0.86	2.05	2.899	169.3
N4−H8···O4^{x}	0.86	2.02	2.865	165.9
N5−H9···O3^{x}	0.86	2.11	2.957	163.0
N5−H10···O2^{vi}	0.86	2.10	2.955	169.8
N6−H11···O1^{ix}	0.86	2.08	2.928	169.3
N6−H12···O1^{xi}	0.86	2.19	2.981	152.3
		VETQIU		
N1−H1···O4^{xii}	0.86	2.18	2.951	149.5
N1−H2···O2^{xiii}	0.86	2.47	3.207	144.1
N2−H3···O2^{xiii}	0.86	2.23	3.020	152.8
N2−H4···O3	0.86	2.21	2.962	145.5
N3−H5···O4^{xi}	0.86	2.28	3.030	145.2

[a] Symmetry codes: i = 1/2+*x*, *y*, 1/2−*z*; ii = *x*, 1/2−*y*, 1/2+*z*; iii = 1/2−*x*, 1/2+*y*, *z*; iv = 1/4−*x*, 5/4+*y*, −5/4+*z*; v = −1/4+*x*, 1/4−*y*, 5/4+*z*; vi = 1/2−*x*, 1−*y*, 1/2+*z*; vii = *x*, *y*, 1+*z*; viii = 1−*x*, 1/2+*y*, −1/2−*z*; ix = 1/4−*x*, −3/4+*y*, 3/4+*z*; x = −1/2+*x*, 1/2−*y*, −1−*z*; xi = 1/4+*x*, −1/4+*y*, 1/4−*z*; xii = 1−*x*, −1−*y*, 1−*z*; xiii = −1+*x*, *y*, *z*.

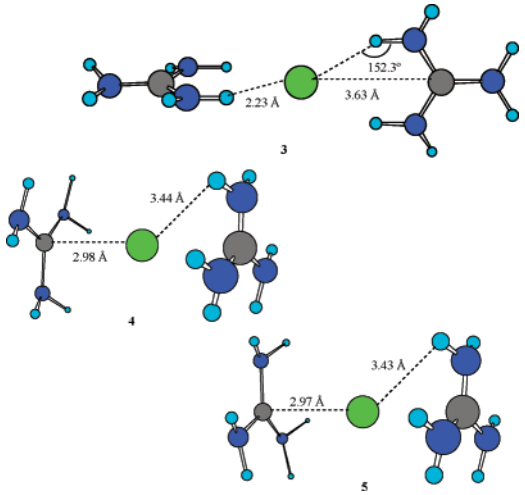


Figure 2. MP2/6-31+G* optimized structures of the 2:1 guanidinium−chloride trimers **3**−**5**.

At the MP2 level of theory, the dimer resulting from the interaction of the Cl anion with the H atoms (complex **1**) is more stable than the C···Cl one (complex **2**) by 10.60 kcal mol^{-1} (see Table 3). This difference in energy is extremely

Table 3. Total Energies (E_T, au) and Interaction Energies (E_I, kcal mol^{-1}) for the Guanidinium (G^+)−Chloride Complexes Calculated at the MP2/6-31+G* Computational Level

	E_T	E_I
$G^+\cdots Cl$ (**1**)	−664.9761235	−109.32
$G^+\cdots Cl$ (**2**)	−664.9592279	−98.71
$G^+\cdots Cl\cdots G^+$ (**3**)	−870.1729693	−150.78

Table 4. Electron Density and Laplacian at the BCP (au) of the HB Found in the Guanidinium (G^+)−Chloride Dimers and Trimers Optimized at the MP2/6-31+G* Level[a]

	X···Cl	ρ(BCP)	$\nabla^2\rho$ (BCP)	d(X···Cl)	<NH···Cl
$G^+\cdots Cl$ (**1**)	C···Cl			3.49	
	H···Cl	0.0369	0.0892	2.08	153.3
$G^+\cdots Cl$ (**2**)	C···Cl	0.0307	0.0742	2.63	
	H···Cl			3.13	78.5
$G^+\cdots Cl\cdots G^+$ (**3**)	C···Cl			3.63	
	H···Cl	0.0235	0.0713	2.23	152.3

[a] The interaction distances (Å) and N−H···Cl angles (°) are also shown.

high, and, therefore, dimer **2** will be probably nonexistent in the gas phase. However, the relevance of orthogonal interactions have been recently put into prespective by Diederich et al.,[17] and, hence, we consider that the characterization of this dimer **2** still deserves some attention, since this type of approach could be present in other related systems such as supramolecular systems or higly sterically hindered guanidiniums.

In the case of the trimers, only complex **3** should be considered since trimers **4** and **5** show six imaginary frequencies, and when they were further optimized they evolved to complex **3** (see Figure 2). The MP2 interaction energies, gathered together in Table 3, are clustered from ca. 98 to 109 kcal mol^{-1} for the dimers **1** and **2** and ca.151 kcal mol^{-1} for trimer **3**, showing that the interactions within all these complexes are extremely strong. In the case of dimer **1**, the minimized structure is consistent with that found in the crystal structure of guanidinium chloride (GANIDC).

The AIM analysis of these complexes shows electron density values around 10^{-2} au and positive Laplacians at the bond critical points (BCP), indicating HB or ionic interactions between the chloride and the electron acceptor atoms (H or C, respectively), Table 4. The position of the BCP is approximately in the middle of the interaction distance only in the case of dimer **2**. Short interaction distances, where the BCP is much closer to the H atom than to the chloride, are found in **1** and **3** and correspond to the large value of the electron density at the BCP as shown in Table 4. The value of the electron density at the BCP depends on the interaction distance as has been shown previously for other systems.[18]

Thus, the shorter the distance is the larger the electron density at the BCP. In all cases the electron density values found correspond to strong HB (complexes **1** and **3**) or strong 'closed-shell' interactions (complex **2**). Two ring critical points (RCP) have been found for the dimer **1** and its corresponding trimer **3** (see Table 4). No BCP was found between the chloride and the H atoms in dimmer **2**, which is in agreement with the small N−H···Cl angles found, which are incompatible with the formation of a HB.

Table 5. Charge Transfer (e), Dipole Moment Enhancement, and Orbital Interaction Energy (kcal mol^{-1}) Calculated for the Guanidinium (G^+)−Chloride Complexes at the B3LYP/6-31+G* Level with the NBO Method

	charge transfer	$\Delta\mu$		E(2)
$G^+\cdots Cl$ (**1**)	0.231	11.54	H···Cl	28.31
$G^+\cdots Cl$ (**2**)	0.251	4.60	H···Cl	137.67
$G^+\cdots Cl\cdots G^+$ (**3**)	0.098		H···Cl	9.09

The dipole moment enhancement, charge transfer, and nature of the interaction using the NBO method at the B3LYP/6-31+G* level are gathered together in Table 5. Large dipole moment enhancement and charge transfer (from the chloride toward the guanidinium cation) are due to the formation of the complexes and are observed in those cases where the Cl···(H,C) distances are short and in both dimers and trimer.

The NBO analysis (Table 5) shows the formation of a very strong ionic Cl−C bond, with over 80% contribution from the chloride, in complex **2** (d(Cl···C) = 2.66 Å). The molecular orbital responsible for this bond shows the important contribution of the *p*-orbitals of the chloride and the carbon atoms. In dimer **1** and trimer **3**, a lone pair from the chloride interacts with antibonding N−H orbitals, from one or two guanidiniums respectively, which is typical of HB interactions.

Guanidinium−Sulfate Complexes. The crystal structure of guanidinium sulfate (GUANSL01) was retrieved from the CSD. It shows an arrangement of guanidinium molecules surrounding each sulfate anion (see Figure 3) with typical S−O···H−N hydrogen bond distances ranging from 2.02 and 2.45 Å. The hydrogen bond interactions within GUANSL01 are brought together in Table 2. Using this structure as a basis, different dimers were considered, and the final MP2 optimized structures are shown in Figure 3.

In the first approximation, a proton transfer from one of the N−H groups to the sulfate oxygen atom occurred yielding a guanidine−bisulfate complex, **6**, possessing a double HB interaction and no negative frequencies. Furthermore, it was possible to optimize structure **7** where no proton transfer has occurred, yielding a complex between the sulfate anion and the guanidinium cation, but this complex showed one negative frequency.

In the case of a dimer with two N−H groups interacting with one O atom, when undergoing optimization, a change in the configuration occurred yielding a guanidine−bisulfate complex, **8**, with three HBs and no negative frequencies.

When considering a putative C···O orthogonal interaction between both ions, after optimization a new bond appeared between the C atom of the cation and an O atom of the sulfate anion to give compound **9**. This molecule **9** occurs at an energy minimum with no negative frequencies.

The O···H distances within the optimized structure **6** are shorter than those found within the crystal structure of guanidinium sulfate (GUANSL01) and guanidinium bisulfate

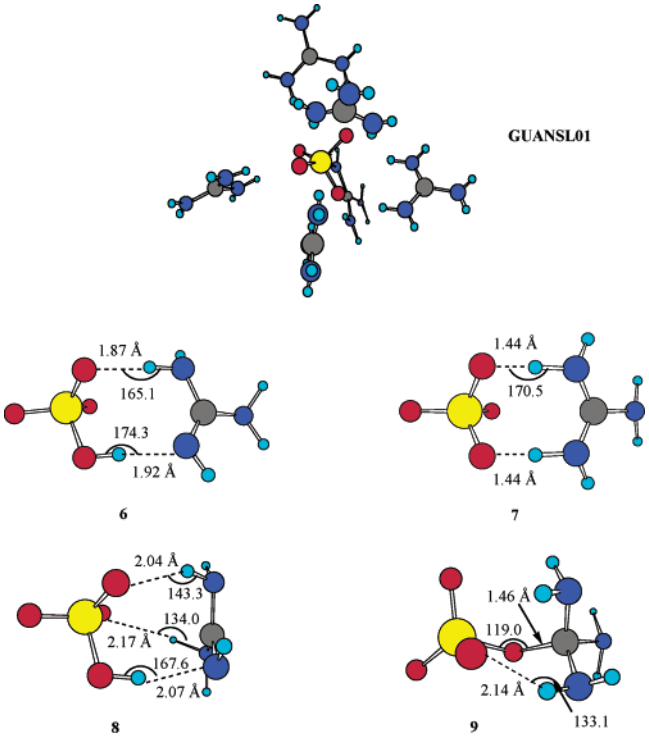


Figure 3. MP2/6-31+G* optimized structures of the 1:1 guanidinium−sulfate dimers **6**−**9** and crystal structure of guanidinium sulfate (GUANSL01) as derived from the CSD.

Table 6. Total Energies (au) and Interaction Energies (kcal mol^{-1}) for the Guanidinium (G^+), Guanidine (G)-Sulfate, Bisulfate Complexes Calculated at the MP2/6-31+G* Computational Level

	E_T	E_I
G···HSO_4 (**6**)	−903.2088034	−15.27
G···HSO_4 (**8**)	−903.2188189	−21.56
G^+−OSO_3 (**9**)	−903.2094419	
G^+···SO_4···G^+ (**10**)	−1108.5093054	−327.84
G^+···HSO_4···G(**11**)	−1108.5110424	−122.87
G^+···SO_4···G^+ (**12**)	−1108.5068796	−326.32
G^+···SO_4···G^+ (**13**)	−1108.5072762	−326.57
G^+−OSO_3···G^+ (**14**)	−1108.5028902	−102.08
G^+−OSO_3···G^+ (**15**)	−1108.5055215	−103.73

(VETQIU, see Table 2), although the O···H distances found for the optimized structures **8** and **9** are closer to those observed in the solid-state structures (between 2.04 and 2.17 Å).[19] These short HB distances and the occurrence of proton transfer is most probably a consequence of the mismatch in charge between the dinegative anion and the monocation within the 1:1 species and the desire for the system to offset this situation.

The total energies of these structures at the MP2/6-31+G* level and the interaction energies are presented in Table 6. Thus, the most stable dimer is **8**, being 6.28 kcal mol^{-1} more stable than **6**. The presence of an extra N−H···O HB interaction in complex **8** could justify the additional stability. The difference in energy between **6** and **8** is quite high, but structure **6** is the most related to that found in the crystal structure, and, thus, subsequent analyses of the interactions established within both complexes will be performed. Compound **9** cannot be compared with complexes **6** and **8** since is a monomer and will not be further discussed.

In the case of the trimers, the starting systems were built from the optimized dimers, and the MP2 optimized structures are shown in Figure 4. Thus, complex **10**, which was formed by the double interactions observed in the GUANSL01 crystal structure, shows no negative frequencies, and no proton transfer occurred. Considering dimer **6**, in which proton transfer did occur, trimer **11** was built and optimized, showing no negative frequencies at the B3LYP/6-31G* level. It was not possible to obtain any complexes as a result of a double proton transfer (a putative 1:2 sulfuric acid−guanidine complex).

Two symmetrical trimers were obtained. One of them, **12**, is at a minimum energy at the B3LYP/6-31G* level, and in this structure one of the O atoms participates in two HB interactions whereas another participates in four. As well, trimer **13** was found to be an energy minimum at the B3LYP level. Finally, two quite irregular dimer complexes, **14** and **15** (derived from compound **9**), were optimized and displayed no negative frequencies.

The total and interaction energies of these complexes are shown in Table 6. In general, these trimers exhibit very large interaction energies and larger still when all of the species involved are charged. This could indicate that a high electrostatic component is involved in the interactions within the trimers

At the MP2 level, the difference in stability among the trimers is small. The most stable of these complexes is **11**, which is 1.09, 2.61, and 2.36 kcal mol^{-1} more stable than **10** (which is a similar complex but lacks the proton resulting from transfer), **12**, and **13**, respectively. Complexes **6**, **8**, and

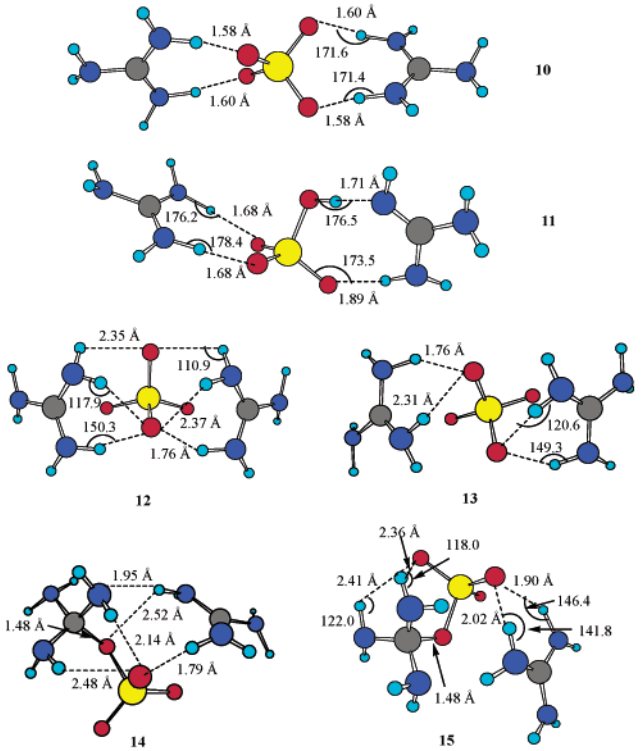


Figure 4. MP2/6-31+G* optimized structures of the 2:1 guanidinium−sulfate trimers **10**−**15**.

Table 7. Electron Density and Laplacian at the BCP (au) of the HB Found in Guanidinium (G^+), Guanidine (G)-Sulfate, Bisulfate Complexes Optimized at the MP2/6-31+G* Level[a]

	DH···A	ρ(BCP)	$\nabla^2\rho$ (BCP)	d(H···A)	<DH···A
G···HSO_4 (**6**)	NH···O	0.0322	0.0984	1.87	165.1
	OH···N	0.0305	0.0891	1.92	174.3
G···HSO_4 (**8**)	NH···O	0.0232	0.0736	2.04	143.3
	NH···O	0.0174	0.0601	2.17	134.0
	OH···N	0.0223	0.0654	2.07	167.6
G^+···SO_4···G^+ (**10**)	NH···O	0.0603	0.1807	1.58	171.8
	NH···O	0.0565	0.1745	1.60	171.5
G^+···HSO_4···G (**11**)	NH···O	0.0480	0.1498	1.68	176.2
	NH···O	0.0481	0.1482	1.68	178.4
	NH···O	0.0282	0.0943	1.89	173.5
	OH···N	0.0493	0.1285	1.71	176.5
G^+···SO_4···G^+ (**12**)	NH···O	0.0423	0.1293	1.76	150.3
	NH···O	0.0153	0.0540	2.34	111.0
	NH···O	0.0118	0.0540	2.37	117.9
G^+···SO_4···G^+ (**13**)	NH···O	0.0423	0.1299	1.76	149.2
	NH···O	0.0161	0.0538	2.31	120.6
G^+−OSO_3···G^+ (**14**)	C−O[b]	0.2298	−0.4981	1.48	118.6
	NH···O[b]	0.0122	0.0442	2.48	119.1
	NH···O[b]	0.0209	0.0734	2.14	119.4
	NH···O	0.0381	0.1202	1.79	156.3
	NH···O	0.0128	0.0438	2.52	102.1
	NH···N	0.0311	0.0830	1.95	164.1
G^+−OSO_3···G^+ (**15**)	C−O[b]	0.2282	−0.4847	1.48	118.2
	NH···O[b]	0.0123	0.0427	2.41	122.0
	NH···O[b]	0.0114	0.0456	2.36	118.0
	NH···O	0.0310	0.0942	1.90	146.4
	NH···O	0.0245	0.0768	2.02	141.8

[a] The interaction distances (Å) and D−H···A angles (°) are also shown. [b] Interactions within the H_6N_3C−OSO_3^- molecule.

11 are the only ones with a noncharged monomer (guanidine), and this may be the reason for the relatively lower interaction energies. However, the interaction energies of these complexes are smaller than those of the complexes with all charged monomers (guanidinium and sulfate). Again, a highly attractive polar interaction cannot be dismissed.

In terms of the electron density analysis, several HB have been identified for the dimers based on the values of the electron density (around 10^{-2} au), the Laplacian (all positive), and bond distances and angles (see Table 7).

Thus, in the guanidine−bisulfate complex **6**, two strong intermolecular HB have been identified, with large electron density values and short HB distances. One of these HB is established between the N−H of the guanidine and one of the O atoms of the bisulfate, with the other between the N of the guanidine and the O−H of the bisulfate. In complex **8**, three different BCP have been identified, and by considering the values of the electron density and the medium HB distances (∼2 Å), all of them correspond to medium strength HBs. In addition to S−O···H−N and O−H···N interactions (similar to those seen in **6**), a third S−O···H−N interaction is established. This explains the larger interaction energy found for **8** compared with **6**.

Four BCP were found for trimers **10** and **11** (see Table 7 and Figure 4). In the case of **10** two of the BCP are symmetrical with short H···O distances and a very large ρ(BCP), which corresponds to two very strong HBs. Trimer **11** exhibits three strong O···H HBs and one strong N···H HB, all with short bond distances and large electron density values at their BCPs. The symmetrical trimer **12** shows a total of six medium-to-strong HB interactions (three on each side of the symmetry plane) involving two NH_2 groups from each guanidinium cation (see Figure 4 and Table 7). Trimer **13** is also symmetrical and shows one strong and one medium HB involving two of the NH_2 groups of each guanidinium cation (see Table 7). Finally, **14** and **15** exhibit a complicated network of inter- and intramolecular interactions in which different BCPs were found. Their electron densities and interaction bond distances are characteristic of medium-to-strong HBs (see Figure 4 and Table 7). In complexes **14** and **15** a BCP was found between the C atom of the guanidinium and one of the O atoms of the sulfate moiety. The ρ(BCP) values were very high, and the Laplacian of those ρ(BCP) was always negative, clearly indicating the covalent nature of those bonds.

As in our previous work[18] an excellent logarithmic correlation was found between the HB distances and the ρ(BCP) values for each of the HB interactions in both dimer and trimer species. Considering all of the N−H···O, N−H···N, and O−H···N interactions found within the present study the following relationship was derived: $d(\mathrm{H}\cdots\mathrm{X}) = -0.53\mathrm{Ln}(\rho(\mathrm{BCP})) + 0.06$ ($R^2 = 0.97$, $n = 25$). This equation not only ratifies the strong dependence between ρ(BCP) and the HB distance but also allows for the prediction of ρ(BCP) values if a certain HB distance is known.

With regard to the NBO analysis, both the charge transferred and the orbital interaction energies were calculated for each interaction found within the complexes (see Table 8).

Strong charge transfer is observed ($\sim 10^{-1}$ e) in those complexes in which each of the interacting species is charged and compares with a charge transfer of around $\sim 10^{-2}$ e for those complexes in which one of the interacting species is neutral (complexes **6**, **8**, and **11**). This observation corroborates the high E_I found in the former complexes (all charged) and indicates that a significant electrostatic force is present and that the HBs are not the only forces responsible for the interactions.

In terms of the orbital interaction energy ($E(2)$) calculated, the NBO analysis shows that in all the N−H···O HBs observed, a lone pair of one of the O atoms of the sulfate or bisulfate anions interacts with antibonding N−H orbitals (from one or two guanidinium molecules). In the case of the O−H···N and N−H···N connections the interaction takes place between a lone pair of one of the N atoms and the antibonding O−H or N−H orbitals, respectively. All of the antibonding molecular orbitals show a very large contribution from the *p*-orbitals based on the N and C atoms. All of the computed $E(2)$ values are in agreement with the medium-to-strong nature of the HB interactions and considering the different nature of the HBs found within these complexes, it is logical that the values are spread over a wide range (from 0.37 to 48.98 kcal mol^{-1}).

The strength of any HB is indicated by considering the electron density in the BCP, and this value should correlate with the orbital interaction energy, which itself gives an idea

Table 8. Charge Transfer (e) and Orbital Interaction Energy (kcal mol^{-1}) Calculated for the Guanidinium (G^+), Guanidine (G)-Sulfate, Bisulfate Complexes at the B3LYP/6-31+G* Level with the NBO Method

	charge transfer		$E(2)$	
G···HSO_4 (**6**)	$SO_4H \rightarrow G$	0.00541	N−H···O	9.71
			O−H···N	15.50
G···HSO_4 (**8**)	$SO_4H \rightarrow G$	0.01085	N−H···O	6.10
			N−H···O	2.86
			O−H···N	3.44
G^+···SO_4···G^+ (**10**)			N−H···O	37.22
	$SO_4H \rightarrow G^+$	0.15584	N−H···O	26.87
G^+···HSO_4···G (**11**)			N−H···O	27.35
	$SO_4H \rightarrow G$	0.0579	N−H···O	8.56
			O−H···N	39.61
	$SO_4 \rightarrow G^+$	0.11638	N−H···O	13.38
G^+···SO_4···G^+ (**12**)			N−H···O	1.96
			N−H···O	0.96
	$SO_4H \rightarrow G$	0.00541	N−H···O	9.71
G^+···SO_4···G^+ (**13**)	$SO_4 \rightarrow G^+$	0.11731	N−H···O	21.29
			N−H···O	2.68
G^+−OSO_3···G^+ (**14**)	G^+−$OSO_3 \rightarrow G^+$	0.12488	N−H···O[a]	2.11
			N−H···O[a]	0.84
			N−H···O	12.76
			N−H···O	0.37
			N−H···N	19.00
	G^+−$OSO_3 \rightarrow G^+$	0.11143	N−H···O[a]	1.11
G^+−OSO_3···G^+ (**15**)			N−H···O[a]	1.52
			N−H···O	7.33
			N−H···O	11.32
	$SO_4 \rightarrow G^+$	0.11731	N−H···O	21.29

[a] Interactions within the H_6N_3C−OSO_3^- molecule.

of the nature of the bond. We first observed such a correlation in previous studies.[20] Thus, all of the $E(2)$ orbital energy values were correlated to those of the corresponding ρ(BCP), and the following relationship was found: $E(2) = 815.92\rho(\text{BCP}) - 11.46$ ($R^2 = 0.90$, $n = 25$). As before, two different trends were observed in the linear plot, and therefore two separate correlations were established. One correlation was found for those interactions involving N as the HB acceptor, and the following relationship was established: $E(2) = 1310.60\rho(\text{BCP}) - 24.26$, $R^2 = 0.99$; it must be mentioned that there were only four cases to be considered. The other correlation took into account the remaining 21 cases in which an O atom was the HB acceptor, and this yielded the following relationship: $E(2) = 763.63\rho(\text{BCP}) - 10.65$, $R^2 = 0.92$.

Conclusions

From our computations at a higher level of theory, it may be concluded that the structure of the guanidinium cation is not planar but can be perceived as two interchangeable mirror image propellers passing through a planar transition state with a very small cost of energy. In this case, DFT calculations performed similarly to other higher and more expensive levels of computation.

Structures of minimum energy have been found for the guanidinium−chloride and guanidinium−chloride−guanidinium dimer and trimer complexes. The geometrical, AIM and NBO analysis of these complexes indicated the presence of strong HB between two H of the guanidinium cations and the chloride anion in a pincerlike disposition. The interactions observed in the dimer are in agreement with the crystal structures found in the CSD, and this kind of arrangement might be anticipated within a putative crystal structure of the trimer.

In the case of the sulfate−guanidinium dimers and trimers, several energy minima have been localized, and in all the cases numerous HB interactions have been found to be formed between both charged species making very strongly bound complexes. The tetrahedral disposition of the sulfate anion opens a wide range of possible interactions. This is reflected by the complex crystal structure found in the CSD where a unique sulfate molecule is surrounded by six guanidinium cations each one interacting in a different way with the dianion. Thus, compounds bearing guanidinium cations would be able to form a variety of very stable complexes with this anion, which augurs well for its future use in structural supramolecular chemistry.[21]

Computations yielded compound **9**, which is a gem-triamine that has, to the best of our knowledge, no known precedent within the literature. However, as a theoretical tool, it gives a good idea of all the possibilities of how the two ionic species may potentially interact.

In general, the interaction energies of all these dimers and trimers are extremely large, while the electron densities at the BCP and the $E(2)$ values are those expected for a HB. This could indicate that in these charged complexes a large polar contribution is being added to that of the formation of HBs resulting in extremely strong interactions. Furthermore, we have verified that a correlation exists between the electron density in the BCP and the orbital interaction energy and that this gives an idea of both the nature and strength of any HB.

Acknowledgment. Funds from the Irish Higher Education Authorities Program for Research in Third Level Institutions (PRTLI) through the Centre for Synthesis and Chemical Biology supported this research. We also thank Prof. Jose Elguero, Dr. Raj S. Chari, and Dr. Anthea C. Lees for helpful discussions.

References

(1) (a) de Silva, A. P.; Gunaratne, H. Q. N.; Gunnlaugsson, T.; Huxley, A. J. M.; McCoy, C. P.; Rademacher, J. T.; Rice, T. E. *Chem. Rev.* **1997**, *97*, 1515. (b) Suksai, C.; Tuntulani, T. *Chem. Soc. Rev.* **2003**, *32*, 192. (c) Lee, D. H.; Lee, H. Y.; Hong, J.-I. *Tetrahedon Lett.* **2002**, *43*, 7273. (d) Beer, P. D. *Chem. Commun.* **1996**, 689. (e) Schmidtchen, F. P.; Berger, M. *Chem. Rev.* **1997**, *97*, 1609. (f) Beer, P. D.; Smith, D. K. *Prog. Inorg. Chem.* **1997**, *46*, 1. (g) Atwood, J. L.; Holman, K. T.; Steed, J. W. *Chem. Commun.* **1996**, 1401. (h) Gunnlaugsson, T.; Kruger, P. E.; Jensen, P.; Pfeffer, F. M.; Hussey, G. M. *Tetrahedron Lett.* **2003**, *44*, 8909. (i) Gunnlaugsson, T.; Kruger, P. E.; Lee, T. C.; Parkesh, R.; Pfeffer, F. M.; Hussey, G. M. *Tetrahedron Lett.* **2003**, *44*, 6575. (j) Kruger, P. E., Mackie, P. R.; Nieuwenhuyzen, M. *J. Chem. Soc., Perkin Trans. 2* **2001**, 1079.

(2) (a) Gale, P. A. *Coord. Chem. Rev.* **2001**, *213*, 79. (b) Gale, P. A. *Coord. Chem. Rev.* **2000**, *199*, 181. (c) Beer, P. D.; Gale, P. A. *Angew. Chem., Int. Ed.* **2001**, *40*, 486. (d) Bowman-James, K.; Garcia-España, E. In *Supramolecular Chemistry of Anions*; Bianchi, A., Ed.; Wiley-VCH: New York, 1997.

(3) Best, M. D.; Tobey, S. L.; Anslyn, E. V. *Coord. Chem. Rev.* **2003**, *240*, 3, and references therein.

(4) Rekharsky, M.; Inoue, Y.; Tobey, S.; Metzger, A.; Anslyn, E. V. *J. Am. Chem. Soc.* **2002**, *124*, 14959.

(5) (a) Weber, D. J.; Meeker, A. K.; Mildvan, A. S. *Biochemistry* **1990**, *29*, 8632. (b) Judice, J. K.; Gamble, T. R.; Murphy, E. C.; de Vos, A. M.; Schultz, P. G. *Science* **1993**, *261*, 1578. (c) Hannon, C. L.; Anslyn, E. V. In The Guanidinium Group: Its Biological Role; Synthetic Analogues. In *Bioorganic Chemistry Frontiers;* Dugas, H., Schmidtchen, F. P., Eds.; Springer: Heidelberg, 1993; p 193.

(6) Keegan, J.; Kruger, P. E.; Nieuwenhuyzen, M.; O'Brien, J.; Martin, N. *Chem. Commun.* **2001**, 2192.

(7) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(8) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(9) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(10) Hariharan, P. A.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(11) Krishnam, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1984**, *80*, 3265.

(12) (a) Bader, R. F. W. In *Atoms in Molecules. A Quantum Theory;* Oxford University: New York, 1990. (b) Bieger-Konig, F. W.; Bader, R. F. W.; Tang, T. H. *J. Comput. Chem.* **1980**, *27*, 1924.

(13) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(14) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187.

(15) Gobbi, A.; Frenking, G. *J. Am. Chem. Soc.* **1993**, *115*, 2362.

(16) (a) Angell, C. L.; Sheppard, N.; Yamaguchi, A.; Shimanouchi, T.; Miyazawa, T.; Mizushima, S. *Trans. Faraday Soc.* **1957**, *53*, 589. (b) Drenth, J.; Drenth, W.; Vos, A.; Wiebenga, E. H. *Acta Crystallogr.* **1953**, *6*, 424. (c) Otvos, J. W.; Edsall, J. T. *J. Chem. Phys.* **1939**, *7*, 632. (d) Kellner, L. *Proc. R. Soc. London* **1941**, *A177*, 456.

(17) Paulini, R.; Muller, K.; Diederich, F. *Angew. Chem., Int. Ed.* **2005**, *44*, 2.

(18) (a) Alkorta, I.; Rozas, I.; Elguero, J. *J. Struct. Chem.* **1998**, *9*, 243. (b) Alkorta, I.; Barrios, L.; Rozas, I.; Elguero, J. *J. Mol. Struct. (THEOCHEM)* **2000**, *496*, 131. (c) Espinosa, E.; Souhassou, M.; Lachekar, H.; Lecomte, C. *Acta Crystallogr.* **1999**, *B55*, 563.

(19) It should be noted that the geometry of H-bonds taken from X-ray crystal data that the positions of H-atoms correspond to electron density maxima which are translated toward the heavier atom to which they are connected, and hence bond lengths are usually shorter than the corresponding distances between nuclei.

(20) Rozas, I.; Alkorta, I.; Elguero, J. *J. Phys. Chem. B* **2004**, *108*, 3335.

(21) Grossel, M. C.; Merckel, D. A. S.; Hutchings, M. G. *Cryst. Eng. Comm.* **2003**, *5*, 77.

CT050009X