JCTC Journal of Chemical Theory and Computation

A New Algorithm for Efficient Direct Dynamics Calculations of Large-Curvature Tunneling and Its Application to Radical Reactions with 9−15 Atoms

Antonio Fernández-Ramos

Departmento de Quimica Fisica, Facultade de Quimica, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S. E., Minneapolis, Minnesota 55455-0431

Received June 15, 2005

Abstract: We present a new algorithm for carrying out large-curvature tunneling calculations that account for extreme corner-cutting tunneling in hydrogen atom, proton, and hydride transfer reactions. The algorithm is based on two-dimensional interpolation in a physically motived set of variables that span the space of tunneling paths and tunneling energies. With this new algorithm, we are able to carry out density functional theory direct dynamics calculations of the rate constants, including multidimensional tunneling, for a set of hydrogen atom transfer reactions involving 9−15 atoms and up to 7 nonhydrogenic atoms. The reactions considered involve the abstraction of a hydrogen atom from hydrocarbons by a trifluoromethyl radical, and in particular, we consider the reactions of CF_3 with CH_4, C_2H_6, and C_3H_8. We also calculate several kinetic isotope effects. The electronic structure is treated by the MPWB1K/6-31+G(d,p) method, which is validated by comparison to experimental results and to CBS-Q, MCG3, and G3SX(MP3) calculations for $CF_3 + CH_4$. Harmonic vibrational frequencies along the reaction path are calculated in curvilinear coordinates with scaled frequencies, and anharmonicity is included in the lowest-frequency torsion.

1. Introduction

Direct dynamics is "the calculation of rates or other dynamical observables directly from electronic structure information, without the intermediacy of fitting the electronic energies in the form of a potential energy function."[1] Although most direct dynamics calculations are based on classical mechanics for the nuclear motion,[2−6] there has also been progress in including quantized vibrations and tunneling.[1,7−12]

To make direct dynamics practical, one often uses inexpensive electronic structure methods, as in dynamics calculations based on semiempirical valence bond configuration interaction,[3] semiempirical molecular orbital theory,[1,2,8,9,12] or tight-binding molecular dynamics.[5] Using an affordable electronic structure method allows one to combine multidimensional tunneling calculations with variational transition-state theory (VTST/MT) for relatively large systems. Therefore, one can calculate thermal rate constants at a low computational cost without building an analytical potential energy surface. For example, one can use semiempirical molecular orbital theory with specific-reaction parameters (SRPs) fit to experimental or selected higher-level calculations.[1,9,12] This approach, although very successful in some cases, has the drawback that SRP surfaces do not form a model chemistry[13] and, hence, cannot be broadly validated. Another approach to lowering the cost is to use automatic and efficient fitting methods.[14] The third approach, which is

* Corresponding author. Phone: (612) 624-7555. Fax: (612) 624-9390. E-mail: editor@chem.umn.edu.

10.1021/ct050153i CCC: $30.25
Published on Web 08/25/2005

the one considered in the present article, is to improve the dynamical algorithms.

For reliable calculations of rate constants in reactions involving the transfer of a hydrogen atom, hydride ion, or proton, it is essential to include corner-cutting tunneling[7,9,15−19] in the calculation. The VTST/MT method is a well-validated[20,21] method for calculating reaction rate constants, including the contribution of corner-cutting tunneling, and when reaction-path curvature is small, it is very efficient because all required data can be obtained from a harmonic expansion of the potential around one[22,23] or more[12] minimum energy paths (MEPs). However, especially for bimolecular reactions involving the transfer of a light particle between two heavy atoms (the so-called heavy−light−heavy systems), additional information is required in the reaction swath (tunneling swath), which is defined[7,24] as the region on the concave side of the MEP where corner-cutting tunneling occurs. Large-curvature tunneling[9,17,19] (LCT) is tunneling that passes through the swath at a distance too far from the MEP to be uniquely or realistically represented by the aforementioned harmonic expansion. Proton, hydride, and hydrogen transfer reactions, especially those with symmetric or nearly symmetric barriers, often involve LCT in which the corner cutting is due to the reaction coordinate being strongly coupled to the hydrogen stretching vibration mode, resulting in a highly curved MEP. In fact, the most probable tunneling path is the best compromise between low-energy but long tunneling paths, near the MEP, and paths that are less favorable energetically but shorter, such as straight-line paths. The optimum path is the one with the least imaginary action.[18] It has been shown[20] that a good approximation to the tunneling probabilities calculated from the least-action path can be obtained by the microcanonical optimized multidimensional tunneling (μOMT) approximation,[9] in which tunneling probability, at each tunneling energy, is the maximum between the small-curvature tunneling (SCT) probability[8,25] evaluated with information obtained along the MEP and the LCT probability[9,17,19,25,26] evaluated with information from the reaction swath. One of the difficulties in evaluating the μOMT transmission coefficients by direct dynamics is the evaluation of the LCT probabilities, because, at least with current algorithms, they are very demanding in terms of the number of electronic structure calculations required for evaluations of the potential. In the present paper, we develop an approach based on a 2D-spline under tension that reduces the computer time by about 2 orders of magnitude without a loss of accuracy.

The gas-phase reactions in which the trifluoromethyl radical (CF_3) abstracts a hydrogen atom from hydrocarbons[27−32] are good examples of heavy−light−heavy systems with corner-cutting tunneling.[9] In this manuscript, we study the following reactions:

(A) CF_3 radical with methane:

$$CF_3 + CH_4 \rightarrow CF_3H + CH_3 \quad \text{(R1)}$$

$$CF_3 + HCD_3 \rightarrow CF_3H + CD_3 \quad \text{(R2)}$$

$$CF_3 + HCD_3 \rightarrow CF_3D + CD_2H \quad \text{(R3)}$$

$$CF_3 + CD_4 \rightarrow CF_3D + CD_3 \quad \text{(R4)}$$

Table 1. Vibrational-Frequency Scaling Factors and Energies (in kcal/mol Relative to Reactants) of Products and Saddle Point of Reaction R1

method	S	ΔE	ΔH_0	$V^{\ddagger}$	$\Delta V_a^{\ddagger}$
experiment		−0.7[a]	(−1.5)[b]		
G3SX(MP3)//MC-QCISD/3	0.9930	**−1.41**	**−2.42**	14.80	12.49
CBS-Q	0.9184[c]	**−0.91**	**−2.27**	**14.34**	**12.23**
MCG3/3//MC-QCISD/3	0.9930	**−1.53**	**−2.54**	14.69	12.38
MC3BB	0.9675[d]	0.84	−0.53	**15.09**	**12.68**
MC3MPW	0.9669[d]	−0.26	−1.46	**14.74**	**12.32**
MC-QCISD	0.9930	−1.33	−2.34	16.63	14.42
consensus[e]		−1.28	−2.41	14.72	12.41
MPWB1K/MG3S	0.9567[f]	0.84	−0.38	15.09	12.68
BB1K/MG3S	0.9590	1.00	−0.17	15.68	13.41
MPW1K/MG3S	0.9581	−0.01	−1.33	14.97	12.69
MPWB1K/DIDZ	0.9537	0.32	−0.85	14.39	12.06
BB1K/DIDZ	0.9561	0.51	−0.59	14.99	12.82
MPW1K/DIDZ	0.9515	−0.72	−1.83	14.42	12.23

[a] Liu et al.[23] [b] 298 K.[41] [c] HF/6-31G(d′).[33] [d] Zhao et al.[40] [e] Average of three bold entries. [f] Zhao and Truhlar.[39]

(B) CF_3 radical with ethane:

$$CF_3 + C_2H_6 \rightarrow CF_3H + C_2H_5 \quad \text{(R5)}$$

$$CF_3 + C_2D_6 \rightarrow CF_3D + C_2D_5 \quad \text{(R6)}$$

(C) CF_3 radical with propane:

$$CF_3 + C_3H_8 \rightarrow CF_3H + C_3H_7 \quad \text{(R7)}$$

$$CF_3 + C_3D_8 \rightarrow CF_3D + C_3D_7 \quad \text{(R8)}$$

We further subdivide R7 and R8 into R7p and R8p, in which a primary H or D is abstracted, and R7s and R8s, in which a secondary H or D is abstracted.

We first show that the new algorithm proposed to evaluate the LCT probabilities allows the evaluation of μOMT transmission coefficients for this kind of system with a reasonable computing time. Then we show that density functional theory (DFT) methods can be used for direct dynamics calculations of the thermal rate constants with optimized multidimensional tunneling even for reactions as complex as CF_3 with propane, which is a difficult case because it involves seven heavy (i.e., nonhydrogenic) atoms (the cost of the electronic structure calculations increases rapidly with the number of heavy atoms).

Section 2 presents the electronic structure calculations used to select a level of theory for the direct dynamics calculations. Section 3 presents the dynamical theory including the new 2D interpolation scheme. Section 4 gives details of the calculations. Section 5 gives results, and Section 6 gives a discussion.

2. Electronic Structure Calculations

The direct dynamics calculations in this paper are based on DFT because it provides a good compromise of affordability and accuracy. Several types of electronic structure calculations were performed for reaction R1 to find a reliable DFT method for the set of reactions R1, R5, and R7; all the methods employed are listed in Table 1. For bond energies, the most accurate methods employed here are CBS-Q,[33]

G3SX(MP3),[34] MC-QCISD/3,[35] and MCG3/3.[35] In tests against a data set of 109 atomization energies, these methods had mean unsigned errors per bond of 0.30, 0.22, 0.37, and 0.22 kcal/mol, respectively, and in a test against 44 barrier heights for H-atom transfer, they had mean unsigned errors of 0.87, 0.94, 1.33, and 1.01 kcal/mol, respectively.[36] In the present article, we employed MCG3/3 and G3SX(MP3) with geometrics and frequencies optimized at the MC-QCISD/3 level, whereas CBS-Q geometries were obtained as specified[33] by the developers of that method. In calculating zero-point energies in Table 1, all frequencies were scaled with empirical factors S, determined previously;[33,37−40] these factors are given in the table. (The use of the harmonic expressions with scaled frequencies may be called the quasiharmonic approximation.) For barrier heights, two more recently developed methods, MC3BB[40] and MC3MPW,[40] have even smaller mean unsigned errors, in particular, 0.78 and 0.75 kcal/mol, respectively. Therefore, G3SX(MP3), CBS-Q, and MCG3 are expected to be the most accurate of the methods employed here for reaction energies, and CBS-Q, MC3BB, and MC3MPW are expected to be the most accurate for barrier heights. We, therefore, used these methods to compute consensus best estimates for the following quantities: ΔE, classical energy of reaction; ΔH_0, zero-point-inclusive energy of reaction; $V^{\ddagger}$, classical barrier height; and $\Delta V_a^{G\ddagger}$, zero-point-inclusive barrier height evaluated at the saddle point. The last-named quantity is given in the harmonic or quasiharmonic approximation for bimolecular reactions by

$$\Delta V_a^{G\ddagger} = V^{\ddagger} + \frac{1}{2}\hbar \sum_{m=1}^{3N-7} [\omega_m(s=0) - \omega_m(s=-\infty)] \quad (1)$$

where $\hbar$ is Planck's constant divided by 2π, $\omega_m(s)$ is a generalized normal mode vibrational frequency in s^{-1} (that is, radians per second), and $\omega_m(s=0)$ and $\omega_m(s=-\infty)$ are the vibrational frequencies of mode m at the saddle point and the reactants, respectively. (For a unimolecular reaction, one would need to subtract $^1/_2\hbar\omega_{3N-6}(s=0)$ at the reactants, but this is zero for a bimolecular reaction. Note that when we quote numerical values of $\omega_m(s)$, we always give $\omega_m(s)/(2\pi c)$ in cm^{-1}, where c is the speed of light.) The consensus values of these four quantities are listed in Table 1, and they may be more accurate than the available experimental[41] values.

All of the levels discussed so far are more expensive than the DFT levels (given in the next sentence) that we considered for direct dynamics, and they only serve as benchmark calculations. The three DFT methods given in Table 1, namely, MPW1K,[37] BB1K,[38] and MPWB1K,[42] have previously been optimized for kinetics. In each, they have been applied with two basis sets: an augmented polarized triple-ζ basis set, MG3S,[43] and an augmented polarized double-ζ basis set, 6-31+G(d,p),[44] also called DIDZ. Table 1 indicates that the MPW1K/MG3S, MPWB1K/DIDZ, and MPW1K/DIDZ results are all reasonably accurate for this reaction. Although MPW1K is slightly closer to the experimental value (which is very approximate) and the high-level calculations, we chose MPWB1K/6-31+G(d,p) for use in the VTST dynamics calculations for all reactions studied here because of MPWB1K's better performance on broader test sets, as shown in Tables 6 and 12 of ref 42. In any event, the difference in the zero-point-inclusive barrier heights evaluated at the saddle point for these two methods of calculating the surface is only 0.17 kcal/mol. The values of the four energetic parameters calculated at the MPWB1K/DIDZ level are listed in Table 2 for reactions R1, R5, and R7.

Table 2. Energetic Parameters (in kcal/mol) for Reactions R1, R5, and R7 Calculated at the MPWB1K/DIDZ Level[a]

	R1	R5	R7s	R7p1	R7p2
ΔE	0.32	−4.12	−7.66	−3.88	−3.88
ΔH_0	−0.85	−5.43	−8.98	−5.08	−5.08
	(−1.49)[b]	(−5.73)	(−7.74)	(−5.35)	(−5.35)
$V^{\ddagger}$	14.39	11.67	9.47	11.90	11.62
$\Delta V_a^{G\ddagger}$	12.06	8.94	6.52	9.16	9.21

[a] In the case of reaction R7, different energetic parameters are obtained for the abstraction of secondary, R7s, or primary, R7p1 and R7p2, hydrogens and also for each of the two possible transition-state conformers (R7p1 and R7p2) for the primary case (see Section 5.3 of text). [b] Numbers in parentheses are ΔH_{298} from experiment.[41]

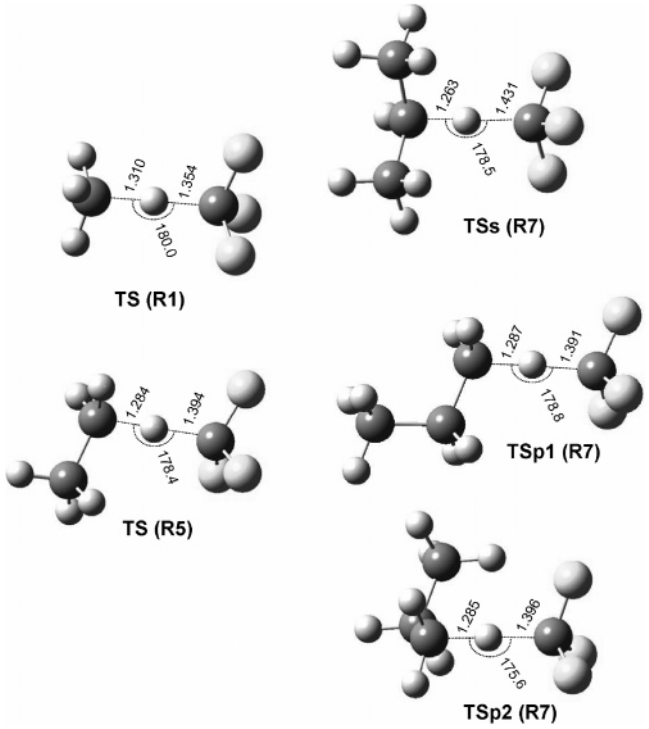


Figure 1. Geometries of the transition states of the R1, R5, and R7 reactions calculated at the MPWB1K/DIDZ level. Distances are in Å, and angles are in degrees.

The geometries of the transition states of reactions R1, R5, and R7 obtained at the MPWB1K/G-31+G(d,p) level are displayed in Figure 1. The transition state for reaction R1 is the only one of the three in which the hydrogen and carbon atoms directly involved in the abstraction are collinear. From the energetic point of view, reactions R5 and R7p are similar, and this is also reflected in the C−H distances of the abstracted hydrogen, which are also similar. Reaction R7s is the one with the lowest barrier and is the most exothermic, and in keeping with Hammond's postu-

late,[45] the transition state is closer to the reactants than it is for the other transition states.

3. Dynamics Calculations

All dynamics calculations were performed by VTST/MT.[9,22,23] This theory minimizes the error due to recrossing trajectories by locating the transition-state dividing surface[46,47] orthogonal to the MEP at a position s along the MEP at which the flux through this surface is minimum, where s is the arc length along the MEP measured from the saddle point. For assigning numerical values of s, all coordinates are scaled to a reduced mass of 1 amu. The set of dividing surfaces over which the flux is minimized is known as the set of generalized transition states[22,23] (GTs). When the flux is minimized for a canonical ensemble, we obtain canonical variational theory (CVT), in which the thermal rate constant, $k^{\mathrm{CVT}}(T)$, at temperature T is the minimum of the generalized transition-state theory rate constant, $k^{\mathrm{GT}}(T,s)$, as a function of s, that is,

$$k^{\mathrm{CVT}}(T) = \min_s k^{\mathrm{GT}}(T,s) = \sigma\frac{k_{\mathrm{B}}T}{h}\frac{Q^{\mathrm{GT}}(T,s_*^{\mathrm{CVT}})}{\Phi^{\mathrm{R}}(T)}\exp[-V_{\mathrm{MEP}}(s_*^{\mathrm{CVT}})/k_{\mathrm{B}}T] \quad (2)$$

where k_{B} is the Boltzmann's constant, h is Planck's constant, σ is the symmetry factor that accounts for the reaction path multiplicity, s_*^{CVT} is the value of s at which the generalized transition-state theory rate constant is a minimum, $V_{\mathrm{MEP}}(s_*^{\mathrm{CVT}})$ is the value of the MEP potential at s_*^{CVT}, $Q^{\mathrm{GT}}(T,s_*^{\mathrm{CVT}})$ is the internal quantum partition function at s_*^{CVT}, and $\Phi^{\mathrm{R}}(T)$ is the product of the relative translational partition function per unit volume and the internal quantum mechanical partition functions of reactants. Both Q^{GT} and Q^{R} are calculated without overall-rotational symmetry numbers, since symmetry is in the σ factor. The internal partition functions are products of the electronic, rotational, and vibrational partition functions, and their zero of energy is at the minimum of the potential, not at the zero-point level. The rotational partition function is approximated as classical, and the vibrational one is a product of separable-mode contributions:

$$Q_{\mathrm{vib}}^{\mathrm{GT}}(T,s) = \prod_{m=1}^{F-1} Q_{\mathrm{vib},m}^{\mathrm{GT}}(T,s) \quad (3)$$

where $F = 3N - 6$ and N is the number of atoms. The partition functions for all the vibrational degrees of freedom except the lowest-frequency mode are quantal harmonic oscillators. For the reactions studied here, the lowest-frequency mode (labeled below as $F - 1$) corresponds to the 3-fold internal rotation of the methyl, ethyl, or propyl group with respect to the CF_3 group and is treated as a hindered rotor. Specifically, the following function is used:

$$V_{\mathrm{vib},F-1}^{\mathrm{GT}}(T,s) = Q_{F-1}^{\mathrm{HO}}(T,s)\tanh\left[\frac{\hbar\omega_{F-1}(s)}{k_{\mathrm{B}}T}Q_{F-1}^{\mathrm{FR}}(T,s)\right] \quad (4)$$

where $\omega_{F-1}(s)$ is the frequency of the internal rotational mode, $Q^{\mathrm{HO}}(T,s)$ is the harmonic oscillator partition function of this mode, given by

$$Q_{F-1}^{\mathrm{HO}}(T,s) = \frac{\exp[-\hbar\omega_{F-1}(s)/2k_{\mathrm{B}}T]}{1-\exp[-\hbar\omega_{F-1}(s)/k_{\mathrm{B}}T]} \quad (5)$$

and $Q^{\mathrm{FR}}(T,s)$ is the free rotor partition function given by

$$Q_{F-1}^{\mathrm{FR}}(T,s) = \frac{[2\pi\, I(s)\, k_{\mathrm{B}}T]^{1/2}}{\hbar\sigma_{F-1}} \quad (6)$$

where σ_{F-1} is the symmetry number for internal rotation (3 for all reactions studied here) and $I(s)$ the reduced moment of inertia of the two counter rotating tops. We use the method[48] (called CW, where C denotes "curvilinear") in which $I(s)$ is obtained from a curvilinear-coordinate description[49] of the torsion and W, which is the barrier for internal rotation, is obtained from electronic structure calculations in which the other degrees of freedom are relaxed while the internal rotation angle is scanned. The frequency, $\omega_{F-1}(s)$ is obtained from the equation

$$\omega_{F-1} = [W_{F-1}/2I(s)]^{1/2}\sigma_{F-1} \quad (7)$$

Because the barrier is very small for these reactions, the partition function for mode $F - 1$ is close to the free-internal-rotor limit.

Tunneling is incorporated into the rate constant by a ground-state transmission coefficient,[9,22,23] $\kappa^{\mathrm{CVT/G}}(T)$, so the final thermal rate constant is given by

$$k^{\mathrm{CVT}/\mu\mathrm{OMT}}(T) = \kappa^{\mathrm{CVT/G}}(T)\, k^{\mathrm{CVT}}(T) \quad (8)$$

Depending on the approximation used in the evaluation of the ground-state tunneling probability, $P^{\mathrm{G}}(E)$, different transmission coefficients for tunneling are obtained. As mentioned in the Introduction, a reliable method to evaluate tunneling effects is the microcanonical optimized multidimensional tunneling transmission factor $\kappa^{\mathrm{CVT}/\mu\mathrm{OMT}}(T)$, for which the probability is obtained as

$$P^{\mu\mathrm{OMT}}(E) = \max\begin{Bmatrix} P^{\mathrm{SCT}}(E) \\ P^{\mathrm{LCT}}(E) \end{Bmatrix} \quad (9)$$

where E is the tunneling energy and $P^{\mathrm{SCT}}(E)$ and $P^{\mathrm{LCT}}(E)$ are the transmission probabilities obtained by the centrifugal-dominant small-curvature semiclassical adiabatic ground state (also called SCT) approximation[8,25] and the large curvature approximation version 4 (LCG4, also called LCT) method,[26] respectively. We will also briefly consider the zero-curvature tunneling[15,22] (ZCT) approximation, which may be considered to be an approximation of SCT in which tunneling occurs along the MEP. An important quantity in all these tunneling theories is the vibrationally adiabatic ground-state potential curve given by

$$V_{\mathrm{a}}^{\mathrm{G}}(s) = V_{\mathrm{MEP}}(s) + \frac{1}{2}\hbar\sum_{m=1}^{F-1}\omega_m(s) \quad (10)$$

where $\omega_m(s)$ are the generalized-normal-mode vibrational frequencies[22,23] at the point that is a distance s along the MEP.

The maximum value of $V_a^G(s)$ is $V_a^{AG}(s)$, and the quantity $\Delta V_a^{G\ddagger}$ discussed above is $V_a^G(s = 0)$ minus $V_a^G(s = -\infty)$.

The LCT probabilities are evaluated along straight-line paths connecting a classical turning point of the vibrationally adiabatic ground-state potential curve on the reactant side ($\tilde{s}_0 < 0$) to a classical turning point of a vibrational potential in the product side ($\tilde{s}_1 > 0$). For ground-state-to-ground-state tunneling, at a given tunneling energy E_{tun}, both turning points satisfy

$$V_a^G(\tilde{s}_i) = E_{tun} \qquad i = 0,1 \tag{11}$$

These points correspond to geometries $\mathbf{x}_0 = \mathbf{x}_{MEP}(\tilde{s}_0)$ and $\mathbf{x}_1 = \mathbf{x}_{MEP}(\tilde{s}_1)$, where $\mathbf{x}$ is a $3N$-dimensional vector specifying the geometry in mass-scaled Cartesian coordinates,[23] also called isoinertial coordinates, and the geometries in isoinertial coordinates along the straight-line tunneling path are

$$\mathbf{x}(\tilde{s}_0,\xi) = \mathbf{x}_0 + \frac{\xi}{\xi_P}(\mathbf{x}_1 - \mathbf{x}_0) \tag{12}$$

where ξ is a progress variable along the tunneling path. The length of the tunneling path is $\xi_P = |\mathbf{x}_1 - \mathbf{x}_0|$. With these definitions, a given tunneling path is specified by $\tilde{s}_0$, and a given point along the straight path is specified by $\tilde{s}_0$ and ξ. At a total energy E, there are contributions from all allowed values of E_{tun} with $E_{tun} \le E$. A value of E_{tun} is not allowed if it is below the zero-point level of either the reactants or products.

In the LCT method, one uses a harmonic approximation to build the effective potential along the straight-line path from the information along the MEP at those points along the tunneling path where (i) all the generalized normal mode coordinates are within their vibrational turning points and (ii) the geometry $\mathbf{x}(\tilde{s}_0,\xi)$ lies within the single-valued region of the reaction-path coordinates. If one of the above conditions is not obeyed, the straight-line path is considered to be in the nonadiabatic region and extrapolation from the MEP is not possible; instead, one evaluates a nonadiabatic effective potential on the basis of the actual potential at a given point of the path and on the contributions of vibrationally adiabatic modes at the adiabatic−nonadiabatic boundaries.[25,26] For a given tunneling path, the nonadiabatic region corresponds to an interval $\xi_I \le \xi \le \xi_{III}$, with ξ_I and ξ_{III} being the boundaries of the adiabatic regions on the reactant and product sides, respectively. The two adiabatic regions, one corresponding to the reactant valley and one to the product valley, are delimited by the intervals $0 < \xi < \xi_I$ and $\xi_{III} < \xi < \xi_P$. The LCG4 transmission coefficients involve an additional condition; that is, the extrapolated vibrationally adiabatic potential obtained from the MEP information should be larger than the calculated nonadiabatic effective potential. The vibrationally adiabatic potential is obtained from information on the MEP, but the evaluation of the nonadiabatic effective potential involves single-point calculations along each tunneling path. The large number of single-point energies required (usually more than 1000) make the evaluation of LCT transmission coefficients time-consuming.

One way of reducing the computational cost is to calculate single-point energies at selected points along the straight-line path at each tunneling energy. The calculated points are fitted to a one-dimensional spline under tension with a subsequent reduction in computer time. This procedure[50] is called one-dimensional spline interpolated large-curvature tunneling (ILCT1D), and it produces converged transmission coefficients with a mean signed percentage error smaller than 4% with only nine calculated points at each tunneling energy.

Another possibility is not only to interpolate in the progress variable ξ at each tunneling energy E_{tun} but also to interpolate in E_{tun}. This two-dimensional interpolation method is presented in the rest of this paragraph and is based on a two-dimensional spline under tension. We called it the ILCT(2D) algorithm. The procedure is as follows: (i) An $M \times N$ two-dimensional grid is defined in terms of the tunneling energies E_i (with $i = 1, ..., M$) at which tunneling is to be evaluated and the points ξ_j (with $j = 1, ..., N$) to be calculated at each of the specified tunneling energies since $\tilde{s}_0$, $\mathbf{x}_0$, and $\mathbf{x}_1$ in eq 12 are functions of E_i. A general point on a straight-line path has coordinates (E_i, ξ_j) and is related to a unique geometry given by eq 12 for each tunneling energy E_i. (ii) The next step is to calculate the potential for the generic point (E_i, ξ_j). If the point is in the adiabatic region, the energy is calculated by a quadratic expansion from the information along the MEP, and so no additional computer time is needed; if the point is in the nonadiabatic region, the classical potential has to be calculated explicitly. It should be noted that a preliminary search is needed in order to know the boundaries between the adiabatic and the nonadiabatic regions. (iii) The initial grid with points (E_i, ξ_j) is transformed in a square grid with points ($\bar{E}_i$, $\bar{\xi}_j$) by performing the following scaling:

$$\bar{E}_i = \frac{E_i - E_{min}}{E_{max} - E_{min}} \text{ and } \bar{\xi}_j = \frac{\xi_j}{\xi_{P,i}} \tag{13}$$

where E_{min} is the lowest tunneling energy at which the particle can tunnel, E_{max} is the maximum V_a^{AG} of the vibrationally adiabatic potential curve, and $\xi_{P,i}$ is the length of the tunneling path at the given tunneling energy E_i. This grid is fitted to a two-dimensional spline under tension. [See Figure 2, where the point corresponding to E_{max} becomes a line when the region from (0, 0) to ($\bar{\xi}_{j_{max}}$, $\bar{E}_{i_{max}}$) is mapped onto a unit square.] (iv) The effective potentials in the swath are then obtained from the fit.

In LCT tunneling calculations by the LCG4 algorithm (or any of the earlier large-curvature algorithms), the tunneling calculation is always carried out in the exoergic direction (as determined for the ground-state-to-ground-state process at 0 K), and transmission coefficients for the other direction are obtained by detailed balance. Tunneling is initiated in the ground state of the reactants (in the exoergic direction) but into a series of ground and excited states of the product (even for thermoneutral regions). The four steps of the previous paragraph are repeated for each of the directly coupled (diabatic) vibrationally excited final states. If only the ground-state product is requested or if the final vibrationally excited states to which the tunneling is directly coupled are not open, the loop is carried out just once.

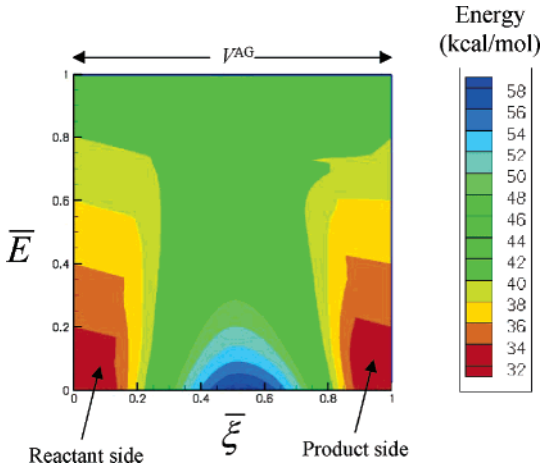


Figure 2. Unit-square potential energy surface evaluated by the ILCT(2D) algorithm using a 9 × 11 grid. The lowest tunneling energy included is 31.11 kcal/mol, and the largest tunneling energy included is 41.73 kcal/mol, which is the maximum of the vibrationally adiabatic ground-state potential curve.

Table 3. Comparison between the Full LCG4 Calculation and Several ILCT(2D) Calculations with $m \times n$ Grids of the LCG4 Transmission Coefficients for Reaction $C_6H_6 + CH_3 \rightarrow C_6H_5 + CH_4$ Using the J1 Potential Energy Surface

$T/m \times n$	full	9 × 11	7 × 9	5 × 9	3 × 5
200	92.1	92.1	88.9	99.3	71.7
250	26.1	26.1	26.1	24.6	21.5
300	12.2	12.2	12.3	10.9	10.3
400	5.07	5.06	5.17	4.50	4.47
500	3.13	3.13	3.20	2.82	2.84
600	2.33	2.33	2.37	2.13	2.16
points[a]	4834	221	205	192	174

[a] This row lists the number of points in the nonadiabatic region that need to be evaluated to obtain the transmission coefficients.

Having obtained the μOMT transmission probability P^G as a function of E, the transmission coefficient is calculated by

$$\kappa^{\mathrm{CVT/Q}} = \frac{\int_{V_a^G(s=-\infty)}^{\infty} e^{-E/k_B T} P^G(E)\, dE}{k_B T \exp[-V_a^G(s_*^{\mathrm{CVT}})/k_B T]} \qquad (14)$$

The energy at which the integrand of the integral in the numerator of this integral has a maximum is called the representative tunneling energy.

To check the accuracy of the ILCT(2D) algorithm, full calculations of the LCT transmission coefficients were carried out, first, for the $C_6H_6 + CH_3 \rightarrow C_6H_5 + CH_4$ reaction using the J1 potential energy surface of the $H_2 + CH_3 \rightarrow H + CH_4$ reaction but with a fictitious mass of 77 amu (mass of C_6H_5) for the donor atom. This reaction is used to show the performance of the ILCT(2D) algorithm when tunneling effects are larger. Results for this reaction are given in Table 3, which shows that excellent performance is obtained when a grid of 99 points is used (11 points along the straight-line tunneling path for tunneling paths corresponding to nine different tunneling energies, with $\bar{\xi}_j$ and $\bar{E}_i$ given by eq 13). Similar tests were carried out for reaction R2, and again, 99 points suffice, as shown in Table 4.

Table 4. LCT Transmission Coefficients for the R2 Reaction Obtained by Using a Grid of 80 × 180 Points (Full-LCT) and the ILCT(2D) Algorithm with a 9 × 11 Grid

T (K)	full LCT	ILCT(2D)
200	75.6	75.9
300	6.82	6.83
400	2.98	2.98
500	2.02	2.02
600	1.63	1.63
700	1.44	1.44

The ILCT(2D) algorithm is implemented in the program POLYRATE 9.3.1.[51] All dynamics calculations were carried out with the program GAUSSRATE 9.1,[52] which was used to provide an interface between GAUSSIAN03[53] and POLYRATE 9.3.1.

4. Dynamics Details

All available experimental measurements of the thermal rate constants of the hydrogen abstraction reactions R1−R8 are relative to the high-pressure limit of the radical recombination rate constant for the reaction $2CF_3 \rightarrow C_2F_6$. To do a direct comparison of our calculated rate constants with those of the experiment, it is necessary to know this rate constant, called $k_{r,\infty}$. Here, we use the value inferred previously,[9] in particular, $k_{r,\infty} = 3.9 \times 10^{-11}$ (cm^3 molecule^{-1}) exp[−1.36 (kcal/mol)/RT].

All dynamics calculations were carried out at the MPWB1K/DIDZ level. To carry out the VTST/MT calculations using μOMT transmission coefficients for tunneling, information about the potential is needed along the MEP and in the reaction swath. The MEP was followed using the Page−McIver algorithm[54] with a step size of 0.01 a_0 and a scaling mass $\mu = 1$ amu; Hessian calculations were saved every nine steps. The dividing surface was defined in terms of redundant internal coordinates.[55] All frequencies were scaled by a factor[38] of 0.9537. The reaction swath was covered by 80 straight-line paths starting and ending at the classical turning points that obey eq 11. The evaluation of the LCT tunneling probabilities for each path (or tunneling energy) required 180 single-point energy calculations to characterize the potential. Since the integral over total energy[56] was carried out by 80-point quadrature, this procedure requires 14 400 single-point calculations for the evaluation of ground-state LCT tunneling probabilities. In addition, for any final vibrational excited-state included, a new set of 14 400 points is needed. As explained above, each set of 14 400 points was obtained by interpolation from a 99-point grid.

The grid for the 2D spline-under-tension calculation on R2 is plotted in Figure 2. For both the full-LCT and the ILCT(2D) methods, single-point calculations are needed only at the boundaries and at points of the grid inside those boundaries. The amount of time that is saved using ILCT(2D) with respect to full-LCT for reactions R2, R5, and R7 is indicated in Table 5. A full-LCT calculation involves 30 times more computer time than the ILCT(2D) calculation but gives similar results. Hereafter, for all reactions, the LCT transmission coefficients were obtained with the ILCT(2D) algorithm.

Table 5. Number of Single-Point Calculations and Computer Time Required for the Evaluation of the LCT Transmission Coefficients by Full Calculation (First Two Numerical Rows) and by the ILCT(2D) Procedure (Third and Fourth Numerical Rows) for Reactions R2, R5, and R7

	R2	R5	R7
full-LCT			
number of points	3370	1827	920
computer time[a] (s)	566 160	678 163	444 960
ILCT(2D)			
number of points	113	70	69
computer time[a] (s)	18 984	25 830	33 672

[a] Calculations carried out on a 3 GHz Pentium 4 PC.

Table 6. Barrier W_{F-1} and Frequency (Obtained from eq 8) ω_{F-1}, Both in cm^{-1}, for the Internal Rotation around the C−C Axis at the Transition State for Reactions R1−R8

	W_{F-1}	ω_{F-1}		W_{F-1}	ω_{F-1}
R1	10.4	22.2	R7s	38.1	12.4
R2	10.4	16.0	R7p1	16.5	10.8
R3	10.4	17.5	R7p2	54.2	14.3
R4	10.4	16.0	R8s	38.1	11.4
R5	23.8	14.0	R8p1	16.5	9.4
R6	23.8	12.2	R8p2	54.2	13.3

5. Calculations, Results, and Discussion

Internal Rotation. The transition state obtained by MPWB1K/DIDZ for reactions R1−R4 have C_3 symmetry, and the fluorine atoms of the CF_3 group and the H atoms of the CH_3 group are in a nearly trans conformation with a FCCH dihedral angle of 178.0 degrees (see Figure 1). The internal rotation of CH_3 with respect to the CF_3 group is almost free and has a rotational barrier of only 10.4 cm^{-1}. The frequency of internal rotation was calculated from eq 7. The reduced moment of inertia is calculated using the CW model.[48] All frequencies for internal rotation for reactions R5−R8 were also calculated by this procedure, and the results are shown in Table 6.

The hindered rotor partition functions for $CF_3 + CH_4$ are 10% less than the free rotor value at 300 K and 5% less than the free rotor value at 700 K. For propane, these percentages range from 8 to 21% at 300 K and 3−10% at 700 K, depending on the conformation.

5.1. CF_3 + Methane. Dynamics. The MEPs for reactions R1−R4 were calculated in the interval $s = -6.5$ to 3.5 a_0. The vibrationally adiabatic ground-state potential energy curves for reactions R1−R4 are plotted in Figure 3. None of the four reactions has a variational effect (i.e., the variational transition state passes through the saddle point), and therefore, the CVT rate constants for these four reactions are equal to those obtained by conventional transition-state theory (TST); that is, eq 8 reduces to

$$k^{\mathrm{TST}/\mu\mathrm{OMT}}(T) = k^{\mathrm{TST/G}}(T)\, k^{\mathrm{TST}}(T) \qquad (15)$$

To obtain $k^{\mathrm{TST}}(T)$, only information at the stationary points (reactants and transition state) is needed. However, the evaluation of the transmission coefficient requires information along the MEP and in the swath.

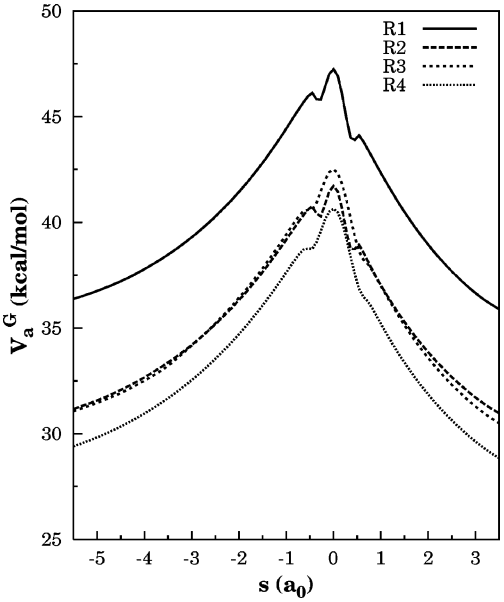


Figure 3. Vibrational adiabatic potential for reactions R1−R4.

The symmetry number of the reaction, which was already introduced in eq 2, is given by the product of the symmetry number of reactants divided by the symmetry number of the transition state. Since $\sigma(CF_3) = 3$, $\sigma(CH_4) = 12$, $\sigma(CD_3H) = 3$, σ(TS of R1, R2, and R4) = 3, and σ(TS of R3) = 1, the symmetry numbers for reactions R1, R2, R3, and R4 are 12, 3, 9, and 12, respectively.

If the internal rotation were treated as a harmonic vibration, the absolute rate constants for R1 would be 2.66 and 3.94 times larger, at temperatures of 300 and 700 K, respectively, than those obtained if eq 4 is used. This shows, in agreement with ref 9, that anharmonicity is significant.

It would be expected that tunneling is more important for reactions R1 and R2 than for reactions R3 and R4, because a hydrogen atom is transferred for the first two reactions whereas, for the latter, the transferred atom is a deuterium, which is heavier. The vibrationally adiabatic ground-state potential energy curves for these four reactions are plotted in Figure 3. A parameter that is sometimes used to indicate the importance of tunneling is the imaginary frequency at the saddle point; a larger imaginary frequency indicates a narrower barrier along the MEP or a lighter tunneling mass. The imaginary frequency is larger for reactions R1 and R2, with values of 1697i cm^{-1} and 1685i cm^{-1}, respectively, than for reactions R3 and R4, which have values of 1247i cm^{-1} and 1243i cm^{-1}, respectively. A more quantitative assessment of tunneling requires multidimensional methods that incorporate the coupling between the reaction coordinate and the $3N - 7$ orthogonal normal modes, that is, coupling to modes transverse to (perpendicular to) the reaction coordinate. Tunneling was evaluated by both the SCT and LCT approximations to obtain the μOMT transmission coefficients, and calculations with the ZCT approximation were carried out for comparison. The ZCT, SCT, LCT, and μOMT multidimensional transmission coefficients are listed in Table 7. The ZCT, SCT, and LCT approximations all include multidimensional effects, but to different extents. The ZCT

Table 7. Tunneling Transmission Coefficients for Reactions R1−R4 Obtained by Multidimensional Tunneling Approximations

reaction	*T*(K)	ZCT	SCT	LCT	μOMT
R1	200	13.8	101	319	329
	300	3.57	9.18	9.60	11.0
	400	2.12	3.69	3.42	3.92
	500	1.64	2.37	2.18	2.43
	600	1.42	1.84	1.71	1.87
	700	1.30	1.57	1.49	1.59
R2	200	10.5	62.7	75.6	84.0
	300	3.14	7.40	6.83	7.90
	400	1.97	3.27	2.98	3.36
	500	1.56	2.18	2.02	2.22
	600	1.37	1.74	1.63	1.76
	700	1.26	1.51	1.44	1.52
R3	200	13.8	72.3	61.9	85.4
	300	3.31	6.52	4.60	6.68
	400	1.97	2.85	2.22	2.86
	500	1.55	1.95	1.64	1.95
	600	1.36	1.59	1.41	1.59
	700	1.25	1.40	1.28	1.40
R4	200	13.1	65.3	58.0	79.4
	300	3.23	6.29	4.49	6.47
	400	1.95	2.80	2.20	2.82
	500	1.54	1.93	1.63	1.93
	600	1.35	1.58	1.40	1.58
	700	1.25	1.40	1.28	1.40

approximation includes coupling of the reaction coordinate to orthogonal degrees of freedom in that it includes the s variation of the generalized normal-mode frequencies $\omega_m(s)$, but it does not include corner cutting. The SCT approximation includes the s variation of $\omega_m(s)$ plus the small-curvature limit of corner cutting. Note that the s variation of $\omega_m(s)$ results from coupling in the potential energy, and corner cutting results from coupling in the kinetic energy. The LCT approximation also includes both of these couplings, and furthermore, it allows for extreme corner cutting.

For reactions R1 and R2, the μOMT transmission coefficients are larger than the SCT transmission coefficients at low temperatures, and this results from extreme corner-cutting contributions at low tunneling energies. For reactions R3 and R4, the SCT and μOMT transmission coefficients are similar, indicating a tunneling path closer to the MEP. It is clear from these results that although the LCT paths may be unimportant for deuterium transfer and for hydrogen transfer at high temperatures (for the reactions under discussion here), they play an important role for hydrogen transfer at low temperatures. The consequence is that, for systems of this type, a tunneling evaluation based exclusively on the SCT transmission coefficients may underestimate the thermal rate for hydrogen transfer and, therefore, may lead to a wrong kinetic isotope effect (KIE). Neglecting corner cutting completely, as in the ZCT approximation, leads to even larger errors, with the largest error being more than a factor of 3 at 300 K and an order of magnitude at 200 K.

Another important aspect to consider is the contribution to the rate constants due to tunneling directly into final

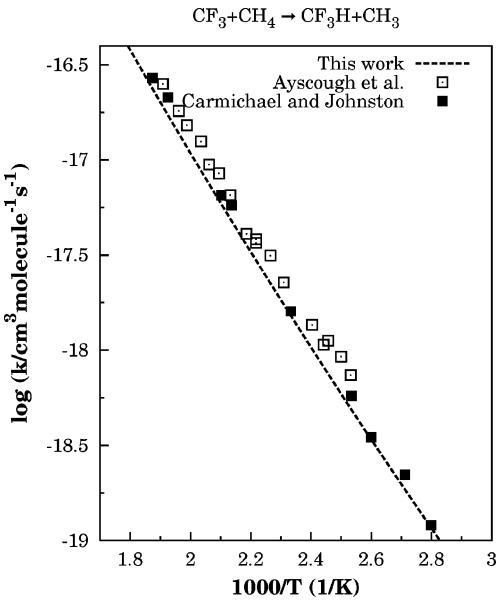


Figure 4. Arrhenius plot for reaction R1.

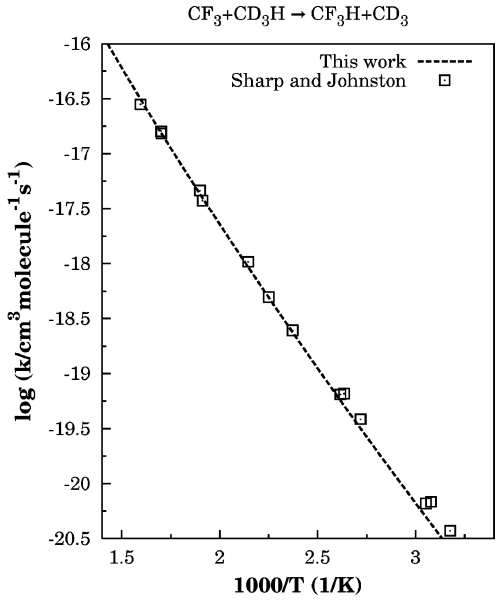


Figure 5. Arrhenius plot for reaction R2.

excited vibrational states, which is allowed only in the LCT approximation. However, in these cases, such contributions due to excited vibrational states are negligible.

The thermal rate constants for reactions R1−R4 are plotted in Figures 4−7 and are given in tables in the Supporting Information. The calculated TST/μOMT thermal rate constants are in good agreement with the experimental values for reactions R1, R2, and R4, but the calculated values for reaction R3 are much too high.

To understand the KIEs better, we factored them as follows. We denote the calculated KIE as η, and we first factor this, on the basis of eq 14, into tunneling and overbarrier contributions:

$$\eta = \eta_{\text{tun}}\eta_{\text{TST}} \tag{16}$$

One should note here, in justifying the subscript "tun", that the transmission coefficients include both tunneling and

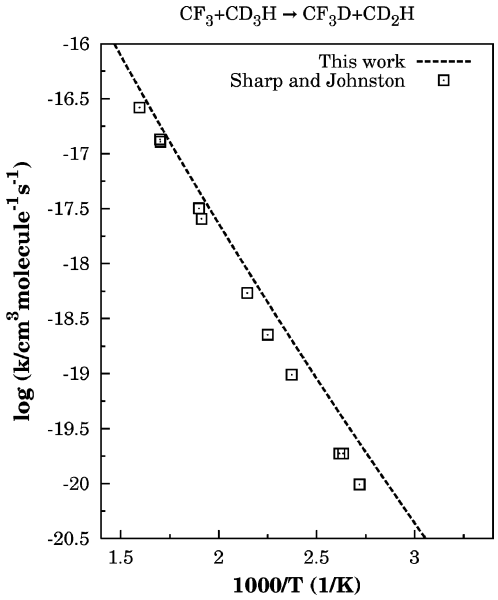


Figure 6. Arrhenius plot for reaction R3.

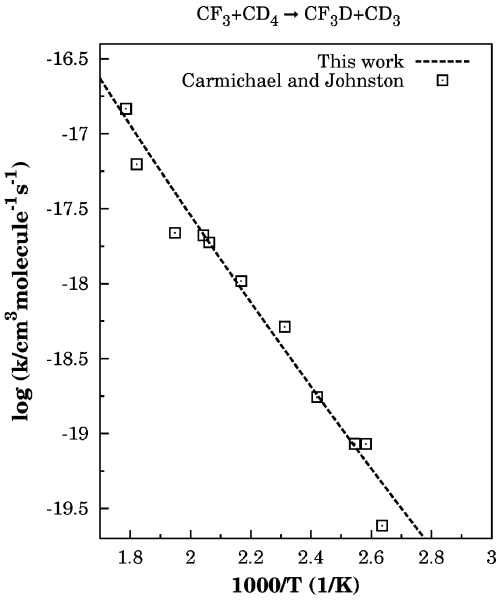


Figure 7. Arrhenius plot for reaction R4.

nonclassical reflections,[23,56] and so technically, they include both these quantum effects on the reaction-coordinate motion, but we usually simply discuss the net effect and call it tunneling, which is usually the dominant of the two effects. The result obtained with classical reaction-coordinate motion and quantized vibrations is also called the quasiclassical (QC) contribution; for the four reactions R1−R4, this is η_{TST}. We factor this into translational, rotational, and vibrational contributions:

$$\eta^{QC} = \eta_{TST} = \eta_{trans}\eta_{rot}\eta_{vib} \qquad (17)$$

Note that η_{trans} and η_{rot} are independent of temperature. Table 8 shows the factors for $\eta = k_{R1}/k_{R4}$ and for $\eta = k_{R2}/k_{R3}$. (Results at more temperatures are in the Supporting Information.) Theory predicts KIEs a little smaller than the experimental values. It is not clear if this presents a challenge for theory or if it indicates inaccuracies in the experiment.

Table 8. Factors in the KIEs

T (K)	η_{tun}	η_{vib}	η^{QC}	η	η_{exp}
		R1/R4[a]			
300	1.71	1.82	5.77	9.83	n.a.[b]
400	1.39	1.22	3.89	5.40	6.2[c]
500	1.26	0.96	3.02	3.80	6.0[c]
700	1.14	0.71	2.25	2.56	n.a
		R2/R3[d]			
300	1.18	4.22	4.39	5.18	n.a.
400	1.17	3.06	3.18	3.72	8.5[e]
500	1.11	2.5	2.61	2.98	5.0[e]
700	1.09	1.98	2.06	2.25	n.a

[a] $\eta_{trans} = 1.31$; $\eta_{rot} = 2.42$. [b] n.a. denotes not available. [c] Interpolated from ref 31. [d] $\eta_{trans} = 1.00$; $\eta_{rot} = 1.04$. [e] Interpolated from ref 29.

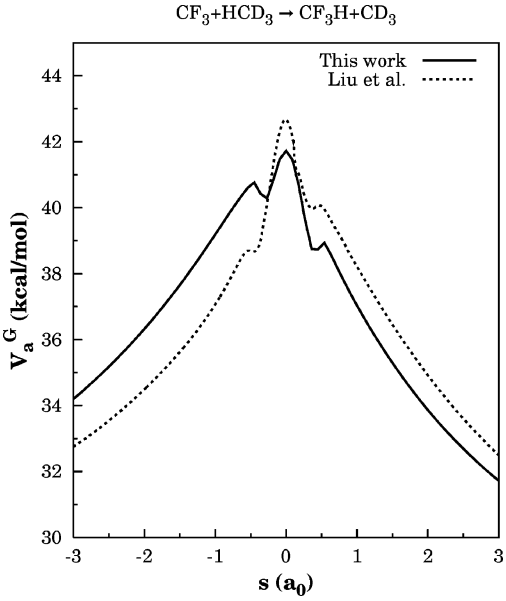


Figure 8. Comparison of vibrationally adiabatic ground-state potential curves obtained in this work and that of Liu et al.[23]

There are some important differences between the present results and those in the previous[9] CF_3 paper. The first difference is that no variational effects are observed in the present work, and the second is that the contribution due to tunneling is much smaller. Figure 8 shows a comparison between the present and previous calculations of the vibrationally adiabatic ground-state potential curve. The MPWB1K and MPW1K levels of calculation lead to imaginary frequencies of 1697i and 1694i cm^{-1}, respectively, whereas the previous AM1-SRP2 calculations led to an imaginary frequency of 2113i cm^{-1}, indicating that the effective potential along the MEP is thinner. An additional factor contributing to the narrower vibrationally adiabatic ground-state potential curve in the previous calculation is that it was obtained in Cartesian coordinates, which usually yields an artificially narrow effective potential as compared to that obtained in the more physically redundant internal coordinates. The present effective potential would be expected to be more realistic not only because of the more physical coordinate system but also because density functionals with one parameter optimized for kinetics are generally more reliable than AM1 for hydrogen atom transfer reactions.[42,57]

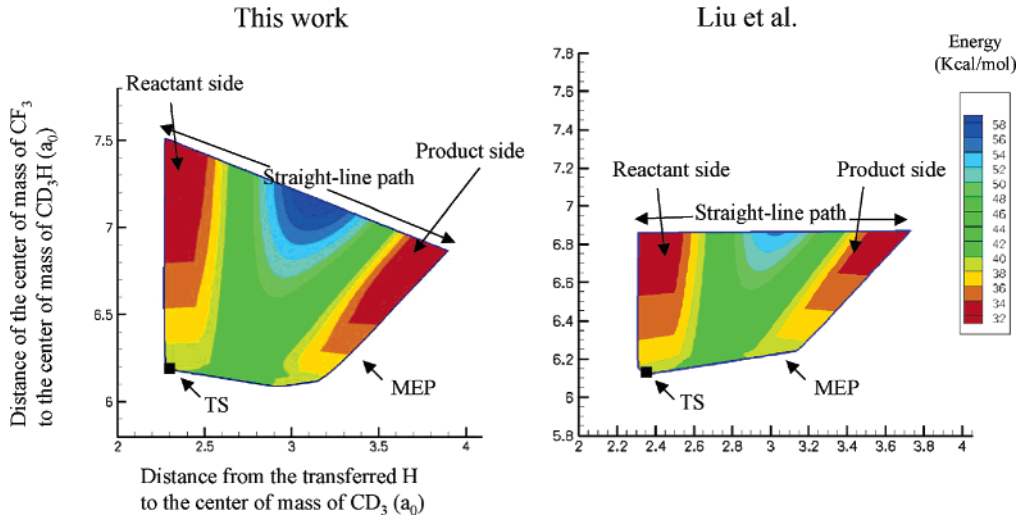


Figure 9. Plot of the effective potential for ground-state tunneling along the MEP and in the reaction "swath" in mass-scaled pseudo-Jacobian coordinates for reaction R2. The effective potential equals the vibrationally adiabatic ground-state potential in the adiabatic region. The left-side contours were obtained in the present work at the MPWB1K/6-31+G(d,p) level, and the lowest tunneling energy included is 31.11 kcal/mol. The right-side contours are those obtained using the AM1-SRP2 method described by Liu et al.,[23] and the lowest tunneling energy included is 32.11 kcal/mol. In both cases, the location of the TS is marked with a black square.

Table 9. Imaginary Frequency, $\omega^{\ddagger}$ (in cm^{-1}), and Energetic Parameters (in kcal/mol) for Reaction R2 Calculated at the AM1-SRP2 and MPWB1K/DIDZ Levels

	AM1-SRP2	MPWB1K
$V^{\ddagger}$	14.63	14.39
ΔE	−0.74	0.32
$\Delta V_a^{G\ddagger}$	12.40	12.06
ΔH_0	−1.19	−0.85
$\omega^{\ddagger}$	2113i	1697i

Table 10. Transmission Coefficients for Reactions R2 and R3

		Liu et al.[a]			this work[b]		
reaction	T (K)	SCT	LCT	μOMT	SCT	LCT	μOMT
R2	367.8	8.0	17.0	18.0	3.98	3.62	4.13
	626.8	2.2	2.6	2.7	1.66	1.57	1.68
R3	367.8	4.0	6.7	7.8	3.46	2.61	3.49
	626.8	1.8	1.7	1.8	1.53	1.36	1.53

[a] AM1-SRP2.[9] [b] MPWB1K/DIDZ.

Nevertheless, one cannot discount the possibility that the MPWB1K/DIDZ method overestimates the width of the barrier (and that such an overestimate adversely affects the agreement with experimental results). The surface properties are further compared in Table 9. Figure 9 compares the effective potentials in the swath. Note that for both cases shown in the figure, the lowest energy shown is the lowest energy actually used in the calculations, and the tunneling probability is negligible at this energy. In the MPWB1K case, V_a^{AG} is 41.73 kcal/mol, and the lowest energy considered is 10.62 kcal/mol below this. In the AM1-SRP2 case, V_a^{AG} is 42.65 kcal/mol, and the lowest tunneling energy considered is 10.54 kcal/mol below this. Tables 10 and 11 compare the representative tunneling energies and transmission coefficients for the two sets of calculations; tunneling is significant down to lower energies in the previous study.

Table 11. Representative Tunneling Energies (in kcal/mol) for Reaction R2

		Liu et al.[a]			this work		
reaction	T (K)	SCT	LCT	μOMT	SCT	LCT	μOMT
R2	367.8	40.1	38.7	38.7	40.5	40.8	40.8
	626.8	41.5	40.2	41.5	40.8	40.8	40.8

[a] Ref 9.

Table 12. KIEs and Their Factors for Reactions R2 and R3

	Liu et al.[9]			this work		
T (K)	η_{tun}	η^{QC}	η	η_{tun}	η^{QC}	η
367.8	2.3	4.0	9.1	1.18	3.48	4.11
626.8	1.5	2.0	3.0	1.10	2.22	2.44

The consequences of these differences are shown in Table 12; the present calculations predict much smaller KIEs.

The differences of the transmission coefficients from the two calculations that are compared in Table 12 are not due entirely to the differences in the potential energy surfaces. Another source of difference is that ref 9 used the LCG3 approximation for the LCT transmission probabilities, whereas the present article used the LCG4 approximation. Repeating the AM1-SRP2 calculations with LCG4 lowers the LCT transmission coefficients for R2 to 11.7 and 2.4 at 367.8 and 626.8 K. The results with LCG4 are closer to the present results for R2, but there is less difference between LCG3 and LCG4 for R3. We conclude that both the potential energy surface and the tunneling approximation contribute to the differences of ref 9 from the present work.

5.2. CF_3 + Ethane. As in the case of reactions R1−R4, reactions R5 and R6 do not show variational effects, and the final variational transition-state theory rate constant is given by eq 15. The symmetry numbers are $\sigma(CF_3) = 3$, $\sigma(C_2H_6) = 6$, $\sigma(C_2D_6) = 6$, and σ(TS of R5 and R6) = 1.

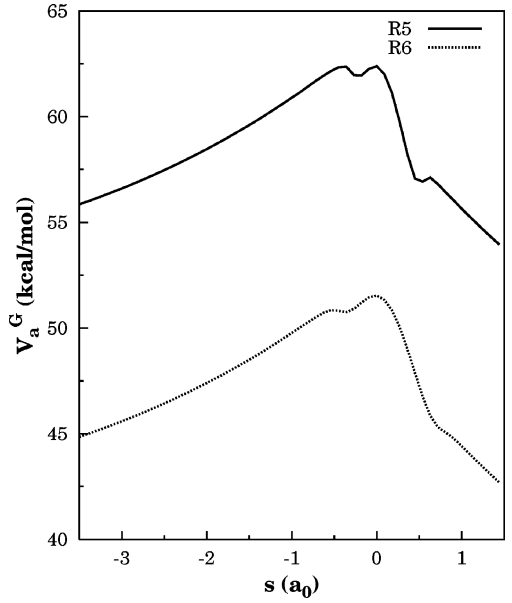


Figure 10. Vibrational adiabatic potential for reactions R5 and R6.

Table 13. SCT, LCT, and μOMT Transmission Factors, κ, for the H and D Abstraction by CF_3 of H(D) Atoms of Ethane

	R5			R6		
T (K)	κ^{SCT}	κ^{LCT}	$\kappa^{\mu OMT}$	κ^{SCT}	κ^{LCT}	$\kappa^{\mu OMT}$
300	3.18	2.19	3.20	3.24	2.34	3.27
400	1.91	1.47	1.91	1.95	1.57	1.95
500	1.51	1.26	1.51	1.54	1.33	1.54
600	1.33	1.17	1.33	1.35	1.21	1.35
700	1.23	1.12	1.23	1.25	1.15	1.25

The lowest-frequency mode, as in the previous case, is treated by the CW approximation, which decreases the rate constant, as compared to a harmonic calculation, by a factor of 1.9 at $T = 300$ K and a factor of 2.7 at $T = 700$ K. The imaginary frequencies for reactions R5 and R6 are 1618i cm^{-1} and 1190i cm^{-1}, respectively. The classical barrier height for this reaction is lower than that for the reaction of CF_3 with methane (see Table 2), and the fractional contribution due to tunneling is smaller, although the reaction still proceeds mainly by tunneling. In particular, the SCT and LCT4 transmission coefficients for reaction R5 at 300 K are 3.17 and 2.18, respectively. The μOMT transmission coefficient at this temperature is 3.20, indicating that tunneling is well-represented by the SCT approximation in the range of temperatures studied. Table 13 shows that the tunneling is mainly of the small-curvature type at 400−700 K as well.

Figure 10 shows that the vibrationally adiabatic potential has two similar maxima, one at $s = -0.40$ and one at $s = 0$ a_0. This indicates that the reaction has two bottlenecks and that both of them should be taken into account in the evaluation of the thermal rate constants. The canonical unified statistical theory[23,58] transmission coefficient Γ_{CUS} incorporates this second bottleneck and is given by

$$\Gamma^{CUS}(T) = [1 + q_{vr}^{CVT}(T)/q_{vr}^{max}(t) - q_{vr}^{CVT}(T)/q_{vr}^{min}(T)]^{-1} \quad (18)$$

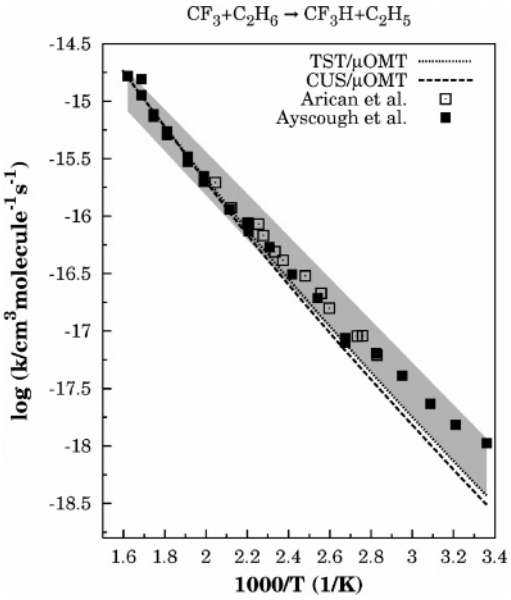


Figure 11. Arrhenius plot for reaction R5. The gray area indicates the standard deviation reported by Arican and Arthur.[32]

where

$$q_{vr}^{CVT}(T) = Q_{vr}^{GT}(T,s_*^{CVT}) \exp[-\beta V_{MEP}(s)] \quad (19)$$

with $Q_{vr}^{GT}(T,s_*^{CVT})$ being evaluated at the absolute maximum of the free energy (which in this case coincides with $s = 0$, because there is no variational effect); $q_{vr}^{max}(T)$ is evaluated at the second highest maximum in the free energy, and $q_{vr}^{min}(T)$ is evaluated at the lowest minimum between the two maxima. The resulting $k^{CUS/\mu OMT}$ rate constant is given by

$$k^{CUS/\mu OMT}(T) = \Gamma^{CUS}(T)\, k^{CVT/\mu OMT} \quad (20)$$

For instance, at $T = 300$ K, the highest maximum in the free energy is located at $s = 0$, where the generalized free energy of activation[22,23] is 45.14 kcal/mol, and the second largest maximum is at $s = -0.37$ a_0 where it is 44.77 kcal/mol, whereas the minimum is located at $s = -0.25$ a_0, where the generalized free energy of activation is 44.50 kcal/mol. From these data, one obtains $R_{CUS}(T = 300\ K) = 0.83$. The value of R_{CUS} becomes unity at temperatures above 510 K. Reaction R6 has only one bottleneck, which is at $s = 0$, and therefore, in this case,

$$k^{CUS/\mu OMT}(T) = k^{CVT/\mu OMT}(T) = k^{TST/\mu OMT}(T) \quad (21)$$

The rate constants are plotted in Figure 11 and tabulated in the Supporting Information.

The TST KIEs, η, were factored using eqs 16 and 17. More generally,

$$\eta = \eta_{trans}\eta_{rot}^{TST}\eta_{vib}^{TST}\eta_{var}\eta^{CUS}\eta_{tun} \quad (22)$$

The product $\eta_{trans}\eta_{rot}^{TST}\eta_{vib}^{TST}$ equals η_{TST}, which is the KIE

Table 14. Factorization of the KIE k(R5)/k(R6)

T (K)	η_{vib}^{TST}	η_{TST}	η^{CUS}	η^{QC}	η_{tun}	η
300	2.79	5.75	0.83	4.76	0.98	4.67
400	1.84	3.82	0.88	3.35	0.98	3.29
500	1.43	2.97	0.91	2.70	0.98	2.65
700	1.07	2.22	1.00	2.22	0.99	2.18

predicted by conventional TST. The KIE due to variational effects, η_{var}, is calculated as

$$\eta_{var} = \frac{k_H^{CVT}(T)\, k_D^{TST}(T)}{k_D^{CVT}(T)\, k_H^{TST}(T)} \tag{23}$$

(Note that η_{rot}^{TST} and η_{vib}^{TST} are not the full rotational and vibrational effects; those would be obtained by also factoring η_{var} into contributions from vibrational and rotational modes.) The KIE due to the recrossing included in CUS theory is

$$\eta_{CUS} = \frac{\Gamma_H^{CUS}(T)}{\Gamma_D^{CUS}(T)} \tag{24}$$

The TST, variational, and CUS contributions may be combined into a quasiclassical contribution such that

$$\eta^{QC} = \eta_{TST}\eta_{var}\eta^{CUS} \tag{25}$$

and

$$\eta = \eta_{tun}\eta^{QC} \tag{26}$$

where the contribution due to quantum effects is included as the ratio between the μOMT transmission coefficients

$$\eta_{tun} = \frac{\kappa_H^{\mu OMT}(T)}{\kappa_D^{\mu OMT}(T)} \tag{27}$$

These factors are given in Table 14, which shows that the temperature dependence of the KIE is dominated by the vibrations and that tunneling nearly cancels out in the KIE for R5 versus R6. (Results for more temperatures are in the Supporting Information.)

5.3. CF_3 + Propane. For reactions R7 and R8, there are two types of hydrogen/deuterium atoms that can be abstracted: primary (from the terminal methyl groups) and secondary (from the central carbon atom). Thus, we make a distinction between the reactions

$$CF_3 + H_3C{-}CH_2{-}CH_3 \rightarrow CF_3H + H_3C{-}CH_2{-}CH_2 \tag{R7p}$$

and

$$CF_3 + H_3C{-}CH_2{-}CH_3 \rightarrow CF_3H + H_3C{-}CH{-}CH_3 \tag{R7s}$$

The transition states for the abstraction of a primary and secondary hydrogen atom are shown in Figure 1.

Propane has C_{2v} symmetry; therefore, $\sigma(C_3H_8) = 2$. Furthermore, $\sigma(CF_3) = 3$, and σ(TS of R7s) = 1; therefore, the symmetry number for the abstraction of a secondary hydrogen, H_s, is 6.

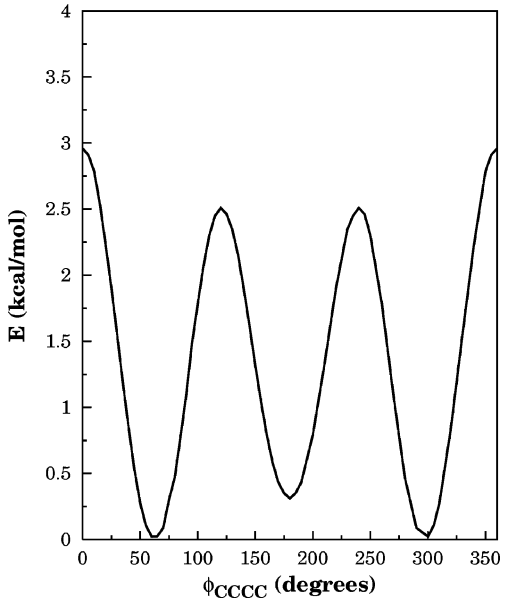


Figure 12. Potential for internal rotation at the TS for reaction R7p.

In the case of abstraction of a hydrogen of the two terminal methyl groups, there are two possibilities: (a) the abstraction of one of the two hydrogen atoms having a dihedral angle $\phi_{HCCC} = 180°$ (we label these hydrogen atoms as H_{p1} and the abstraction reaction as R7p1) and (b) the abstraction of one of the four atoms having dihedral angles of $\phi_{HCCC} = 60°$ (two of them) and $\phi_{HCCC} = -60°$(the other two), and we label these hydrogen atoms as H_{p2} and the abstraction reactions as R7p2. The TS for the hydrogen abstraction of H_{p1} has no symmetry, and therefore, the symmetry number for this reaction is 6. The rotation around the terminal methyl group (see Figure 12) starting from this transition-state structure indicates that there are another two transition states at $\phi_{CCCC} = 62°$and $-62°$, respectively, which cannot be superimposed and which are separated by a barrier of 2.51 kcal/mol. These two transition states correspond to the H_{p2} hydrogen abstractions. Therefore, the symmetry number for the abstraction of the H_{p2} hydrogen abstractions is $\sigma(CF_3) \times \sigma(C_3H_8) \times m^{\ddagger}(TSp) = 12$, with $m^{\ddagger}(TSp) = 2$ because of the two isoenergetic optically active transition states. The imaginary frequencies at the transition states are 1484i, 1630i, and 1619i cm^{-1} for R7s, R7p1, and R7p2, respectively, and 1099i, 1198i, and 1190i cm^{-1} for R8s, R8p1, and R8p2, respectively. The total rate constant for the abstraction of a hydrogen atom from propane by CF_3 can be obtained as the sum of all these contributions, that is,

$$k_{R7}(T) = k_{R7s}(T) + k_{R7p1}(T) + k_{R7p2}(T)$$

The vibrationally adiabatic ground-state potential curves are plotted in Figure 13. As indicated in Table 2, the abstraction of a secondary hydrogen has a barrier height about 2.5 kcal/mol lower than the abstraction of a primary hydrogen. Thus, the secondary hydrogen abstraction contributes substantially more than the primary one, although the latter contribution increases with temperature. As to the relative contribution of H_{p1} and H_{p2} to the primary hydrogen abstraction thermal rate, the abstraction of H_{p1} is more

Table 15. μOMT Transmission Coefficients and CUS/μOMT Thermal Rate Constants (in s^{-1}) for H and D Abstraction from the Primary (p1 and p2) and Secondary (s) H(D) Atoms of Propane by CF_3.

T (K)	$\kappa_{H,p1}$	$k_{H,p1}$	$\kappa_{H,p2}$	$k_{H,p2}$	$\kappa_{H,s}$	$k_{H,s}$	$\kappa_{D,p1}$	$k_{D,p1}$	$\kappa_{D,p2}$	$k_{D,p2}$	$\kappa_{D,s}$	$k_{D,s}$
300	2.73	6.25(−20)	2.95	3.59(−20)	2.89	1.82(−18)	3.12	1.69(−20)	3.01	1.04(−20)	1.81	3.83(−19)
400	1.75	3.39(−18)	1.83	1.84(−18)	1.87	4.55(−17)	1.92	1.23(−18)	1.87	7.18(−19)	1.40	1.44(−17)
500	1.43	4.63(−17)	1.47	2.42(−17)	1.51	3.74(−16)	1.52	2.02(−17)	1.50	1.13(−17)	1.24	1.53(−16)
600	1.28	3.02(−16)	1.31	1.53(−16)	1.34	1.70(−15)	1.34	1.50(−16)	1.33	8.16(−17)	1.16	8.25(−16)
700	1.20	1.26(−15)	1.22	6.24(−16)	1.24	5.40(−15)	1.24	6.87(−16)	1.23	3.64(−16)	1.12	3.31(−15)

favorable even when the barrier height and the vibrational adiabatic potential are very similar for the two cases.

The vibrationally adiabatic potentials of reactions R7s, R8s, R7p1, R8p1, R7p2, and R8p2 present two maxima with similar energies, and so CUS calculations are carried out. Some of these reactions also have variational effects. For instance, variational effects are important in reactions R7s and R7p2.

The transmission coefficients are listed in Table 15. The rate constants for reaction R7 are plotted in Figure 14 and tabulated in Table 15 and the Supporting Information. Although the series of reactions studied here was chosen (on the basis of our previous results with an AM1 potential energy surface) to illustrate the new method for efficient calculation of LCT contributions, with the implicit surface of the new direct dynamics calculations, the tunneling contributions to the rate constants are relatively small and well accounted for the SCT approximation. Nevertheless, it is important to have an efficient algorithm for LCT because one does not know a priori which kind of tunneling approximation will be more appropriate, and therefore, it is advisable, when possible and in the absence of experience on similar systems, to carry out both SCT and LCT calculations and to use the μOMT approximation.

Arican et al.[30] also reported the ratio between the abstraction of a secondary and a primary hydrogen. They obtain a ratio k_{R7s}/k_{R7p} of 4.52 and 2.69 at $T = 400$ and 500 K, respectively. The values obtained by CUS/μOMT theory at the same two temperatures are 8.70 and 5.30, respectively. Figure 15 shows that the CUS/μOMT values overestimate the importance of the abstraction of a secondary hydrogen versus the primary hydrogens as compared to the experiment.

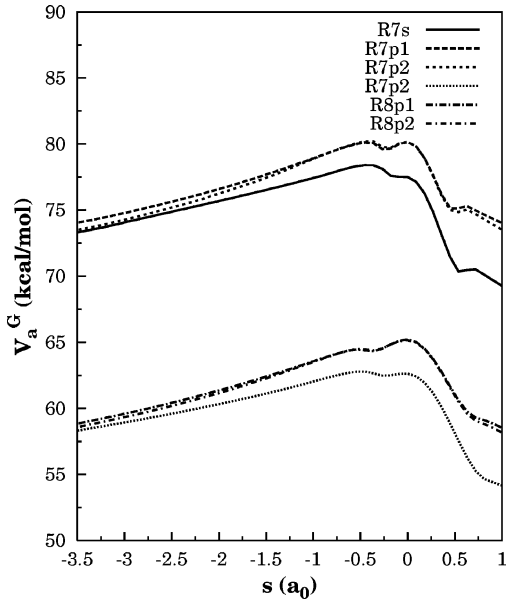


Figure 13. Vibrationally adiabatic potential for reactions R7 and R8.

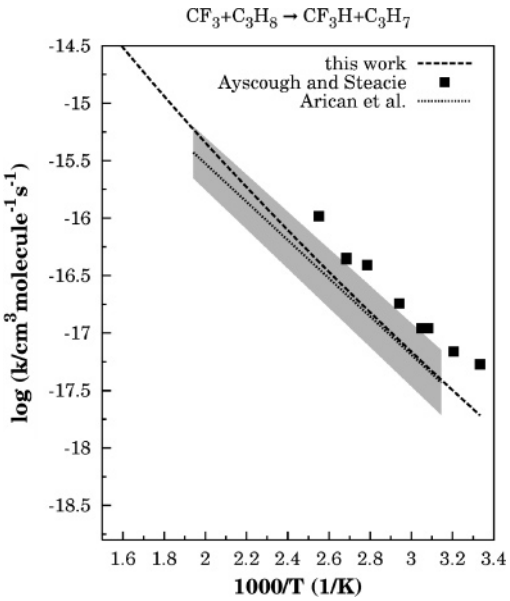


Figure 14. Arrhenius plot for reaction R7. The gray area indicates the standard deviation reported by Arican et al.[30]

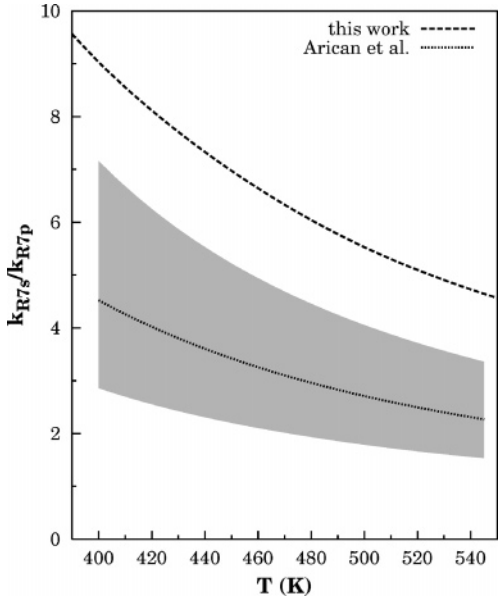


Figure 15. Ratio between the secondary and primary rate constants. The gray area indicates the standard deviation reported by Arican et al.[30]

Table 16. Factors in KIEs for Reactions R7s and R8s[a]

T (K)	η_{tun}	η_{vib}^{TST}	η_{TST}	η_{var}	η^{CUS}	η^{QC}	η
300	1.60	3.35	5.60	0.43	1.24	2.98	4.76
314.9	1.54	3.10	5.18	0.45	1.23	2.87	4.42
367.8	1.40	2.48	4.15	0.51	1.18	2.52	3.52
400	1.34	2.23	3.72	0.54	1.17	2.36	3.16
466.1	1.25	1.87	3.13	0.59	1.15	2.12	2.64
500	1.22	1.74	2.91	0.61	1.14	2.02	2.45
522.9	1.20	1.67	2.78	0.62	1.14	1.96	2.34
600	1.15	1.48	2.46	0.64	1.14	1.79	2.06
626.8	1.14	1.42	2.38	0.64	1.14	1.74	1.98
700	1.11	1.31	2.19	0.59	1.14	1.47	1.63

[a] $\eta_{trans} = 1.16$ and $\eta_{rot}^{TST} = 1.44$.

Table 17. Experimental and Theoretical (CUS/μOMT) Arrhenius Parameters (in cm^3 molecule^{-1} s^{-1} and in kcal/mol) for Reactions R1−R5 and R7 Obtained in the Experimental Range of Temperatures[a,b]

	CUS/μOMT values		experimental values		
reaction	log A	E_a	log A	E_a	T range
R1	−12.0	11.5	−12.1	11.0± 0.5	395−524[c]
			−11.8	11.7	357.3−533.8[d]
R2	−12.4	11.8	−12.7	11.2 ± 0.3	314.9−626.8[e]
R3	−12.1	12.7	−12.0	13.4 ± 0.3	314.9−626.8[e]
R4	−11.8	13.0	−12.6	12.8 ± 2.1	379.4−560[d]
R5	−11.4	9.8	−12.1	8.2 ± 0.5	353.5−489[c]
			−11.9	8.6 ± 0.2	297.6−617[f]
R7	−11.9	8.1	−12.1	7.2 ± 0.5	300−392[g]
			−12.2	7.6 ± 0.2	318−505[h]

[a] CUS/μOMT reduces to CVT/μOMT when there is only one local maximum in the free energy of activation profile and to TST/μOMT when the single maximum occurs at the saddle point. [b] The experimentalists' stated error bar is given when it is more than one in the last digit. [c] Ayscough et al.[27] [d] Carmichael and Johnston.[31] [e] Sharp and Johnston.[29] [f] Arican and Arthur.[32] [g] Ayscough and Steacie.[28] [h] Arican et al.[30]

The KIEs for propane are dominated by the abstraction of secondary hydrogens. The factorization for the secondary abstraction is given in Table 16, which shows that quasiclassical vibrational effects are more important than tunneling and that variational effects lower the KIE by about a factor of 2.

5.4. Arrhenius Parameters. The experimentalists fit their results to an Arrhenius form

$$k = A \exp(-E_a/RT) \tag{28}$$

and for comparison, we made the same kind of fits to the theoretical data. Table 17 shows good agreement in the magnitudes of the Arrhenius parameters for reactions R1−R4; furthermore, theory and experiment agree that the activation energy E_a is considerably smaller for R5 and R7 than for R1−R4, but the preexponential factor A is almost the same. The lowering of the activation energy follows directly from the lower barrier heights in Table 2. It will be interesting to confirm whether MPWB1K gets the substituent effects wrong or whether the discrepancy is due to experimental error.

6. Concluding Remarks

Heavy−light−heavy bimolecular reactions are perhaps the most difficult class of reactions for quantum mechanical calculations of rate constants. We have presented a new, efficient algorithm for LCT calculations, and we illustrate it by straight direct dynamics calculations based on DFT for reactions with five to seven heavy (nonhydrogenic) atoms. This algorithm allowed us to do direct dynamical calculations on a reaction with up to 7 heavy atoms and 15 total atoms (45 degrees of freedom) with quantized generalized normal-mode vibrational frequencies based on curvilinear coordinates, torsional anharmonicity, and optimized multidimensional tunneling. This allowed tests of implicit density functional theory potential energy surfaces against the experiment.

Acknowledgment. The authors are grateful to Ben Ellingson, Steve Mielke, and Yan Zhao for assistance. This work was supported, in part, by the Office of Basic Energy Sciences of the U.S. Department of Energy. A.F. thanks the Ministerio de Educación y Ciencia for a Ramón y Cajal Research Contract and for the Project No. BQU2003-01639.

Supporting Information Available: Tables of theoretical and experimental rate constants, KIEs for abstraction of H from primary and secondary positions in propane, and factors in various KIEs. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618.

(2) Wang, I.; Karplus, M. *J. Am. Chem. Soc*. **1973**, *95*, 8160.

(3) Truhlar, D. G.; Duff, J. W.; Blais, N. C.; Tully, J. C.; Garrett, B. C. *J. Chem. Phys*. **1982**, *77*, 764.

(4) Car, R.; Parrinello, M. *Phys. Rev. Lett*. **1985**, *55*, 2471. Galli, G.; Parrinello, M. *NATO Adv. Sci. Inst. Ser., Ser. E* **1991**, *205*, 283.

(5) *Tight-Binding Approach to Computational Materials Science*; Turchi, P. E. A., Gonis, A., Colombo, L., Eds.; Materials Research Society: Warrendale, PA, 1998.

(6) LeForestier, C. *J. Chem. Phys*. **1978**, *68*, 4406. Helgaker, T.; Uggerud, E.; Jensen, H. J. A. *Chem. Phys. Lett.* **1990**, *173*, 145. Uggerud, E.; Helgaker, T. *J. Am. Chem. Soc*. **1992**, *114*, 4265. Millam, J. M.; Bakken, V., Chem, W.; Hase, W. L.; Schlegel, H. B. *J. Chem. Phys*. **1999**, *111*, 3800. Yu, H.-G.; Muckerman, J. T.; Francisco. J. S. *J. Phys. Chem. A* **2005**, in press.

(7) Truhlar, D. G.; Gordon, M. S. *Science* **1990**, *249*, 491.

(8) Liu, Y.-P.; Lynch, G. C.; Truong, T. N.; Lu, D.-h.; Truhlar, D. G.; Garrett, B. C. *J. Am. Chem. Soc*. **1993**, *115*, 2408.

(9) Liu, Y.-P.; Lu, D.-h.; Gonzàlez-Lafont, A.; Truhlar, D. G.; Garrett, B. C. *J. Am. Chem. Soc*. **1993**, *115*, 7806.

(10) Baldridge, K. K.; Gordon, M. S.; Steckler, R.; Truhlar, D. G. *J. Phys. Chem*. **1989**, *93*, 5107. Truhlar, D. G. *Understanding Chem. React*. **1995**, *16*, 229. Pu, J.; Truhlar, D. G. *J. Chem. Phys*. **2002**, *116*, 1468.

(11) Tuckerman, M. E. *Neumann Inst. Comput. Ser*. **2002**, *10*, 299.

(12) Truhlar, D. G.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sanchez, M. L.; Villà, J. *Acc. Chem. Res*. **2002**, *35*, 341. Truhlar, D. G.; Gao, J.; Garcia-Viloca, M.; Alhambra, C.; Corchado, J.; Sanchez, M. L.; Poulsen, T. D. *Int. J. Quantum Chem*. **2004**, *100*, 1136.

(13) Pople, J. A. In *Energy, Structure, and Reactivity*; Smith, D. W., McRae, W. B., Eds.; Wiley: New York, 1973; p 51.

(14) Castillo, J. F.; Aoiz, F. J.; Bañares, L.; Collins, M. A. *J. Phys. Chem. A* **2004**, *108*, 6611. Lin, H.; Pu, J.; Albu, T. V.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 4112.

(15) Truhlar, D. G. Kuppermann, A. *J. Chem. Phys*. **1972**, *56*, 2232 and references therein.

(16) Marcus, R. A.; Coltrin, M. E. *J. Chem. Phys*. **1977**, *67*, 2609. Babamov, V. K.; Marcus, R. A. *J. Chem. Phys*. **1978**, *74*, 1790. Garrett, B. C.; Truhlar, D. G. *Proc. Natl. Acad. Sci. U.S.A*. **1979**, *76*, 4755. Skodje, R. T.; Truhlar, D. G.; Garrett, B. C. *J. Chem. Phys*. **1982**, *77*, 5955.

(17) Garrett, B. C.; Truhlar, D. G.; Wagner, A. F.; Dunning, T. H., Jr. *J. Chem. Phys*. **1983**, *78*, 4400. Bondi, D. K.; Connor, J. N. L.; Garrett, B. C.; Truhlar, D. G. *J. Chem. Phys*. **1983**, *78*, 5989.

(18) Garrett, B. C.; Truhlar, D. G. *J. Chem. Phys*. **1983**, *79*, 4931.

(19) Kreevoy, M. M.; Ostovic, D.; Truhlar, D. G.; Garrett, B. C. *J. Chem. Phys*. **1986**, *90*, 3766.

(20) Allison, T. C.; Truhlar, D. G. In *Modern Methods for Multidimensional Dynamics Computations in Chemistry*; Thompson, D. L, Ed.; World Scientific: Singapore, 1998; p 618.

(21) Pu, J.; Truhlar, D. G. *J. Chem. Phys*. **2002**, *117*, 1479.

(22) Garrett, B. C.; Truhlar, D. G. *J. Chem. Phys*. **1980**, *72*, 3460. Garrett, B. C.; Truhlar, D. G.; Grev, R. S.; Magnuson, A. W. *J. Phys. Chem*. **1980**, *84*, 1730, **1983**, *87*, 4554(E).

(23) Truhlar, D. G.; Isaacson, A. D.; Garrett, D. C. In *Theory of Chemical Reaction Dynamics*; Baer, M., Ed.; CRC Press: Boca Raton, FL, 1985; p 65.

(24) Truhlar, D. G.; Brown, F. B.; Steckler, R.; Isaacson, A. D. *NATO Adv. Sci. Inst. Ser., Ser. C* **1986**, *170*, 285.

(25) Lu, D.-h.; Truong, T. N.; Melissas, V. S.; Lynch, G. C.; Liu, Y.-P.; Garrett, B. C.; Steckler, R.; Isaacson, A. D.; Rai, S. N.; Hancock, G. C.; Lauderdale, J. G.; Joseph, T.; Truhlar, D. G. *Comput. Phys. Commun*. **1992**, *71*, 235.

(26) Fernandez-Ramos, A.; Truhlar, D. G. *J. Chem. Phys*. **2001**, *114*, 1491.

(27) Ayscough, P. B.; Polanyi, J. C.; Steacie, E. W. *Can. J. Chem*. **1955**, *33*, 743.

(28) Ayscough, P. B.; Steacie, E. W. R. *Can. J. Chem*. **1956**, *34*, 103.

(29) Sharp, T. E.; Johnston, H. S. *J. Chem. Phys*. **1962**, *37*, 1541.

(30) Arican, H.; Potter, E.; Whytock, D. A. *J. Chem. Soc*., *Faraday Trans*. **1973**, *69*, 1811.

(31) Carmichael, H.; Johnston, H. S. *J. Phys. Chem*. **1964**, *41*, 1975.

(32) Arican, H.; Arthur, N. L. *Int. J. Chem. Kinet*. **1986**, *18*, 437.

(33) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *J. Chem. Phys*. **1996**, *104*, 2598.

(34) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys*. **2001**, *114*, 108.

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

(36) Lynch, B. J.; Truhlar, D. G. *ACS Symp. Ser*. **2005**, in press.

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

(38) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715.

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

(40) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 4786.

(41) Afeefy, H. Y.; Liebman, J. F.; Stein, S. E. *NIST Standard Reference Database* **2003**, *69* (http://webbook.nist.gov).

(42) Zhao, Y.; Lynch, B. J.; Truhlar D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.

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

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

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

(46) Tucker, S. C.; Truhlar, D. G. *NATO Adv. Sci. Inst. Ser., Ser. C* **1989**, *267*, 291.

(47) Jackels, C. F.; Gu, Z.; Truhlar, D. G. *J. Chem. Phys*. **1995**, *102*, 3188.

(48) Chuang, Y.-Y.; Truhlar, D. G. *J. Chem. Phys*. **2000**, *112*, 1221; **2004**, *121*, 7036(E); **2005**, another erratum in preparation.

(49) Pitzer, K. S. *J. Chem. Phys*. **1946**, *14*, 239.

(50) Fernandez-Ramos, A.; Truhlar, D. G. *J. Phys. Chem*. **2002**, *106*, 4957.

(51) Corchado, J. C.; Chuang, Y.-Y.; Fast, P. L.; Villà, J.; Hu, W.-P.; Liu, Y.-P.; Lynch, G. C.; Nguyen, K. A.; Jackels, C. F.; Elllingson, B.; Melissas, V. S.; Lynch, B. J.; Rossi, I.; Coitiño, E. L.; Fernandez-Ramos, A.; Pu, J.; Albu, T. V.; Steckler, R.; Garrett, B. C.; Isaacson, A. D.; Truhlar, D. G. *POLYRATE*, version 9.3.1; University of Minnesota: Minneapolis, MN, 2005.

(52) Corchado, J. C.; Chuang, Y.-Y.; Coitiño, E. L.; Truhlar, D. G. *GAUSSRATE*, version 9.1; University of Minnesota: Minnespolis, MN, 2003.

(53) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., 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.; Amado, 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.; Nanyakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(54) Page, M.; McIver, J. W., Jr. *J. Chem. Phys*. **1988**, *88*, 922.

(55) Chuang, Y.-Y.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 242.

(56) Garrett, B. C.; Truhlar, D. G. *J. Phys. Chem*. **1979**, *83*, 2921.

(57) Dybala-Defratyka, A.; Paneth, P.; Pu, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2475.

(58) Hu, W.-P.; Truhlar, D. G. *J. Am. Chem. Soc*. **1995**, *117*, 10726.

CT050153I

JCTC Journal of Chemical Theory and Computation

Does a Sodium Atom Bind to C_{60}?

Jose Pitarch-Ruiz,[†] Stefano Evangelisti,*[,‡] and Daniel Maynau[‡]

Departamento de Quimica-Fisica, Institut de Ciencia Molecular, Universitat de Valencia, Dr. Moliner, 50, E-46100 Burjassot, Valencia, Spain, and Laboratoire de Physique Quantique, UMR 5626, Université Paul Sabatier, 118, Route de Narbonne, F-31062 Toulouse, Cedex, France

Received February 22, 2005

Abstract: A Multi-Reference Configuration-Interaction study of the NaC_{60} system is presented. It is shown that the experimentally measured dipole moment of this system can be explained by the existence of a charge-transfer state of $Na^+C_{60}^-$ nature. Moreover, the present work shows that Configuration-Interaction techniques based on local orbitals permit a Multi-Reference treatment of systems containing several tens of atoms.

The interaction of fullerenes with alkaline atoms deserved a special attention in the past decade.[1−5] In the case of sodium, a state with a dipole moment of 16.3 D has been detected for the NaC_{60} system.[6,7] Theoretical calculations predict a minimum for the sodium atom approaching one of the hexagonal faces of the fullerene.[8,9] However, Unrestricted Hartree−Fock (UHF) and Density-Functional Theory (DFT) calculations give rather different values for the binding energy D: the system is virtually unbounded at the UHF level[10] (D=0.10 eV), while DFT-LDA gives a much larger value[11] (D=2.10 eV). The DFT-B3LYP result is placed between these two extrema[8] (D=0.65 eV). The distance values between Na and the closest carbon atoms are close at the DFT level (2.69 Å for LDA and 2.74 Å for B3LYP), while UHF gives a very different result (5.08 Å). The fullerene π system is mainly concentrated on the hexagon bonds. For this reason, although the predicted energy difference between the hexagon site and the other positions is very small, it seems likely that the hexagon sites play a special role among the different adsorption positions.

If one neglects a possible Jahn−Teller distortion, the system has a C_{3v} symmetry in the case of a hexagon-approach path. In this case, three different states are in competition: (1) a neutral $^2A_1$ state, C_{60} + Na, which is essentially *repulsive* (except for a possible van der Waals minimum at large distance) and (2) a ionic ^2E (doubly degenerated) state, C_{60}^- + Na^+, which is *attractive*.

To treat all these states on an equal foot, a Multi-Reference (MR) approach is strongly recommended. We used our recently developed local CAS-SCF formalism,[12,13] which is able to concentrate the active space in the region of interest of a molecular system. According to our approach,[14,15] we used a *minimal* active space at the CAS-SCF level, while the effect of dynamical correlation is included via a subsequent Configuration-Interaction (CI) treatment. For the present case, the active space is composed of three orbitals: the 3s sodium orbital (3s(**Na**)) and the two degenerated π^* orbitals which are delocalized on the hexagonal ring (π^*_{hex}(**C_{60}**)).

Even with a small active space, the local MR-CI treatment on such a large system is extremely heavy. A possibility to reduce the size of the problem is to freeze some orbitals whose effect on the studied properties is expected to be small. For this reason, taking benefit of the use of localized orbitals, we froze most of the σ orbitals at the SCF level. Only the σ and σ^* orbitals nearest to Na (12 bonds, i.e., 24 orbitals) were correlated at the CI level.

The choice of the atomic basis set is a very sensible point. It is clear that a minimal basis set alone has not enough flexibility to describe the orbital relaxation and the correlation effects of this system. In particular, because of the competition between neutral and charged forms, the experimental ionization potential (IP) of Na and the electron affinity (EA) of C_{60} must be accurately reproduced. Even with a minimal basis set, this is not a problem for Na (with the (3s2p) basis

* Corresponding author phone: +33-(0)561-558593; fax: +33-(0)561-556065; e-mail: stefano@irsamc.ups-tlse.fr.
[†] Universitat de Valencia.
[‡] Université Paul Sabatier.

10.1021/ct050035v CCC: $30.25
Published on Web 08/25/2005

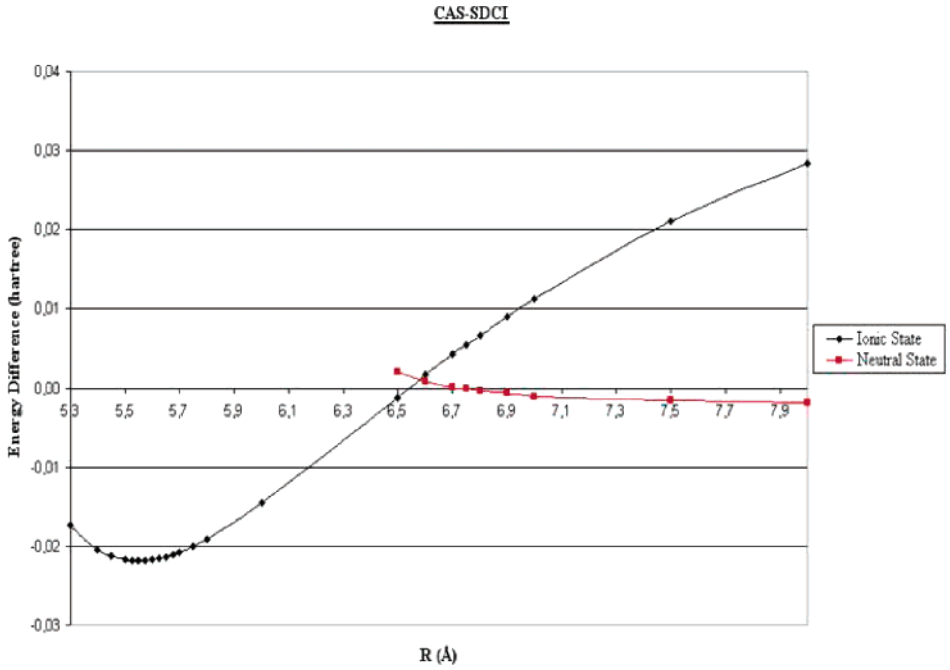


Figure 1. Energy curves of the $^2A_1$ and ^2E states as a function of the distance between the Na atom and the C_{60} center. The energy zero corresponds to the value of the neutral $^2A_1$ state at infinite distance.

we obtain 4.90 eV to be compared with the experimental IP of about 5.14 eV). On the other hand, the minimal basis set is certainly too poor for C_{60}, since it leads to an EA = 4.10 eV, while the experimental value is substantially lower (EA_{exp}=2.65 eV).

Therefore, to keep the basis-set size reasonable, we used an ANO (2s1p) basis on each carbon and a (3s2p) basis on the sodium atom,[16] supplemented with a set of uncontracted Gaussians centered on the C_{60} center. We used a (1s1p1d1f1g1h1i) set of 49 functions, having a mean radius $\langle r \rangle$ equal to the C_{60} radius. These additional orbitals are intended to provide the required angular correlation to the mobile π electrons on the C_{60} surface. With this choice of the basis set, one has EA = 2.28 eV for C_{60}, in a reasonable accord with the experimental value. Therefore, this was our final basis-set choice.[17]

The MR-CI results (single and double excitations on the local CAS-SCF wave function) are reported in Figure 1. The neutral A_1 state is the lowest one at an infinite distance at the MR-CI level, and the ionic E state lies about 2.6 eV above the A_1 one at dissociation and becomes the lowest one at about R = 6.5 Å. At about R = 5.55 Å the E state has a minimum, 0.59 eV below the asymptotic energy of the A_1 state (see Table 1, E_D). The dipole moment (15.6 D) is in a good agreement with the experimental value of 16.3 D.[6,7] The MR-CI results have been also corrected by using the Counter-Poise (CP) procedure,[18] to take into account the Basis-Set Superposition Error (BSSE) (see Figure 2). The ionic minimum becomes less deep, and it is almost degenerated with the asymptotic energy of the neutral state at dissociation. The neutral state is less affected by the correction (that comes essentially from the effect of the sodium basis set on C_{60}^-), and a *significant barrier* separates the two minima (see Table 1, E_B). On the other hand, the position of the minimum and the corresponding value of the dipole moment are very little affected by the BSSE.

Table 1. MR-CI Results (with and without BSSE Correction) and Experimental Dipole Moment

	R_e[a] (Å)	E_D[b] (eV)	E_B[c] (eV)	μ[d] (D)
NaC_{60}	5.55	0.59	0.61	15.6
NaC_{60} + BSSE	5.68	0.02	0.37	16.5
experiment				16.3

[a] R_e, equilibrium distance between the sodium atom and the center of C_{60}. The distances between Na and the closest C atoms are 2.72 and 2.83 Å (without and with BSSE correction, respectively) in good accord with the DFT values. [b] E_D, binding energy with respect to the neutral system at infinite distance. [c] E_B, dissociation barrier with respect to the ionic-state minimum. [d] μ, dipole moment.

The small energy difference between the ionic minimum and the neutral asymptote raises the question of the stability of the ionic form. Since entropy favors the dissociated species, one can expect that, at thermodynamic equilibrium, the neutral form should dominate. However, the barrier to dissociation of the ionic species is not negligible. If the ionic curve is fitted with a Morse potential, the lowest vibrational level is only 237.3 cm^{-1} higher than the Morse ionic minimum, and about 20 vibrational levels are predicted below the barrier to dissociation (E_B=0.37 eV with BSSE, see Table 1). For these reasons, one can expect the ionic minimum to have (at least) a metastable character.

It should also be noticed that several effects are expected to contribute to an ionic-minimum stabilization. One is geometry relaxation, including a symmetry lowering due to a Jahn−Teller effect. Moreover, the BSSE is often overestimated by the CP correction. Finally, the difference between the ionic and neutral asymptotes is 4.90 − 2.28 = 2.62 eV with our calculations, while the experimental value is 5.14 − 2.65 = 2.49 eV.

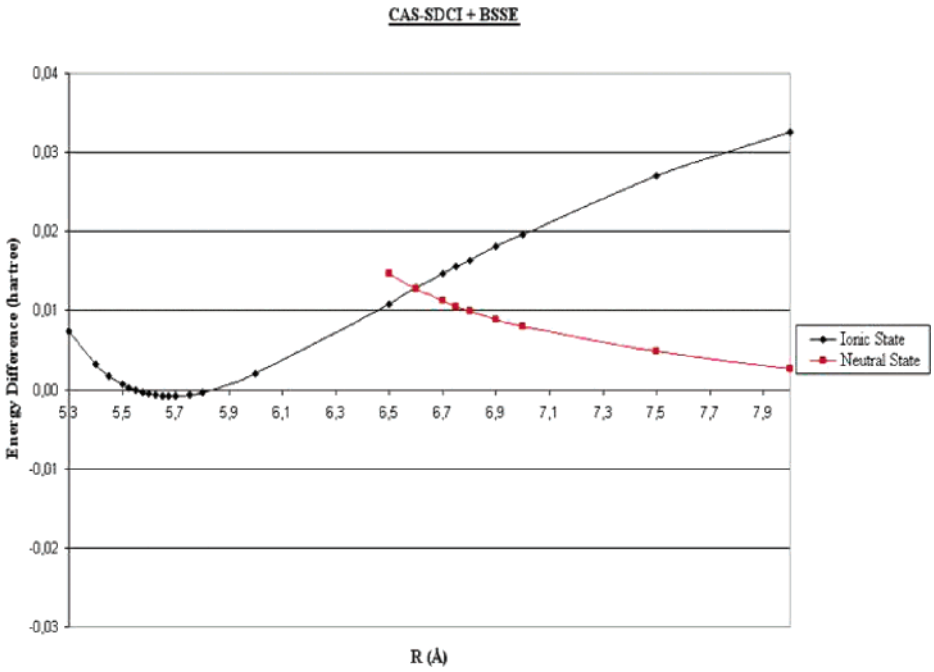


Figure 2. Same as Figure 1, with the inclusion of the BSSE correction.

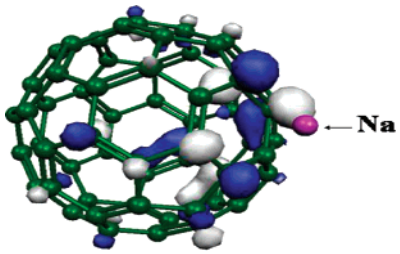


Figure 3. One of the active π^* orbitals of e symmetry, computed at the equilibrium distance of the ionic state.

In any case, our MR-CI calculations clearly indicate the presence of an ionic local minimum, having a dipole moment in excellent agreement with the experimental results. It seems likely to associate this local minimum (either an absolute one or a metastable minimum having a significant barrier to dissociation) to the experimentally detected species.

It is instructive to plot the active orbitals obtained through our local CAS-SCF algorithm. At the equilibrium distance of the attractive E state, the two π^* CAS-SCF orbitals are localized on the hexagon near to the sodium atom but show significant tails on the closest carbon atoms. This partially local behavior that was observed in local CAS-SCF treatment of other systems[19,20] is due to the effect of the charged Na^+ ion on the π^* electron, that cannot freely delocalize on the whole fullerene. One of the two degenerated π^* orbitals is shown in Figure 3.

In conclusion, the present study shows the existence of a charge-transfer local minimum on the Na + C_{60} PES. The recently observed dipole moment of this system can be interpreted as due to this state, although further investigations are needed in order to discriminate between a global minimum and a meta-stable state having a significant barrier to dissociation. As a more general conclusion, the present work proves that local-orbital techniques permit a MR-CI treatment of systems as large as a fullerene. These techniques can be a useful tool for those cases where several electronic states are in competition and for which DFT still presents open problems.

Acknowledgment. We wish to thank Fernand Spiegelman and Jose Sánchez-Marín for helpful discussions and suggestions. One of us (J.V.P.R.) would like to acknowledge financial support from the European Community through the TMR activity "Marie Curie Research Training Grants", Grant No. HPMF-CT-2002-01521. Financial support from the European Community (COST D26) and the French CNRS are also gratefully acknowledged.

References

(1) Wang, L. S.; Cheshnovsky, O.; Smalley, R. E.; Carpenter, J. P.; Hwo, S. J. *J. Chem. Phys.* **1992**, *96*, 4028−4031.

(2) Wang, L. S.; Alford, J. M.; Chai, Y.; Diener, M.; Zhang, J.; McClure, S. M.; Guo, T.; Scuseria, G. E.; Smalley, R. E. *Chem. Phys. Lett.* **1993**, *207*, 354−359.

(3) Tegelmann, R.; Krawez, N.; Lin, S. H.; Hertel, I. V.; Campbell, E. E. *Nature (London)* **1996**, *382*, 407.

(4) Aree, T.; Kerdcharoen, T.; Hannongbua, S. *Chem. Phys. Lett.* **1998**, *285*, 221−225, and references therein.

(5) Palpant, B.; Negishi, Y.; Sanekata, M.; Miyajima, K.; Nagao, S.; Judai, K.; Rayner, D. M.; Simard, B.; Hackett, P. A.; Nakjima, A.; Kaya, K. *J. Chem. Phys.* **2001**, *114*, 8459−8466, and references therein.

(6) Antoine, R.; Rayane, D.; Benichou, E.; Dugourd, Ph.; Broyer, M. *Eur. Phys. J. D* **2000**, *12*, 147−151.

(7) Dugourd, Ph.; Antoine, R.; Rayane, D.; Compagnon, I.; Broyer, M. *J. Chem. Phys.* **2001**, *114*, 1970−1973.

(8) Roques, J.; Calvo, F.; Spiegelman, F.; Mijoule, C. *Phys. Rev. B* **2003**, *68*, 205412.

(9) Roques, J.; Calvo, F.; Spiegelman, F.; Mijoule, C. *Phys. Rev. Lett.* **2003**, *90*, 75505.

(10) Shira, A. S.; Ray, A. K. *Phys. Rev. A* **1995**, *52*, 141−148.

(11) Hamamoto, N.; Jitsukawa, J.; Satoko, C. *Eur. Phys. J. D* **2002**, *19*, 211−221.

(12) Maynau, D.; Evangelisti, S.; Guihéry, N.; Calzado, C. J.; Malrieu, J. P. *J. Chem. Phys.* **2002**, *116*, 10060−10068.

(13) Angeli, C.; Evangelisti, S.; Cimiraglia, R.; Maynau, D. *J. Chem. Phys.* **2002**, *117*, 10525−10533.

(14) Pitarch-Ruiz, J.; Evangelisti, S.; Maynau, D. *J. Mol. Struct. (THEOCHEM)* **2004**, *681*, 203−207.

(15) Pitarch-Ruiz, J.; Calzado, C. J.; Evangelisti, S.; Maynau, D. *Int. J. Quantum Chem.*, in press.

(16) Widmark, P.-O.; Malmqvist, P.-Å.; Roos, B. O. *Theor. Chim. Acta* **1990**, *77*, 291−306.

(17) Different strategies can also be imagined. We tested, for instance, the Locally-Dense-Basis-Set (LDBS) approach of di Labio and co-workers, that corresponds to use a larger basis set on the atoms close to the point of attach of Na. This has as a consequence, however, to induce a spurious permanent dipole moment on *isolated* C_{60} of 3.13 D. Since the dipole moment is precisely the physical quantity to be reproduced, the LDBS formalism does not seem appropriate in the present case.

(18) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.

(19) Pitarch-Ruiz, J.; Evangelisti, S.; Maynau, D. *Chem. Phys. Lett.* **2003**, *372*, 22−27.

(20) Pitarch-Ruiz, J.; Evangelisti, S.; Maynau, D. *Int. J. Quantum Chem.* **2004**, *97*, 688−699.

CT050035V

JCTC Journal of Chemical Theory and Computation

Chemical Notions from the Electron Density

Jaime Fernández Rico, Rafael López,* Ignacio Ema, and Guillermo Ramírez

Departamento de Química Física Aplicada, Facultad de Ciencias C-XIV, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Received April 11, 2005

Abstract: The study of density and the role played by its atomic representation is proposed as a way for the rationalization of chemical behavior. As this behavior has been long rationalized in terms of the basic concepts of empirical structural chemistry, a direct link between both approaches is searched for by using the exact representation of the density provided by the deformed atoms in molecules method (Rico, J. F.; López, R.; Ema, I.; Ramírez, G.; Ludeña, E. *J. Comput. Chem.* **2004**, *25*, 1355−1363). Noting that the spherical terms of the pseudoatoms cannot be mainly responsible for the chemical behavior, we study the small nonspherical deformations and find that they reflect and support all basic concepts of empirical structural chemistry. Lone pairs; single, double, and triple bonds; different classes of atoms; functional groups; and so forth are paralleled by the density deformations in a neat manner. These facts are illustrated with several examples.

1. Introduction

Because one-electron density plays an increasingly central role[1−9] in both the conceptual and practical developments of theoretical chemistry, the methods for its meaningful analysis are of paramount importance.[10−26]

In the Born−Oppenheimer approximation (the paradigm in the study of molecular structure), the electronic energy including the nuclear repulsion is the potential energy for the movement of the nuclei, and as a consequence, the components of the force acting on a nucleus are determined by their derivatives with respect to its coordinates. The Hellmann−Feynman (electrostatic)[27] theorem states that these derivatives are equal to the components of the electrostatic force generated by the electron cloud plus the remaining nuclei. Thus, the forces can be obtained in two ways: from the electronic energy through its derivatives or from the electron density using classical electrostatics. The first way is expensive and hardly provides chemical insight. The second one is very cheap to apply and provides plenty of chemical insight.

Nonetheless, although the electrostatic theorem has been known for more than 60 years, the possibilities that it opens up have been scarcely exploited, and today the theorem is mostly regarded as a scientific curiosity. There have been two main reasons for this. The first one is that the fulfillment of the theorem requires high quality densities. In particular, it leads to disastrous results for densities computed with commonly used poor basis sets, whereas energy is less sensitive to the quality of the basis set. The second reason comes from the fact that, for extracting chemical information from the theorem, one needs a representation of the density that brings insight to the chemist.

These reasons no longer hold. As it has been recently proved,[28] densities computed with good Slater basis sets, and with very high quality Gaussian basis sets too, fulfill the electrostatic theorem with an accuracy that is sufficient for most quantitative applications and, a fortiori, for the qualitative ones as well. Moreover, a representation of the density[29−31] aimed to retain the identity of the atoms in a molecule as much as possible has also been reported, and that, in turn, facilitates the application of the electrostatic theorem.

This representation was originally intended as an aid for the calculation of several functionals of electron density such as the molecular electrostatic potential, molecular force field, forces on nuclei, and so forth, and the usefulness in those applications was proved.[32] The method was also applied to the analysis of binding forces, to the calculation of bonding

* Corresponding author. E-mail: rafael.lopez@uam.es. Phone: +34−91−4978642.

10.1021/ct0500951 CCC: $30.25
Published on Web 08/27/2005

energies from density in diatomics,[33] and to the explanation of rotational barriers in terms of the density.[34]

These studies render evidence that, when this representation of the density is combined with the electrostatic theorem, basic concepts of chemistry can be regarded from a novel perspective that may help to build a bridge between electron density and the classical notions of empirical structural chemistry. The exploration of this possibility is the aim of this work.

In Section 2, the electrostatic theorem is rederived in a way oriented to stress that chemical forces and energy variations are determined by the density alone. It follows that the relevant physical effects must be reflected in the density and must affect the forces and energies through density, a fact that, in turn, sets the study of density and its relationship with these forces as the central problem in the rationalization of chemical behavior. Section 3 deals with this relationship and with the role played by the atomic representation of the density in this regard.

It is noted there that the part of the density spherically distributed around the nuclei, although being largely dominant, is not responsible for the molecular stability. The chemical behavior of the molecules is mainly determined by small nonspherical deformations caused by the molecular environment on the density of the atoms.

On the other hand, chemical behavior has been largely rationalized by means of the concepts of empirical structural chemistry. As noted in Section 4, as much as they have been supported by many years of successful usage, these concepts must have some physical ground, and thus, the results of Section 2 imply that they must be somehow reflected in the density, and the arguments of Section 3 suggest that this should be searched for mainly in the density deformations.

Section 5 contains a summary of the results of this search, and some comments about the implications of these results are included in the final section.

2. Density, Chemical Forces, and Bonding

Because the bonding of atoms is usually explained in terms of electron pairing or exchange, while the electrostatic approach and its implications are often disregarded, a brief review on this subject is pertinent.

The electrostatic approach has its roots in the Hellmann−Feynman theorem,[27] which has been considered by Slater as *one of the most powerful theorems applicable to molecules*.[35] Herein, we will rederive the theorem in a way oriented to stress that chemical forces and bonding are related to either the whole electronic energy or the electron density through fundamental equations, whereas the relation with the components of energy is indirect.

Let $\hat{\mathscr{H}}$ be the electronic Hamiltonian of a molecule with a fixed number of electrons and an external potential $v(\mathbf{r})$, and let Ψ_I be its normalized eigenfunctions. Therefore

$$\langle \Psi_I | \hat{\mathscr{H}} | \Psi_I \rangle = E_I \tag{1}$$

and

$$\langle \Psi_J | \hat{\mathscr{H}} | \Psi_I \rangle = 0 \tag{2}$$

As it is well-known, these equations can be written as

$$E_I = \int d\mathbf{r}\, \rho^{II}(\mathbf{r})\, v(\mathbf{r}) + \int d\mathbf{r} \left[-\frac{1}{2}\nabla'^2\, \hat{\rho}^{II}(\mathbf{r},\mathbf{r}') \right]_{\mathbf{r}=\mathbf{r}'} + \int d\mathbf{r} \int d\mathbf{r}' \frac{\Gamma^{II}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{3}$$

and

$$0 = \int d\mathbf{r}\, \rho^{JI}(\mathbf{r})\, v(\mathbf{r}) + \int d\mathbf{r} \left[-\frac{1}{2}\nabla'^2\, \hat{\rho}^{JI}(\mathbf{r},\mathbf{r}') \right]_{\mathbf{r}=\mathbf{r}'} + \int d\mathbf{r} \int d\mathbf{r}' \frac{\Gamma^{JI}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{4}$$

where $\hat{\rho}^{II}(\mathbf{r},\mathbf{r}')$, $\rho^{II}(\mathbf{r})$, and $\Gamma^{II}(\mathbf{r},\mathbf{r}')$ are, respectively, the density matrix, the density function, and the pair density function and $\hat{\rho}^{JI}(\mathbf{r},\mathbf{r}')$, $\rho^{JI}(\mathbf{r})$, and $\Gamma^{JI}(\mathbf{r},\mathbf{r}')$ are the transition density matrix, transition density, and transition pair density functions, respectively.

Since a first-order change in the wave function Ψ_I can be expressed in terms of the Ψ_J:

$$\Psi' = \frac{\Psi_I + \sum_{J\neq I} \lambda_J \Psi_J}{(1 + \sum_{J\neq I} |\lambda_J|^2)^{1/2}} \tag{5}$$

Equation 2 implies

$$0 = \int d\mathbf{r}\, \delta\rho(\mathbf{r})\, v(\mathbf{r}) + \int d\mathbf{r} \left[-\frac{1}{2}\nabla'^2\, \delta\hat{\rho}(\mathbf{r},\mathbf{r}') \right]_{\mathbf{r}=\mathbf{r}'} + \int d\mathbf{r} \int d\mathbf{r}' \frac{\delta\Gamma(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{6}$$

where $\delta\hat{\rho}(\mathbf{r},\mathbf{r}')$, $\delta\rho(\mathbf{r})$, and $\delta\Gamma(\mathbf{r},\mathbf{r}')$ are, respectively, the transition density matrix, density function, and pair density function associated with the wave function and its first-order change. Because this equation is valid for every electronic state, here and in subsequent equations, the indices referring to states will be suppressed for simplicity.

In the study of the electronic structure of molecules, the external potential is a continuous function $v(\mathbf{r},\lambda)$ of a set of parameters, $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)$ (the nuclear charges and the degrees of freedom associated with the coordinates of the nuclei), so that all of the terms of eq 3 are functions of these parameters.

Derivation with respect to λ_p yields

$$\frac{\partial E(\lambda)}{\partial \lambda_p} = \int d\mathbf{r}\, \rho(\mathbf{r},\lambda) \frac{\partial v(\mathbf{r},\lambda)}{\partial \lambda_p} + \int d\mathbf{r}\, v(\mathbf{r},\lambda) \frac{\partial \rho(\mathbf{r},\lambda)}{\partial \lambda_p} + \int d\mathbf{r} \left[-\frac{1}{2}\nabla'^2 \frac{\partial \hat{\rho}(\mathbf{r},\mathbf{r}',\lambda)}{\partial \lambda_p} \right]_{\mathbf{r}=\mathbf{r}'} + \int d\mathbf{r} \int d\mathbf{r}' \frac{\partial}{\partial \lambda_p} \frac{\Gamma(\mathbf{r},\mathbf{r}',\lambda)}{|\mathbf{r}-\mathbf{r}'|} \tag{7}$$

and since eq 6 holds for every λ, the three last terms must cancel out. Thus

$$\frac{\partial E(\lambda)}{\partial \lambda_p} = \int d\mathbf{r}\, \rho(\mathbf{r},\lambda) \frac{\partial v(\mathbf{r},\lambda)}{\partial \lambda_p} \tag{8}$$

which is the Hellmann−Feynman theorem.

In the study of a particular molecule, the nuclear charges, ζ_A, are fixed, and the external potential only depends on the

nuclear coordinates, that is, $\lambda = \lambda(\mathbf{R}_1, \mathbf{R}_2, ...)$. Furthermore, the electronic energy plus the nuclear repulsion plays the role of potential energy, $E_T(\lambda)$, for the movement of the nuclei, and hence, the force acting upon every nucleus is determined by its gradient. Therefore, one has

$$\mathbf{F}_A = -\vec{\nabla}_A E_T = \zeta_A \left[\int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r},\lambda) \frac{\mathbf{r} - \mathbf{R}_A}{|\mathbf{r} - \mathbf{R}_A|^3} - \sum_{B \neq A} \zeta_B \frac{\mathbf{R}_B - \mathbf{R}_A}{|\mathbf{R}_B - \mathbf{R}_A|^3} \right] \quad (9)$$

where

$$\vec{\nabla}_A = \mathbf{i}\frac{\partial}{\partial X_A} + \mathbf{j}\frac{\partial}{\partial Y_A} + \mathbf{k}\frac{\partial}{\partial Z_A} \quad (10)$$

which is the electrostatic theorem.

In response to a previous criticism by Coulson and Bell[36] on the Hellmann−Feynman theorem, Berlin noted[37] that the nuclear attraction term

$$V(\lambda) = \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r},\lambda)\, v(\mathbf{r},\lambda) \quad (11)$$

can be considered as the sum of two different components: such that

$$V(\lambda) = V_1(\lambda) + V_2(\lambda) \quad (12)$$

$$\frac{\partial V_1(\lambda)}{\partial \lambda_p} = \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r},\lambda) \frac{\partial v(\mathbf{r},\lambda)}{\partial \lambda_p} \qquad \forall p \quad (13)$$

and

$$\frac{\partial V_2(\lambda)}{\partial \lambda_p} = \int \mathrm{d}\mathbf{r}\, v(\mathbf{r},\lambda) \frac{\partial \rho(\mathbf{r},\lambda)}{\partial \lambda_p} \qquad \forall p \quad (14)$$

Thus, if $T(\lambda)$ and $G(\lambda)$ are, respectively, the kinetic energy and the two-electron repulsion, and eq 6 is taken into account, one has

$$\frac{\partial E(\lambda)}{\partial \lambda_p} = \frac{\partial V_1(\lambda)}{\partial \lambda_p} \quad (15)$$

$$\frac{\partial}{\partial \lambda_p}[V_2(\lambda) + T(\lambda) + G(\lambda)] = 0 \quad (16)$$

Berlin used these equations to invalidate Coulson and Bell's objection to the electrostatic theorem, but one can go beyond this point analyzing their implications. Thus, eqs 15 and 16 imply

$$E(\lambda_2) - E(\lambda_1) = V_1(\lambda_2) - V_1(\lambda_1) = \int_{\lambda_1}^{\lambda_2} \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r},\lambda)\, \mathrm{d}v_\lambda(\mathbf{r},\lambda) \quad (17)$$

where

$$\mathrm{d}v_\lambda(\mathbf{r},\lambda) = \sum_p \frac{\partial v(\mathbf{r},\lambda)}{\partial \lambda_p} \mathrm{d}\lambda_p \quad (18)$$

and

$$V_2(\lambda) + T(\lambda) + G(\lambda) = \mathrm{const} \quad (19)$$

Equation 17 is the integrated Hellmann−Feynman theorem,[5,38] which establishes an explicit relationship between density and energy.

Equation 19 is the extremum condition of eq 6 written in a different way and implies that throughout a chemical process, the changes in kinetic energy, $T(\lambda)$, and in electron repulsion, $G(\lambda)$, are just canceled by changes in $V_2(\lambda)$.

Berlin's partition of eq 12 is not the only way to look at chemical bonding. Since energy and its components are scalars, instead of dividing the nuclear attraction term into two parts, one could divide the kinetic energy, or the electron repulsion, or some combination of both, and define constants involving parts of these quantities, as it has been implicitly done in studies of the chemical bond based on the kinetic energy density[39−46] or in conventional explanations based on the electron pair density.

Nonetheless, it seems that Berlin's partition of $V(\lambda)$ is the most natural choice and, probably, the most simple and useful. This partition is grounded in two fundamental equations: the extremum condition—eq 6—and the Hellmann−Feynman theorem—eq 8. It introduces two functions endowed with very clear insight: $V_1(\lambda)$ is associated with the total electron energy, and $V_2(\lambda)$ is associated with the universal functional[47] $F = T + G$ of the density functional theory.[1−4] Note, in this respect, that

$$-\mathrm{d}V_2(\lambda) = \mathrm{d}T(\lambda) + \mathrm{d}G(\lambda) = \mathrm{d}F(\lambda) \quad (20)$$

$$V_2(\lambda) = \mathrm{const} - F(\lambda) \quad (21)$$

Finally, it provides relations for linking these functions with the basic physical variables: electron density and external potential,[48−52] which are exact, simple, and easy to apply in practice.

Bearing in mind these arguments, our search for the relationship between the classical notions of chemistry and electron density will be based on this approach.

It is noteworthy that, from this point of view, kinetic energy and electron repulsion only appear in the extremum condition, eq 6, which fixes the density for every conformation, and therefore, they influence the energy through density.

In this context, the physical effects appearing in the kinetic energy or in the two-electron term cannot be directly related to bond energies, because these quantities are part of the constant in eq 19. One must consider, instead, using eq 6, how they affect the density and, next, through eq 17, whether or not the induced changes favor the decrease in energy. In summary, from this perspective, one must be aware that every physical effect must put its mark in the density, and it is through this mark that it affects the energy.

3. Density Partition and Forces: A Way For Rationalizing Chemistry

The arguments of the previous section prove that the rationalization of chemical behavior must deal with the study of density and its evolution along the chemical process. This study provides us a complete theoretical description because, from density, it is possible not only to follow the changes of the electronic energy but also to analyze the forces acting

upon the several parts of the system and to know how they contribute to changing the energy.

In particular, it must be noted that the direct relation between density and forces ensures that fragments with similar densities in different molecules will contribute in a similar way to the driving forces of their chemical processes and that this fact sets the description of the molecular electron densities as the key for the comparative study of their chemical properties.

As it is well-known, the electron density of a molecule in its ground state is highly peaked and strongly concentrated around the nuclear positions for every conformation. Thus, it is convenient to partition the electronic density of a molecule into fragments centered at the nuclei and such that every fragment accompanies its nucleus in the displacement to the largest possible extent.

Denoting the fragments as $\rho^A(\mathbf{r}_A)$, one can write

$$\rho(\mathbf{r}) = \sum_A \rho^A(\mathbf{r}) \tag{22}$$

and following the usual convention, one can consider the fragments $\rho^A(\mathbf{r})$ as atoms in molecules (or pseudo-atoms), though being aware that atoms in molecules are no longer quantum mechanical observables and, therefore, that this denomination is just a mere resort to endow them with chemical insight.

There are several ways to carry out the decomposition of eq 22, but irrespective of their merits or limitations, all of them can be combined with the electrostatic theorem to yield a model of forces in the system. Thus, all of them imply a separation of the force on each nucleus into two distinct contributions. Combining eq 22 with eq 9, one has

$$\mathbf{F}_A = \zeta_A \int \mathrm{d}\mathbf{r}\, \rho^A(\mathbf{r}_A) \frac{\mathbf{r}_A}{r_A^3} + \zeta_A \sum_{B\neq A} \left[\int \mathrm{d}\mathbf{r}\, \rho^B(\mathbf{r}_B) \frac{\mathbf{r}_A}{r_A^3} - \zeta_B \frac{\mathbf{R}_B - \mathbf{R}_A}{|\mathbf{R}_B - \mathbf{R}_A|^3} \right] \tag{23}$$

where ζ_I denotes the nuclear charges and $\mathbf{R}_I$ denotes the nuclei positions.

The first contribution on the right-hand side of eq 23 is the internal or *self-pulling* force, namely, the force exerted on nucleus *A* by its own electronic cloud. The second one is the *external force* exerted on *A* by the clouds and nuclei of the remaining atoms.

As it has been remarked above, the electron density of a molecule is strongly concentrated around its nuclei and, for every reasonable partition, the spherical averages of the atomic densities:

$$f_{00}^A(r_A) = \frac{1}{4\pi} \int_0^{2\pi} \mathrm{d}\phi_A \int_0^{\pi} \mathrm{d}\theta_A \sin\theta_A\, \rho^A(\mathbf{r}_A) \tag{24}$$

are, by far, the largest components of the density. Therefore, it is convenient to distinguish this dominant part from the small remaining parts:

$$\Delta^A(\mathbf{r}_A) = \rho^A(\mathbf{r}_A) - f_{00}^A(r_A) \tag{25}$$

and to analyze separately the roles of the spherical components and the rest in the force model of eq 23. For this purpose, one can write

$$\mathbf{F}_A = \zeta_A \int \mathrm{d}\mathbf{r}\, f_{00}^A(r_A) \frac{\mathbf{r}_A}{r_A^3} + \zeta_A \sum_{B\neq A} \left[\int \mathrm{d}\mathbf{r}\, f_{00}^B(r_B) \frac{\mathbf{r}_A}{r_A^3} - \zeta_B \frac{\mathbf{R}_B - \mathbf{R}_A}{|\mathbf{R}_B - \mathbf{R}_A|^3} \right] + \zeta_A \sum_B \int \mathrm{d}\mathbf{r}\, \Delta^B(r_B) \frac{\mathbf{r}_A}{r_A^3} \tag{26}$$

Symmetry considerations are sufficient to see that the first term on the right-hand side is null. Moreover, the second term can be written in terms of the effective charges, ζ_B^{eff}, given by the Gauss theorem

$$\zeta_B^{\mathrm{eff}} = \zeta_B - 4\pi \int_0^{|\mathbf{R}_B - \mathbf{R}_A|} \mathrm{d}r_B\, r_B^2\, f_{00}^B(r_B) \tag{27}$$

so that the expression (eq 26) of the force can be rewritten as

$$\mathbf{F}_A = -\zeta_A \zeta_B^{\mathrm{eff}} \frac{\mathbf{R}_B - \mathbf{R}_A}{|\mathbf{R}_B - \mathbf{R}_A|^3} + \zeta_A \sum_B \int \mathrm{d}\mathbf{r}\, \Delta^B(r_B) \frac{\mathbf{r}_A}{r_A^3} \tag{28}$$

As one could expect, *the spherical terms yield no contribution to the internal forces and only participate in the external forces through the partial screening of the nuclear charges*. It should be noted, in this respect, that the electronic charge of an atom is given by

$$Q_{00}^B = 4\pi \int_0^{\infty} \mathrm{d}r_B\, r_B^2\, f_{00}^B(r_B) \tag{29}$$

and, hence, in atoms with positive density, the second term of eq 27 will be smaller than the electronic charge. As a consequence, neutral atoms or cations have a positive ζ_B^{eff} and give a net repulsive force on *A* at all distances. For anions, ζ_B^{eff} is negative at long distances but it turns to positive at short distances. Thus, the forces exerted by negative ions on nucleus *A* will be attractive when they are far away from *A*, but they will repel it at short distances.

In view of this, it is very illustrative to consider what the forces between a pair of atoms or ions with positive defined and strictly spherical densities would be. In this case, the self-pulling forces are obviously zero and only external forces remain. As proved before, these forces are repulsive for couples of neutral atoms, an atom and a cation, or pairs of cations. In the case of a system consisting of a cation (or a neutral atom) and an anion, the force on the anion nucleus will be repulsive at all distances whereas the force on the cation nucleus will be attractive at long distances and repulsive at short ones. Finally, for a couple of negative ions, the forces on both nuclei will be attractive at sufficiently long distances and will become repulsive when the separation decreases. Moreover, the forces on both nuclei will be, in general, different, and the reversions will occur at different distances.

Certainly, the absence of interactions between the clouds in the context of the electrostatic theorem makes these forces very different from the ones expected when electrostatics is applied to spherical ions in the usual way, and this leads to an evident conclusion: *none of these couples can conform a stable system*. To form a stable system, densities cannot have spherical symmetry or, in other words, *density deformations are essential for the appearance of stable systems*.

Let us complete this analysis by considering the contribution of the atomic deformations to the forces. To do this, atomic densities can be expanded in spherical harmonics centered at the nuclei:

$$\rho^A(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} z_l^m(\mathbf{r}_A)\, f_{lm}^A(r_A) \tag{30}$$

where $f_{lm}^A(r_A)$ are atomic radial factors:

$$f_{lm}^A(r_A) = \frac{2l+1}{2(1+\delta_{m0})\pi r^l}\frac{(l-|m|)!}{(l+|m|)!}\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \sin\theta\, z_l^m(\mathbf{r}/r)\, \rho^A(\mathbf{r}) \tag{31}$$

and $z_l^m(\mathbf{r})$ are the regular harmonics:

$$z_l^m(\mathbf{r}) = r^l(-1)^m P_l^m(\cos\theta)\cos(m\phi) \qquad \text{for } m \geq 0$$
$$z_l^m(\mathbf{r}) = r^l(-1)^{|m|} P_l^{|m|}(\cos\theta)\sin(|m|\phi) \qquad \text{for } m < 0 \tag{32}$$

$P_l^{|m|}$ being the associated Legendre functions.[53]

The terms with $l = 0$ are just the spherical averages whose roles were discussed above. The remaining terms ($l > 0$) are the atomic deformations:

$$\Delta^A(\mathbf{r}) = \rho^A(\mathbf{r}) - f_{00}^A(r_A) = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} z_l^m(\mathbf{r}_A)\, f_{lm}^A(r_A) \tag{33}$$

but decomposed so that one can identify dipole-type contributions ($l = 1$), quadrupole ($l = 2$), octapole ($l = 3$), and so on.

From this expansion, one easily finds that

$$\mathbf{F}_{\mathrm{int}}^A = \zeta_A \int \mathrm{d}\mathbf{r}\, \Delta^A(\mathbf{r})\frac{\mathbf{r}_A}{r_A^3} = -\mathbf{i}q_{11}^A - \mathbf{j}q_{1-1}^A - \mathbf{k}q_{10}^A \tag{34}$$

where

$$q_{1m}^A = \frac{4\pi}{3}\zeta_A \int_0^{\infty} \mathrm{d}r\, r\, f_{1m}^A(r) \qquad m = -1, 0, 1 \tag{35}$$

It is clear that *the self-pulling force on an atom is only determined by the dipole-type deformation of its own density*. Moreover,

$$\mathbf{F}_{\mathrm{ext}}^A = \zeta_A \sum_{B\neq A}\left[\int \mathrm{d}\mathbf{r}\, \rho^B(\mathbf{r})\frac{\mathbf{r}_A}{r_A^3} - \zeta_B\frac{\mathbf{R}_B - \mathbf{R}_A}{|\mathbf{R}_B - \mathbf{R}_A|^3}\right] = \zeta_A\left[\zeta_B^{\mathrm{eff}}\frac{\mathbf{R}_B - \mathbf{R}_A}{|\mathbf{R}_B - \mathbf{R}_A|^3} + \vec{\nabla}_A \sum_{B\neq A}\sum_{l=1}^{\infty}\sum_{m=-l}^{l}\frac{z_l^m(\mathbf{R}_B - \mathbf{R}_A)}{|\mathbf{R}_B - \mathbf{R}_A|^{2l+1}} V_{lm}^B(|\mathbf{R}_B - \mathbf{R}_A|)\right] \tag{36}$$

where

$$V_{lm}^B(\mathbf{r}) = \frac{4\pi}{2l+1}\left[\int_0^r \mathrm{d}r'\, r'^{2l+2} f_{lm}^B(r') + r^{2l+1}\int_r^{\infty} \mathrm{d}r'\, r'\, f_{lm}^B(r')\right] \tag{37}$$

so that the external force contains the forces associated with the density deformation beside the contribution of the screened nuclear charges.

We stress that the atomic deformations caused by the molecular environment play a basic role in the study of chemical behavior supported by the electrostatic theorem. As proved above, they completely determine the self-pulling forces and a part of the external forces, being essential for the existence of stable systems.

We finally note that, though these conclusions are valid for every *exact* partition of the density, the detailed form of both the spherical averages and the atomic deformations depend on the definition of the atomic fragments, $\rho^A(\mathbf{r})$, and thus, one is faced with the choice of the partition criterion. We will return to this point in the following section.

4. Empirical Structural Chemistry: The Conventional Rationalization Of Chemistry

Let us consider, now, the main ideas employed by experimentalists in the description of chemical behavior. As it is well-know, they need neither to determine or analyze the density nor to invoke the electrostatic theorem. Instead of it, they use some simple concepts that have their roots in empirical structural chemistry, mostly developed before the advent of quantum mechanics, from the rationalization of empirical data.

In this thought, a structural hypothesis is basic, according to which molecules are composed by bonded atoms that are distributed in space adopting characteristic conformations. Moreover, there are groups of atoms (functional groups) that share conformations and properties in distinct molecules.

Atoms are bonded by their valence electrons, whose structures can be described by Lewis' rules[54] in terms of lone pairs (owned by individual atoms), bond pairs (shared by pairs of atoms), and eventually by lone electrons, all of them lying around the ionic core formed by the nuclei and their core electrons.

Bonds can be single, double, or triple according to the number of the bond pairs shared by a couple of atoms. The set of bonds conforms the skeleton of the molecule, and chemical reactions are regarded as reorganizations of the skeleton in which some bonds are broken and others are formed.

These basic assumptions are too simple to embrace all chemical behavior, but complemented with some ideas extracted from valence theory (resonance, hyperconjugation, etc.), they compose a powerful tool that allows experimentalists to describe macroscopic substances and chemical processes at the molecular level and to carry out temptable predictions, in summary, to rationalize most of the chemical behavior.

One has, thus, two different approaches to the rationalization of chemistry: the one based on the study of electron density and, in particular, on its small deformations and that of empirical structural chemistry, mainly based on Lewis' structures.

The first one is firmly grounded in theory. The second is supported by many years of successful usage. In view of this, it seems licit to ask ourselves whether they are as different as they appear at first sight or, on the contrary, they are simply two different comprehensions of the same objects. Note, in this respect, that according to the discussion in Section 2, *as far as Lewis structures and the remaining concepts of empirical structural chemistry have some physical basis, they must be somehow reflected in electron density*.

To analyze whether, and eventually how, they are reflected, we have carried out a systematic study of density oriented to analyze its relationship with Lewis' structures and other basic concepts of empirical structural chemistry.

In the analysis, we have used *atomic deformations*, which are defined as the atomic densities minus their spherical averages, that is, the Δ^A functions introduced in eq 25; *group deformations*, which are the sum of the atomic deformations of a given group of atoms:

$$\Delta^G(\mathbf{r}) = \sum_A^{\text{group}} \Delta^A(\mathbf{r}_A) \tag{38}$$

and, finally, *molecular deformations*, in which the sum extends over all the atoms of a given molecule:

$$\Delta(\mathbf{r}) = \sum_A^{\text{molec}} \Delta^A(\mathbf{r}_A) \tag{39}$$

Molecular deformations have an antecedent in (but should not be confused with) the so-called charge density difference functions,[55] which are the differences between the molecular density and the density of a hypothetical molecule built with the isolated atoms, in suitable valence states, and placed in the corresponding sites with appropriate orientations.

Note, in this respect, that the spherical averages defined in eq 24 are obtained from the very molecular density itself, without invoking the existence of reference atoms in ad hoc valence states.

Thus, conceptual distinctions apart, the density difference functions, $D(\mathbf{r})$, contain terms dependent on the angular moments of the reference atoms, which do not appear in the molecular deformations as defined above. In fact, the density of an isolated atom in a state with angular moment L has the expansion

$$\rho'^A(\mathbf{r}_A) = \sum_{l=0}^{L} \sum_{m=-2l}^{2l} z_{2l}^m(\mathbf{r}_A)\, f'_{2l\ m}(r_A) \tag{40}$$

If the molecular density is partitioned and expanded—according to eqs 22, 30, and 31—and the densities of the isolated atoms are subtracted, one has

$$D(\mathbf{r}) = \Delta(\mathbf{r}) + \sum_A [f_{00}^A(r_A) - \sum_{l=0}^{L} \sum_{m=-2l}^{2l} z_{2l}^m(\mathbf{r}_A)\, f'^A_{2lm}(r_A)] \tag{41}$$

If, as usual, the reference atoms are spherical, the $D(\mathbf{r})$ function will contain spherical contributions, $f_{00}^A(r_A) - f'^A_{00}(r_A)$, that are not present in molecular deformations. If they are in a P state, a further contamination with quadrupole-type terms appear, and so on.

Atomic and group deformations do not have, to our knowledge, direct antecedents, yet Hirshfeld and Rzotkiewicz[56] have researched this type of information on diatomics in the context of the charge density difference functions.

It was remarked in Section 3 that the detailed form of the atomic deformations depends on the particular partition adopted to define the atomic fragments, which, as it is well-known, fall into two categories. In space-partition schemes,[10,12−17] the molecular density is separated into pieces, each one contained in an atomic domain Ω_A (satisfying $\cup_{A=1}^M \Omega_A = R^3$ and $\Omega_A \cap \Omega_B = \phi$, $\forall A \neq B$) so that

$$\rho(\mathbf{r} \in R^3) = \sum_{A=1}^{M} \rho(\mathbf{r} \in \Omega_A) \tag{42}$$

Alternatively, the density can be decomposed in overlapping fragments centered at the nuclei, each one extending over the whole space R^3:[11,18,19,21,29−31,57]

$$\rho(\mathbf{r} \in R^3) = \sum_{A=1}^{M} \rho(\mathbf{r}_A \in R^3) \tag{43}$$

Though the general features of the forces model discussed above are valid for every decomposition in which the fragments add up exactly to the whole density (note that this only excludes approximations based on spherical atoms and the like[18,57]), not all of the partition schemes are equally useful in this context. It must be recalled, in this regard, that conventional studies of density are aimed at analyzing the dependence of $\rho(\mathbf{r},\lambda)$ with respect to $\mathbf{r}$ for fixed λ, while for the study of forces and energy, the essential point is how the density evolves when λ varies.

From a practical standpoint, the methods of the second category are much more suitable in this respect. Among them, the method of *deformed atoms in molecules* (DAM)[29−31] offers clear advantages because it allows for the direct application of the approach discussed in Section 3, and therefore, we will use it in this study. In the DAM method, the partition of the density is intended to retain, as much as possible, the identity of the atoms within the molecule at every conformation. Basically, in this method, the minimal distortion of each atomic density is achieved by assigning to it the one-center charge distributions centered at its nucleus as well as the parts of the two-center distributions nearest to

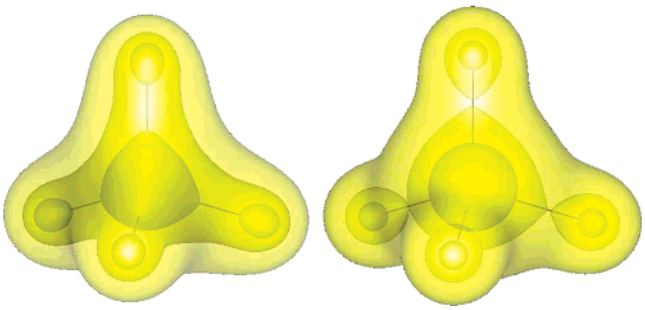


Figure 1. Electron density surfaces in methane. Total density (left) and summation of the spherical atomic contributions (right). Contour values: 0.3 (innermost), 0.2, and 0.1 (outermost) au.

it. With such a partition, the densities of all the fragments exactly retrieve (on summation) the whole density and, furthermore, can be accurately expanded when regular harmonics are multiplied by analytical radial factors, a fact that greatly facilitates the application of the electrostatic theorem. Though this method has been described in detail in previous works,[29−31] a short review is given in the Appendix for the sake of completeness.

5. Results

We have studied the density deformations in a sizable sample consisting of several tens of molecules at their equilibrium geometries. For all of them, the densities were computed at the Hartree−Fock (HF) level using the reasonably good VB1 Slater basis set[58] (composition: [5,3,1] for B to Ne and [3,1] for H). The calculation of integrals and HF optimization were carried out with the SMILES program.[59−63] The atomic expansions were carried out with the DAM method using the DAM program.[29−31] For testing the effects of the basis set on the results, some calculations were repeated using the CVB2 Slater basis set[58] and Dunning's cc-pVTZ and cc-pVQZ Gaussian basis sets[64] with the Gaussian version of DAM (G-DAM[65]). The effect of correlation was also examined by performing configuration interaction with single and double excitation (CISD) calculations with MOLPRO.[66] No qualitative changes were observed in the test samples (see below), and quantitative changes were so small that HF calculations with the VB1 basis sets were considered sufficiently good for our purposes.

The density, the atomic and accumulated spherical terms, and its deformations were visualized by depicting their constant value surfaces with gOpenMol.[67] As an example, we consider the methane molecule, whose density varies from more than 100 au (e/b^3), in the nearness of the C nucleus, to 0 au, far away from the molecular region, while its deformation varies only from 0.08 to −0.03 au.

Because of the smallness of the density deformation, the contour surfaces of the whole density of this molecule and those of the accumulated spherical terms are very similar, as it is illustrated in Figure 1, so that the details of the difference are difficult to appreciate. On the contrary, the structure of these differences is evident in the contour surfaces of the density deformations depicted in Figure 2. This figure shows the density deformations obtained at Roothaan−Hartree−Fock (RHF) and CISD levels using Slater VB1[58] and Gaussian pVTZ[64] basis sets and illustrates the changes in the density deformations produced by changes in the computational method or the basis set.

Because the deformations bear no net charge (they integrate to zero), physical space is partitioned into regions where the electronic charge is accumulated (positive deformation) and into others in which it is depleted (negative). In every region, surfaces of a given absolute value are enclosed by surfaces of *lower* absolute value.

For methane, the contours corresponding to positive deformations (red surfaces in Figure 2) show four electron charge accumulations with nearly cylindrical symmetry between the carbon and hydrogens, just giving a clear image of the four σ bonds of this molecule. The pictures of negative deformations (blue surfaces in Figure 2) complete the description of the charge redistribution, showing the regions where the electronic charge is drawn from.

Looking at these type of pictures for several molecules, one finds that the concepts of empirical chemistry appear in a beautiful and neat, though somewhat unexpected, manner. Lone pairs and bonds are paralleled by charge accumulations placed where one would expect them to be; single, double, and triple bonds are clearly distinguished; concepts such as aromaticity and delocalization can be directly visualized. The different chemical classes of a given atom can be identified by glancing at their atomic density deformations. At the same time, the individuality of each atom can be recognized from the features of the dominant spherical term. Functional groups are identified as a common group of atoms of given classes that retain their structure and density in different molecules, and so on.

For obvious reasons, the literally hundreds of pictures that have been drawn cannot be included in this paper. We present a minimal selection of them in order to support and clarify our previous comments. Further pictures can be found at http://www.uam.es/departamentos/ciencias/qfa/DAM.

5.1. Lone Pairs. Because the notion of a localized electron pair is essential in Lewis'[54] and other[68,69] successful models for geometry and reactivity, its physical support has been extensively researched in the context of the pair density function[70] and associated[6] with the maxima of the density Laplacian, $\nabla_r^2\rho(\mathbf{r},\lambda)$.

Certainly, the interactions between electrons should be studied in terms of pair density, but as stressed in Section 2, its chemical relevance is determined by its effect on the density, which, according to Section 3, can be searched for in the density itself through the density deformations.

Lone pairs compose a nice example of the close relationship between chemical notions and density deformations. Figure 3 shows the atomic deformations of nitrogen in ammonia (left plate) and the full molecular deformations (right plate). In ammonia, besides the three charge accumulations along the N−H bonds, there is a large, and nearly spherical, charge accumulation in the vicinity of the N atom, just as it would be expected from the presence of a single lone pair there ($:NH_3$).

Figure 4 shows the atomic density deformations of oxygen (left plates) and the full molecular deformations (right plates)

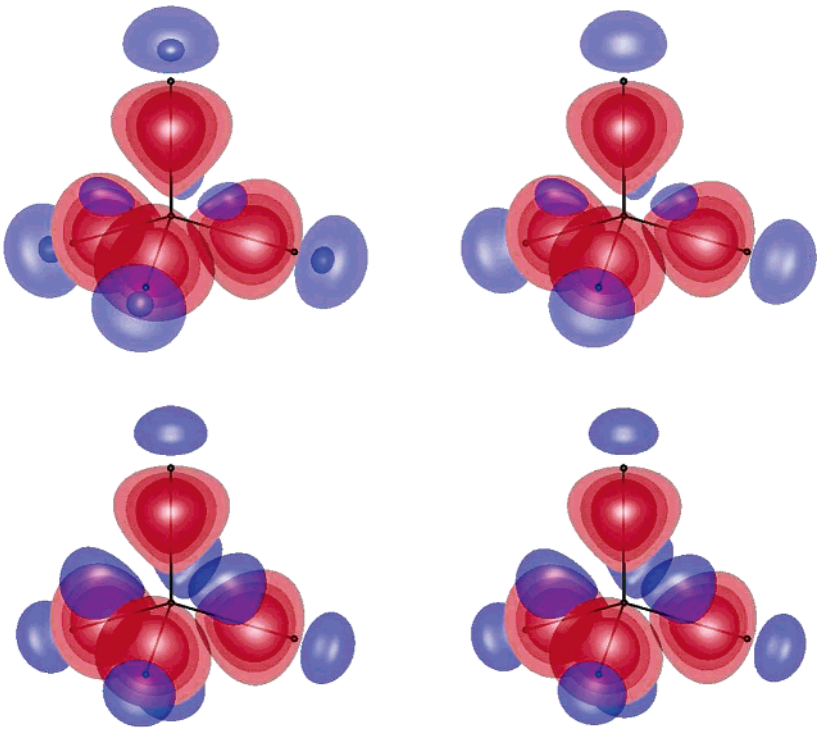


Figure 2. Molecular deformation of methane. Red: positive deformation. Blue: negative deformation. Contour values: 0.045 (innermost), 0.030, 0.015 (outermost), −0.030, and −0.015 au. Upper: Slater VB1 basis set. Lower: Gaussian pVTZ basis set. Left: RHF density. Right: CISD density.

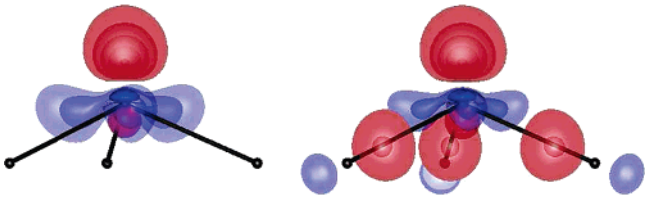


Figure 3. Atomic density deformations of nitrogen (left) and the molecular deformation (right) in ammonia. Contour values in au: 0.150, 0.125, 0.100, 0.075, 0.050 (red); −0.075, −0.050 (blue).

in water (:$\ddot{O}H_2$), formaldehyde (:$\ddot{O}$═CH_2), and carbon monoxide (:O≡C:). In water, the charge accumulations associated with the two lone pairs of oxygen extend above and below the molecular plane, whereas in formaldehyde, they are placed in the molecular plane. Finally, in carbon monoxide, the two charge accumulations placed outside the bond region show the existence of a lone pair on each atom. Moreover, the characteristic deformations of the lone pairs appear in the atomic density, supporting the idea that they are owned by atoms.

It is noteworthy that these structures appear almost unchanged everywhere the empirical chemistry evinces this class of atom. Thus, the distortion of nitrogen in ammonia is reproduced practically unchanged in amines; that of oxygen in water appears in alcohols, ethers, carboxylic acids, esters, and so forth. The distortion of oxygen in formaldehyde is typical of the keto oxygen, and it can be observed in aldehydes, ketones, carboxylic acids, and so on.

As an example, we present in Figure 5 the density deformations of the oxygen atom and the full molecular deformations in methanol, dimethyl ether, and phenol, where

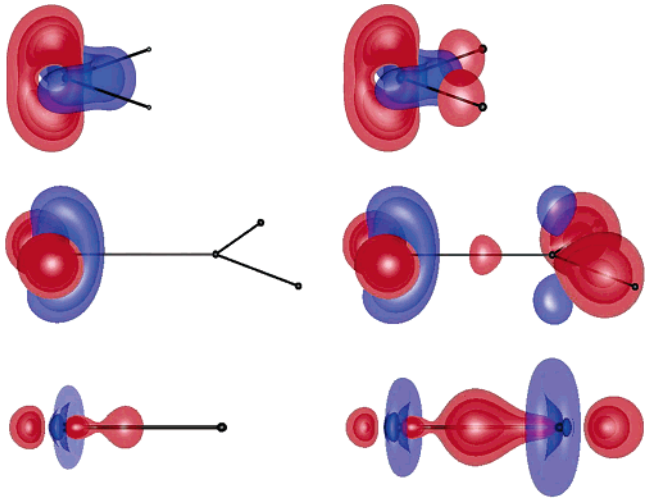


Figure 4. Oxygen deformation (left plates) and molecular deformation (right plates) in water (upper), formaldehyde (middle), and carbon monoxide (lower). Contour values in au: 0.150, 0.125, 0.100, 0.075, 0.050 (red); −0.075, −0.050 (blue).

the presence of two lone pairs such as those of water is clear, despite the strongly different molecular environments. Certainly, the presence of a charge accumulation, everywhere lone pairs are predicted to be, shows that this notion has an actual support in the density.

5.2. Single, Double, and Triple Bonds. Figure 6 illustrates how the density deformations reflect the chemical notions of single, double, and triple bonds. It shows several contour surfaces corresponding to ethane (first plate), ethylene (second plate), and acetylene (third plate). In ethane, the σ-bond skeleton is clearly reflected by charge accumulations

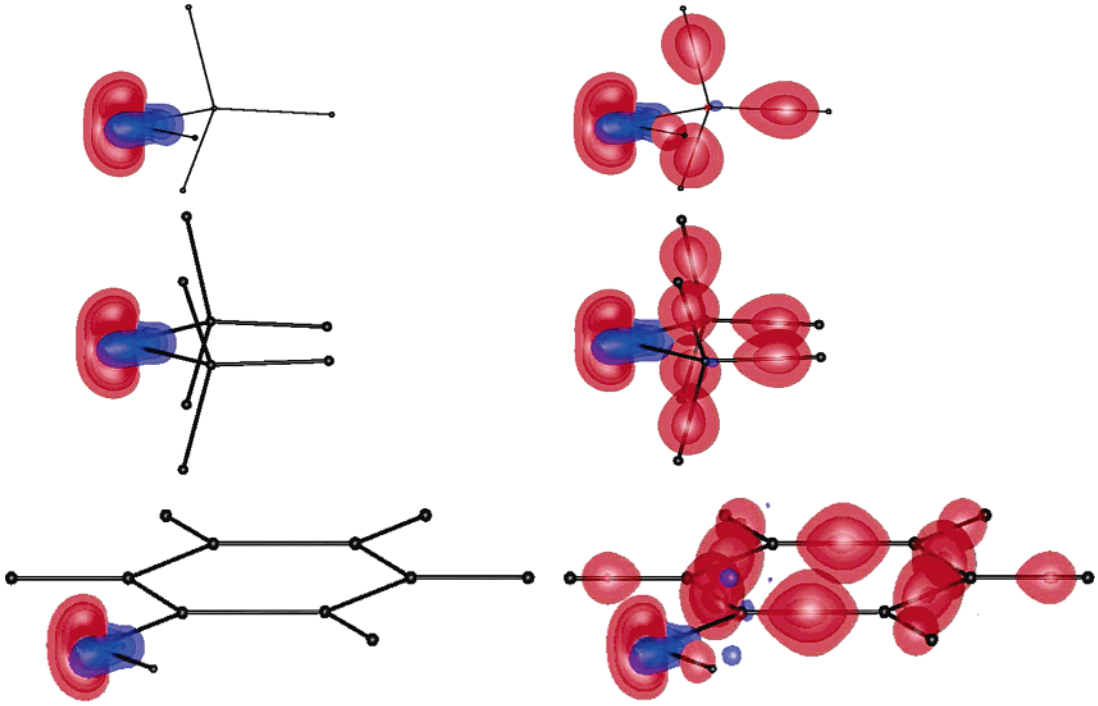


Figure 5. Atomic deformation of oxygen (left plates) and molecular deformation (right plates) in methanol (upper), dimethyl ether (middle), and phenol (lower). Contour values: ± 0.150, ± 0.125, ± 0.100, ± 0.075, ± 0.050 au (red: positive. blue: negative).

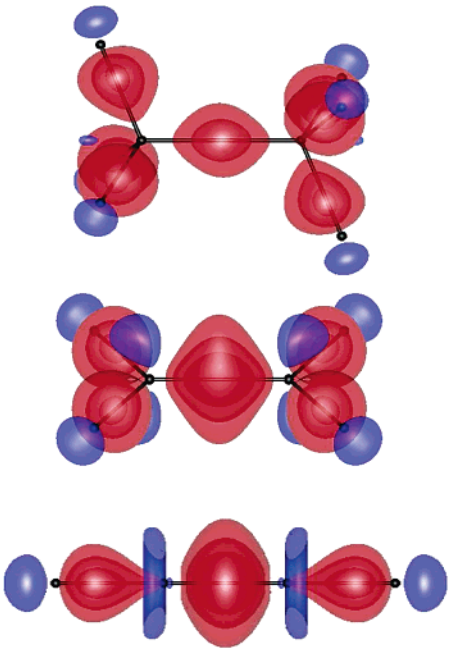


Figure 6. Density deformation of ethane (upper), ethylene (middle), and acetylene (lower). Contour values: 0.075, 0.050, ± 0.025 au (red: positive. blue: negative).

with nearly cylindrical symmetry placed where the C−H and C−C bonds are expected to be. Moreover, a comparison between Figures 2 and 6 renders evident the close similarity of the C−H σ-type charge deformation in methane and ethane.

In ethylene, besides the σ-type deformation characteristic of the C−H bonds, there is a strong charge accumulation, with upward and downward distortions, in the middle of the C−C line as well as charge depletions above and below the molecular plane near the carbons. Again, the density deformation parallels known properties. The π character is evident from the asymmetry (ellipticity[6]) of the charge distribution, and the large rotational barrier around the C−C bond can be attributed both to this distortion and to the charge depletions, which must generate forces that hinder the movements of the hydrogens out of the plane.

The density deformation of acetylene shows the two C−H σ bonds and a strong charge accumulation of cylindrical symmetry in the C−C line, accompanied by two perpendicular rings of charge depletions centered in the carbons. This structure is characteristic of the C≡C triple bond, but it also appears (with quantitative but not qualitative changes) in triple bonds involving other atoms, as it can be seen, for instance, in case of carbon monoxide, Figure 4.

5.3. Aromaticity. The chemical notions of a delocalized cloud and aromaticity are also reflected in the density deformations. For high deformation values, contour surfaces of aromatic molecules show only the charge concentrations associated with the skeleton of conjugated double bonds (intermediate between single and double), as well as those of their σ bonds. If that contour value is lowered in order to follow the corresponding evolution of the charge accumulation, one observes the growth of upward and downward protuberances in the middle of the double bond. Additional lowering shows that the protuberances tend to acquire a shape like two mushrooms placed at opposite sides. In aromatic compounds, the tops of adjacent mushrooms tend to join each other, forming a sandwich cloud that encloses the deformations of the internal skeleton. On the contrary, in nonaromatic compounds, this joint does not occur.

Figure 7 shows the contour surfaces of values ± 0.001 for the deformations of benzene, anthracene, and diphenylene

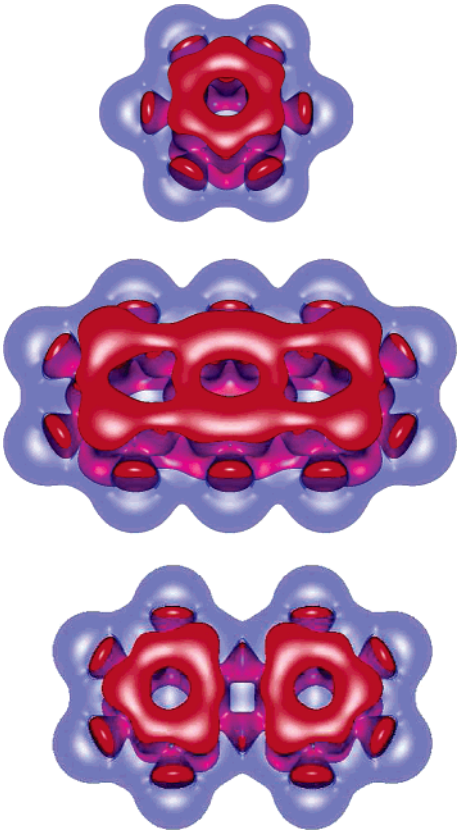


Figure 7. Molecular density deformations in benzene (upper), anthracene (middle), and diphenylene (lower). Contour values ± 0.001 au.

(note that this contour value is much lower than those previously chosen). It can be seen that the external cloud in benzene and anthracene is delocalized over all the molecule, while in diphenylene, it extends separately over each benzene ring. Clearly, the density deformations allow us to visualize this chemical notion in a neat manner.

5.4. Atoms in Molecules. The strong differences between the shapes of the density deformations associated with the different *classes* of a given atom may induce (see Figure 4, for instance) one to think that atoms, in the usual meaning of the term, cannot be distinguished in molecules, contradicting the intuitive notion of molecules as composed by slightly deformed atoms. The point to be stressed here is that the values of density deformations (typically about 10^{-2} au or less) are several orders of magnitude smaller than those of the spherical terms and that it is in these latter where the *identity* of the atoms resides. To illustrate this fact, we depict, in Figure 8a, the spherical radial factors, $f_{00}(r_O)$, of oxygen in a set of molecules covering the different classes of oxygen discussed above together with the radial factor of the isolated atom in its ground state. Note that the differences are so small in the scale of the figure that all the oxygen atoms seem to have the same spherical radial factor. To make perceptible these differences, we also present, in Figure 8b, the usual radial distributions, $4\pi r_O^2 f_{00}(r_O)$, that enhance the resolution in the valence zone.

Now, one can observe that all of the atomic radial distributions of oxygen are greater than that corresponding

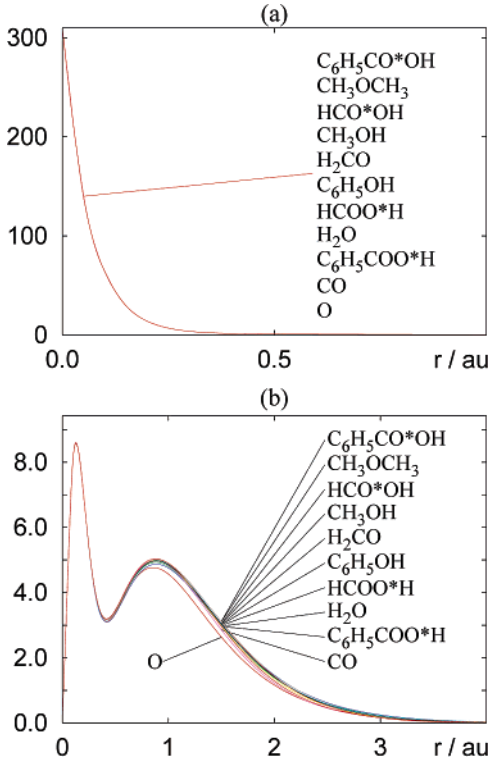


Figure 8. Comparison of the radial factors of spherical terms of oxygen in different molecules. (a) $f_{00}(r)$, (b) radial charge densities, $4\pi r^2 f_{00}(r)$.

to the isolated atom (evincing the charge transfer toward the O), but the differences among them are still very small.

5.5. Functional Groups. A notion related to the *class* and *identity* of an atom is that of the *functional group*. As remarked above, atoms of the same class can be identified from the resemblance of both their spherical averages and atomic deformations. In functional groups, this resemblance extends to the whole set of atoms forming the group, thus giving fragments with similar densities in different molecules. According to Sections 2 and 3, these fragments will contribute in almost the same way to the chemical behavior of the molecules. Figure 9 illustrates the density deformation of the carboxylic group of formic and benzoic acids. These pictures render evident the resemblance of the group deformations of −COOH in these significantly different environments, the very small differences between both pictures being illustrative of the smallness of the environmental effects.

Once again, another basic notion of empirical chemistry is supported by the density structure in a nice manner.

6 Final Comments

When the total one-electron density distribution of a molecule is examined, nothing about classic chemical notions is apparent.[6,7] There are no lone pairs, no distinct chemical bonds, no π clouds, and so forth. This occurs because the molecular density is so largely dominated by the spherical terms of its constituent atoms that the chemically relevant structure is completely masked. However, when the spherical terms are removed, the fine structure of the molecular density becomes evident and, from its analysis, there emerges a

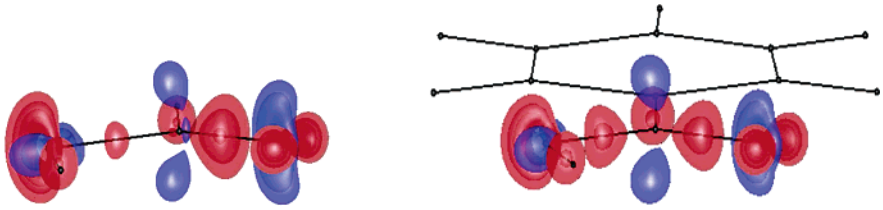


Figure 9. Accumulated density deformations of the −COOH group in formic acid (left) and benzoic acid (right). Contour values in au: 0.150, 0.125, 0.100, 0.075, 0.050 (red); −0.075, −0.050 (blue).

picture that reflects and supports all the intuitive notions of the empirical chemistry.

The reason for this is clear from the standpoint of the Hellmann−Feynman (electrostatic) theorem: chemical forces only depend on electron density, and the physical effects determine the chemical behavior through their influence on the density. Moreover, the spherical parts of the atomic density, along with the nuclei charges, cannot lead to stable systems, and consequently, these cannot play the main role in the explanation of the chemical behavior. This behavior is determined by the small nonspherical deformations of the density at the same time that it is described by the empirical notions of chemistry. Chemical notions and nonspherical density deformations must be related to each other, and they actually are.

In this work, a systematic study of the density deformations of several tens of molecules has been summarized. This study has been aimed at analyzing whether, and eventually how, these deformations reflect Lewis' structures and other basic concepts of empirical structural chemistry.

The theoretical support of this analysis lies on the extremum condition and the Hellmann−Feynman theorem. It has been remarked that the explicit relationship between density, force, and energy, previously reported in the framework of the DAM method, can be extended in a straightforward manner to almost any other method for the analysis of density. In particular, we note that, for every molecular partition of the electron density into atomic contributions, the spherical averages can be separated from the rest. In this way, atomic, group, and molecular deformations can be defined without invoking any external reference such as the density of isolated atoms. Moreover, it was also noted that this separation facilitates both the description of the energy changes in terms of forces and the forces in terms of density. Bearing all this in mind, we have carried out a detailed discussion on the interatomic forces, in the context of the electrostatic theorem, that evinces the basic role played by small density deformations in chemical behavior, and which supports the idea that these deformations are related to the basic notions of chemistry.

For the practical determination of these relationships, reasonably good densities were computed for a number of representative molecules, using Slater basis sets[58] with SMILES,[63] and these densities were analyzed with the DAM package.[31]

As illustrated in Section 5, molecules were classified according to their conventional structures (having lone pairs; single, double, or triple bonds; etc.), and their density deformations were depicted and compared. The density deformations thus obtained show clear and characteristic structures associated with every classical notion, the correspondence being so tight that it seems possible to accurately predict (with a suitable parametrization of the deformations) the density of molecules from their conventional chemical formulas.

The appearance of such a close relationship between chemical notions and density deformations gives us sensible grounds to regard the traditional chemical language as a symbolic representation of the density deformations responsible for chemical behavior. From this perspective, it is both gratifying and amazing to confirm that chemists, on the basis of macroscopic data alone and being guided just by their intuition, were able to create this symbolic language almost a century before these deformations were studied.

This work has been circumscribed to the study of density at the equilibrium conformation, but as remarked above, even more important is how to characterize the evolution of density along conformational and reactive changes. We are, at present, carrying out an analysis of the derivatives of density with respect to nuclear coordinates and exploring the fit of multidimensional energy surfaces and forces in terms of the density, with practical applications in molecular dynamics, spectroscopy, and molecular mechanics. The preliminary results are encouraging and will soon be reported.

A beta version of the DAM codes used in this work is available upon request at rafael.lopez@uam.es.

Appendix

In the linear combination of atomic orbitals context, one-electron density has the expression

$$\rho(\mathbf{r}) = \sum_A \sum_a \sum_{a'} \rho_{aa'}\, \chi_a(\mathbf{r}_A)\, \chi_{a'}(\mathbf{r}_A) + \sum_A \sum_{B\neq A} \sum_a \sum_b \rho_{ab}\, \chi_a(\mathbf{r}_A)\, \chi_b(\mathbf{r}_B) \quad (44)$$

where A, B, and so forth label the nuclei, centered respectively at $\mathbf{R}_A$, $\mathbf{R}_B$, and so forth; a, a', b, and b' label the subsets of the basis functions, $\chi_a(\mathbf{r}_A)$, $\chi_b(\mathbf{r}_B)$, and so forth; and $\mathbf{r}_A = \mathbf{r} - \mathbf{R}_A$, $\mathbf{r}_B = \mathbf{r} - \mathbf{R}_B$.

In the DAM method, the atomic fragments are defined by

$$\rho^A(\mathbf{r}_A) = \sum_a \sum_{a'} \rho_{aa'}\, \chi_a(\mathbf{r}_A)\, \chi_{a'}(\mathbf{r}_A) + 2 \sum_{B\neq A} \sum_a \sum_b \rho_{ab}\, d^A_{ab}(\mathbf{r}_A) \quad (45)$$

The first contribution contains the one-center distributions placed at $\mathbf{R}_A$. The second one is obtained by partitioning the

two-center distributions into minimally deformed components assigned, respectively, to *A* and *B*:

$$\chi_a(\mathbf{r}_A)\,\chi_b(\mathbf{r}_B) = d^A_{ab}(\mathbf{r}_A) + d^B_{ab}(\mathbf{r}_B) \tag{46}$$

The minimal deformation criterion used in this partition and its practical implementation have been detailed elsewhere.[29−31]

Finally, each atomic fragment of eq 45 is expanded in regular harmonics centered in its nucleus, so that the whole density becomes

$$\rho(\mathbf{r}) = \sum_A \sum_{l=0}^{\infty} \sum_{m=-l}^{l} z^m_l(\mathbf{r}_A)\, f^A_{lm}(r_A) \tag{47}$$

If one regards these fragments as (pseudo-) atoms, it is clear that the terms with $l = 0$ can be associated with the spherical parts of the atomic clouds and the remaining ones with the deformations caused by the environment. The spherical parts contain the whole electronic charge of the molecule and are largely dominant. The remaining terms determine the atomic dipoles ($l = 1$), quadrupoles ($l = 2$), octapoles ($l = 3$), and so on. As proved in our previous works,[29−31] these terms are small and quickly decreasing with *l*.

From these equations, it is evident that the DAM method is basis-set-dependent, and in fact, it will lead to meaningless results for highly unbalanced basis sets, such as strictly one-center ones. Nevertheless, when reasonably good basis sets are used, the results are qualitatively independent of the particular basis set chosen, the quantitative changes in density deformations being rather small, as it was illustrated in Figure 2.

Acknowledgment. The authors thankfully acknowledge Prof. E. Ludeña for his helpful criticisms and encouraging comments. They also acknowledge Prof. L. Laaksonen for the availability of gOpenMol. Financial support from the Dirección General de Investigación Científica y Técnica (CTQ2004-06615/BQU) is gratefully acknowledged.

References

(1) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136B*, 864.

(2) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(3) Kryachko, E. S.; Ludeña, E. V. *Energy Density Functional Theory of Many Eletron Systems*; Kluwer: Dordrecht, The Netherlands, 1990.

(4) Ayers, P. W. *Theor. Chem. Acc.* **2003**, *110*, 267−275.

(5) Deb, B. M. *The Force Concept in Chemistry*; Van Nostrand: New York, 1981.

(6) Bader, R. F. W. *Atoms in Molecules: Quantum Theory*; Clarendon: Oxford, U. K., 1990.

(7) Bachrach, S. M. *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, B. B., Eds.; VCH Publications: New York, 1994; Vol. 5, Chapter 3.

(8) Murray, J. S.; Sen, K. *Molecular electrostatic potentials: Concepts and Applications*; Elsevier: Amsterdam, 1996.

(9) Wiener, J. J. M.; Murray, J. S.; Grice, M. E.; Politzer, P. *Mol. Phys.* **1997**, *90*, 425−430.

(10) (a) Politzer, P.; Harris, R. R. *J. Am. Chem. Soc.* **1970**, *92*, 6451. (b) Politzer, P.; Stout, E. W. *Chem. Phys. Lett.* **1971**, *8*, 519−522. (c) Politzer, P. *Theor. Chim. Acta* **1971**, *23*, 203. (d) Politzer, P.; Elliot, S. D.; Meroney, B. F. *Chem. Phys. Lett.* **1973**, *23*, 331−334.

(11) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129.

(12) (a) Daudel, R. *Acad. Sci., C. R.* **1953**, *237*, 601. (b) Aslangul, C.; Constanciel, R.; Daudel, R.; Esnault, L.; Ludeña, E. V. *Int. J. Quantum Chem.* **1974**, *8*, 499−522. (c) Aslangul, C.; Constanciel, R.; Daudel, R.; Kottis, P. *Adv. Quantum Chem.* **1972**, *6*, 93. (d) Daudel, R. *Quantum Theory of the Chemical Bond*; Reidel: Dordrecht, The Netherlands, 1974.

(13) (a) Bader, R. F. W.; Nguyen-Dang, T. T.; Tal, Y. *J. Chem. Phys.* **1979**, *70*, 4316−4329. (b) Bader, R. F. W.; Nguyen-Dang, T. T. *Adv. Quantum Chem.* **1981**, *14*, 63. (c) Bader, R. F. W. *Atoms in Molecules*; Clarendon Press: Oxford, U. K., 1990.

(14) Mazziotti, A.; Parr, R. G.; Simons, G. *J. Chem. Phys.* **1973**, *59*, 939−942.

(15) Mujica, V.; Squitieri, E.; Nieto, P. *THEOCHEM* **2000**, *501*, 115−123.

(16) (a) Martín Pendás, A.; Costales, A.; Luaña, V. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *55*, 4275−4284. (b) Martín Pendás, A.; Blanco, M. A.; Costales, A.; Mori Sánchez, P. *Phys. Rev. Lett.* **1999**, *83*, 1930−1933. (c) Mori Sánchez, P.; Martín Pendás, A.; Luaña, V. *J. Am. Chem. Soc.* **2002**, *124*, 14721−14723. (d) Martín Pendás, A.; Luaña, V.; Pueyo, L.; Francisco, E. *J. Chem. Phys.* **2002**, *117*, 1017−1023.

(17) Kosov, D. S.; Popelier, P. L. A. *J. Chem. Phys.* **2000**, *113*, 3969−3974.

(18) Stewart, R. F. *Isr. J. Chem.* **1977**, *16*, 124−131.

(19) Hall, G. G.; Martin, D. *Isr. J. Chem.* **1980**, *19*, 255−259.

(20) (a) Hurley, A. C. *Proc. Phys. Soc. London, Ser. A* **1955**, *68*, 149−155. (b) Hurley, A. C. *Proc. Phys. Soc. London, Ser. A* **1956**, *69*, 49−56. (c) Hurley, A. C. *Proc. Phys. Soc. London, Ser. A* **1956**, *69*, 301−309.

(21) (a) Rychlewski, J.; Parr, R. G. *J. Chem. Phys.* **1986**, *84*, 1696−1703. (b) Li, L.; Parr, R. G. *J. Chem. Phys.* **1986**, *84*, 1704−1711. (c) Parr, R. G. *J. Chem. Phys.* **1986**, *85*, 3135.

(22) Nalewajski, R. F.; Parr, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8879−8882.

(23) Ayers, P. W. *J. Chem. Phys.* **2000**, *113*, 10886−10898.

(24) Bader, R. F. W.; Bayles, D. *J. Chem. Phys.* **2000**, *104*, 5579−5589.

(25) Walker, P. O.; Mezey, P. G. *J. Am. Chem. Soc.* **1993**, *115*, 12423−12430.

(26) Chodkiewicz, M. L.; Howard, S. T.; Wozniak, K. *Int. J. Quantum Chem.* **2004**, *98*, 1−10.

(27) (a) Feynman, R. P. *Phys. Rev.* **1939**, *56*, 340. (b) Hellmann, H. *Einführung in die Quantemchemie*; Deutiche: Leipzig, Germany, 1937.

(28) Fernández Rico, J.; López, R.; Ema, I.; Ramírez, G. *Int. J. Quantum Chem.* **2004**, *100*, 221−230.

(29) Fernández Rico, J.; López, R.; Ramírez, G. *J. Chem. Phys.* **1999**, *110*, 4213−4220.

(30) Fernández Rico, J.; López, R.; Ema, I.; Ramírez, G. *J. Chem. Phys.* **2002**, *117*, 533−540.

(31) Fernández Rico, J.; López, R.; Ema, I.; Ramírez, G.; Ludeña, E. V. *J. Comput. Chem.* **2004**, *25*, 1355−1363.

(32) Fernández Rico, J.; López, R.; Ema, I.; Ramírez, G. *J. Comput. Chem.* **2004**, *25*, 1347−1354.

(33) Fernández Rico, J.; López, R.; Ema, I.; Ramírez, G. *J. Chem. Phys.* **2002**, *116*, 1788−1799.

(34) Fernández Rico, J.; López, R.; Ema, I.; Ramírez, G. *J. Chem. Phys.* **2003**, *119*, 12251−12256.

(35) Slater, J. C. *J. Chem. Phys.* **1972**, *57*, 2389−2396.

(36) Coulson, J. A.; Bell, R. P. *Trans. Faraday Soc.* **1945**, *41*, 141−149.

(37) Berlin, T. *J. Chem. Phys.* **1951**, *19*, 208−213.

(38) Epstein, S.; Hurley, A. C.; Wyatt, R. E.; Parr, R. G. *J. Chem. Phys.* **1967**, *47*, 1275−1286.

(39) Koga, T. *Theor. Chim. Acta* **1981**, *58*, 173.

(40) Schmider, H. L.; Becke, A. D. *THEOCHEM* **2000**, *527*, 51−61.

(41) Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **2002**, *116*, 3184−3193.

(42) Tachibana, A. *J. Chem. Phys.* **2001**, *115*, 3497−3517.

(43) Tachibana, A. *Int. J. Quantum Chem.* **2004**, *100*, 981−993.

(44) Ayers, P. W.; Parr, R. G.; Nagy, A. *Int. J. Quantum Chem.* **2002**, *90*, 309−326.

(45) Thakkar, A. J.; Sharma, B. S.; Koga, T. *J. Chem. Phys.* **1986**, *85*, 2845−2849.

(46) Sharma, B. S.; Thakkar, A. J. *Int. J. Quantum Chem.* **1986**, *29*, 323.

(47) Nakatsuji, H.; Parr, R. G. *J. Chem. Phys.* **1975**, *63*, 1112−1117.

(48) Wilson, E. B., Jr. *J. Chem. Phys.* **1962**, *36*, 2232−2233.

(49) Politzer, P.; Parr, R. G. *J. Chem. Phys.* **1974**, *61*, 4258−4262.

(50) Delgado-Barrio, G.; Prat, R. F. *Phys. Rev. A: At., Mol., Opt. Phys.* **1975**, *12*, 2288−2297.

(51) Politzer, P. In *Fundamental World of Quantum Chemistry: A Tribute to the Memory of Per-Olov Lowdin*; Brandas, E. J., Kryachko, E. S., Eds.; Kluwer: Dordretch, The Netherlands, 2003; p 631.

(52) Politzer, P. *Theor. Chem. Acc.* **2004**, *111*, 395−399.

(53) Gradshteyn, I. S.; Ryzhik, I. M. *Table of Integrals, Series, and Products*, 5th ed.; Academic Press: New York, 1994; Equations 8.7 and 8.8.

(54) Lewis, G. N. *J. Am. Chem. Soc.* **1916**, *38*, 762.

(55) (a) Roux, M.; Besnainou, S.; Daudel, R. *J. Chim. Phys.* **1958**, *54*, 218. (b) Roux, M.; Besnainou, S.; Daudel, R. *J. Chim. Phys.* **1958**, *55*, 754. (b) Roux, M.; Cornille, M.; Burnelle, L. *J. Chem. Phys.* **1962**, *37*, 933−936.

(56) Hirshfeld, F. L.; Rzotkiewicz, S. *Mol. Phys.* **1974**, *27*, 1319−1343.

(57) Gilbert, A. T. B.; Gill, P. M. W. *J. Chem. Phys.* **2004**, *120*, 7887−7893.

(58) Ema, I.; García de la Vega, J. M.; Ramírez, G.; López, R.; Fernández Rico, J.; Meissner, H.; Paldus, J. *J. Comput. Chem.* **2003**, *24*, 859−868.

(59) Fernández Rico, J.; García de la Vega, J. M.; Fernández Alonso, J. I.; Fantucci, P. *J. Comput. Chem.* **1983**, *4*, 33−40.

(60) Fernández Rico, J.; Paniagua, M.; Fernández Alonso, J. I.; Fantucci, P. *J. Comput. Chem.* **1983**, *4*, 41−47.

(61) Fernández Rico, J.; López, R.; Aguado, A.; Ema, I.; Ramírez, G. *J. Comput. Chem.* **1998**, *19*, 1284−1293.

(62) Fernández Rico, J.; López, R.; Aguado, A.; Ema, I.; Ramírez, G. *Int. J. Quantum Chem.* **2001**, *81*, 148−153.

(63) Fernández Rico, J.; López, R.; Ema, I.; Ramírez, G. *J. Comput. Chem.* **2004**, *25*, 1987−1994.

(64) (a) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358−1371. (b) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806. (c) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(65) Fernández Rico, J.; López, R.; Ema, I.; Ramírez, G. *THEOCHEM* **2005**, *727*, 115−121.

(66) (a) Werner, H.-J.; Knowles, P. J.;Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO*, version 2002.5; University College Cardiff Consultants Limited: Wales, U. K. (b) Werner, H.-J.; Knowles, P. J. *J. Chem. Phys.* **1985**, *82*, 5053−5063. (c) Knowles, P. J.; Werner, H.-J. *Chem. Phys. Lett.* **1985**, *115*, 259−267. (d) Werner, H.-J.; Knowles, P. J. *J. Chem. Phys.* **1988**, *89*, 5803−5814. (e) Knowles, P. J.; Werner, H.-J. *Chem. Phys. Lett.* **1988**, *145*, 514−522. Parts b and c relate to the MCSCF/CASSCF codes, and d and e relate to the internally contracted MRCI codes.

(67) (a) Laaksonen, L. *J. Mol. Graphics* **1992**, *10*, 33−34. (b) Bergman, D. L.; Laaksonen, L.; Laaksonen, A. *J. Mol. Graphics* **1997**, *15*, 301−306. (c) www.csc.fi/gopenmol/.

(68) Bader, R. F. W.; Gillespie, R. J.; Macdougall, P. J. *J. Am. Chem. Soc.* **1988**, *110*, 7329−7336.

(69) Gillespie, R. J. *Molecular Geometry*; Van Nostrand: London, 1972.

(70) Fradera, X.; Austen, M. A.; Bader, R. F. W. *J. Phys. Chem. A* **1999**, *103*, 304−314.

CT0500951

JCTC Journal of Chemical Theory and Computation

Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules

M. A. Blanco,* A. Martín Pendás, and E. Francisco

Departamento de Química Física y Analítica, Facultad de Química, Universidad de Oviedo, 33006-Oviedo, Spain

Received April 22, 2005

Abstract: We make use of the Quantum Theory of Atoms in Molecules (QTAM) to partition the total energy of a many-electron system into intra- and interatomic terms, by explicitly computing both the one- and two-electron contributions. While the general scheme is formally equivalent to that by Bader et al., we focus on the separation and computation of the atomic self-energies and all the interaction terms. The partition is ultimately performed within the density matrices, in analogy with McWeeny's Theory of Electronic Separability, and then carried onto the energy. It is intimately linked with the atomistic picture of the chemical bond, not only allowing the separation of different two-body contributions (point-charge-like, multipolar, total Coulomb, exchange, correlation, ...) to the interaction between a pair of atoms but also including an effective many-body contribution to the binding (self-energy, formally one-body) due to the deformation of the atoms within the many-electron system as compared to the free atoms. Many qualitative ideas about the chemical bond can be quantified using this scheme.

I. Introduction

Quantum chemistry is a successful theory, able to predict global properties of molecular species, such as binding energies, molecular geometries, spectra, etc., very accurately. However, its main results, the wave function of the molecule and its associated energy, are difficult to correlate to such cornerstone chemical concepts as individual atoms, functional groups, and the bonds and interactions among them. To build a connection between quantum results and chemical concepts, there have been proposed many different a posteriori analyses of these results that extract chemical information with different degrees of success: concepts such as atomic charges, bond orders, bond energies, ... have already reached textbook status.[1] Particularly, the partition of the total molecular energy into chemically meaningful components has deserved a lot of attention over the years[2−8] and is still today a rather active research topic.[9−15]

Among the many energy partitioning schemes already proposed in the literature, some rely on orbital descriptions and the characteristics of the LCAO method to define the fragments, so that the components of the energy can be either directly extracted or else computed from the intermediate properties that have to be evaluated in this type of calculation. However, this makes them dependent on the particular model elements used on the calculations (basis sets, for example). Some partitions define unphysical intermediate states to be used as a reference to compare with the global calculation, so that the binding process can be conceptually divided into different steps.[3,4,6,9,10] There is an ongoing controversy about this point, whether to focus on the final, bonded state or on the evolution from an initial, unbonded state, to understand the chemical bond. They can be considered to provide complementary views of the same concept, although it is difficult to compare them and understand the origin of the different terms on physical grounds.

These analyses suggest that the energy partition of a molecular system should fulfill some key conditions to be of general interest. We propose the following.

A good energy partition should be a useful tool for chemists and must be theoretically sound. It has to provide

* Corresponding author phone: +34 985105017; e-mail: miguel@carbono.quimica.uniovi.es.

10.1021/ct0501093 CCC: $30.25
Published on Web 08/25/2005

a means to identify the atoms and functional groups within the molecule, which have to be transferable. This requirement will make it consistent with the fact that atoms and functional groups retain, albeit partially, their identity in the molecule. This fact also points to the molecular binding energy as a better chemical indicator than the total energy, which contains large atomic contributions. To describe the chemical bond, the partition should also give detailed definitions of the interactions among atoms, functional groups, and molecules. Finally, the partition should be derived from the molecular wave function without resort to the approximations involved in its calculation. It must be unique and exhaustive and should recover the exact quantum chemical energy of the system.

In this article, we propose an energy partition scheme according to the requirements posed above. Thus, we partition the energy of the molecule with atoms as chemically meaningful fragments. Each atom consists of its nucleus and its 3D atomic basin as defined in the Quantum Theory of Atoms in Molecules (QTAM) of Bader and co-workers.[8] This induces a partition of the one- and two-electron density matrices, which is carried onto an exhaustive partition of the total energy. The elements of this partition were already proposed by Bader[8] and have been used in different ways by others.[16,14] However, the recent development of a fast algorithm to compute the two-electron integrals ($\mathcal{O}(N^4)$ instead of $\mathcal{O}(N^6)$ for the numerical quadrature of the two-center case),[17] together with a means to simplify them for correlated calculations ($\mathcal{O}(N^2)$ instead of $\mathcal{O}(N^4)$ in the two-electron density matrix expansion),[18] have made it now less computationally intensive to perform this analysis in practice. In addition, the arrangement of the terms in our partition differs from the previous ones: in order to avoid having large, mutually canceling energy terms, we have rearranged them following the spirit of McWeeny's Theory of Electronic Separability (TES),[19] a Hilbert-space (as opposed to real-space) approximate partitioning of the energy into strongly orthogonal interacting electron groups. We note here that we need not restrict ourselves to QTAM's partition, not even to space partitions, and we can talk about general 1- and 2-matrix partitions, as pointed by the work of Ruedenberg.[2] Such a generalization and the comparison of different partitioning schemes will be the subject of a forthcoming article.

The scheme of this paper is as follows. First, we introduce the energy partition in section II. Then, we present a thorough energy analysis of the hydrogen molecule, to introduce the concepts within a particular example, followed by a comparison of the different contributions in several molecules representative of different traditional bonding types. Finally, in section IV we give our conclusions and a plan for future work.

II. Energy Partition: Theory

In this section we present the theoretical aspects of the QTAM energy partition that we propose. First, we partition the density matrices and show how the different contributions can be evaluated. Then, we show how the rearrangement of these contributions into one-atom and two-atom terms gives an appealing chemical image with strong physical grounds. Later, we examine the one-atom terms and their relation to binding, and we will finish this section with a further decomposition of the two-atom terms through the second-order density matrix.

A. Density Matrices and Energy Partition. Let us assume that we know the many-electron wave function Ψ_e with the desired accuracy, and from it we construct the first- (nondiagonal) and second-order (diagonal) density matrices

$$\rho_1(1;1') = N_e \int \Psi_e(1, \cdots, N_e)\Psi^*_e(1', 2, \cdots, N_e)\, d\mathbf{x}_2 \cdots d\mathbf{x}_{N_e} \quad (1)$$

$$\rho_2(1, 2) = N_e(N_e - 1) \int \Psi_e(1, \cdots, N_e)\Psi^*_e(1, \cdots, N_e)\, d\mathbf{x}_3 \cdots d\mathbf{x}_{N_e} \quad (2)$$

where the $\mathbf{x}$ vectors include spatial and spin coordinates, 1, 2, ... represent $\mathbf{x}_1$, $\mathbf{x}_2$, ...within density matrices and Ψ_e, and N_e is the number of electrons. These density matrices suffice to compute the expectation value of the electronic wave function within the Coulomb Hamiltonian scheme (atomic units)

$$E_e = h + V_{ee} = \int_\infty \hat{h}\rho_1(1;1')\, d\mathbf{x}_1 + \frac{1}{2}\int_\infty\int_\infty \rho_2(1, 2) r_{12}^{-1}\, d\mathbf{x}_1\, d\mathbf{x}_2 \quad (3)$$

Here is understood that the prime superscript is removed after the operators act on functions but before performing the integrations. $\hat{h}$ is the usual monoelectronic operator containing kinetic energy ($\hat{T}$) and nuclear attraction ($\hat{U}$) operators, while r_{12}^{-1} is the interelectronic repulsion, thus defining the global properties h and V_{ee}. To obtain the total Born−Oppenheimer energy of the molecule, we have to add the internuclear repulsion, $E = E_e + V_{nn}$, which can be easily computed as

$$V_{nn} = \frac{1}{2}\sum_{A\neq B} V_{nn}^{AB} = \frac{1}{2}\sum_A \sum_{B\neq A} \frac{Z^A Z^B}{r_{AB}} \quad (4)$$

where Z^X is the nuclear charge of atom X and r_{AB} is the distance between atoms A and B.

Since V_{nn} has a quite clear two-body partition, the partitioning of E requires that of E_e, which we will carry on through a partition of the density matrices. To this end, we introduce QTAM's partition of real space:[8] atomic basins Ω are defined as the 3D attraction basins of the gradient field of the electron density, $\nabla\rho(\mathbf{r})$, which are bounded by a zero local flux surface of this field ($\nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = 0$ for $\mathbf{r} \in S(\Omega)$, where $\mathbf{n}(\mathbf{r})$ is a vector normal to the surface $S(\Omega)$). These basins usually contain one and only one nucleus and are easier to handle by defining a Heaviside-like basin step function

$$\Theta_\Omega(1) = \begin{cases} 1 & \text{if } r_1 \in \Omega \\ 0 & \text{elsewhere} \end{cases} \quad (5)$$

Notice that, since QTAM's partition is exhaustive, $\sum_A \Theta_A = 1$. Using this identity, we partition the first-order density

matrix into atomic contributions

$$\rho_1(1;1') = \sum_{\mathrm{A}} \rho_1^{\mathrm{A}}(1;1') = \sum_{\mathrm{A}} \rho_1(1;1')\Theta_{\mathrm{A}}(1') \tag{6}$$

This induces a partition in all monoelectronic properties, since

$$O = \int_\infty \hat{O}\rho_1(1;1')\,\mathrm{d}\mathbf{x}_1 = \sum_{\mathrm{A}} \int_\infty \hat{O}\rho_1{}^{\mathrm{A}}(1;1')\,\mathrm{d}\mathbf{x}_1 = \sum_{\mathrm{A}} \int_\infty \Theta_{\mathrm{A}}(1')\hat{O}\rho_1(1;1')\,\mathrm{d}\mathbf{x}_1 = \sum_{\mathrm{A}} \int_{\Omega_{\mathrm{A}}} \hat{O}\rho_1(1;1')\,\mathrm{d}\mathbf{x}_1 = \sum_A O^{\mathrm{A}} \tag{7}$$

In particular, since $\hat{h} = \hat{T} + \hat{U}$

$$h = \sum_{\mathrm{A}} h^{\mathrm{A}} = T + U = \sum_{\mathrm{A}} T^{\mathrm{A}} + \sum_{\mathrm{A}} U^{\mathrm{A}} \tag{8}$$

The atomic kinetic energy, T^{A}, is at the origin of QTAM's partition.[8] The zero-flux definition of the atomic basins ensures that both kinetic energy operators ($-\nabla^2/2$ and $\nabla\nabla'/2$) have the same expectation value (by making the momentum operator hermitian within the bounded region),[20] giving a kinetic energy, T^{A}, which is transferable[21] and that fulfills an atomic virial theorem.[22] Bader[8] used this theorem to partition the total energy in atomic (one-body) contributions, $E = -T = -\sum_{\mathrm{A}} T^{\mathrm{A}} = \sum_{\mathrm{A}} E(A)$ in his notation. The use of this one-body partition, while successful, is at the core of many criticisms of the theory. Particularly, this is only valid at equilibrium (where $2T + V = 0$), unnecessarily restricting the scope of the partition. It also hides the interatomic interactions within a single additive contribution, interpreting atomic energies as kinetic energies only. Thus, we will not use this scheme in what follows, allowing for nonequilibrium geometries and interaction contributions.

Regarding the electron−nucleus attraction, its operator is defined as

$$\hat{U} = \sum_{\mathrm{B}} \hat{V}_{\mathrm{en}}^{\mathrm{B}} = -\sum_{B} Z^{\mathrm{B}}/r_{1\mathrm{B}} \tag{9}$$

and thus

$$U = \sum_{\mathrm{A}} U^{\mathrm{A}} = \sum_{\mathrm{A}} \sum_{\mathrm{B}} V_{\mathrm{en}}^{\mathrm{AB}} \tag{10}$$

where

$$V_{\mathrm{en}}^{\mathrm{AB}} = \int_{\Omega_{\mathrm{A}}} \hat{V}_{\mathrm{en}}^{\mathrm{B}} \rho_1(1;1')\,\mathrm{d}\mathbf{x}_1 = -\int_{\Omega_{\mathrm{A}}} \frac{\rho(\mathbf{r}_1) Z^{\mathrm{B}}}{r_{1\mathrm{B}}}\,\mathrm{d}\mathbf{r}_1 \tag{11}$$

is the electrostatic interaction energy between electrons in basin A and the nucleus of atom B. When A = B, this is an intra-atomic (one-center/body) nuclear attraction, whereas when A ≠ B this is an interatomic (two-center/body) nuclear attraction, monoelectronic in both cases. It is to be noticed that, since it involves different types of particles, it does not include the 1/2 prefactor of eq 4 and also that $V_{\mathrm{en}}^{\mathrm{AB}} \neq V_{\mathrm{en}}^{\mathrm{BA}}$ in general, since both Z and $\rho(\mathbf{r}_1)$ within each basin can be different for the two atoms. Note that the order of sub- and superscripts does matter, i.e., $V_{\mathrm{en}}^{\mathrm{AB}} \equiv V_{\mathrm{ne}}^{\mathrm{BA}} \neq V_{\mathrm{en}}^{\mathrm{BA}} \equiv V_{\mathrm{ne}}^{\mathrm{AB}}$.

Unfortunately, a monoelectronic $\rho_1(1;1')$ partition does not provide a partition for the bielectronic $\rho_2(1, 2)$. However, Li and Parr[23] showed that if $\rho = \rho_{\mathrm{A}} + \rho_{\mathrm{B}} = \Theta_{\mathrm{A}}\rho + \Theta_{\mathrm{B}}\rho$ (with arbitrary Θ_{X} weighting factors in their general case), then a partition for the second-order density matrix that preserves the physical meaning of all energy contributions in the global and both in the intra- and interatomic cases is $\rho_2^{\mathrm{AB}}(1, 2) = \rho_2(1, 2)\Theta_{\mathrm{A}}(1)\Theta_{\mathrm{B}}(2)$ or, symmetrizing, $\rho_2^{\mathrm{AB}}(1, 2) = \rho_2(1, 2)\cdot(1/2)[\Theta_{\mathrm{A}}(1)\Theta_{\mathrm{B}}(2) + \Theta_{\mathrm{B}}(1)\Theta_{\mathrm{A}}(2)]$. Both expressions give the same symmetric energy results, since the $r_{12}{}^{-1}$ operator is hermitian, but only the latter fulfills $\rho_2^{\mathrm{AB}}(1, 2) = \rho_2^{\mathrm{BA}}(1, 2)$. Let us notice that within McWeeny's scheme,[19] $\rho_2^{\mathrm{AA}}(1, 2)$ is allowed to include as much correlation as needed, but $\rho_2^{\mathrm{AB}}(1, 2)$ is restricted to Coulomb and exchange terms only, making it approximately valid only in the case of weakly correlated electron groups. In this regard, our partition makes no approximation or assumption beyond those needed to obtain the wave function, and thus $\rho_2^{\mathrm{AB}}(1, 2)$ can be fully correlated.

Using this partition, $\rho_2 = \sum_{\mathrm{A,B}} \rho_2^{\mathrm{AB}}$, we can partition the two-electron energy as

$$V_{\mathrm{ee}} = \frac{1}{2}\int_\infty \int_\infty \sum_{\mathrm{A,B}} \rho_2^{\mathrm{AB}}(1, 2) r_{12}{}^{-1}\,\mathrm{d}\mathbf{x}_1\,\mathrm{d}\mathbf{x}_2 = \sum_{\mathrm{A}} V_{\mathrm{ee}}^{\mathrm{AA}} + \frac{1}{2}\sum_{A}\sum_{\mathrm{B}\neq\mathrm{A}} V_{\mathrm{ee}}^{\mathrm{AB}} \tag{12}$$

Here we have defined, on one hand

$$V_{\mathrm{ee}}^{\mathrm{AA}} = \frac{1}{2}\int_{\Omega_{\mathrm{A}}}\int_{\Omega_{\mathrm{A}}} \rho_2(1, 2) r_{12}{}^{-1}\,\mathrm{d}\mathbf{x}_1\,\mathrm{d}\mathbf{x}_2 \tag{13}$$

the total intrabasin interelectronic repulsion, coincident with Bader's $V_{\mathrm{ee}}(\Omega_{\mathrm{A}},\Omega_{\mathrm{A}})$, but on the other hand

$$V_{\mathrm{ee}}^{\mathrm{AB}} = \int_{\Omega_{\mathrm{A}}}\int_{\Omega_{\mathrm{B}}} \rho_2(1, 2) r_{12}{}^{-1}\,\mathrm{d}\mathbf{x}_1\,\mathrm{d}\mathbf{x}_2 \tag{14}$$

the total interbasin interelectronic repulsion interaction, which is twice the $V_{\mathrm{ee}}(\Omega_{\mathrm{A}},\Omega_{\mathrm{B}})$ defined by Bader (see ref 8). In this way, intrabasin components represent total contributions, but the interbasin components represent interaction contributions, which have to be halved afterward to avoid double-counting in the global properties. This completes the partition of the total energy using QTAM's zero-flux atomic basins, which we have recast into a density matrix partitioning scheme so that it can be easily compared with other atomic partitions.[19,23] The next subsection is devoted to rearrange all of these energy terms into chemically meaningful contributions.

B. QTAM Energy Partition. We have several contributions to the total energy: T^{A}, $V_{\mathrm{nn}}^{\mathrm{AB}}$, $V_{\mathrm{en}}^{\mathrm{AB}}$, and $V_{\mathrm{ee}}^{\mathrm{AB}}$ (A = B or A ≠ B). All of them have in fact been previously described in the literature, although only the first three have been widely used due to the extremely high computational resources needed to evaluate the two-electron integrals. The situation has changed with the recent development of a fast algorithm to evaluate these integrals,[17] and thus all of the contributions can now be used in the partition of the total energy. However, they convey a much finer partition than needed and present an important drawback: these properties

have very large values when compared with the binding energy of the molecule, which can be several orders of magnitude smaller. This comes from an almost exact cancellation of the different contributions and has a clear physical meaning, as we shall shortly see.

Instead of using these terms as such, we will reorganize them according to McWeeny's ideas. Thus, we first gather all of the monatomic (effective one-body) contributions in what we call the atomic net energy

$$E_{\mathrm{net}}^{\mathrm{A}} = T^{\mathrm{A}} + V_{\mathrm{en}}^{\mathrm{AA}} + V_{\mathrm{ee}}^{\mathrm{AA}} \tag{15}$$

containing all of the contributions from the particles ascribed to atom A. Then, we define an interatomic interaction energy by collecting all of the two-atom (two-body) contributions

$$E_{\mathrm{int}}^{\mathrm{AB}} = V_{\mathrm{nn}}^{\mathrm{AB}} + V_{\mathrm{en}}^{\mathrm{AB}} + V_{\mathrm{ne}}^{\mathrm{AB}} + V_{\mathrm{ee}}^{\mathrm{AB}} \quad (\mathrm{A} \neq \mathrm{B}) \tag{16}$$

This includes all interaction potential energies of particles ascribed to atom A with particles ascribed to atom B and is symmetric with respect to the A $\leftrightarrow$ B interchange (notice that $V_{\mathrm{ne}}^{\mathrm{AB}} = V_{\mathrm{en}}^{\mathrm{BA}}$).

Using these net and interaction energies, the total energy of the molecule can be written as

$$E = \sum_{\mathrm{A}} E_{\mathrm{net}}^{\mathrm{A}} + \frac{1}{2}\sum_{\mathrm{A}}\sum_{\mathrm{B}\neq\mathrm{A}} E_{\mathrm{int}}^{\mathrm{AB}} \tag{17}$$

This is the main equation in our partition scheme, which states that the total energy of a molecule can be exactly partitioned in net energy contributions, one-body self-energy terms which carry all of the intra-atomic contributions (kinetic, nuclear attraction, and two-electron repulsion within the atom), and interaction energy contributions, two-body pairwise additive terms including all interparticle potentials (nucleus−nucleus, nucleus−electron, electron−nucleus, and electron−electron). Although they are not going to be used in this study, it is also interesting to define, following McWeeny, the effective energy of a given atom

$$E_{\mathrm{eff}}^{\mathrm{A}} = E_{\mathrm{net}}^{\mathrm{A}} + \sum_{\mathrm{B}\neq\mathrm{A}} E_{\mathrm{int}}^{\mathrm{AB}} \tag{18}$$

which contains all the A-dependent terms in eq 17, and also what we call the additive energy of an atom

$$E_{\mathrm{add}}^{\mathrm{A}} = E_{\mathrm{net}}^{\mathrm{A}} + \frac{1}{2}\sum_{\mathrm{B}\neq\mathrm{A}} E_{\mathrm{int}}^{\mathrm{AB}} \tag{19}$$

whose value is in the recovery of the total energy by a simple sum, $E = \sum_{\mathrm{A}} E_{\mathrm{add}}^{\mathrm{A}}$.

Equation 17 has many interesting properties. First, it is equally valid for molecules and extended systems, where the energy per unit formula will be just computed by summing up the additive energies of the atoms in this unit formula; thus, the only infinite (but convergent) summation involved is the one over interactions with all B $\neq$ A atoms in eq 19. Moreover, eq 17 is the same one successfully used in semiempirical atomistic simulations (solid and liquid state, but also gas phase) over the years,[24−26] in which the focus is in the pairwise terms, assuming that the self-energies of the atoms do not change much within the particular simulation and can thus be ignored as a constant. In fact, this partition includes also the many-body effects through the electron density matrices, derived from the many-body wave function, so that the net and interaction terms should be more properly called effective one- and two-body. It is also behind the size extensiveness of the total energy, which is at the heart of order-*N* scaling algorithms: the net energies are local and simply additive, while the interaction energies are usually of short-range, mainly with the nearest neighbors, and thus half of their sum (their contribution to the additive energies) for a given atom also represents a local term to the energy of an extensive system. This equation is also equivalent to eq 2 in ref 14, although these authors used a Hartree−Fock scheme.

Compared with other partition schemes, eq 17 has also the same form as those derived from McWeeny's energy of interacting electron groups[19] when they are variationally optimized,[27] and it can also be shown to be equivalent to that by Li and Parr,[23] as we shall see below. Let us also recall that, unlike Bader's original scheme based in kinetic energies only, our scheme is not linked to the equilibrium nuclear geometry and can be used to analyze any point in the potential energy surface, thus being very valuable in understanding kinetics and reaction barriers, for example.

Given the simple form of eq 17, we can also group together several atoms to form functional groups. Let us consider the system formed by several groups of atoms, $\mathscr{G}$, $\mathscr{H}$, ... We can define the net energy of a group by adding the net energies of its constituents and their intragroup interactions (counted once each)

$$E_{\mathrm{net}}^{\mathscr{G}} = \sum_{A\in\mathscr{G}} E_{\mathrm{net}}^{\mathrm{A}} + \frac{1}{2}\sum_{\mathrm{A}\in\mathscr{G}}\sum_{\substack{\mathrm{B}\neq\mathrm{A}\\ \mathrm{B}\in\mathscr{G}}} E_{\mathrm{int}}^{\mathrm{AB}} \tag{20}$$

We can also define the interaction between two groups by adding all the interactions of atoms in one group with atoms in the other

$$E_{\mathrm{int}}^{\mathscr{G}\mathscr{H}} = \sum_{\mathrm{A}\in\mathscr{G}}\sum_{\mathrm{B}\in\mathscr{H}} E_{\mathrm{int}}^{\mathrm{AB}} \tag{21}$$

In this way, the total energy can be recovered from these net and interaction group energies as

$$E = \sum_{\mathscr{G}} E_{\mathrm{net}}^{\mathscr{G}} + \frac{1}{2}\sum_{\mathscr{G}}\sum_{\mathscr{H}\neq\mathscr{G}} E_{\mathrm{int}}^{\mathscr{G}\mathscr{H}} \tag{22}$$

an expression completely equivalent to that for atomic contributions, eq 17. Thus, inasmuch as a functional group is unchanged from system to system, its net energy will remain constant, and it will be its interaction energy with other groups the one governing the combined system behavior. In covalent, organic molecules, the intergroup interaction will be typically dominated by a single bond between the group and the backbone of the molecule, and whenever this bond is more or less equivalent to those of other possible substituents, we have the classical additive character of the energy with respect to the interchange of functional groups.[28] Another important possibility is for the

groups to be entire molecules, in which case the net energies will be the proper energies of each molecule, and the interaction will be the intermolecular interaction, thus treating on equal footing the intra- and intermolecular interactions.

Up to this point, we have mostly fulfilled the program stated in the Introduction for a good energy partition. We are able to identify atoms and functional groups, which are transferable (as shown by Bader[8]). The energy contains a clear description of interactions, very intuitive since it is given in terms of the Coulombic interactions of the constituent particles of each atom, and which is equally capable of describing intermolecular interactions. In practice, the partition only needs the nuclear geometry and a wave function (or, equivalently, the nondiagonal first-order and the diagonal second-order density matrices), and thus it is a physical property of the system, as accurate as the wave function being used. A final point remains to be fully justified, that is, how the binding energy arguments can be cast into this new language. This will be addressed in the next subsection.

C. Atomic Self-Energy and Deformation Energy. It is clear that the net energy is the quantity carrying the atomic identity from system to system: the core and the internal valence are not going to change much for a given atom in different systems, and thus their kinetic energies, the attraction by its atomic nucleus, and the electron−electron repulsion (including correlation) are going to be largely the same. Thus, we could take the net energy of an atom (or functional group) in a given environment as a reference when comparing the same atom in different environments, and its changes will be very small, while those in the interaction energy will be larger. This can be beautifully explained using an argument by Li and Parr.[23] For a given atom or functional group, we can see the influence of the rest of the system as a nonlocal potential acting over it. When the chemical environment changes (provided there is not a large charge transfer), we can estimate the effective energy change through the Hellmann−Feynman theorem as

$$\Delta E_{\mathrm{eff}}^{\mathrm{A}} \simeq \int_{\mathrm{A}} \Delta \hat{V} \rho_0^{\mathrm{A}} \tag{23}$$

up to first order, where ρ_0^{A} is the atomic density in the original environment. Since the potential change only affects the interaction energy, the net energy will remain mostly constant (up to first order) for small and moderate changes in the environment of a given atom or functional group. However, if the environment change is large, there are several factors through which the net energy can change: (i) charge transfer, (ii) electronic reorganization within the atomic basin, and (iii) a change in the interatomic surface.

The net energy contains all of the energy contributions that are already present in an isolated atom, and so the free atomic energies are comparable in order of magnitude. This fact can be exploited by taking our net energy reference as that of the atoms in vacuo, thus defining what we call the atomic deformation or reorganization energy

$$E_{\mathrm{def}}^{\mathrm{A}} = E_{\mathrm{net}}^{\mathrm{A}} - E_{\mathrm{vac}}^{\mathrm{A}} \tag{24}$$

We label this as deformation since it takes into account the deformations in the atomic density on going from the free state into the combined system. This labeling is appropriate when there is a change in shape of the density or the atomic basin, but its larger component, by far, comes from charge transfer when this is present. For very ionic systems, the reference can be set in the isolated ions, for example, but for intermediate systems (partially ionic) there is usually no isolated-atom system that can be taken as a faithful reference. Nevertheless, the net energy does not change between similar environments even for cases where there is no close enough reference, and this constancy can always be exploited by taking as reference a nonisolated atomic net energy. However, there is an important property to be drawn from the deformation energy with respect to the isolated atoms: using eq 17, the binding energy can be computed as

$$E_{\mathrm{bind}} = E - \sum_{\mathrm{A}} E_{\mathrm{vac}}^{\mathrm{A}} = \sum_{\mathrm{A}} E_{\mathrm{def}}^{\mathrm{A}} + \frac{1}{2} \sum_{\mathrm{A}} \sum_{\mathrm{B} \neq \mathrm{A}} E_{\mathrm{int}}^{\mathrm{AB}} \tag{25}$$

In this way, all the contributions are usually of the same order of magnitude as the binding energy itself, since the large cancellation of the atomic contributions is already included in the deformation energy, and so the origin of stability can be easier to rationalize. Thus, one can see the binding of the molecule as the sum of a deformation energy, which can be shown to be necessarily positive in homodiatomics and is usually positive (depending on the reference) in any other case, and an interaction energy, that is usually negative. This can be described as the following virtual process: the atoms have to raise their energy in order to form the molecule, but this energy gain is compensated by a larger energy lowering due to their interaction.

Equation 25 is equivalent to that given in Li and Parr's partition scheme,[23] where they label the deformation energy as a promotion energy, in view of the mental process of molecule formation just mentioned. However, our scheme differs mainly in two points from theirs. First, their Θ_{A} weighting factors defining the density matrices' partition are diffuse, continuous functions in the general case, instead of being step functions (they can be analytically shown to coincide with QTAM's partition for the H_2^+ case, but that is no longer true when the electron−electron interaction comes in). Second, they put an a priori requirement into the atomic partition, namely that it gives the lowest possible sum of the promotion energies (deformation energies in our language). This is done so that the atoms in the molecule are energetically as close as possible to the free atoms, but it really amounts to minimizing the net energy sum (or maximizing the interaction energy). This requirement defines the partition into atomic densities, in the same way that the zero-flux condition defines ours. Many other partitioning schemes can be introduced by selecting different definitions for the atomic density weighting factors. In the case of Li and Parr's scheme, while the partition can be reasonable for homonuclear and low polarity systems, it will fail to meet chemical intuition for highly polar or ionic compounds: since the promotion energy into ions is very large, the minimization of the net energies will tend to make all of the atoms in the molecule neutral, and so the traditional understanding of the chemistry of these compounds as coming from a large charge-transfer compensated by a larger electrostatic energy

lowering will be lost. This is not the case in our partition, where we choose the atomic weighting functions according to the topology of the electron density. Thus, a charge transfer will be predicted whenever the electron density rearranges in that way and not through an energy criterion.

A final word on the energy references is necessary. It is the behavior of the net energy, which should not change much for an atom across similar compounds, that gives the partition its meaning, and not the reference used to define the deformation energy. The effect of the reference is to shift the atomic energies by a constant, and it is only introduced for convenience in comparing the net energies of an atom in different compounds with a common level and in understanding binding energies.

D. Interatomic Interaction Energy. As already mentioned, the interaction energy provides a very intuitive definition of the interaction between two atoms, including all of the pairwise interparticle contributions: V_{nn}^{AB}, V_{ne}^{AB}, V_{en}^{AB}, V_{ee}^{AB}. It is thus based on the physical interaction between the two systems, fully taking into account the quantum nature of the electrons through the use of the second-order density matrix in the electron−electron repulsion. It also does not make any a priori distinction between bonded (i.e. atoms connected by a chemical bond), nonbonded, or even intermolecular interactions, treating all of them on the same basis. Of course, the ranges of interaction energy values will be different for each of these cases and also for different kinds of bonds. To better understand the interactions, it is desirable to make a different arrangement of its components: although the final atomic charges are usually small, and close to zero in many cases, the nuclear charges are usually much higher, and also the total number of electrons. Thus, the above-mentioned terms will be several orders of magnitude larger than the interaction energy, while an approximate application of the Gauss theorem ensures that they will mostly cancel out for neutral systems.

The second-order density matrix admits a natural partition that facilitates a meaningful rearrangement, as presented in ref 18. We define the Coulomb term of ρ_2 as $\rho_2^C(\mathbf{r}_1,\mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)$ and the Fock-Dirac exchange term as $\rho_2^X(\mathbf{r}_1,\mathbf{r}_2) = -\rho_1(2;1)\rho_1(1;2)$. The remainder then defines the correlation density, $\rho_2^{corr} = \rho_2 - \rho_2^C - \rho_2^X$, and we can also define an exchange-correlation term as $\rho_2^{xc} = \rho_2^X + \rho_2^{corr}$, so that $\rho_2 = \rho_2^C + \rho_2^{xc}$. The total electron−electron interaction in eq 14 can be then partitioned in Coulomb, exchange, and correlation terms

$$V_\tau^{AB} = \int_{\Omega_A}\int_{\Omega_B}\rho_2^\tau(1,2)r_{12}^{-1}\,d\mathbf{x}_1\,d\mathbf{x}_2 \tag{26}$$

where $\tau = C, X,$ corr. In this way

$$V_{ee}^{AB} = V_C^{AB} + V_X^{AB} + V_{corr}^{AB} \tag{27}$$

where V_C^{AB} coincides with the J^{AB} integral and V_X^{AB} with $-K^{AB}$ in eqs 10 and 11 of ref 17. Now we can collect all the classical electrostatic terms in what we call the classical interaction energy

$$V_{cl}^{AB} = V_{nn}^{AB} + V_{ne}^{AB} + V_{en}^{AB} + V_C^{AB} \tag{28}$$

and thus the interaction energy becomes

$$E_{int}^{AB} = V_{cl}^{AB} + V_{xc}^{AB} \tag{29}$$

since $V_{xc}^{AB} = V_X^{AB} + V_{corr}^{AB}$.

The V_{cl}^{AB} classical term is the one where the interaction cancellations occur. Our atoms include both nucleus and electrons, and they tend to preserve their core and only partially share, donate, or accept their valence electrons, thus being neutral or having small charges. In this way, V_{cl}^{AB} for neutral atoms will be negligible if they are well separated, and it will be positive and larger if they are closer together (bonded, one could say). Of course, if the atoms are heavily charged, the charge−charge classical interaction will be the leading contribution to V_{cl}^{AB}, positive or negative for same or oppositely charged atoms, respectively. In any case, V_{cl}^{AB} will always be much smaller than the individual terms in eq 28.

It should be recalled here that, as shown in ref 17, all of the two-electron interactions can be efficiently computed through an exact and convergent multipolar expansion of the energy, and the same scheme is used here for consistency for the one-electron terms. So, each of the potential energy contributions, one- and two-electron but also one- and two-body, can always be written as a multipolar sum. This is most important for the classical electrostatic terms, where we can define approximate energy contributions based on the point multipoles (Q_l^A are the order-l spherical multipoles of the electron density as defined in eq 26 of ref 17). In this way, given the total charge of the atom, $Q^A = Z^A - Q_0^A$, the leading point-charge (monopolar) term is $V_Q^{AB} = Q^AQ^B/r_{AB}$. Joining all of the electronic point multipole contributions, we define the long-range Coulomb term $V_{C,lr}$ (equivalent to J_{lr}^{AB} as defined in eq 24 of ref 17). If we substitute $Q_0{}^A \rightarrow Q^A$ within $V_{C,lr}$, thus including the point multipole contributions to the electron−nucleus and the nucleus−nucleus interaction, we obtain the long-range part of the classical interaction, $V_{cl,lr}^{AB}$, an approximation to V_{cl}^{AB}. These long-range approximations will be especially useful in understanding intermolecular interactions, where they account for most of the interaction, but they are also valuable when discussing some intramolecular interactions.

The exchange-correlation contribution to the AB interaction, V_{xc}^{AB}, contains all of the quantum terms in the interaction, and it will typically be smaller than the Coulomb term, V_C^{AB}. However, its importance when compared with the classical interaction, V_{cl}^{AB}, will depend on the type of interaction between A and B. Although we have further split the exchange-correlation interaction into exchange and correlation terms, it should be kept in mind that any such separation is arbitrary and can produce misleading results. Nevertheless, our Fock-Dirac-like exchange, based on the first-order density matrix that can be derived from the second-order one, is uniquely defined, independent of any orbital representation, and coincides with the usual exchange (with zero correlation) in the monodeterminantal limit.

The partition into classical and quantum terms, and into exchange- and correlation-only terms, can also be carried on within the intra-atomic contributions by splitting

ρ_2 as above, thus defining

$$V_\tau^{AA} = \frac{1}{2}\int_{\Omega_A}\int_{\Omega_A} \rho_2^\tau(1,2) r_{12}^{-1}\, d\mathbf{x}_1\, d\mathbf{x}_2 \qquad (30)$$

where $\tau = C, X$, corr. This is analogous to eq 13, but it should be noticed that V_C^{AA} and V_X^{AA} so defined are half of the integrals J^{AA} and $-K^{AA}$ as defined in eqs 10 and 11 of ref 17 when A = B. In this way, $V_{ee}^{AA} = V_C^{AA} + V_X^{AA} + V_{corr}^{AA}$, and we can define $V_{cl}^{AA} = V_{ne}^{AA} + V_C^{AA}$ and $V_{xc}^{AA} = V_X^{AA} + V_{corr}^{AA}$, so that $E_{net}^A = T^A + V_{cl}^{AA} + V_{xc}^{AA}$. However, this partitioning will not be as useful as that in the interaction energy, since there will not be any cancellation within the single basin. Furthermore, the net energy also contains the electronic kinetic energy, another quantum term with a much larger value than the exchange-correlation one.

III. Energy Analysis

In this section, we will present the results of our energy partition scheme for several chemically representative systems. In the first subsection, we will give a thorough analysis of the hydrogen molecule, a well understood benchmark with which to compare any new idea. Then, we will perform a comparative analysis of N_2, H_2O, LiF, and He_2, molecules representing apolar covalent, polar covalent, ionic, and van der Waals bonding types, respectively. Bonding in the molecules of metallic atoms is not qualitatively different from covalent bonding, and so we will not include any example, while hydrogen bonding is a complicated enough issue as to warrant a separate study.

All of the calculations have used the *gamess* code[29] to obtain the wave function and our code *promolden* to do the QTAM analysis and energy partition. The wave functions have been computed using Hartree−Fock (HF, for H_2 and N_2), different levels of complete active space multiconfiguration calculations (CAS[*n*, *m*], *n* active electrons and *m* active orbitals: CAS[2,2] for H_2, CAS[10,10] for N_2, CAS[6,5] for H_2O, and CAS[10,10] for LiF), and full configuration interactions (FCI, for H_2 and He_2). The basis sets used where Pople's 6-311G(d,p), except for He, in which a cc-PVTZ Dunning basis set was used. The numerical integrations in *promolden* used β-spheres for all of the atoms, with radii between 0.3 a_0 (H in H_2O) and 1.6 a_0 (F in LiF). Inside these spheres, a 600 points Gauss-Chebychev second kind radial quadrature and a Lebedev angular quadrature with 74 points were used, considering generalized energy multipoles up to $L = 6$. Outside the β-spheres, a trapezoidal radial quadrature with 500 points and a 1202 angular points Lebedev quadrature were used, with generalized energy multipoles up to $L = 10$ (see ref 17 for the definition of the computational parameters and the integration algorithm). These integrations are enough to obtain the so-called chemical accuracy, since the errors with respect to the analytical energies as given by *gamess* were always smaller than 0.33 kcal/mol. All of the calculations were performed at the respective experimental equilibrium geometries (except for He_2, where an accurate theoretical value is used instead[30]), to better compare calculations with different approximate wave functions.[31] We want to stress once more that our QTAM energy partition does not rely on the virial theorem, and thus can be used outside of the equilibrium geometry of the corresponding Hamiltonian. This allows us to obtain the partition at any point of the potential energy surface, and we will use it to follow the behavior of the different components when varying the distance in the H_2 molecule in the next subsection.

Table 1: QTAM Energy Partition for H_2 at the Experimental Equilibrium Distance (0.74144 Å), as Described by Three Different Wave Functions[a]

Prop	HF	CAS[2,2]	FCI
Q^A	16 μe	15 μe	16 μe
Q_1^A	−0.1021	−0.1068	−0.1064
Q_2^A	−0.3617	−0.3452	−0.3488
T^A	0.5608	0.5805	0.5849
V_{en}^{AA}	−1.2153	−1.2277	−1.2251
V_{ee}^{AA}	0.1979	0.1628	0.1532
V_{cl}^{AA}	−0.8196	−0.8283	−0.8267
V_{xc}^{AA}	−0.1979	−0.2366	−0.2452
V_C^{AA}	0.3957	0.3994	0.3984
V_X^{AA}	−0.1979	−0.1988	−0.1967
V_{corr}^{AA}	0.0000	−0.0378	−0.0486
E_{net}^A	−0.4566	−0.4844	−0.4871
ΔT^A	0.0610	0.0807	0.0851
ΔV_{en}^A	−0.2157	−0.2281	−0.2255
ΔV_{ee}^A	0.1979	0.1628	0.1532
E_{def}^A	0.0432	0.0154	0.0128
V_{nn}^{AB}	0.7137	0.7137	0.7137
V_{en}^{AB}	−0.5974	−0.5975	−0.5975
V_{ee}^{AB}	0.2619	0.2993	0.2871
V_{cl}^{AB}	0.0426	0.0423	0.0423
V_{xc}^{AB}	−0.2619	−0.2244	−0.2365
V_C^{AB}	0.5238	0.5237	0.5236
V_X^{AB}	−0.2619	−0.2522	−0.2510
V_{corr}^{AB}	0.0000	0.0279	0.0145
E_{int}^{AB}	−0.2193	−0.1820	−0.1942
E (integ)	−1.1324	−1.1509	−1.1683
E (analy)	−1.1325	−1.1509	−1.1683
ΔE	41μE_h	40μE_h	42μE_h
E_{bind}	−0.1328	−0.1513	−0.1687

[a] Atomic units are used throughout.

A. The Hydrogen Molecule. All of the H_2 QTAM energy partition contributions and their energy components as defined in section II are gathered in Table 1. The different columns refer to wave functions of increasing accuracy: single-determinant Hartree−Fock, two-determinant CAS[2,2], and multideterminantal full CI. The first block corresponds to the atomic charge (Q^A), dipole (Q_1^A), and quadrupole (Q_2^A) of atom A (the one in the negative *z* axis). The second one contains the different one-atom contributions to the net energy, E_{net}^A, together with the atomic deformation energy, E_{def}^A (the change in net energy of the atom with respect to the in vacuo reference), and its $\Delta X^A = X^A(H_2) - X^A(H_{vac})$ ($X^A = T^A$, V_{en}^{AA}, V_{ee}^{AA}) components. The third block lists the interatomic interaction energy, E_{int}^{AB}, and its decomposition according to the various schemes proposed in the previous subsection. Finally, the last block lists the total energy as

recovered by the sum of the integrated QTAM contributions, the analytical value as obtained from *gamess*, the total energy integration error $\Delta E = E(\text{integ}) - E(\text{analy})$, and the binding energy with respect to the free atom reference.

Let us first address the precision of the numerical integration: the total charge is always smaller than 2×10^{-5} *e*, while the error in the total energy sum is on the order of 4×10^{-5} E_h. We have found that both absolute errors are usually related, as one should expect. The main source of error in the energy lies in its largest term, the intra-atomic nuclear attraction, and it is related to the discontinuity in the integrand at the atomic surface: the angular integration of a discontinuous function introduces a noise related to the jump in the value of the function at the discontinuity. We have observed that any error in the estimation of the charge introduces a proportional error in the nuclear attraction integral. Thus, one could perform previous exploratory charge integrations, of order N^3 complexity, to estimate the order of magnitude of the error in the energy, whose calculation has an N^4 complexity with a much higher prefactor (also growing with the square of the number of atoms).

As could also be expected for H_2, in which the HF approximation gives qualitatively correct geometrical results, the three calculations give a qualitatively consistent picture. However, HF fails to give accurate enough values for the kinetic and electron−electron contributions. Simply allowing the mixing with the $(\sigma_u)^2$ configuration (introducing the antibonding orbitals coming from 1*s* AOs in the simple orbital picture) generates a wave function in much better agreement with the FCI result for this basis set. The deformation energy in HF is three times larger than in CAS and FCI, due to two complementary facts: the well-known dissociation error of HF, which is partially translated into the equilibrium situation, and the lack of correlation energy (which in fact mainly corrects for the latter error), of the same order of magnitude of the overestimation.

The atomic dipoles (Q_1^A) correspond to an outward displacement of the electronic charge, while the quadrupolar deformation (Q_2^A) shifts charge outside from the internuclear axis. The values are fairly insensitive to the type of calculation, being almost identical for the CAS and FCI calculations.

Let us focus on the second block of Table 1, including intra-atomic contributions. First of all, let us notice that $V_{ee}^{AA} \neq 0$: although there is one electron on average on each atom, there is a non-negligible probability of having two electrons on a given atomic basin. Regarding the net energy (see eq 15) it is clearly seen that it stems from the balance of the large, negative electron−nuclear attraction, and the somewhat smaller kinetic (on the order of half the absolute value of $V_{en}^{AA} + V_{ee}^{AA}$, even though no virial theorem applies to them) and electron repulsion positive contributions, all of them of the same order of magnitude as the net energy itself. The situation does not change if we define the classical and exchange-correlation contributions: both are negative but still of the same order of magnitude as the kinetic energy. This is inherent to the intra-atomic energies, which always bear large cancellations. However, as pointed out in the previous subsection, these large contributions are similar to the free atom ones, as we can see from the ΔX^A contributions: they are an order of magnitude smaller than the $X^A(H_2)$ values, except for ΔV_{ee}^A. This is due to the special character of the monoelectronic H atom, with no electron−electron repulsion, and does not happen in many-electron atoms, as we shall see below. The increase in kinetic energy, decrease in the nucleus−electron attraction, and increase in the electron−electron repulsion are consistent with Slater's ideas on covalent bonding:[32] electrons in the molecule go faster and are closer to the nuclei, and the sharing puts them closer together so that they repel each other more (this will also be true when the free atom V_{ee}^{AA} is not zero). The intra-atomic balance still gives a positive deformation energy, which is almost an order of magnitude smaller than the ΔX^A values. This is general and reveals that the deformation energy is a hard to partition property, coming from a very delicate interplay of large magnitudes.

Regarding the third block in Table 1, containing the interatomic interaction contributions, we can see that the positive and negative (there are two, symmetrical in this homonuclear case, electron−nucleus interactions) contributions have all of them the same order of magnitude, and they mostly cancel to give the interaction energy, smaller than any of them (see eq 16). However, we see that V_{cl}^{AB} is quite small and very similar in the three calculations (as they are its components, V_{nn}^{AB}, V_{en}^{AB}, and V_{C}^{AB}), while the exchange-correlation quantum contribution is larger in absolute value and varies with the amount of correlation energy included. In fact, it is remarkable that the Fock-Dirac exchange contribution in correlated wave functions is within 1% (intra-) and 5% (interatomic, and almost equal for CAS and FCI) the HF one. Due to this, the correlation interaction energy, V_{corr}^{AB}, is the main contribution to the differences in the interaction energies when varying the correlation energy scheme. This did not happen with the intra-atomic contributions, where the variations in the density matrices introduced changes in the kinetic energy and nuclear attraction.

Finally, the total energy, the sum of the net energies of both H atoms plus their interaction (see eq 17), is clearly a large number, mainly dominated by the net energies sum. However, when the binding energy is computed (eq 25), its value is an order of magnitude smaller than the total energy, and we also see that its main contribution comes from the interaction: E_{int}^{AB} is about 2.5 times the sum of deformation energies in the HF case, and about 6 times larger for the correlated cases. Thus, binding in H_2 can be understood through a small deformation of the H atoms (8 kcal/mol each in the FCI calculation), compensated by a much larger interaction energy (−122 kcal/mol), which in turn is dominated by the pure quantum mechanical exchange-correlation part (−148 kcal/mol versus 26 kcal/mol for the classical interaction).

It is interesting to compare our HF results with those of Salvador et al. in ref 14. Although they used a somewhat poorer basis set (6-31G(d,p), double-ζ in the valence instead of our triple-ζ), the main results are equivalent within the basis set errors. The numerical precision is similar as well as the net, interaction, and total energies, differing in less

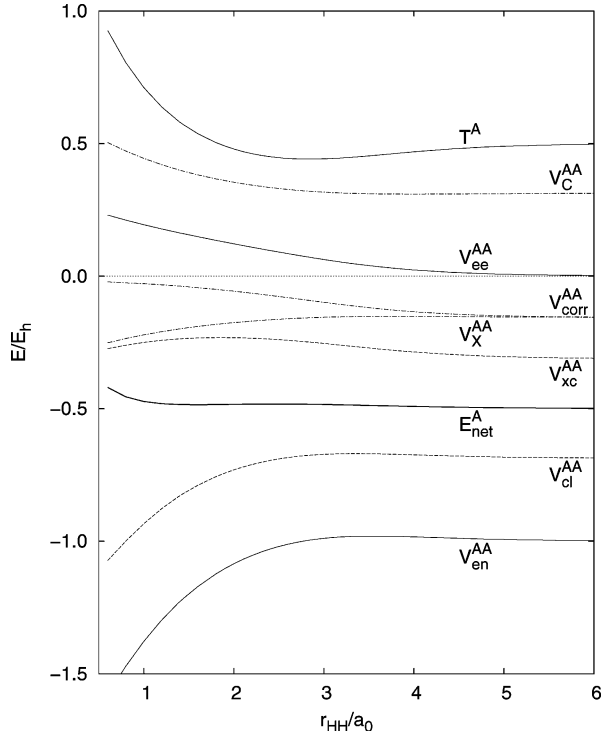


Figure 1. CAS[2,2] net energy (E^{A}_{net}, intra-atomic) and its components for H_2 as functions of internuclear distance (r_{HH}).

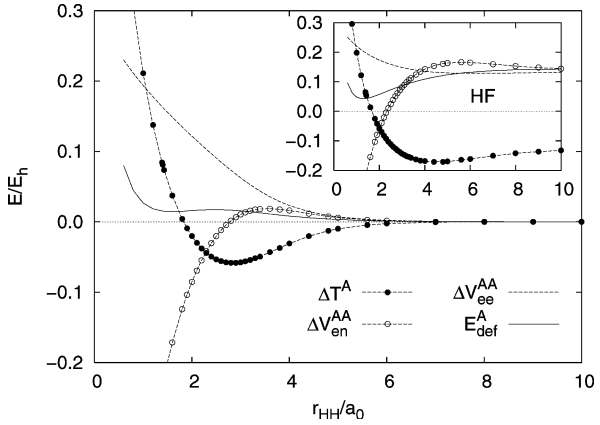


Figure 2. CAS[2,2] deformation energy (E^{A}_{def}) and its contributions for H_2 as functions of internuclear distance (r_{HH}). The inset shows the corresponding HF values.

than 2 mE_h. The main difference lies in the kinetic energies, since our calculation is away from equilibrium to better compare with the correlated cases, but is nevertheless of about 4 mE_h.

Let us now examine the behavior of the partition for a wide range of distances using the CAS wave function. In Figure 1 we have included the net energy, together with its several components and their possible rearrangements, as functions of the internuclear distance. The net energy appears to be flat in this energy scale, with a value of $-0.5\ E_h$ for distances above 1 a_0. It is clearly seen that all of its components are large, and not much insight can be extracted from them. In fact, the main components T^A and V^{AA}_{en} stand out from the rest above and below, respectively, while V^{AA}_{ee} is only important at short distances.

It is interesting to follow the independent behavior of exchange and correlation: the latter seems to go to zero at short distances, while at large distances exchange and correlation tend to the same limit. This result is clearly related to the simple, two-determinantal wave function. At short distances, with the He atom as limit, the antibonding orbital goes into a united atom p state, with a quite small contribution to correlation and hence the wave function will behave as a single-determinant one, with $V^{AA}_{corr} \simeq 0$. On the other hand, at large distances the singlet is in fact a 50% superposition of two independent doublets, leading to this artificial exchange-correlation partition. So, the separation into exchange and correlation only makes sense when there is a dominant determinant, and the correlation contribution comes from a partial (but not negligible) mixing with other configurations. In any case, one must remember that this separation is not uniquely defined. Also, the meaning of the exchange-correlation term is distorted in the long-distance regime of covalent (i.e., electron pairing) bonds. In this regime the system should be properly treated as two separate and independent subsystems. The dependence introduced to make a singlet from the two doublets is related to the quantum entanglement between them: if one of the H atoms is spin-up, the other is certainly spin-down. It should be stressed that this is not solvable with a better wave function: the CAS[2,2] gives the exact solution for this system in the $r_{HH} \rightarrow \infty$ limit within the basis set chosen, since the 1s HF orbital is the exact solution for H.

Although the individual components of E^{A}_{net} are widely separated, subtracting the infinite distance limit values (free atoms) transforms these components into the ΔX^A values, and the net energy into E^{A}_{def}, as shown in Figure 2. Notice that the energy scale is much smaller than that in Figure 1 and that all of the curves go to zero as $r_{HH} \rightarrow \infty$, but E^{A}_{def} is still almost an order of magnitude smaller than these components. Let us stress again that E^{A}_{def} is very small whenever the charge transfer is small, and in general that E^{A}_{net} tends to maintain its value for a given atom up to first order in the variations of its environment (this fails under large compression here and also when there is a significant charge transfer). Figure 2 shows that, at long distances, the kinetic energy decreases, while the nuclear attraction increases, with respect to the free atom limit. Both trends are reversed at smaller distances, so that at the equilibrium geometry (1.4 a_0) the kinetic energy is larger and the nuclear attraction smaller than the free atom limit. On the other hand, the electron−electron repulsion increases monotonically from zero (one-electron separated atoms) when decreasing the distance. Adding up all of these contributions, the deformation energy varies as follows: for distances larger than 5 a_0, there are large cancellations and its value is almost zero; in going from 5 to 3 a_0, its value increases slightly; between 3 and 1 a_0 it is almost flat; and finally, below 1 a_0 it increases sharply due to the increase of the kinetic energy (that becomes even steeper than ΔV^{AA}_{en}) and the steady increase of the electron repulsion energy. The HF results, shown in the inset, have a completely different behavior, owing to the wrong dissociation limit.

Figure 3 represents the H−H interaction energy, E^{AB}_{int}, and its main contributions. All of the V^{AB}_{nn}, $-V^{AB}_{en}$, and V^{AB}_{ee}

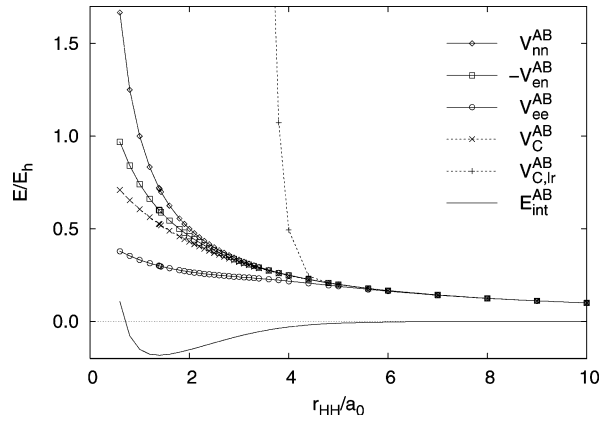


Figure 3. CAS[2,2] interaction energy (E_{int}^{AB}, interatomic) and its components for H_2 as functions of internuclear distance (r_{HH}).

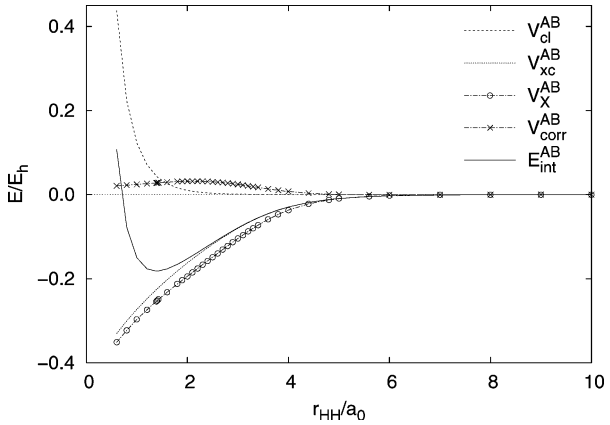


Figure 4. Alternative partition of the CAS[2,2] interaction energy (E_{int}^{AB}) into classical (V_{cl}^{AB}) and exchange-correlation (V_{xc}^{AB}) contributions for H_2 as functions of internuclear distance (r_{HH}).

components are monotonic and much larger than the interaction energy itself, with a large cancellation that is most evident at long distances: all of them fall as $1/r_{HH}$ with distance, as expected, but their sum goes to zero much faster, being the interaction energy negligible above 5 a_0. This is precisely the limit in which V_{nn}^{AB} is equivalent to V_{ee}^{AB} and to $-V_{en}^{AB}$ and also to V_C^{AB}: at long distance, the point multipolar approach is exact, and the long-range (point multipole) component of the Coulomb interaction, $V_{C,lr}^{AB}$, becomes meaningful and equivalent to the overall V_C^{AB} value, its leading monopole term being the product of the number of electrons of each atom (1 e) over their distance, $1/r_{HH}$.

Figure 4 shows the alternative partition of the interaction energy into classical and exchange-correlation terms, which gives more chemical insight: the V_{cl}^{AB} term is almost zero for distances larger than 2 a_0 and largely repulsive below this point, while the exchange-correlation term is monotonically attractive. Thus, the interaction energy resembles V_{xc}^{AB} on its attractive part, while its repulsive contribution is dominated by the classical part at small distances. At the experimental equilibrium distance of 1.4 a_0, almost at the minimum of the E_{int}^{AB} curve, the classical part is still small, and the interaction is dominated by the exchange-correlation part.

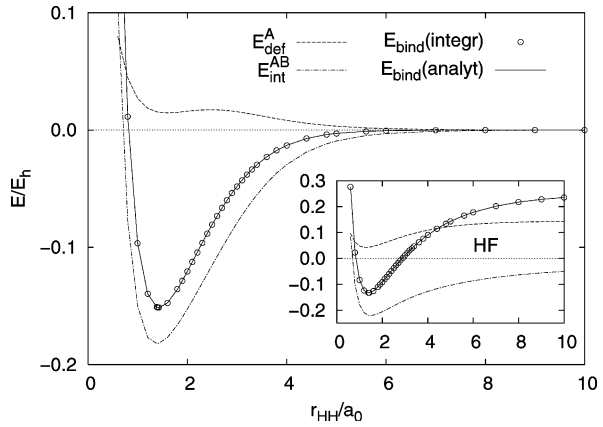


Figure 5. CAS[2,2] binding energy (E_{bind}) partition for H_2 as a function of internuclear distance (r_{HH}). The inset shows the corresponding HF values. Notice the coincidence of the numerically integrated (circles) and analytically computed (solid line) values in both cases.

Moreover, Figure 4 also includes the separation of the exchange-correlation interaction energy into exchange-only and correlation-only contributions. Since the correlation part is much smaller than the exchange part, it is the latter which dominates V_{xc}^{AB}, and through it the attractive part of E_{int}^{AB} and also of E_{bind} (since E_{def}^{A} is positive and almost flat, Figure 2). Thus, we may say that H_2 is mainly bound by the exchange interaction between the two atoms, in the same spirit of the Heitler-London original treatment and in accordance with the traditional ideas of covalent bonding. It is important to notice that the V_X^{AB} integral is in fact equivalent to the Heitler-London resonance integral: it is an exchange integral in which each of the electrons is restricted to a different atom (eq 26), through the density matrix contribution $\rho_2^{X,AB}(1, 2) = -\rho_1(2;1)\rho_1(1;2)\Theta_A(1)\Theta_B(2)$. The energy lowering acquired through this delocalization is responsible for the attractive force, compensating the electrostatic classical repulsion and the intra-atomic (deformation) energy increase.

Finally, let us compare the two main components of binding, the deformation and the interaction energies. Figure 5 presents the binding energy of H_2 as computed by numerical integration of its components within QTAM (symbols) and analytically within *gamess* (solid line). They are almost indistinguishable, the integration error being always smaller than 0.1 kcal/mol and mostly independent of the distance. Notice the depth of the well, about 0.15 E_h, and that the energy scale is much smaller than in previous figures. As already mentioned, the deformation (net) energy is almost flat between 1 and 3 a_0, precisely the region of the minimum, and thus the binding energy minimum is almost coincident with the minimum of the interaction energy. On the other hand, since the deformation energy is smaller, interaction dominates the overall binding behavior above 1 a_0, while the repulsion below this point is the sum of both deformation and interaction.

This picture is not clear at all in the HF results, shown in the inset of Figure 5. This is again due to the dissociation problem, which makes the three curves to have different $r_{HH} \rightarrow \infty$ limits. Notice how the deformation energy is always

Table 2: QTAM Energy Partition on Several Molecules at Their Equilibrium Geometries: Experimental for N_2 (1.09768 Å), H_2O (0.95781 Å, 104.4776°), and LiF (1.563864 Å), Theoretical for He_2 (5.6 ± 0.05 a_0)[a]

properties	N_2(HF)	N_2(CAS)	H_2O(CAS)		LiF(CAS)		He_2(FCI)
atomic	N	N	O	H	F	Li	He
Q^A	1.4*e*−4	1.3*e*−4	−1.0903	0.5448	−0.9299	0.9299	0.000009
$Q_1{}^A$	−0.6218	−0.6269	0.1828	−0.1705	0.3415	−0.0147	−0.000727
$Q_2{}^A$	−0.7014	−0.6608	0.5876	0.0084	−0.1726	−0.0330	−0.004062
T^A	54.3857	54.5993	75.3874	0.3968	99.8543	7.3589	2.900972
V_{en}^{AA}	−129.3892	−129.6904	−184.4199	−0.7517	−243.8116	−16.3471	−6.751391
V_{ee}^{AA}	20.8630	20.7466	34.3837	0.0447	44.3978	1.7518	0.950394
V_{cl}^{AA}	−102.4366	−102.5803	−141.5102	−0.6397	−188.8937	−12.9159	−4.702448
V_{xc}^{AA}	−6.0895	−6.3635	−8.5259	−0.0673	−10.5201	−1.6794	−1.098549
V_C^{AA}	26.9525	27.1101	42.9096	0.1120	54.9179	3.4312	2.048943
V_X^{AA}	−6.0895	−6.0899	−8.4137	−0.0519	−10.2325	−1.6483	−1.013100
V_{corr}^{AA}	0.0000	−0.2736	−0.1122	−0.0154	−0.2876	−0.0310	−0.085449
ΔT^A	−0.0140	0.1852	0.5678	−0.1030	0.2985	−0.0732	0.000052
ΔV_{en}^A	−1.0352	−1.3133	−6.3206	0.2479	−5.1238	0.7989	−0.000189
ΔV_{ee}^A	1.3036	1.1861	5.9043	0.0447	4.7557	−0.5300	0.000464
E_{net}^A	−54.1405	−54.3445	−74.6487	−0.3102	−99.5595	−7.2364	−2.900025
E_{def}^A	0.2543	0.0579	0.1515	0.1896	−0.0696	0.1956	0.000326
interaction	N−N	N−N	O−H	H−H	Li−F		He−He
V_{nn}^{AB}	23.6222	23.6222	4.4199	0.3494	9.1362		0.712697
V_{en}^{AB}	−22.1174	−22.1327	−4.8922	−0.1434	−10.0462		−0.712554
V_{ne}^{AB}	−22.1174	−22.1327	−1.6458	−0.1434	−6.2778		−0.712554
V_{ee}^{AB}	19.9235	20.1843	1.6449	0.0539	6.8579		0.711748
V_{cl}^{AB}	0.2348	0.2294	−0.2751	0.1225	−0.2836		0.000001
V_{xc}^{AB}	−0.9239	−0.6883	−0.1981	−0.0059	−0.0463		−0.000664
V_C^{AB}	20.8474	20.8726	1.8430	0.0598	6.9042		0.712412
$V_{C,lr}^{AB}$	1.51*e*3	1.29*e*3	4.96*e*4	0.1238	6.8241		0.712412
V_X^{AB}	−0.9239	−0.8642	−0.2112	−0.0012	−0.0456		−0.000646
V_{corr}^{AB}	0.0000	0.1760	0.0131	−0.0048	−0.0007		−0.000018
E_{int}^{AB}	−0.6891	−0.4589	−0.4732	0.1166	−0.3299		−0.000664
global							
E (integ)	−108.9700	−109.1479	−76.0991		−107.1258		−5.800713
E (analy)	−108.9696	−109.1480	−76.0986		−107.1258		−5.800702
ΔE	−0.0004	0.0001	−0.0005		1.8*e*−5		−0.000011
E_{bind}	−0.1800	−0.3432	−0.2988		−0.2039		−1.323*e*−7

[a] Dipoles are positive when the electrons displace toward the internuclear region, quadrupoles correspond to the z^2 spherical component along the bond (the only nonzero one) except for water in which the main component is written ($x^2 - y^2$ for O with z the molecular axis, yz for H with z the H−O line and yz the molecular plane). Atomic units are used for all properties.

positive, as mentioned in subsection IIC for homodiatomics, but it falls below its apparent long-distance energy. This is because its value is worse at a very long distance, but not so much near equilibrium, where the HF determinant dominates within the exact wave function. On the other hand, the interaction energy goes to zero much more slowly than with the CAS wave function: this is due to the exchange part, which forces a $1/r_{HH}$ behavior instead of the cancellation shown by the CAS wave function results. These problems are the main source of error for the HF wave function, which nevertheless can give qualitatively correct results when the depth of the binding energy well is larger (masking the singlet limit dissociation error) or for noncovalent bonding cases. Also, the dominant contributions to either the deformation or the interaction energies can usually be correctly guessed within the HF scheme.

B. Representative Molecules. We will now analyze the QTAM energy partition on some molecules representative of different kinds of bonding. The results for N_2 (both HF and CAS[10,10]), H_2O, LiF (at their experimental equilibrium distances), and He_2 (at its theoretically predicted[30] equilibrium distance) are gathered in Table 2. Once again, the total charge integration errors are small (the largest is the one in water, with a 0.7 m*e* excess), as are the total energy integration errors (the largest corresponds again to water, with a 0.5 mE_h defect).

The HF dissociation problem displayed by H_2 persists in N_2: the HF binding energy is half its CAS value, the deformation more than four times larger, and the interaction is 50% larger in absolute value. In addition, the experimental geometry is away from the HF equilibrium geometry, and so the HF calculation is far from being virial-consistent. This is the reason behind the negative value of ΔT^N (−0.0140 E_h), since the quantum mechanical molecular calculation gives $2T + V = -0.2\ E_h$, with a clear lack of kinetic energy, whereas the calculation of the isolated atom gives $2T_{vac} +$

$V_{vac} = 0.005\ E_h$. On the other hand, the rest of the HF values are surprisingly good, being that the correlation energy (both intra- and interatomic) is the main source of error for this method. Thus, the HF calculation can indeed be used as a qualitative guide of the partition results. Overall, the binding in N_2 shows again a relatively small deformation energy (notice that, in the CAS calculation, it is 36 kcal/mol, 3 orders of magnitude smaller than the net energy), with a large interaction energy (−288 kcal/mol). This interaction is again dominated by the exchange-correlation term, although in this case the classical term is of comparable magnitude (144 kcal/mol versus −432 kcal/mol for the former). This is probably due to the much larger polarization of the N atoms when compared with the H atoms (see the corresponding Q_1^A dipole values), which is also responsible for the fact that $E_{def}^N \gg E_{def}^H$.

Compared with Salvador et al. results,[14] our integration accuracy is here an order of magnitude better, due to the considerably denser integration grids we used, but the overall results again agree within the basis set differences (less than 26 mE_h for net, interaction, and total energies) for the HF case.

Let us turn to the H_2O molecule. Here, the charge transfer as obtained by QTAM is quite large, each H atom having donated more than half an electron to the O atom (notice that there is a 0.7 m*e* integration excess, the largest on these calculations, accompanied by a −0.5 mE_h total energy error, also the largest in absolute value). Surprisingly, atomic dipoles are not very large, nor the Q_2^H quadrupole, but the $Q_2^O\ x^2 - y^2$ component is quite large due to the asymmetry of the molecule. The large charge transfer is responsible for the large deformation energies (about three times the CAS entry for N_2, for example). In the case of O, the electron increase induces a large energy lowering through ΔV_{en}^{OO}, but this is compensated by a large increase in the interelectronic repulsion and in the kinetic energy. As mentioned before, this behavior is consistent with the traditional ideas of covalent bonding, where there is a charge concentration in the bonding region, and is also displayed here by the O atom, which increases its total charge. However, the behavior is much different in the H atoms, which lose charge. For them, the kinetic energy decreases and the electron−nucleus attraction increases from the free atoms situation, owing to the decrease in the number of electrons, while the electron repulsion increases slightly from its separated atoms zero limit.

Regarding the interaction contributions, there is a clear distinction between the bonded, polar O−H interaction and the nonbonded, nonpolar H−H interaction. The first one displays a large and negative exchange-correlation contribution, as the previous ones considered, but here the classical interaction is also negative and even larger, in contrast with the results for H_2 and N_2. Clearly, this classical interaction arises from the charge transfer, with a leading point-charge term $V_Q^{OH} = -0.328\ E_h$. However, the long-range multipolar series does not converge, as shown by the $V_{C,lr}^{OH}$ unphysically large value: there is a multipolar overlap (MPOV) contribution,[33] since this is a covalent bond, which makes necessary the use of the proper short-range accounting in the r_{12}^{-1} multipolar expansion. The overall O−H interaction energy is thus negative and very large, as much as the one for the traditionally much stronger N−N bond. However, one must remember that, to get the strongly binding classical contribution, the atoms have to suffer a charge transfer and pay a price in an also high deformation energy, so that overall the N−N bond is indeed stronger even than the two O−H bonds together.

On the other hand, the nonbonded H−H interaction has a very different behavior. The exchange-correlation contribution is almost negligible, while the classical term is larger and positive. Once again, this is due to the charge transfer, with the point charge term ($V_Q^{HH} = 0.104\ E_h$) accounting for most of the interaction. Nevertheless, the short distance between the H atoms (1.51 Å) still induces a poor convergence of the long-range series, and $V_{C,lr}^{HH}$, although not divergent, is still twice the real value of the interatomic V_C^{HH}. Of course, nonbonded intramolecular interactions at larger distances will eventually display a pure long-range behavior. The sum of the positive classical term and the small negative exchange-correlation term gives a moderately positive H−H interaction, as one would expect.

The final balance of binding in the water molecule is then as follows. The charge transfer induces large and positive deformation energies (95 kcal/mol for O and 119 kcal/mol for H) and a repulsion between the positively charged H atoms (73 kcal/mol) that is mostly of classical origin. However, these positive contributions are more than compensated by the two very strong O−H negative interactions (−297 kcal/mol each): to a relatively large negative exchange-correlation contribution (−124 kcal/mol), proper of covalent systems, it adds a larger negative classical contribution (−173 kcal/mol), coming from the polar nature of the bond.

To increase further the polarity of the bond, we have considered LiF as a molecule widely accepted as ionic. Here, the QTAM charge transfer is almost the purely ionic one, with 0.93 *e*. The fluorine atom (almost a fluoride anion) polarizes toward the lithium atom (almost a Li^+ cation), which in turn has a very slight outward polarization. This is consistent with the classical images of polarizable anions and hard, spherical-like cations.[27] In addition, F^- is the least polarizable anion, hence its not so large dipole, and Li^+ is one of the least polarizable cations. There is also a small quadrupolar distortion along the bond.

The large charge-transfer puts so many electrons into the F atom that its deformation energy is *negative*: the electron−nucleus energy decrease more than compensates for the small kinetic energy increase and the electron−electron repulsion increase. However, E_{def}^F is small (−44 kcal/mol) when compared with the much larger positive deformation energy of the Li atom (123 kcal/mol). The atom losing charge has negative ΔT^{Li} and ΔV_{ee}^{LiLi} but a much larger and positive ΔV_{en}^{LiLi} value (in analogy with the H results in water). In contrast, both deformation energies are small if we take the separated ions as reference: the deformation of the fluoride will be $-0.019\ E_h$ (−12 kcal/mol), that for Li^+ will be 0.013 E_h (8 kcal/mol), and the binding energy will be $-0.3361\ E_h$ (−211 kcal/mol, which is larger than the nonionic reference

$E_{bind} = -128$ kcal/mol, since the separated ions are higher in energy than the separated atoms). In fact, their roles are inverted: F^- loses charge in entering the molecule, and so ΔT^{F^-} and $\Delta V_{ee}^{F^-F^-}$ are negative for it, with positive $\Delta V_{en}^{F^-F^-}$, while Li^+ gains charge upon formation of the molecule, and the increments are opposite. The usefulness of each of these two energy references will depend, of course, on the context, but the ionic image is more appealing, with smaller deformations that hint at a more faithful description of the binding when considering ions as the components to bind. Notice, however, that the interaction energy (−207 kcal/mol) does not depend on this choice, since it is computed for the atoms already within the molecule.

Regarding the partition of the interaction energy, the exchange-correlation contribution is very small (−29 kcal/mol), while the classical interaction is large and negative (−178 kcal/mol). This is the expected behavior in an ionic compound, and in fact the purely point-charge interaction, $V_Q^{LiF} = -0.2926\ E_h$ (−184 kcal/mol), is within 3% of the total classical value. In this case, this is indeed because of a correct long-range behavior, since $V_{C,lr}^{LiF}$ is also within 2% of V_C^{LiF}, and points out that both atoms are almost non-MPOV and also almost spherical, as seen by their small dipoles. Hence, the point charge term is a very good approximation for the classical interaction energy. Overall, the picture for this molecule is that of mostly spherical ions, with a large classical interaction very well approximated by the point charge term. The large energy penalty due to ionization (paid through the deformation energy if we use the neutral atom reference and through the already high reference in the ionic case) is still smaller than the interaction energy gain.

Finally, as an example of a van der Waals bonded system, we have performed a full CI calculation on He_2 and partitioned its energy. The FCI calculation, using a moderate-size correlation-consistent basis set, displays a very shallow minima, with a binding energy of 0.042 K (the FCI, complete basis set extrapolation in ref 30 is 11 K). This very low binding energy, 0.13 μE_h, is beyond the precision attainable with the numerical integration in a reasonable time. However, using the same numerical integration parameters as in the rest of the calculations (but taking much more time, given the number of active orbitals and determinants), we obtain an error in the total energy of 11 μE_h, the lowest in this series of compounds. This is an order of magnitude smaller than most of the partition components, particularly deformation and interaction, and hence we can reliably say that the low binding energy comes from the cancellation of these much larger components. This reliability has two causes: the density matrices are essentially non-MPOV (faster convergence), and the interatomic surface discontinuity lies in a low-ρ region.

He_2 is possibly the worst-case scenario for the cancellation of properties problem: the monatomic components are of the order of units of E_h, while the deformation is 4 orders of magnitude smaller; the interaction components are of the order of 0.7 E_h (even for this large distance, 2.97 Å, they still decrease as $1/r$), while the total interaction is 3 orders of magnitude smaller. Nevertheless, the ΔX^{He} intra-atomic components display the familiar pattern for homonuclear systems, being of the same order of magnitude as the deformation energy. In the case of the interaction, V_{cl}^{HeHe} is smaller than 1 μE_h, displaying a 6 orders of magnitude cancellation of the electrostatic interactions. This is not so surprising, if we look at the accuracy of the long-range multipolar approximation $V_{C,lr}^{HeHe}$, exactly matching the complete V_C^{HeHe} integral: both He atoms are non-MPOV and neutral, and hence the V_Q^{HeHe} point-charge contribution has to be zero, while the other multipolar components fall much faster with distance (the classical point dipole−point dipole interaction will be about 3 nE_h). On the other hand, the exchange-correlation component comprises most of the interaction energy in He_2. The image is, then, that of a very small polarization of the two atoms, having a negligible effect in the classical energy, with a small but non-negligible deformation energy; the binding component in the interaction is the exchange-correlation, as expected for a dispersion-bound system, although the binding energy value comes after the 3 orders of magnitude cancellation of these two (interaction and deformation) terms.

IV. Conclusions and Perspectives

An energy partition based on QTAM but using ideas from McWeeny's TES is proposed that splits the total energy exhaustively into atomic contributions, both intra- and interatomic. The partition recovers many chemical concepts and an atomistic intuition about binding, by recognizing the importance of the large contributions of the intra-atomic energy, which should show little variation from compound to compound for a given atom, and the smaller interatomic components that vary with the different types of interactions. In particular, this partition can fully account for the definition and maintained identity of functional groups and can treat on the same footing bonded and nonbonded interactions and even intermolecular ones. Being based on QTAM, its atoms have well-defined kinetic energies and are transferable while at the same time making the partition independent of any arbitrariness in the wave function calculation: everything (including the atomic basins) is derived from the first- and second-order density matrices, and hence it is truly a physical property of the system. It can be applied both to single-determinant (HF) and multideterminant approximate wave functions.

The partition acknowledges the existence of large cancellations to recover the binding energy, both in the intra-atomic and in the classical interatomic terms, coming from the essentially unchanged and neutral atoms that form the molecules in many cases of interest. The resulting terms are not so small when there is charge transfer, since varying the number of electrons undoubtedly changes the intra-atomic energies, while the electrostatic interaction between charged systems is quite large. Usually, the intra-atomic deformation terms are positive, and they are larger in absolute value for atoms with stronger and shorter bonds and whenever there is a large charge transfer. Homopolar-bonded and negatively charged atoms generally suffer a kinetic and electron repulsion energy increase which is partially compensated by a corresponding decrease in the nucleus−electron attraction energy, while the opposite happens with atoms losing

electron density. The interaction energy is usually larger than the deformation terms, and hence binding can be seen as the tradeoff of a small intra-atomic deformation energy increase by a much larger decrease due to the interatomic interaction energy. The interaction energies in bonded pairs are generally negative and large, while nonbonded interactions (either intramolecular or intermolecular) are much smaller and can often be approximated by point multipolar series, neglecting both short-range classical and exchange-correlation contributions.

Regarding the different patterns for different kinds of bonds, we have found the covalent homopolar bonds to have a small deformation energy, with a large and negative exchange-correlation interaction contribution as mainly responsible for binding. Their classical interaction is usually small and positive, as it should be for neutral, nonpenetrating charge systems; however, they display MPOV, and hence the point-multipole long-range contributions diverge and cannot be properly used. Ionic bonds have atoms with large deformation energies due to the charge transfer, but they are more than compensated by the large and negative classical interactions, where the point-charge term already accounts for 90% of the interaction energy. For them, the long-range multipolar approach is convergent and accurate to reproduce the classical energy, due to their small MPOV, and the exchange-correlation interaction is very small. The case of polar covalent bonds is intermediate: the deformation is dependent on the charge transfer extent, and both the covalent-like exchange-correlation and the ionic-like classical attraction terms are present and comparable in value. For them, the point-charge estimation gives the correct trend for the classical interaction, but the long-range multipolar series does not converge due to the covalent-like MPOV. Finally, van der Waals interactions are exceedingly weak when compared with bonded ones; in a dispersion-bound system like He_2, the classical interaction is negligible, and only exchange-correlation constitutes an attractive term.

In summary, we have presented a powerful tool to quantitatively analyze both the intra- and interatomic components of the molecular energy and relate them to the chemical concepts of bonding. Our results in selected molecules give plausible trends, which we have also seen in other preliminary calculations, making a consistent and appealing image of the physical atomistic components in the different chemical bonds. We plan to expand this study to other chemical systems of interest, on one hand, and to generalize it to different atomic partitions and compare their results, on the other.

Acknowledgment. Financial support from the MCyT, Project No. BQU2003-06553, is hereby acknowledged. We also want to thank Prof. L. Pueyo for his suggestions.

References

(1) Jensen, F. *Introduction to Computational Chemistry*; Wiley: Chichester, 1999.

(2) Ruedenberg, K. *Rev. Mod. Phys.* **1962**, *34*, 326.

(3) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325.

(4) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1558 and 1755.

(5) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W., Jr. *J. Chem. Phys.* **1984**, *80*, 4378.

(6) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(7) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules;* Oxford University Press: New York, 1989.

(8) Bader, R. F. W. *Atoms in Molecules*; Oxford University Press: Oxford, 1990.

(9) Glendening, E. D.; Streitwieser, A. *J. Chem. Phys.* **1994**, *100*, 2900−2909.

(10) Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530.

(11) Pophristic, V.; Goodman, L. *Nature* **2001**, *431*, 565.

(12) Bickelhaupt, F. M.; Baerends, E. J. *Angew. Chem., Int. Ed.* **2003**, *115*, 4315.

(13) Frenking, G.; et al. *Coord. Chem. Rev.* **2003**, *238*−*239*, 55.

(14) Salvador, P.; Duran, M.; Mayer, I. *J. Chem. Phys.* **2001**, *115*, 1153−57.

(15) Salvador, P.; Mayer, I. *J. Chem. Phys.* **2004**, *120*, 5046.

(16) Popelier, P. L. A.; Kosov, D. S. *J. Chem. Phys.* **2001**, *114*, 6539−47.

(17) Martín Pendás, A.; Blanco, M. A.; Francisco, E. *J. Chem. Phys.* **2004**, *120*, 4581−92.

(18) Martín Pendás, A.; Francisco, E.; Blanco, M. A. *J. Comput. Chem.* **2005**, *26*, 344−351.

(19) McWeeny, R. *Methods of Molecular Quantum Mechanics*, 2nd ed.; Academic Press: London, 1992.

(20) Bader, R. F. W.; Preston, H. J. T. *Int. J. Quantum Chem.* **1969**, *3*, 327.

(21) Bader, R. F. W.; Beddall, P. M. *J. Chem. Phys.* **1972**, *56*, 3320.

(22) Srebrenik, S.; Bader, R. F. W. *J. Chem. Phys.* **1975**, *63*, 3945.

(23) Li, L.; Parr, R. G. *J. Chem. Phys.* **1986**, *84*, 1704.

(24) Born, M.; Huang, K. *Dynamical Theory of Crystal Lattices*; Oxford University Press: Oxford, 1954.

(25) Hirschfelder, J. O.; Curtis, C. F.; Bird, R. B. *Molecular Theory of Gases and Liquids*; Wiley: New York, 1954.

(26) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.

(27) Luaña, V.; Pueyo, L. *Phys. Rev. B* **1990**, *41*, 3800−3814.

(28) Cohen, N.; Benson, S. W. *Chem. Rev.* **1993**, *93*, 2419.

(29) Schmidt, M. W.; et al. *J. Comput. Chem.* **1993**, *14*, 1347.

(30) van Mourik, T.; Dunning, T. H., Jr. *J. Chem. Phys.* **1999**, *111*, 9248.

(31) Costales, A.; Blanco, M. A.; Mori-Sánchez, P.; Luaña, V.; Martín Pendás, A. *J. Phys. Chem. A* **2004**, *108*, 2794−801.

(32) Slater, J. C. *Quantum Theory of Molecules and Solids*; McGraw-Hill: New York, 1963; Vol. 1.

(33) As discussed in ref 17, an overlap contribution in the sense of the multipolar expansion occurs when the sum of the maximum radii of both atomic basins surpasses the interatomic distance, $R_A + R_B > r_{AB}$, making invalid the long-range multipolar approximation; see Figure 2 of that paper.

CT0501093

JCTC Journal of Chemical Theory and Computation

A Theoretical Investigation of the Geometries and Binding Energies of Molecular Tweezer and Clip Host−Guest Systems

Maja Parac,[†] Mihajlo Etinski,[‡] Miljenko Peric,[‡] and Stefan Grimme*[,†]

Organisch-Chemisches Institut der Universität Münster, Corrensstrasse 40, D-48149 Münster, Germany, and Faculty of Physical Chemistry, University of Belgrade, Studentski Trg 12-16, 11000 Belgrade, Serbia

Received May 3, 2005

Abstract: A quantum chemical study of host−guest systems with dimethylene-bridged clips and tetramethylene-bridged tweezers as host molecules and six different aliphatic and aromatic substrates as guests is presented. The geometries and binding energies of the complexes are investigated using the recently developed density functional theory with empirical corrections for dispersion interactions (DFT-D) in combination with the BLYP functional and basis sets of TZVP quality. It is found that the DFT-D method provides accurate geometries for the host−guest complexes that compare very favorably to experimental X-ray data. Without the dispersion correction, all host−guest complexes are unbound at the pure DFT level. Calculations of the clip complexes show that the DFT-D binding energies of the guests agree well with those from a more sophisticated SCS-MP2/aug-cc-pVTZ treatment. By a partitioning of the host into molecular fragments it is shown that the binding energy is clearly dominated by the aromatic units of the clip. An energy decomposition analysis of the interaction energies of some tweezer complexes revealed the decisive role of the electrostatic and dispersion contributions for relative stabilities. The calculations on the tweezer complexes show that the benzene spaced tweezer is a better receptor for aliphatic substrates than its naphthalene analogue that has a better topology for the binding of aromatic substrates. The tweezer with a OAc substituent in the central spacer unit is found to favor complex formation with both aliphatic and aromatic substrates. The theoretical results are qualitatively in very good agreement with previous experimental findings although direct comparison with experimental binding energies which include solvent effects is not possible. The good results obtained with the DFT-D-BLYP method suggest this approach as a standard tool in supramolecular chemistry and as the method of choice for theoretical structure determinations of large complexes where both electrostatic and dispersive interactions are crucial.

1. Introduction

The noncovalent interactions between atoms and molecules play an important role in structural biology and supramolecular chemistry. They control the structures of proteins and DNA, host−guest systems, enzyme−substrate binding, antigen−antibody recognition, or the orientation of molecules on surfaces or in molecular films.[1,2] Because of their ubiquitous role in diverse fields the investigation and understanding of these weak interactions has become one of the major goals of modern chemistry. Besides the relatively strong hydrogen bonding,[3−5] ion pairing,[6−8] and

* Corresponding author phone: (+49)-251-8336512; e-mail: grimmes@uni-muenster.de.

[†] Organisch-Chemisches Institut der Universität Münster.

[‡] University of Belgrade.

10.1021/ct050122n CCC: $30.25
Published on Web 09/20/2005

cation-π interactions,[8] the nonspecific van der Waals (vdW, dispersive) interactions seem much more important than first expected.[9,10]

Molecular tweezers and molecular clips as synthesized and investigated in the group of Klärner[11−18] are simple hydrocarbons containing only nonconjugated benzene and/or naphthalene rings arranged in a convex-concave molecular topology. Experimentally, it was observed that the molecular tweezers and clips serve as selective receptors for electron deficient aromatic and aliphatic substrates.[11,13−18] Computational methods at different levels of theory (e.g. AM1, HF/6-31G*, and standard density functional theory, DFT) have been used to generate electrostatic potential surfaces on the convex and concave side of the hosts.[12] The results show considerably larger negative potential within the tweezers and clips interior than at the exterior which could explain binding of electron-deficient aromatic and aliphatic substrates into the host cavity.

Despite the growing interest in host−guest complexation processes relatively few quantum chemical computations on tweezer and clip systems have been reported so far.[12,19−21] Studies that employ reliable quantum chemical methods and perform complete geometry optimizations of the complexes are still lacking, and thus, important structural information is often missing. From the computational point of view, ab initio methods that accurately account for the important dispersive interactions [e.g. CCSD(T)] suffer from very demanding computation times even for small- to medium-sized systems. On the other hand, the cheaper second-order Møller−Plesset perturbation theory (MP2)[22,23] systematically overestimates the binding energies for dispersive $\pi-\pi$ interactions (see, e.g. refs 24−26 and references therein). Another drawback for the application of correlated ab initio methods to weakly bound systems is the usually large basis set superposition error (BSSE). When dispersion is dominant for the interaction, the BSSE is often several times larger than with mean-field methods such as Hartree−Fock or DFT. Thus, additional calculations of the so-called counterpoise correction (CP[27]) are mandatory.

Recently one of us introduced an empirical correction scheme for density functional theory calculations (termed DFT-D[28]) that accounts for vdW interactions by pairwise additive C_6/R^6 potentials. Calculations with this very efficient method have shown promising results for a wide variety of weakly bonded systems[28−30] (for related methods see e.g. refs 31−33, for other recent attempts to the DFT/vdW problem see refs 34−37). Standard DFT with current functionals of GGA of hybrid-type must strictly be avoided for vdW complexes because it fails completely not providing any binding at all (this also holds for recently proposed functionals such as e.g. X3LYP, see ref 38).

In the present study we have investigated a series of typical host−guest complexes with the DFT-D method. Our systems consist of dimethylene-bridged clips and tetramethylene-bridged tweezers as host molecules and six different aliphatic and aromatic substrates as guest molecules. The "real" systems investigated experimentally are considered without any modification. The DFT-D binding energies and structures as obtained from full geometry optimizations will be presented. Although we are aware that solvent effects are important to interpret and understand the experimental condensed phase data (for a force-field approach to this problem see ref 19), we think that it is also important to investigate the intrinsic properties of the systems in the gas phase. Note, that the present DFT-D approach represents the first attempt to model both the electrostatic and dispersive interactions in such large complexes at a similar level of accuracy (for a CCSD(T) study of this aspect for smaller benzene model complexes see ref 39). Force-field methods that are usually applied for such complexes suffer from an oversimplified description (often of point-charge type) of the electrostatic interactions.

After consideration of some technical details in section 2, the accuracy of the DFT-D method for binding energies and geometries is considered in section 3.1. In the following sections (3.2 and 3.3) the complexes are grouped according to their host structure. The dimethylene-bridged clips (named CLIP1 and CLIP2 in Figure 1) and tetramethylene-bridged tweezers (named TWEEZER1-TWEEZER3 in Figure 1) as the host molecules and the following substrates as guest molecules are considered: cyanomethane (**1**), dicyanomethane (**2**), benzene (**3**), 1,4-dicyanobenzene (**4**), 1,4-benzenecarboaldehyde (**5**), and quinone (**6**). The detailed analysis of the host−guest interactions in the framework of an energy decomposition scheme[40,41] will be discussed in section 3.4.

2. Technical Details

All density functional theory (DFT)[42,43] and second-order Møller−Plesset perturbation theory (MP2)[22,23] calculations have been performed with the TURBOMOLE package of programs.[44] The optimum geometries and binding energies were obtained using the DFT-D method[28] together with the B-LYP[45,46] functional. Additionally, the spin-component-scaled MP2 approach (termed SCS-MP2[26]) which improves in general the accuracy of MP2 has been employed. In all DFT-D, MP2, and SCS-MP2 calculations the resolution of identity (RI) approximation for the two-electron integrals[47,48] is used. The RI auxiliary basis sets[49,50] are taken from the TURBOMOLE library.[51]

All geometries have been fully optimized without any symmetry restrictions. Gaussian valence-triple-ζ AO basis set augmented with polarization functions on all atoms (TZVP[52]) have been used in the DFT-D calculations. For all MP2 and SCS-MP2 calculations Dunnings aug-cc-pVTZ basis set[53] has been employed. Diffuse basis functions that are necessary in MP2 calculations to accurately describe the fragment polarizabilities do not necessarily improve the description of weak complexes when DFT approaches are applied.[28] As long as larger monomers (with more than 4−6 atoms) are considered and AO basis sets of at least TZVP quality are used, BSSE effects are quite small (<10% of ΔE) at the DFT level, and the laborious calculation of the counterpoise-correction (CP)[27] can be avoided (these small BSSE effects have been absorbed into the dispersion potential, for details see ref 28). The CP correction is mandatory when correlated wave function methods are used, and thus all MP2 and SCS-MP2 calculations have been CP corrected and only these values are reported.

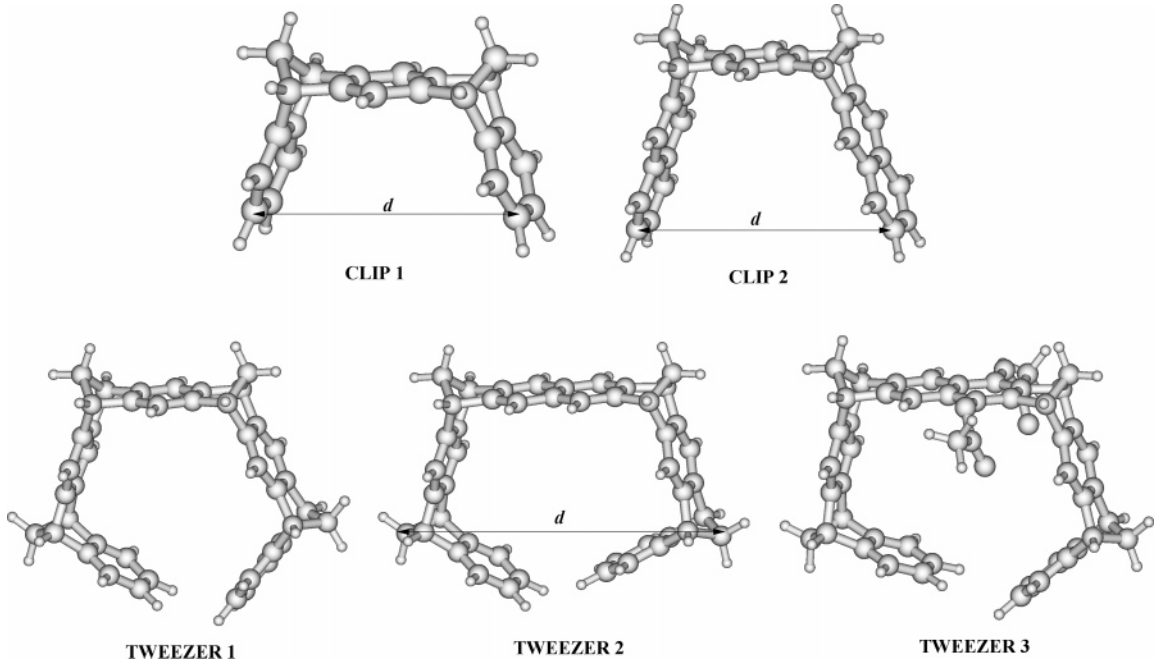


Figure 1. Optimized structures (DFT-D-BLYP) of the host molecules CLIP1, CLIP2, TWEEZER1, TWEEZER2, and TWEEZER3.

Table 1. Comparison of Selected Calculated and Experimental Geometric Parameters (*r* in Å, Angles in Degree) for the Tetramethylene-Bridged Tweezer as Host and 1,4-Dicyanobenzene as Guest (4@TWEEZER2)

geometric parameter	DFT-D	exp
r C2−C6	3.59	3.61
r C3−C7	3.49	3.59
r C9−C10	12.50	12.62
r C11−C12	3.91	4.16
angle C2−C6−C5	86.7	82.8
dihedral angle C1−C4−C5−C8	93.4	95.7

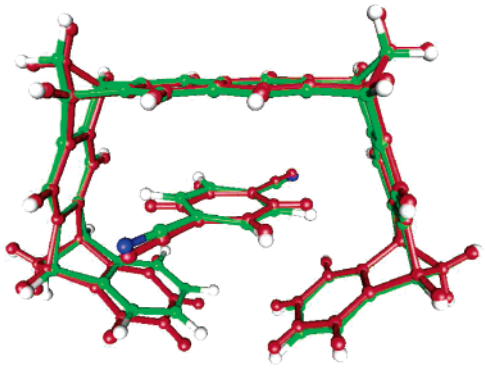


Figure 2. Comparison of DFT-D-BLYP (green) and experimental (red) structures for the host−guest complex 4@TWEEZER2.

3. Results and Discussion

3.1. Accuracy of the DFT-D Method. In this section we compare the DFT-D-BLYP binding energies for the smallest system to those from MP2 and SCS-MP2 calculations. A comparison of calculated and experimental X-ray[11] parameters for some important geometrical variables of the tetramethylene-bridged TWEEZER2 as host and 1,4-dicyanobenzene as guest molecule are also presented (see Table 1, the atom numbering scheme is given in Figure 4). A graphical overlay of calculated and experimental structures is given in Figure 2. Both comparisons should provide some impression what accuracy can be expected from the DFT-D-BLYP method for the most important properties.

In the 4@TWEEZER2 complex, the guest molecule is placed almost parallel to the central naphthalene spacer unit with calculated distances of 3.59 Å (C2−C6) and 3.49 Å (C3−C7), respectively. This is in good agreement with the experimentally found values of 3.61 and 3.59 Å, respectively. Note, that for typical van der Waals complexes differences between theory and experiment for intermolecular distances are often about 0.1−0.2 Å even for very sophisticated theoretical treatments due to the flatness of the corresponding potentials.[28] The distance C9−C10 describes the width of the host cavity and is calculated to be 12.50 Å which is only 0.12 Å shorter than found experimentally (12.62 Å). The distance C11−C12 is the shortest nonbonded contact within the host cavity, and the calculated value of 3.91 Å agrees again well with the experimental value of 4.16 Å. The position of the guest molecule within the cavity can be described by the angle C2−C6−C5 and the dihedral angle C1−C4−C5−C8. The calculated and experimental values differ by only about 4 and 1.3 degrees, respectively. This can be considered as very satisfactory especially if one keeps in mind that the potential energy surface of the host−guest interactions are very flat (which, unfortunately, results in many necessary geometry optimization cycles).

In a successful theoretical investigation[20,21] of the NMR chemical shifts of the same host−guest complex, the structure has been optimized at the Hartree−Fock level. However, in this work it was necessary to employ constraints in the geometry optimization in order to preserve the shape of the complex as observed in the X-ray structure. This underlines the importance of the dispersive interactions that are not accounted for in standard HF or DFT calculations. With DFT-D, full geometry optimizations resulting in force-free

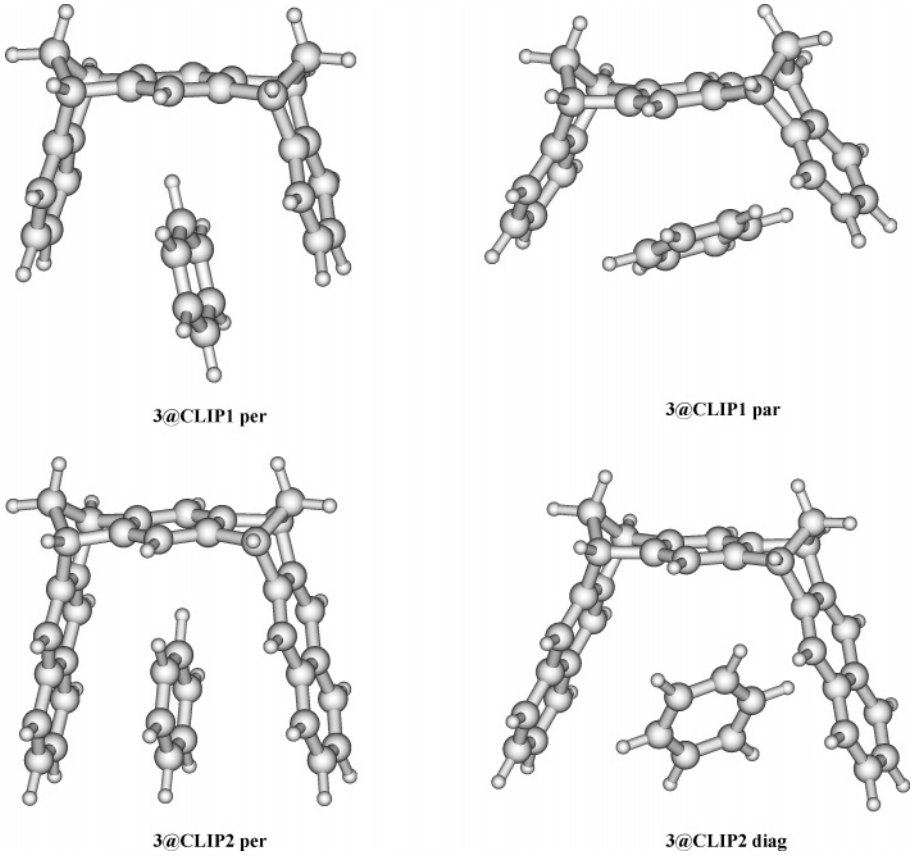


Figure 3. Optimized structures (DFT-D-BLYP) of the host−guest complexes 3@CLIP1PAR, 3@CLIP1PER, 3@CLIP2DIAG, and 3@CLIP2PER.

structures are possible that compare favorably with experimental data.

The 3@CLIP1PER complex is relatively small compared to the other systems that will be discussed further, and, therefore, computationally more demanding treatments such as MP2 and SCS-MP2 can be performed. The MP2 and SCS-MP2 binding energies have been obtained by rigid monomer calculations on the DFT-D-BLYP optimized structures. Opposed to a previous MP2 treatment of 4@TWEEZER1,[21] we employed the large aug-cc-pVTZ AO basis in these benchmark calculations that should provide a quite realistic account of the dispersive host−guest interactions. Note that as mentioned above, the BSSE is much larger for the MP2 type methods, and thus these data include the counterpoise correction. The MP2 and SCS-MP2 binding energies in the 3@CLIP1PER complex are computed to be −12.1 kcal/mol and −7.8 kcal/mol, respectively. As found previously[26,30,54] the SCS-MP2 results for systems with $\pi-\pi$ interactions (e.g. benzene dimers) are much better and more consistent compared to standard MP2 which tends to overbind considerably. In our calculations, the SCS-MP2 values compare favorably with DFT-D results with differences of about 1.2 kcal/mol. This also holds for the 3@CLIP1PER complex where DFT-D and SCS-MP2 agree to within 1.8 kcal/mol, while MP2 is off by almost a factor of 2. In conclusion one can be quite certain that the DFT-D absolute interaction energies are accurate to about 10−20% of the true values (slightly underbinding), while the errors on a relative scale (i.e. when comparing different guests and/or hosts) may be even smaller.

3.2. Clip Complexes. First, complexes of benzene with the dimethylene-bridged clip (named CLIP1 in Figure 1) are discussed.

The geometry optimizations lead to two different structures 3@CLIP1PER and 3@CLIP1PAR (see Figure 3). In the 3@CLIP1PER complex the guest molecule benzene is oriented parallel to the sidewalls and almost perpendicular to the central benzene unit of the host molecule. The DFT-D-BLYP binding energy of this complex is computed to be −6.6 kcal/mol (see Table 2).

The binding energy of the 3@CLIP1PAR complex, where the substrate molecular plane is nearly parallel to the central unit, is computed to be lower by 1.4 kcal/mol. This relatively small energy difference between the 3@CLIP1PER and 3@CLIP1PAR complexes indicates that the associated potential energy surface (PES) is quite flat. In the complexes 3@CLIP1PER and 3@CLIP1PAR the distances between the benzene sidewalls (named *d* in Figure 1) are computed to be 7.43 and 9.35 Å, respectively, while without any substrates bound it is computed to be 7.98 Å. The increase in steric strain resulting from an expansion of the cavity in 3@CLIP1PAR (that has more CH-πs than $\pi-\pi$ interactions) is certainly the main reason that the complex 3@CLIP1PER with a parallel arrangement of the substrate is more stable.

Furthermore, complexes of benzene with dimethylene-bridged clips including naphthalene sidewalls (CLIP2 in Figure 1) are discussed. Two different structures are obtained: 3@CLIP2PER, with the substrate molecular plane oriented parallel to the receptor sidewalls, and 3@CLIP2DIAG, where the substrate molecule is placed parallel to the

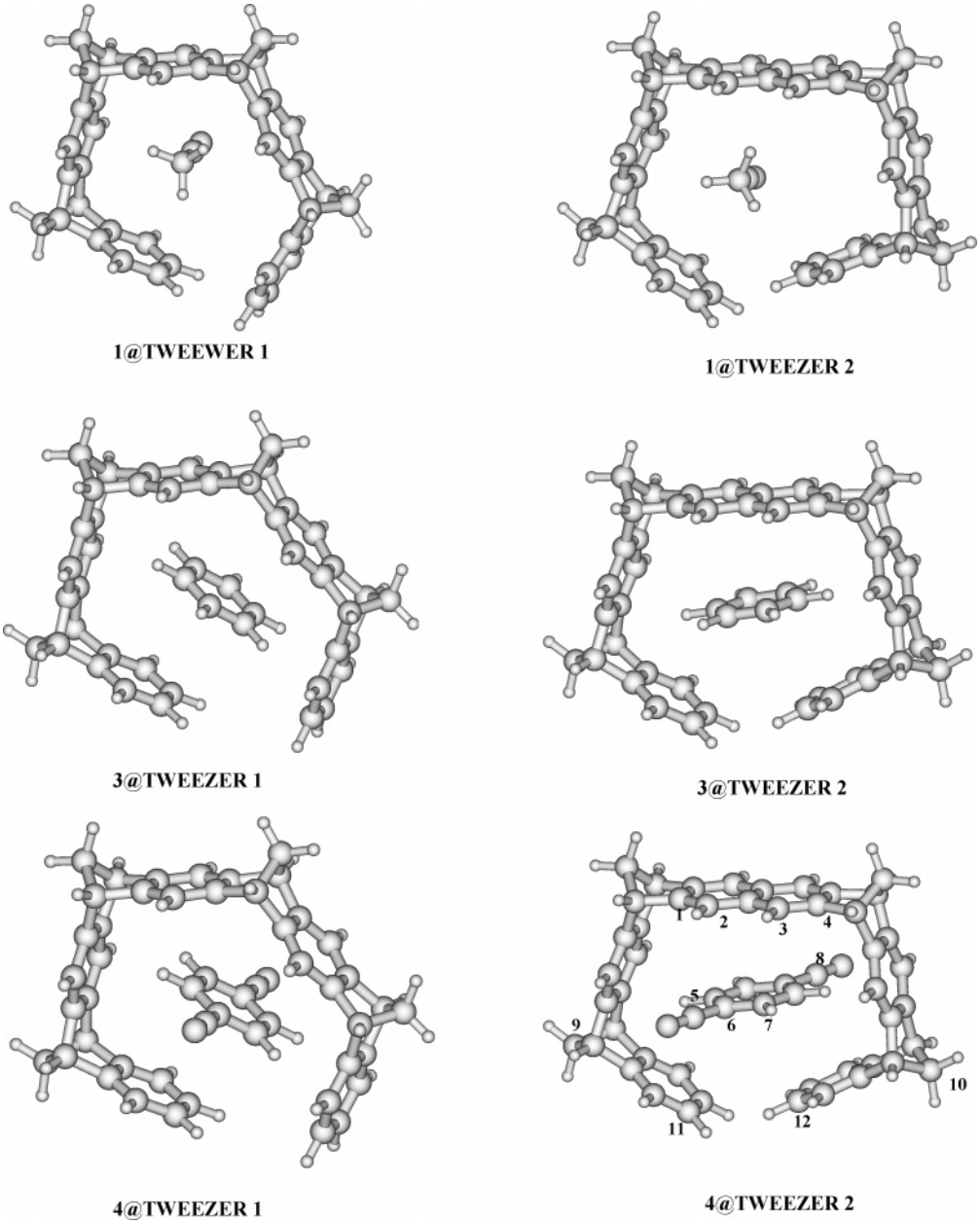


Figure 4. Optimized structures (DFT-D-BLYP) of the host−guest complexes 1@TWEEZER1, 1@TWEEZER2, 3@TWEEZER1, 3@TWEEZER2, 4@TWEEZER1, and 4@TWEEZER2.

Table 2. Calculated Binding Energies (in kcal/mol) for the Investigated Clip Complexes[b]

	ΔE		
molecule	DFT-D-BLYP	MP2[a]	SCS-MP2[a]
3@CLIP1PER	−6.6 (7.0)	−12.1	−7.8
3@CLIP1PAR	−5.2 (7.6)		
3@CLIP2PER	−9.5 (11.2)	−17.8	−11.3
3@CLIP2DIAG	−7.9 (6.4)		

[a] aug-cc-pVTZ AO basis. [b] Values in parentheses refer to a pure DFT-BLYP treatment, i.e. without the dispersion correction.

methylene bridged groups and diagonal to the sidewalls and central unit (see Figure 3). A structure with benzene orientated parallel to the central benzene unit similar to those obtained with CLIP1 could not be found and instead converged to 3@CLIP2DIAG structure. The results for the binding energies in Table 2 show that the 3@CLIP2PER complex is more stable than 3@CLIP2DIAG, and the binding energies are −9.5 kcal/mol and −7.9 kcal/mol, respectively. The characteristic distances *d* are computed to be 7.62 Å for 3@CLIP2PER, 10.24 Å for 3@CLIP2DIAG, and 9.67 Å for the host molecule without the substrate.

Inspection of Table 2 also reveals that CLIP2 with naphthalene sidewalls forms more stable complexes with benzene than CLIP1 with benzene sidewalls. This can be explained by more favorable van der Waals contacts with the naphthalene sidewalls. Similar effects for hosts with naphthalene and anthracene sidewalls are discussed in ref 16.

3.3. Tweezer Complexes. In the following we discuss the structures and binding energies of the complexes with the tetramethylene-bridged molecular tweezer (TWEEZER1 in Figure 1) as receptor. It was found that the bond angle distortions in molecular tweezer systems require little energy

Table 3. DFT-D-BLYP Binding Energies (in kcal/mol) of the Investigated Tweezer Complexes[a]

	ΔE		
guest	@TWEEZER1	@TWEEZER2	@TWEEZER3
1	−13.7 (1.5)	−12.0 (1.6)	−15.0 (0.4)
2	−19.0 (1.8)	−16.6 (1.1)	−20.4 (0.2)
3	−9.5 (12.2)	−12.7 (13.5)	−15.2 (13.5)
4	−18.3 (8.7)	−26.7 (8.2)	−29.3 (8.0)
5	−20.2 (11.4)	−23.5 (11.6)	−27.7 (10.3)
6	−17.2 (9.6)	−22.2 (8.3)	−23.7 (9.6)

[a] Values in parentheses refer to a pure DFT-BLYP treatment, i.e. without the dispersion correction.

meaning that these systems are flexible enough to allow the receptor sidewalls to adjust their geometry specifically for different substrate binding.[16] This should be investigated in this study by consideration of the six substrates.

The binding energies (see Table 3) for small aliphatic substrates such as acetonitrile (**1**) in the 1@TWEEZER1 complex and malononitrile (**2**) in the 2@TWEEZER1 complex are already for these small guests quite large, i.e., −13.7 kcal/mol and −19.0 kcal/mol, respectively. For the complexes with the naphthalene-spaced analogue (TWEEZER2) with the same aliphatic substrates (1@TWEEZER2 and 2@TWEEZER2) the corresponding values are slightly smaller, i.e., −12.0 kcal/mol and −16.6 kcal/mol, respectively (for the optimized structures see Figure 4).

In the experimental study[11,16] it was shown that aliphatic substrates will be bound only inside the benzene-spaced host. Evidently, our calculations agree with this observation and show that the benzene-spaced tweezer is indeed a better receptor for aliphatic substrates than its naphthalene-spaced analogue. The differences $\Delta\Delta E$ between TWEEZER1 and TWEEZER2 are 1.7−2.4 kcal/mol which is large enough to explain the observed selectivity in solution (in the gas phase, of course, most reasonably sized molecules will have negative binding energies, and so the absolute values cannot be directly compared with experimental findings).

The binding energies of the aromatic substrates **3**−**6** with TWEEZER1 are calculated to be −9.5 kcal/mol, −18.3 kcal/mol, −20.2 kcal/mol, and −17.2 kcal/mol, respectively. This underlines the importance of ES interactions because benzene as a substrate is bound by about 10 kcal/mol less strongly than the other guests. Inspection of the complex structures shows that the distances between the methylene groups of the sidewalls in the receptor molecule (named *d* in Figure 1) are expanded from 11.47 Å in the empty TWEEZER1 to 12.1 Å, 12.0, 11.9, and 12.4 Å in the host−guest complexes, respectively. According to these results the binding of aromatic substrates in the benzene-spaced tweezer requires a substantial distortion of the receptor geometry which is only partly compensated by the attractive substrate−receptor interactions. Moreover, we find that all aromatic substrates bound into TWEEZER1 have the molecular plane aligned parallel to the sidewalls of the complex, which is not the case for the naphthalene-spaced TWEEZER2 as the host molecule.

As mentioned above, the benzene-spaced TWEEZER1 is a better receptor for aliphatic substrates than its naphthalene-spaced analogue TWEEZER2, and experimentally it was found that the TWEEZER2 has an almost ideal topology for the binding of aromatic derivatives.[14,16] The binding energies for the aromatic substrates **3**−**6** are calculated to be −12.7 kcal/mol, −26.7 kcal/mol, −23.5 kcal/mol, and −22.2 kcal/mol, respectively. These values are 3.1−8.1 kcal/mol lower than for TWEEZER1. Inspection of the complex geometries shows that the distances between the sidewalls in the aromatic TWEEZER2 complexes (named *d* in the Figure 1) has to expand only very slightly, i.e., from 12.4 to 12.6 Å, 12.5, 12.4, and 12.5 Å, respectively. This smaller increase in steric strain compared to the analogue TWEEZER1 complexes where the distance between sidewalls change up to 0.9 Å clearly leads to an additional stabilization of the TWEEZER2 complexes. Opposed to the TWEEZER1 complexes, in all investigated complexes of TWEEZER2 with aromatic substrates the plane of the guest molecule is aligned parallel to the central spacer unit. The optimized structures of the 1@TWEEZER1, 1@TWEEZER2, 3@TWEEZER1, 3@TWEEZER2, 4@TWEEZER1, and 4@TWEEZER2 complexes are shown in Figure 4. Previous MP2/SVP calculations of 4@TWEEZER1[21] reported a binding energy of −38.8 kcal/mol. However, this study suffers from the small AO basis set employed, and the important BSSE effects that have not been considered seem to be responsible for the large difference to our DFT-D result (−18.3 kcal/mol).

We extend our study to the molecular tweezer (named TWEEZER3 in Figure 1) substituted with OAc groups in the central space unit. Our discussion focuses on the effect of the substituent on the stability of the complexes with aliphatic and aromatic substrates. Although, the OAc groups can have several different conformations in TWEEZER3, we only investigated the conformation in which both OAc groups point toward the molecular cavity. This conformation is expected to allow more favorable ES interactions between the negatively charged carbonyl oxygen atoms and the electron deficient substrate molecules.

The binding energies of the aliphatic substrates such as acetonitrile (**1**) in 1@TWEEZER3 and malononitrile (**2**) in the 2@TWEEZER3 complex are calculated to be −15.0 kcal/mol and −20.4 kcal/mol, respectively. The comparison of these binding energies with the corresponding values of the parent complexes 1@TWEEZER2 and 2@TWEEZER2 shows that the OAc substituent favors complex formation with aliphatic substrates. The complex formation with the aromatic substrates is also favored due to presence of the OAc groups. The calculated binding energies for **3**−**6** are −15.2 kcal/mol, −29.3 kcal/mol, −27.7 kcal/mol, and −23.7 kcal/mol, respectively, and are in the range of 1.4−4.2 kcal/mol more negative than for TWEEZER2. This is in line with the results of recent experimental work[18] where the effect of the several substituents in the molecular tweezers on the complex formation was examined.

3.4. Analysis of the Host−Guest Interactions. To obtain more detailed information about which parts of a molecular host system are involved in binding a substrate, the host in the 3@CLIP1PER complex is portioned into three different parts: the central benzene unit and the two benzene rings forming the sidewalls. The relative orientation of the guest

Table 4. Energy Decomposition Analysis for the Binding Energies (in kcal/mol) of Some of the Investigated Tweezer Complexes

	ΔE_{Pauli}	ΔE_{ES}	ΔE_{POCT}	ΔE_{DFT}	ΔE_{disp}	ΔE_{DFT-D}
1@TWEEZER1	13.5	−9.0	−4.1	0.4	−14.5	−14.1
1@TWEEZER2	11.5	−7.7	−3.5	0.2	−12.6	−12.4
1@TWEEZER3	16.2	−12.0	−5.2	−1.0	−14.9	−16.0
3@TWEEZER1	25.0	−8.0	−5.0	12.0	−24.0	−12.0
3@TWEEZER2	26.8	−8.2	−5.3	13.3	−26.3	−13.0
3@TWEEZER3	29.0	−10.3	−6.0	12.7	−28.9	−16.2

with respect to the host fragments is kept constant and dangling bonds are saturated with hydrogens. The interaction energies with the substrate molecule are computed to be −2.1 kcal/mol, −2.1 kcal/mol, and −1.9 kcal/mol for the three parts, respectively. This sums up to a total binding energy of −6.1 kcal/mol which is only slightly less than from the full calculation that yielded −6.6 kcal/mol. This indicates that the interactions influencing the binding energies are clearly dominated by the aromatic units of the clip and that the bridging units only place the aromatic rings into some more or less fixed position with respect to the guest. In the same spirit it was shown that individual aromatic units dominate the NMR chemical shifts of the 1,4-dicyanobenzene guest in the tetramethylene-bridged tweezer host.[20,21]

To get even more insight into the host−guest interactions in the tweezers, an energy decomposition analysis (EDA) of the binding energies has been performed for the tweezer complexes. The EDA has proven to give detailed information about the nature of chemical bonding[41] as well as for the interactions in hydrogen bonded systems.[55] The method has been described in detail before and is only briefly described here. The formation of bonding between two fragments (guest and host in our case) is divided into three physically plausible steps. In the first step, the fragment electronic densities (in the frozen geometry of the supermolecule) are superimposed which yields the quasiclassical electrostatic interaction energy ΔE_{ES}. Renormalization and orthogonalization of the product of monomer wave functions yield a repulsive energy term that is usually called Pauli (exchange) repulsion (ΔE_{Pauli}). In the final step, the molecular orbitals are allowed to relax to their final form which yields the (usually stabilizing) polarization, orbital and charge-transfer interaction energy ΔE_{POCT}. The sum $\Delta E = \Delta E_{ES} + \Delta E_{Pauli} + \Delta E_{POCT} + \Delta E_{disp}$, that also includes the dispersion energy term from the DFT-D approach, differs from the true interaction energy by the energy necessary to bring the optimum monomer geometries into the form they have in the supermolecule (ΔE_{prep}). The latter term is very small (<1.5 kcal/mol) in our case and not discussed further. Preliminary investigations of the EDA/DFT-D scheme for weakly bonded complexes have shown that the individual terms compare very well to those from a more sophisticated SAPT analysis.[56,57]

As can be seen from Table 4, the guests acetonitrile (1) and benzene (3) behave very differently, while the hosts TWEEZER1-TWEEZER3 are more similar. With acetonitrile, the uncorrected DFT interaction energies are close to zero meaning that repulsive Pauli and attractive electrostatic/CT/orbital interactions almost cancel. Thus, the total binding energy (and also the differences between the three hosts) are entirely given by the dispersion correction. For benzene as a guest the situation is completely different: here, the repulsive Pauli contribution is about a factor of 2 larger and overcompensates the other terms that are similar to the guest (1). However, in this case also the dispersion correction is much larger leading overall to very similar interaction energies for acetonitrile and benzene. In conclusion one can thus say that the binding of acetonitrile is dominated by the electrostatics while it is mostly dispersion for benzene. Turning now to the different hosts it is clearly seen that all individual terms are (absolutely) larger for the TWEEZER3 with OAc groups which leads in summary to a lower total binding energy. This can be explained by considering ΔE_{ES} as the driving force that yields short intermolecular contacts and thus larger (absolute) Pauli and dispersion terms for TWEEZER3 compared to TWEEZER1 and TWEEZER2 (similar arguments have been put forward to deconvolute the role of potential and kinetic energy in the formation of a chemical bond, see e.g. ref 58). Note, however, that the ordering of the binding energies for TWEEZER1-TWEEZER3 (3 > 1 > 2 and 3 > 2 > 1 for acetonitrile and benzene, respectively) is not correctly described at the pure DFT level and entirely dominated by the small differences in the vdW terms.

4. Conclusion

The recently developed DFT-D method with empirical corrections for long-range dispersion effects has been used to predict the structures and binding energies of host−guest systems consisting of dimethylene-bridged clips and tetramethylene-bridged tweezers as receptors and six aliphatic and aromatic guest molecules. On the basis of the presented results a number of useful conclusions considering the strength of $\pi-\pi$ and π-H interactions in the host−guest complexes with molecular tweezers and clips and the applicability of the DFT-D-BLYP method can be drawn. This study has shown that the DFT-D-BLYP method provides accurate geometries of the investigated complexes compared to the experimental X-ray data. It can therefore serve as a reliable method for fully geometry optimization of the large host−guest complexes.

For two structures of the 3@CLIP2 complex we have demonstrated that the DFT-D method provides accurate binding energies compared to computationally more demanding SCS-MP2 calculations. By an energy partitioning into different parts of the host we have shown that the binding in the 3@CLIP2 complex is clearly dominated by the aromatic units of the clip and that the bridging units only place the aromatic rings into the right position with respect to the guest. The calculations for CLIP2 with naphthalene sidewalls revealed that more stable complexes with benzene are formed compared to CLIP1 which can be explained by better van der Waals contacts in CLIP2.

The calculations on the tweezer complexes have shown that the benzene-spaced TWEEZER1 is a better receptor for aliphatic substrates than its naphthalene-spaced analogue TWEEZER2. On the other hand, TWEEZER2 has an almost

ideal topology for the binding of aromatic substrates. Moreover, our calculations have shown that all aromatic substrates bound into TWEEZER2 have the molecular plane aligned parallel to the central spacer unit, whereas in the TWEEZER1 they are aligned parallel to the molecular sidewalls. The calculations on TWEEZER3 with a OAc substituent in the central spacer demonstrate that the polar OAc group favors complex formation with aliphatic and aromatic substrates. The conclusions discussed above are in very good qualitative agreement with previous experimental findings. We thus finally conclude that the DFT-D method represents an important tool for large scale applications in supramolecular chemistry where the complex geometries are not known and where both electrostatic and dispersive interactions are important.

Acknowledgment. M. Parac thanks M. Piacenza for helpful discussions. This work was supported by the International Graduate School for Chemistry in Münster by a scholarship to M. Etinski.

References

(1) Lehn, J. M. *Supramolecular Chemistry. Concepts and Perspectives*; VCH: Weinheim, 1995.

(2) Atwood, J. L.; Davies, J. E. D.; MacNicol, D. D.; Vögtle, F.; Suslick, K. S. *Comprehensive Supramolecular Chemistry*; Elsevier: Oxford, 1996.

(3) Jeffrey, G. A. *An Introduction of Hydrogen Bonding*; Oxford University Press: New York, 1997.

(4) Cho, Y. L.; Rudkevich, D. M.; Shivanyuk, A.; Rissanen, K.; Rebek, J. *Chem. Eur. J.* **2000**, *6*, 3788−3796.

(5) Prins, L. J.; Reinhoudt, D. N.; Timmerman, P. *Angew. Chem., Int. Ed.* **2001**, *40*, 2383−2426.

(6) Kim, E.; Paliwal, S.; Wilcox, C. S. *J. Am. Chem. Soc.* **1998**, *120*, 11192−11193.

(7) Raposo, C.; Wilcox, C. S. *Tetrahedron Lett.* **1999**, *40*, 1285−1288.

(8) Gallivan, J. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 870−874.

(9) Hobza, P.; Šponer, J. *J. Chem. Rev.* **1999**, *99*, 3247−3276.

(10) Šponer, J.; Hobza, P. In *Encyclopedia of Computational Chemistry*; von Rague-Schleyer, P., Ed.; Wiley: New York, 2001.

(11) Klärner, F. G.; Burkert, U.; Kamieth, M.; Boese, R.; Benet-Buchholz, J. *Chem. Eur. J.* **1999**, *5*, 1700−1707.

(12) Klärner, F. G.; Panitzky, J.; Preda, D.; Scott, L. T. *J. Mol. Model.* **2000**, *6*, 318−327.

(13) Klärner, F. G.; Burkert, U.; Kamieth, M.; Boese, R. *J. Phys. Org. Chem.* **2000**, *13*, 604−611.

(14) Klärner, F. G.; Panitzky, J.; Bläser, D.; Boese, R. *Tetrahedron* **2001**, *57*, 3673−3687.

(15) Ruloff, R.; Seelbach, U. P.; Merbach, A. E.; Klärner, F. G. *J. Phys. Org. Chem.* **2002**, *15*, 189−196.

(16) Klärner, F. G.; Kahlert, B. *Acc. Chem. Res.* **2003**, *36*, 919−932.

(17) Klärner, F. G.; Lobert, M.; Naatz, U.; Bandmann, H.; Boese, R. *Chem. Eur. J.* **2003**, *9*, 5036−5047.

(18) Klärner, F. G.; Polkowska, J.; Panitzky, J.; Seelbach, U. P.; Burkert, U.; Kamieth, M.; Baumann, M.; Wigger, A. E.; Boese, R.; Bläser, D. *Eur. J. Org. Chem.* **2004**, 1405−1423.

(19) Chang, C. E.; Gilson, M. K. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.

(20) Brown, S. P.; Schaller, T.; Seelbach, U. P.; Koziol, F.; Ochsenfeld, C.; Klärner, F. G.; Spiess, H. W. *Angew. Chem., Int. Ed.* **2001**, *40*, 717−720.

(21) Ochsenfeld, C.; Koziol, F.; Brown, S. P.; Schaller, T.; Seelbach, U. P.; Klärner, F. G. *Solid State Nucl. Magn. Reson.* **2002**, *22*, 128−153.

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

(23) Cremer, D. In *Encyclopedia of Computational Chemistry*; von Rague-Schleyer, P., Ed.; Wiley: New York, 1998.

(24) Tsuzuki, S.; Honda, K.; Uchimura, T.; Mikami, M. *J. Chem. Phys.* **2004**, *120*, 647−659.

(25) Sinnokrot, M. O.; Valeev, W. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887.

(26) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095−9102.

(27) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−557.

(28) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

(29) Zou, B.; Dreger, K.; Mück-Lichtenfeld, C.; Grimme, S.; Schäfer, H. J.; Fuchs, H.; Chi, L. *Langmuir* **2005**, *21*, 1364−1370.

(30) Piacenza, M.; Grimme, S. *Chem. Phys. Chem.* **2005**, *6*, 1554−1558.

(31) Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. *J. Chem. Phys.* **2001**, *115*, 8748.

(32) Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515−524.

(33) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149−5155.

(34) Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **1996**, *76*, 102−105.

(35) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.

(36) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 154104.

(37) Furche, F.; Voorhis, T. V. *J. Chem. Phys.* **2005**, *122*, 164106.

(38) Šponer, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624−1626.

(39) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem.* **2003**, *107*, 8377.

(40) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325.

(41) Bickelhaupt, F. M.; Baerends, E. J. Kohn−Sham Density Functional Theory: Predicting and Understanding Chemistry. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2000; Vol. 15.

(42) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.

(43) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: New York, 2001.

(44) TURBOMOLE (Vers. 5.6); Ahlrichs, R.; Bär, M.; Baron, H.-P.; Bauernschmitt, R.; Böcker, S.; Ehrig, M.; Eichkorn, K.; Elliott, S.; Furche, F.; Haase, F.; Häser, M.; Horn, H.; Huber, C.; Huniar, U.; Kattannek, M.; Kölmel, C.; Kollwitz,

M.; May, K.; Ochsenfeld, C.; Öhm, H.; Schäfer, A.; Schneider, U.; Treutler, O.; von Arnim, M.; Weigend, F.; Weis, P.; Weiss, H. Universität Karlsruhe, 2003. See also: http://www.turbomole.com.

(45) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

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

(47) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283−289.

(48) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331−340.

(49) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2001**, *116*, 3175.

(50) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119.

(51) The basis sets are available from the TURBOMOLE homepage via the FTP server button (in the subdirectories basen, jbasen, and cbasen). See http://www.turbomole.com.

(52) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829−5835.

(53) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(54) Grimme, S. *Chem. Eur. J.* **2004**, *10*, 3423.

(55) Swart, M.; Fonseca, C.; Bickelhaupt, F. M. *J. Am. Chem. Soc.* **2004**, *126*, 16718−16719.

(56) Jeziorski, B.; Szalewicz, K. In *Encyclopedia of Computational Chemisty*; von Rague-Schleyer, P., Ed.; J. Wiley: New York, 1998; Vol. 2, p 1376.

(57) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.

(58) Kutzelnigg, W. *Einführung in die Theoretische Chemie: Band 2, Die chemische Bindung*; Verlag Chemie: Weinheim, 1978.

CT050122N

JCTC Journal of Chemical Theory and Computation

Considerations for Reliable Calculation of ^{77}Se Chemical Shifts

Craig A. Bayse*

Department of Chemistry and Biochemistry, Old Dominion University, Hampton Boulevard, Norfolk, Virginia 23529

Received May 21, 2005

Abstract: The theoretical chemical shifts of a large series of selenium compounds have been calculated using GIAO-MP2 and -DFT methods in several basis sets. Reliable chemical shifts are calculated for many compounds, especially with the mPW1PW91 exchange-correlation functional and either a triple-ζ basis set (tzvp: 13% mean absolute error) or a limited RECP set chosen for practical applications on complex molecules (BSL: 11.8% mean absolute error). Molecules with three-center-four-electron bonding or low-lying n→π* states require additional diffuse functions and nonperturbative methods, respectively, but terminal selenium anions cannot be calculated reliably in the gas phase due to the neglect of solvation. When these cases are excluded, the mean absolute error decreases from 16.5% to 8.9% in GIAO-MP2/BSL but only slightly for DFT methods.

1. Introduction

Chemical shifts can be calculated using ab initio or density functional theory (DFT) through the gauge invariant atomic orbitals (GIAO) method with general applicability and reasonable accuracy.[1] For selenium compounds, early ab initio calculations showed that electron correlation was necessary for computation of accurate shielding constants.[2,3] Møller−Plesset perturbation theory largely corrects the inaccuracy of GIAO-HF[4] but is highly expensive and fails for some systems.[5] Failures at the MP2 level can be corrected at higher levels of theory (CCSD(T)),[6] but these methods are currently too expensive for practical application. DFT promises to be an economical means of calculating chemical shifts[5,7] but systematic errors often result in the overestimation of the paramagnetic contribution to the shielding.[8] Numerical corrections such as scaling factors,[9] adjustments to the HOMO−LUMO gap,[10] and level shifting of virtual orbital energies[8,11] have been employed to improve DFT results.

GIAO shifts have been used successfully to explain experimental observations of complex molecules using simplified models.[12,13] Bayse has recently shown that theoretical ^{77}Se NMR calculations of simple systems correlate well to experimental shifts of selenoproteins.[14] Continued interest in applications of theoretical ^{77}Se chemical shifts led us to examine the reliability of GIAO-MP2 and -DFT methods for a broad range of selenium functionalities. Basis set effects may be important for selenium due to the sensitivity of the nucleus to its electronic environment. Thus far, detailed studies of basis set effects have been limited to GIAO-HF methods or a few small molecules. The Schaefer et al. double-ζ basis set for selenium was selected for modification with diffuse and polarization functions. Chemical shifts were also calculated using Dunning's triple-ζ and correlation consistent basis sets. The reliability of GIAO shifts of geometries derived from effective core potential basis sets and GIAO-MP2//DFT chemical shifts are also examined for application to molecules of biological interest for which MP2 geometry optimizations would be prohibitively expensive.

2. Theoretical Details

Theoretical ^{77}Se NMR chemical shifts have been calculated for 64 selenium compounds using the GIAO-MP2 and -DFT (B3LYP,[15,16] BLYP,[17,16] B3P86,[15,18] and mPW1PW91[19]) methods available in Gaussian98.[20] Configuration interaction wave functions were calculated for selected species using Gaussian98 (CIS) and GAMESS-UK[21] (CISD). The reported

* Corresponding author e-mail: cbayse@odu.edu.

10.1021/ct050136t CCC: $30.25
Published on Web 09/15/2005

chemical shifts are derived from the theoretical isotropic shielding referenced to that of Me_2Se calculated at the same level and basis set (eq 1).

$$\delta_i^{\mathrm{calc}} = \sigma_{\mathrm{ref,Me_2Se}}^{\mathrm{calc}} - \sigma_{i,\mathrm{Se}}^{\mathrm{calc}} \quad (1)$$

For basis set I (BSI), selenium, phosphorus, sulfur, and chlorine basis sets are derived from the double-ζ plus polarization functions basis (DZP) of Schafer et al.[22] The Se DZP basis set includes the diffuse selenium d-functions an early study showed were necessary for accurate energetics and geometries.[23] Hydrogen and first-row atoms were represented by Dunning's split valence triple-ζ (TZV) basis set[24] augmented with polarization functions on all non-carbon first-row atoms and hydrogens bonded to non-carbon heavy atoms. BSII adds a set of even-tempered diffuse s, p, and d functions to all non-carbon heavy atoms. BSIII modifies BSII by adding an additional set of s, p, and d diffuse functions to all heavy atoms. The effects of high angular momentum functions have been examined through the addition of f-type polarization functions to the selenium center ($\zeta_{\mathrm{f,Se}}$=5.8). The exponent of these functions was optimized through MP2 geometry optimizations on SeH_2. Additional f-type polarization functions were derived from this set of tight functions by dividing the exponent by four. GIAO calculations were also performed for a test set of 41 molecules with the mPW1PW91 functional and Dunning's TZVP, cc-pVDZ,[25] and cc-pVTZ[26] basis sets. The latter was modified by removing the f-type polarization functions on heavy atoms and the d-type polarization functions on hydrogen.

A basis set for geometry optimizations of large systems (BSL) was assembled from the relativistic effective core potential (RECP) basis set of Hurley et al.[27] for Se and the Wadt-Hay RECP for P, S, and Cl.[28] Oxygen, nitrogen, and fluorine were represented as in BSI. Hydrogens attached to non-carbon heavy atoms were represented as in BSII. Carbon and hydrogen atoms bonded to carbon were represented by the Dunning DZ basis set.[29] The RECP and carbon basis sets were modified by addition of a set of polarization functions. The selenium basis set also included an additional set of diffuse functions. The RECP basis set for selenium was used only for geometry optimization and was replaced with the all-electron Schafer et al.[22] basis set used in BSI for GIAO calculations.

Relativistic effects have not been explicitly considered in these calculations, but effects due to the contraction of inner shell electrons are uniform and cancel in the calculation of the relative chemical shift.[5,30] Structural changes due to relativistic effects should be incorporated in the BSL calculations through the RECP, but the difference between the BSI and BSL optimized geometries in this study is minor. Chemical shifts were calculated on structures corresponding to the minimum of the potential energy surface; rovibrational and dynamic effects have been ignored. The effects of conformational changes on the selenium chemical shift have been discussed in detail elsewhere.[3,5] The lowest energy conformations were found for use in subsequent GIAO calculations.

Based upon previous studies of solvent dependence, the gas-phase theoretical chemical shifts presented in this work should correlate well with most liquid- and solution-phase experimental results. Luthra et al.[31] examined a series of selenides and diselenides to show a less than 20 ppm solvent-dependent shift in aprotic solvents. Protic solvents gave larger differences in certain cases. In another study, the chemical shifts of Me_2Se and PhSeEt in six protic and aprotic solvents ranging in polarity from cyclohexane to DMSO show standard deviations of roughly 7 ppm.[32] For a selection of cationic selenium compounds and those with selenium adjacent to multiple bonds, the largest solvent-dependent difference between chloroform and trifluoroacetic acid was 21 ppm.[33] Larger solvent and phase dependencies occur in protic solvents with anionic solutes or when the solute can be protonated or deprotonated by the solvent.[33] For example, NaSeH shows an upfield shift with increasing hydrogen bonding ability of the solvent (−447 ppm (DMF), −495 ppm (EtOH), −529 (H_2O)). Alkylselenols are particularly sensitive to pH with shifts ranging from −150 to −200 ppm upon deprotonation.[33] Given these experimental observations of solvent dependence, correlation with gas-phase theoretical calculations should be maximized when weak intermolecular forces dominate solvation, but large discrepancies are expected when solvent effects are strong (i.e., anions, selenols).

3. Results and Discussion

Theoretical chemical shifts for selected basis sets and methods are listed with experimental values in Table 1. Tabulated results for all other basis sets and methods can be found in the Supporting Information. Detailed comparisons to previous theoretical studies will be omitted in the following discussion, but the overlapping data found in this study are generally comparable to previous work.[2−8]

To gauge the overall quality of the calculated shifts, the experimental[33−35] and theoretical data sets shown in Table 1 were plotted against one another. Plots for GIAO-MP2/BSI and GIAO-DFT(B3LYP)/BSI can be found in Figure 1 and are representative of plots of data for other basis sets and functionals. Regression and error data for additional methods and basis sets appear in Table 2. Three outliers appear in the GIAO-MP2 data (Figure 1a). These are $[Se_4]^{2+}$ **41**, a known problem case for MP2,[4,5] and related ions *cis*-(**42**) and *trans*-$[Se_2S_2]^{2+}$ (**43**). Several groups[4,5] have examined the problems of GIAO-MP2 with the square $[Se_4]^{2+}$ cluster and conclude that DFT or multireference methods are required. Tuononen et al. have recently calculated very accurate chemical shifts for **41**−**43** using GIAO-DFT (BPW91 and B3PW91) and GIAO-CAS.[36] Omission of **41**−**43** from the GIAO-MP2 regression provides a near 1:1 correspondence and an overall upfield shift of 24.12 ppm based upon the *y*-intercept. Nakanishi and Hayashi[12] used similar plots of simple systems to calibrate method and basis set for their study of aryl selenides and obtained better correlation (B3LYP/6-311++G(3df,2pd), R^2=0.998) but used a limited reference set (11 compounds) and a single data point per compound. This study compares a test set of 41 compounds to averages of available experimental shifts in various solvents (Table 1).[37] Examination of individual data points shows that GIAO-MP2/BSI performs reasonably

Table 1. Experimental and Theoretical ^{77}Se Chemical Shifts for Various Selenium Compounds

		exp	solvent	GIAO-MP2 MP2/BSI	GIAO-MP2 MP2/BSL	GIAO-DFT B3LYP/ BSI	GIAO-MP2 B3LYP/ BSL	GIAO-DFT mPW1PW91/ BSI	GIAO-DFT mPW1PW91/ BSL	GIAO-DFT mPW1PW91/ tzvp	GIAO-DFT mPW1PW91/ cc-pvdz
1	Me_2Se	0.0		0 (1970[a])	0.0 (2004[a])	0 (1691[a])	0 (1956[a])	0 (1804[a])	0 (1804[a])	0 (1805[a])	0 (1874[a])
2	MeSeEt	108[b]	CH_2Cl_2	67	61	64	61	64	57	52	62
3	$MeSeCHCH_2$	192[c]	$CH_2Cl_2/MeNO_2$	179	179	183	176	218	194	176	168
4	$[Me_3Se]^+$	258,[d] 256[e]		224	252	201	229	238	247	223	243
5	MeSeH	−130,[f] −155[g]	$(CD_3)_2CO$,	−137	−128	−158	−121	−129	−125	−116	−124
6	$[MeSe]^-$	−332,[b,h] −330[d]	D_2O pD 10, H_2O	−215	−274	−186	−243	−243	−290	−303	−243
7	H_2Se	−345(g), −288(l)[d]		−300	−276	−359	−264	−301	−299	−281	−290
8	$[HSe]^-$	−495,[i] −529[g], −518,[h] −447[h]	EtOH, H_2O, EtOH, DMF	−599	−578	−667	−554	−624	−614	−618	−570
9	Me_2SeO	819[d], 812[j]	D_2O pD 7, H_2O	804	852	751	851	775	799	771	712
10	$MeSeO_2H$	1216[b]	H_2O	1272	1321	1237	1307	1248	1260	1232	1099
11	$[MeSeO_3]^-$	1045[k]	H_2O	1093	1136	984	1111	1027	1030	991	889
12	$FSeO_3H$	1001[l]	neat	1034	1069	989	1035	1040	1033	1003	915
13	MeSeSeMe	268,[b] 270,[m] 275,[n] 281,[o,d]	$CDCl_3$	249	223	276	278	217	291	289	282
14	F_2CSe	688[p]	benzene	574	577	673	591	684	725	690	576
15	$AcSe^-$	409[q]	EtOH/benzene	288	336	332	340	355	416	322	258
16	selenophene	605[p,k]	$CDCT_3$	583	602	546	588	581	591	567	539
17	1,2,3-selena-diazole	1502[r]	$CDCl_3$	1349	1360	1591	1484	1539	1534	1494	1458
18	Me_2SeCl_2	448[j]	CH_2Cl_2	339	363	382	376	387	374	327	368
19	MeSeCN	120,[s] 123[t] 125,[o] 133[u]	neat, $CDCl_3$, $CDCl_3$, $CDCl_3$	88	117	70	104	84	114	89	88
20	CF_3SeCl	1077[v]		1028	1048	1344	1150	1238	1246	1218	1130
21	Me_3PSe	−256,[w] −235,[x,y,z] −262	toluene, CH_2Cl_2	−322	−284	−310	−243	−311	−257	−265	−296
22	$Se(SiH_3)_2$	−666[aa]		−633	−662	−666	−671	−634	−670	−691	−606
23	SeO_3	944[j]	neat	891	924	1316	886	1374	1364	1356	1262
24	$[SeO_3]^{2-}$	1263[bb]	H_2O	1385	1396	1297	1384	1303	1275	1268	1118
25	$[SeO_4]^{2-}$	1033,[bb] 1024[l,cc] 1051[d]	H_2O	1185	1222	1098	1202	1131	1122	1090	960
26	$[SeCN]^-$	−322,[h] −273[dd]	EtOH, $CDCl_3$/DMSO	−404	−341	−497	−348	−448	−385	−413	−398
27	$[Se]^{2-}$	−511[f]		−1040	−1040		−1054	−1206	−1206	−1203	−1135
28	$SeOF_2$	1378[l]	neat	1381	1420	1415	1382	1444	1442	1407	1335
29	$SeOCl_2$	1479,[l] 1501[v]	neat	1453	1443	1581	1434	1557	1509	1447	1466
30	$SeCl_2$	1758[ee,ff,gg]	CH_2Cl_2	1785	1796	2548	2017	2275	2248	2162	2104
31	$SeCl_4$	1454[l,ee]	DMF	1192	1187	1591	1341	1424	1368	1251	1353
32	$[SeCl_6]^{2-}$	863[hh]	solid state	926	940	1298	1025	1136	1076	971	1053
33	SeF_4	1092,[l] 1083[ii]	neat	1221	1267	1215	1256	1226	1230	1161	1150

Table 1 (Continued)

34	SeF_6	610,[l] 600[ii]	neat	795	837	732	812	770	768	665	671
35	Se_2Cl_2	1295[ee,ff,gg,jj]	CH_2Cl_2/neat	1214	1243	1588	1397	1444	1466	1433	1371
36	MeSSeSMe	784[kk]		751	753	989	807	952	865	891	930
37	ClSSeSCl	1348[gg]	neat	1018	1054	1276	1143	1164	1170	1115	1110
38	SeCO	−447[p]	benzene	−490	−446	−574	−476	−502	−469	−479	−460
39	SeCS	102[p]	benzene	−17	46	−27	10	47	114	98	63
40	CSe_2	300,[e] 331[p]	CH_2Cl_2/benzene	141	170	189	145	251	290	259	248
41	$[Se_4]^{2+}$	1933[ll]	SO_2	323	744	2434	484	2311	2298	2226	2227
42	*trans*-$[Se_2S_2]^{2+}$	1887[mm]	SO_2	831	917	2352	996	2254	2152	2057	2138
43	*cis*-$[Se_2S_2]^{2+}$	2020[mm]	SO_2	658	898	2594	749	2469	2364	2279	2338

[a] Absolute shielding of the reference Me_2Se. [b] Reference 34a ($NaSeMe$). [c] Reference 34b. [d] Reference 34c (Me_3Sel, $NaSeMe$, Na_2SeO_4). [e] Reference 34d (Me_3Sel). [f] Reference 34e ($(NH_4)_2$ Se). [g] Reference 34f. [h] Reference 34g ($NaSeH$, $KSeCN$, $NaSeMe$). [i] Reference 34h ($KSeH$). [j] Reference 34i. [k] Reference 34j. [l] Reference 34k (K_2SeO_4). [m] Reference 34l. [n] Reference 34m. [o] Reference 34n. [p] Reference 34o. [q] Reference 34p ($AcSeLi$). [r] Reference 34q. [s] Reference 34r. [t] Reference 34s. [u] Reference 34t. [v] Reference 34u. [w] Reference 34v. [x] Reference 34w. [y] Reference 34x. [z] Reference 34y. [aa] Reference 34z. [bb] Reference 34aa (Na_2SeO_4, Na_2SeO_3). [cc] Reference 34bb ($(NH_4)_2SeO_4$). [dd] Reference 34cc ($KSeCN$). [ee] Reference 34dd. [ff] Reference 34ee. [gg] Reference 34ff. [hh] Reference 34gg ($[NH_4]_2[SeCl_6]$). [ii] Reference 34hh. [jj] Reference 34ii. [kk] Reference 12. [ll] Reference 35. [mm] Reference 35b.

well for many systems: 29% of the test set have errors less than 5%; 56% are below 15%. The mean absolute error of 21.0% is partially due to molecules such as **41**−**43** for which MP2 is known to fail. If the three outliers in the correlation plot are eliminated, the average error improves to 17.2%. Addition of a single set of f-type polarization functions to Se in BSI+f reduces the overall upfield shift of the theoretical data relative to experiment to 18.85 ppm and improves correspondence with experiment (slope = 1.001). The percentage of the test set with error below 15% decreases despite the slight improvement in the mean absolute error over BSI (Table 2). Diffuse functions in BSII increase the average error by ∼2% for reasons explained below; errors for larger basis sets BSII+f and BSIII also exceed the error for BSI but by less than BSII.

The B3LYP XC functional commonly used in DFT studies has been documented to produce poor GIAO chemical shifts.[7] Only 22% of the test molecules have error less than 5% and roughly half exceed 20% error. The B3LYP correlation plot (Figure 1b) has a slope of 1.198 consistent with the overestimation of the paramagnetic shielding in DFT. A skewed correlation plot is common for GIAO-DFT and is obtained to varying degrees for each functional in this study. Comparison of the B3LYP shifts with those obtained with BLYP and B3P86 (33.0 and 17.1% mean absolute error, respectively) suggests that the LYP correlation functional performs poorly in comparison to other functionals in the calculation of magnetic shielding properties. The mPW1PW91 functional shown to produce very accurate chemical shifts for ^{13}C[38] gives an average absolute error comparable to B3P86 but has fewer species in excess of 15% error. GIAO-DFT(mPW1PW91) chemical shifts were also calculated using the TZVP, cc-pVDZ, and cc-pVTZ-fd basis sets. Each of these gives a lower mean absolute error than BSI and a slope closer to 1.0 for the correlation plot.

For GIAO studies of molecules of biological or catalytic interest, it may be practical to use RECP basis sets for heavy atoms such as Se, P, and Cl (as in BSL) to reduce the computational effort of the geometry optimization step. (Note that the ECP basis set for selenium must be replaced with an all-electron basis, here that of Schaefer et al. from BSI, for the GIAO calculations.) GIAO-MP2 calculations in BSL give the most ideal correlation plot of any method considered in this study (if $[Se_4]^{2+}$ and its related ions are omitted). The mean absolute error also improves over MP2/BSI: 78% of test molecules have less than 20% error. Chemical shifts were also calculated from the B3LYP/BSL optimized geometry to obtain a mean absolute error ∼2% greater than MP2/BSL likely due to the poor geometries generated with that functional.[7] The mPW1PW91/BSL chemical shifts are even more accurate (11.8% average absolute error). Fifteen of the test molecules have less than 5% error, and only two molecules (MeSeEt and SeO_3) exceed 30% error. Large improvements in the theoretical chemical shift with mPW1PW91/BSL are also obtained for several species. SeCS has only 11.8% error in mPW1PW91/BSL but exceeds 100% error in virtually every other method and basis set. MeSeCN has roughly 30% error in BSI-III but is less than 10% for BSL. Less dramatically, the error for $[Me_3Se]^+$ is reduced

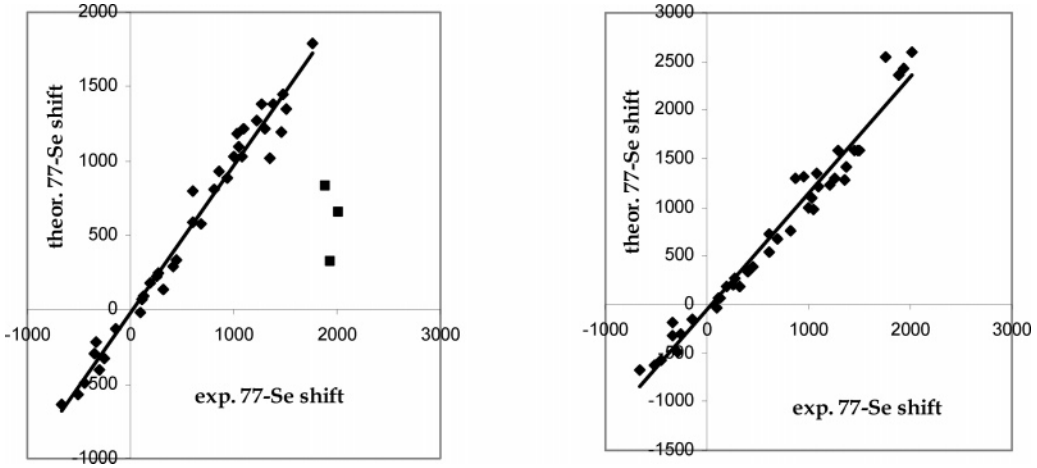


Figure 1. Plot of theoretical ((a) GIAO-MP2/BSI and (b) GIAO-DFT(B3LYP)/BSI) versus experiment ^{77}Se chemical shifts for the compounds in Table 1.

Table 2. Regression and Error Analysis for Experimental and Theoretical Data Presented in Table 1

method	geometry	slope	intercept	R^2	error[b]	error[c]
GIAO-MP2/BSI[a]	MP2/BSI	0.993	−24.116	0.975	20.9	17.2
GIAO-DFT[d]/BSI	B3LYP/BSI	1.198	−53.642	0.966	22.2	21.9
GIAO-DFT[d]/BSI	BLYP/BSI	1.242	−76.400	0.935	33.0	34.2
GIAO-DFT[d]/BSI	B3P86/BSI	1.149	−26.535	0.977	17.1	17.0
GIAO-DFT[d]/BSI	mPW1PW91/BSI	1.133	−24.739	0.975	17.0	16.7
GIAO-DFT[d]/tzvp	mPW1PW91/tzvp	1.078	−26.636	0.977	13.0	13.1
GIAO-DFT[d]/cc-pvdz	mPW1PW91/cc-pvdz	1.058	−33.051	0.972	15.1	15.1
GIAO-DFT[d]/cc-pvtz-fd	mPW1PW91/cc-pvtz-fd	1.072	−36.055	0.976	15.1	15.3
GIAO-MP2/BSI+f[a]	MP2/BSI+f	1.004	−18.847	0.975	20.1	16.7
GIAO-MP2/BSII[a]	MP2/BSII	1.025	−44.622	0.977	22.8	18.5
GIAO-MP2/BSII+f[a]	MP2/BSII+f	1.034	−37.647	0.978	21.0	17.3
GIAO-MP2/BSIII[a]	MP2/BSIII	1.034	−39.007	0.977	21.5[e]	17.5[e]
GIAO-MP2/BSL[a]	MP2/BSL	0.998	−5.567	0.975	16.5	13.4
GIAO-DFT[d]/BSL	mPW1PW91/BSL	1.100	−2.240	0.978	11.8	11.5
GIAO-MP2/BSL[a]	B3LYP/BSL	1.036	−5.888	0.980	18.3	14.9

[a] Regression calculations in MP2 omit $[Se_4]^{2+}$, *cis*- and *trans*-$[Se_2S_2]^{2+}$. [b] Mean absolute error for all 41 molecules in the test set. [c] Mean absolute error for the test set omitting $[Se_4]^{2+}$, *cis*- and *trans*-$[Se_2S_2]^{2+}$. [d] GIAO-DFT calculations use the same functional as for the geometry optimization. [e] MP2/BSIII calculations omit $[SeCl_6]^{2-}$.

below 5% versus error exceeding 15% for shifts calculated in most all-electron basis sets.

In the data shown in Table 1, systems with large error (other than **41**−**43**) tend to be species with terminal anionic selenium (**6**, **8**, **15**, **26**), low-lying n→π* excited states (**14**, **21**, **28**, **39**, **40**), or three-center-four-electron bonding (**18**, **31**, **32**, **33**, **34**). The average error for each of these groups is shown in Table 3. When these groups are excluded, the mean absolute error for the remaining 'well-behaved' species drops to ∼9% with MP2 in BSI-BSIII. The error for systems with low-lying n→π* excited states and 3c4e bonding decreases with the addition of f functions and diffuse functions, respectively, but that for species with anionic Se increases substantially when diffuse functions are added (MP2/BSI: 29.4%, MP2/BSII: 39.3%, MP2/BSIII: 39.9%). The large errors associated with these species lead to a net increase in error for the overall test set. For anions such as $[MeSe]^-$ or $[Se]^{2-}$ where the charge is localized on selenium, solvent interactions will serve to delocalize this charge and deshield the selenium nucleus. Therefore, gas-phase theoretical shifts tend to be upfield of the experimental values, more so when diffuse functions are included. For example, the GIAO-MP2/BSI shift (−574 ppm) of $[SeH]^-$ agrees well with experimental data, but more diffuse basis sets shift it increasingly further upfield (BSII −698) because additional basis functions allow more charge to accumulate on Se (BSI −0.972e; BSII −1.020e). Similar basis set effects are observed for some systems with electronegative atoms. Methylselenenate $[MeSeO]^-$ (Table 5) shows a steady downfield shift with an increased basis set. The Mulliken charges show that as the basis set becomes more diffuse, the electron density on the oxygen increases, pulling charge and electron density toward oxygen (BSI: −0.921e; BSIII: −1.277e) to deshield the selenium nucleus. A similar basis set dependency is observed for MeSeF, MeSeCl, $SeCl_2$, and other molecules with divalent selenium bonded to electronegative atoms.

Errors for compounds with 3c4e bonding are partially due to an incomplete basis set and are reduced with additional diffuse functions or larger basis sets (Table 3). For example, error for SeF_4 is consistently above 10% in all methods in BSI and BSL but decreases to 7.7% in MP2/BSII, 5.0% in

Table 3. Error Analysis for Species with Large Error

method	geometry	error[b]	error[c]	error[d]	error[e]
GIAO-MP2/BSI	MP2/BSI	29.4	45.3	17.2	9.0
GIAO-DFT[a]/BSI	B3LYP/BSI	40.0	44.1	23.1	14.4
GIAO-DFT[a]/BSI	BLYP/BSI	54.4	74.2	41.4	22.0
GIAO-DFT[a]/BSI	B3P86/BSI	29.3	26.3	21.0	12.5
GIAO-DFT[a]/BSI	mPW1PW91/BSI	28.1	22.2	18.4	13.5
GIAO-DFT[a]/tzvp	mPW1PW91/tzvp	22.3	7.0	10.7	12.6
GIAO-DFT[a]/cc-pvdz	mPW1PW91/cc-pvdz	27.4	19.3	11.3	12.7
GIAO-DFT[a]/cc-pvtz-fd	mPW1PW91/cc-pvtz-fd	32.3	23.1	11.1	11.4
GIAO-MP2/BSI+f	MP2/BSI+f	28.7	41.3	18.0	9.1
GIAO-MP2/BSII	MP2/BSII	39.3	48.6	13.8	9.2
GIAO-MP2/BSII+f	MP2/BSII+f	38.7	41.9	14.5	8.9
GIAO-MP2/BSIII	MP2/BSIII	39.9	44.9	12.8[f]	8.8
GIAO-MP2/BSL	MP2/BSL	15.6	26.1	20.5	8.9
GIAO-DFT[a]/BSL	mPW1PW91/BSL	13.8	6.5	17.6	10.5
GIAO-MP2/BSL	B3LYP/BSL	17.2	33.6	19.0	9.9

[a] GIAO-DFT calculations use the same functional as for the geometry optimization. [b] Mean absolute error for terminal anionic Se species. [c] Mean absolute error for molecules with low-lying n→π* excited states. [d] Mean absolute error for species with 3c4e bonding. [e] Mean absolute error for the remainder of the test set also omitting **41**−**43**. [f] MP2/BSIII calculations omit $[SeCl_6]^{2-}$.

MP2/BSIII, and 1.9% in mPW1PW91/cc-pVTZ-fd. Augmentation with f-type Se polarization functions shows only a slight improvement for the molecules with low-lying n→π* excited states. Of this class of molecule, SeCS gives some of the largest errors for any molecule in this study (116% in MP2/BSI). A plot of the MP2/BSI error for the series **38**−**40** versus the first CIS excitation energy is roughly linear suggesting that the excessive error is due to an overestimation of the contribution of the n→π* excitation to the shielding which would be corrected with nonperturbative methods such as DFT or CCSD. The mPW1PW91 functional with the BSL and TZVP basis sets gives the lowest error (6.5 and 7.0% error, respectively).

The addition of f-type polarization functions in BSI+f and BSII+f results in downfield shifts of ∼230 ppm for $[Se_4]^{2+}$, slight decreases in bond length, and significant MP2 energy corrections (Table 3). A diffuse exponent f-type polarization function (BSI+f′: ζ=0.36) was used to give a shift further downfield and an even shorter bond length, but the correlation energy recovered did not compare with that obtained using the tight f function of BSI+f.[39] Geometries of $[Se_4]^{2+}$ were calculated using CISD in BSI and BSI+f, giving Se−Se bond lengths in very close agreement with the X-ray structure (exp: 2.28 Å; CISD/BSI+f: 2.284 Å). The additional f function gives a significant reduction in energy for a variational calculation. Additional f functions in GIAO-MP2 (BSI+2f, BSI+3f) successively move the $[Se_4]^{2+}$ chemical shift downfield and reduce the bond length to within 0.03 Å of the experimental value. These data suggest that the problem with MP2 is fundamentally one of basis set which is partially corrected with f-type polarization functions. An s- or p-type function cannot mix with face-to-face interactions of the filled d orbitals in the core of each selenium. As a result, the ring must expand to relieve repulsions between electrons in these orbitals. Inclusion of f-type functions allow these AOs to distort away from the ring, depleting electron density in the ring and allowing a closer interaction of the d orbitals. The net result is a shorter Se−Se bond length and an improved theoretical chemical shift in MP2.

Table 4. Relative Energies and Chemical Shifts for $[Se_4]^{2+}$

	ΔE, kcal/mol[a]	δ, ppm[b]	d(Se−Se), Å
MP2/BSI	0.0	323	2.360
MP2/BSI+f	−523.70	552	2.355
MP2/BSI+f′	−83.36	889	2.326
MP2/BSI+2f	−698.65	973	2.336
MP2/BSI+3f	−767.16	1236	2.311
MP2/BSII	−270.99	286	2.364
MP2/BSII+f	−575.78	516	2.359
MP2/BSIII	−617.79	359	2.359
CISD/BSI	[0.0]		2.295
CISD/BSI+f	[−332.26]		2.284
B3LYP/BSI	(0.0)	2434	2.347
B3LYP/BSI+f	(−0.22)	2432	2.347
exp		1958	2.283[c]

[a] Relative to BSI, DFT results are in parentheses, CISD in brackets. [b] Relative to Me_2Se. [c] Reference 35a.

The B3LYP chemical shifts and bond lengths in Table 4 (BSI, BSI+f) do not display large basis set dependence; the bond length remains the same, and there are only minor changes in the theoretical chemical shift. Schreckenbach et al.[5] performed GIAO-DFT calculations on $[Se_4]^{2+}$ with the addition of one and two sets of f functions with no change in the absolute ^{77}Se shielding constant. The hybrid functional B3 consistently underestimates the theoretical shielding (compare B3LYP to BLYP; B3PW91 to BPW91 in ref 36). The hybrid functionals (B3 and mPW1PW91) underestimate the shielding by ∼200 ppm, suggesting that the inclusion of the exact HF exchange results in excessive deshielding of these aromatic clusters.

The theoretical data in Table 5 for compounds that do not have experimental data available agree well with other known analogues. The quality of these comparisons is important if simplified model compounds[14] are to be used for more complicated chemical systems. For the series of increasingly

Table 5. Theoretical ^{77}Se Chemical Shifts for Various Selenium Compounds

		GIAO-MP2 MP2/BSI	GIAO-MP2 MP2/BSL	GIAO-DFT B3LYP/BSI	GIAO-MP2 B3LYP/BSL	GIAO-DFT mPW1PW91/BSI	GIAO-DFT mPW1PW91/BSL
44	MeSeCCH	52	85	32	69	75	83
45	Me_2SeO_2	973	1047	884	1020	931	944
46	MeSeOH	1153	1178	1291	1188	1212	1218
47	$[MeSeO]^-$	836	828	868	869	815	819
48	$[MeSeO_2]^-$	1199	1251	1122	1250	1136	1157
49	$MeSeO_3H$	1083	1129	995	1100	1040	1043
50	$Me_2Se(OH)_2$	486	520	457	512	471	480
51	$Me_2Se(OMe)_2$	519	629	591	629	612	598
52	Me*Se*Se(O)Me	408	393	630	476	507	507
	MeSe*Se*(O)Me	990	1029	966	1070	957	998
53	MeSeSMe	333	348	423	386	385	388
54	MeSSe(O)Me	994	1038	979	1062	976	997
55	$MeSeNH_2$	508	534	560	546	526	535
56	$MeSeNMe_2$	726	717	775	742	746	737
57	H_2CSe	2763	2729	3259	2720	3210	3247
58	Me_2CSe	2173	2142	2358	2180	2340	2351
59	$Me_2Se(OH)Cl$	461	510	456	517	468	485
60	MeSeF	1961	1991	2261	2000	2131	2141
61	MeSeCl	920	959	1289	1097	1171	1187
62	$SeCl_6$	446	472	638	499	546	506
63	SeF_2	3414	3476	3957	3492	3734	3834
64	Se_2F_2	1743	1584	1921	1819	1817	1817

unsaturated alkenes **2**, **3**, and **44** there is a roughly 110 ppm downfield shift from the alkane to the alkene and a ~120 ppm upfield shift from the alkene to the alkyne. A similar phenomenon has been observed in the experimental shifts of a series of 2-phenyl asymmetric selenides ($MeSeCH_2CH_2$-Ph 90 ppm, MeSeCHCHPh 186 ppm, MeSeCCPh 78 ppm). The experimental shift of methylphenylselenone is 977 ppm,[40] very similar to that of **45**. The theoretical shift of selenoacetone **58** (2173 ppm) corresponds well to the simplest known selenoketone (^{t}Bu)$_2$C═Se (2131 ppm).[41] The upfield shift of methylselenonic acid **49** relative to methylseleninic acid **10** mirrors observations for CF_3SeO_2H (1231 ppm) and CF_3SeO_3H (981 ppm).[33] Our previous study reported that the theoretical shifts of the simple diselenide **13** and selenenyl sulfide **53** compare well to the experimental shifts of biologically relevant analogues.[14] Known selenenyl sulfide oxides are found around 1050 ppm,[42] very similar to the GIAO-MP2/BSIII data (1027 ppm).

Unfortunately, broad comparisons with experimental analogues break down when some of the simple systems in Table 1 are compared to their aryl derivatives. Lower oxidation state selenium compounds are more affected by aryl substitution than highly oxidized centers because the aromatic ring can delocalize electron density located on selenium. For example, methylselenenamide **55** and *N*,*N*-dimethylmethylselenenamide **56** are upfield of aryl selenenamides which typically appear between 600 and 950 ppm.[43] Two known organoselenium dichlorides, Ph_2SeCl_2 (586 ppm)[44] and $MePhSeCl_2$ (488 ppm),[40] appear downfield of **33** due to phenyl effects. In comparison to **51**, a bisbenzyl derivative, $MePhSe(OBz)_2$, has a ^{77}Se chemical shift of 663−672 ppm.[40,44] Related selenoxide hydrates such as **50** have been proposed as intermediates in the racemization of selenoxides but have not been reported in NMR.

Conclusions

Theoretical chemical shifts have been calculated in several methods and basis sets to demonstrate the overall reliability of GIAO for a wide range of selenium functionalities. For 'well-behaved' systems, the mean absolute error is ~9.0% at the MP2 level and 10−15% for B3LYP, B3P86, and mPW1PW91. The mPW1PW91 hybrid functional gives the lowest mean error for the entire test set of compounds (11.8% BSL, 13.0% tzvp) and is the recommended functional based upon the results of this limited analysis of XC functionals. The larger error for DFT for this somewhat arbitrarily defined set of 'well-behaved' molecules is due to the established difficulties inherent in nonexact DFT.[8] However, the successful calculation of chemical shifts for systems for which MP2 fails or performs poorly (**41**−**43** and species with low-lying n→π* excited states) makes DFT an important tool nonetheless. The reliability of GIAO-DFT is limited in many of the same cases as MP2 but outperforms for $[SeO_3]^{2-}$, $[SeO_4]^{2-}$, $SeCF_2$, 1,2,3-selenadiazole, and $[Se_4]^{2+}$ and related ions but underperforms in several notable cases (e.g., Me_2SeO, SeO_3, $SeCl_2$, and MeSeSeMe). However, neither DFT nor MP2 was able to calculate an accurate chemical shift for MeSeEt which exceeds 30% error in every method and basis set. The consistency of the theoretical shifts for MeSeEt may indicate that solvent effects contribute significantly to the experimental chemical shift.

Of the basis sets examined in this study, it is difficult to suggest a generally reliable choice due to significant variance in performance with method and functional, but a basis set of at least triple-ζ quality would be recommended. Systems involving 3c4e bonding should be augmented with diffuse functions, and theoretical chemical shifts of terminal selenium anions should be treated with skepticism due to the

absence of solvation effects. The limited RECP basis set (BSL) gives surprisingly good results, a promising development for the reliable calculation of theoretical chemical shifts for the larger molecules of interest to synthetic, biological, and medicinal chemists.

Further applications of theoretical chemical shifts are currently being examined by this research group.

Supporting Information Available: Table of theoretical chemical shifts (GIAO-MP2 (BSI+f, BSII, BSII+f, BSIII, BSI//B3LYP/BSI) and GIAO-DFT (BLYP/BSI, B3P86/BSI, mPW1PW91/cc-pVTZ-fd)). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Ditchfield, R. *Mol. Phys*. **1974**, *27*, 789. (b) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc*. **1990**, *112*, 8251.

(2) (a) Higashioji, T.; Hada, M.; Sugimoto, M.; Nakatsuji, H. *Chem. Phys*. **1996**, *203*, 159. (b) Nakatsuji, H.; Higashioji, T.; Sugimoto, M. *Bull. Chem. Soc. Jpn*. **1993**, *66*, 3235.

(3) Magyarfalvi, G.; Pulay, P. *Chem. Phys. Lett*. **1994**, *225*, 280.

(4) Bühl, M.; Thiel, W.; Fleisher, U.; Kutzelnigg, W. *J. Phys. Chem*. **1995**, *99*, 4000.

(5) Schreckenbach, G.; Ruiz-Morales, Y.; Ziegler, T. *J. Chem. Phys*. **1996**, *104*, 8605.

(6) Bühl, M.; Gauss, J.; Stanton, J. F. *Chem. Phys. Lett*. **1995**, *241*, 248.

(7) Keal, T. W.; Tozer, D. J. *Mol. Phys*. **200**5, *103*, 1007.

(8) Magyarfalvi, G.; Pulay, P. *J. Chem. Phys*. **2003**, *119*, 1350.

(9) Chesnut, D. B. *Chem. Phys*. **2004**, *305*, 237.

(10) Malkin, V. G.; Malkina, O. L.; Casida, M. E.; Salahub, D. R. *J. Am. Chem. Soc*. **1994**, *116*, 5898.

(11) Lee, A. M.; Handy, N. C.; Colwell, S. M. *J. Chem. Phys*. **1995**, *103*, 10005.

(12) (a) Hayashi, S.; Nakanishi, W. *J. Org. Chem*. **1999**, *64*, 6688. (b) Block, E.; Birringer, M.; DeOrazio, R.; Fabian, J.; Glass, R. S.; Guo, C.; He, C.; Lorance, E.; Qian, Q.; Schroeder, T. B.; Shan, Z.; Thiruvashi, M.; Wilson, G. S.; Zhang, X. *J. Am. Chem. Soc*. **2000**, *122*, 5052. (c) Nakanishi, W.; Hayashi, S. *J. Phys. Chem. A* **1999**, *103*, 6074.

(13) Fleischer, H.; Glang, S.; Schollmeyer, D.; Mitzel, N. W.; Bühl, M. *Dalton Trans*. **2004**, 3764.

(14) Bayse, C. A. *Inorg. Chem*. **2004**, *43*, 1208.

(15) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648.

(16) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev*. **1988**, *B37*, 785. (b) Colle, R.; Salvetti, O. *Theor. Chim. Acta* **1975**, *37*, 329.

(17) Becke, A. D. *Phys. Rev*. **1988**, *A38*, 3098.

(18) Perdew, J. P. *Phys. Rev*. **1986**, *B33*, 882.

(19) Adamo C, Barone V. *J. Chem. Phys*. **1998**, *108*, 664.

(20) 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. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.

(21) GAMESS-UK is a package of ab initio programs written by Guest, M. F.; van Lenthe, J. H.; Kendrick, J.; Schoffel, K.; Sherwood, P. with contributions from Amos, R. D.; Buenker, R. J.; Dupuis, M.; Handy, N. C.; Hillier, I. H.; Knowles, P. J.; Bonacic-Koutecky, V.; von Neissen, W.; Harrison, R. J.; Rendell, A. P.; Saunders, V. R.; Stone, A. J. The package is derived from the original GAMESS code due to Dupuis, M.; Spangler, D.; Wendoloski, J. Warrington, U.K, 1995.

(22) Schafer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys*. **1992**, *97*, 2571.

(23) Angyan, J. G.; Csizmadia, I. G.; Daudel, R.; Poirier, R. A. *Chem. Phys. Lett*. **1986**, *131*, 247.

(24) Dunning, T. H. *J. Chem. Phys*. **1971**, *55*, 716.

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

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

(27) Hurley, M. M.; Pacios, L. F.; Christiansen, P. A.; Ross, R. B.; Ermler, W. C. *J. Chem. Phys*. **1986**, *84*, 6840.

(28) Wadt, W. R.; Hay P. J. *J. Chem. Phys*. **1985**, *82*, 284.

(29) Dunning, T. H. *J. Chem. Phys*. **1970**, *53*, 2823.

(30) Yates, J. R.; Pickard, C. J.; Payne M. C. *J. Chem. Phys*. **2003**, *118*, 5746.

(31) Luthra, N. P.; Boccanfuso, A. M.; Dunlap, R. B.; Odom, J. D. *J. Organomet. Chem*. **1988**, *354*, 51.

(32) (a) Luthra N. P.; Dunlap, D. B.; Odom, J. D. *J. Magn. Reson*. **1983**, *52*, 318. (b) Duddeck, H.; Wagner, P.; Biallab, A. *Magn. Reson. Chem*. **1991**, *29*, 248.

(33) Duddeck, H. *Prog. Nucl. Magn. Reson. Spect*. **1995**, *27*, 1 and references therein.

(34) (a) McFarlane, W.; Wood, R. J. *J. Chem. Soc., Dalton. Trans*. **1972**, 1397. (b) Dieden, R.; Hevesi, L. *Bull. Magn. Reson*. **1989**, *11*, 193. (c) Odom, J. D.; Dawson, W. H.; Ellis, P. D. *J. Am. Chem. Soc*. **1979**, *101*, 5815. (d) Pan, W. H.; Fackler, J. P. *J. Am. Chem. Soc*, **1978**, *100*, 5783. (e) Odom, J. D.; Dawson, W. H.; Ellis, P. D. *J. Am. Chem. Soc*. **1979**, *101*, 5815. (f) Ellis, P. D.; Odom, J. D.; Lipton, A. S.; Chen, Q.; Gulick, J. M. In *Nuclear Magnetic Shieldings and Molecular Structure*; NATO ASI Series; Tossell, J. A., Ed.; Kluwer Academics: Amsterdam, 1993. (g) Cusick, J.; Dance, I. *Polyhedron* **1991**, *10*, 2629. (h) Björgvinsson, M.; Schrobilgen, G. J. *Inorg. Chem*. **1984**, *30*, 2540. (i) Nakanishi, W.; Hayashi, S. *J. Phys. Chem. A* **1999**, *103*, 6074. (j) McFarlane, W.; Rycroft, D. S.; Turner, C. J. *Bull. Soc. Chim. Belg*. **1977**, *86*, 457. (k) Birchall, T.; Gillespie, R. J.; Vekris, S. L. *Can. J. Chem*. **1965**, *43*, 1672. (l) Luthra, N. P.; Boccafuso, A. M.; Dunlap, R. B. Odom, J. D. *J. Organomet. Chem*. **1988**, *354*, 51. (m) Dawson, W. H.; Odom, J. D. *J. Am. Chem. Soc*. **1977**, *99*, 8352. (n) Christaens, L.; Piette, J. L.; Laitem, L.; Baiwir, M.; Denoel, J.; Llabres, G. *Org. Magn. Res*. **1976**, *8*, 354. (o) Gombler, W. *Z. Naturforsch., B: Chem. Sci*. **1981**, *36B*, 1561. (p) Kojima, Y.; Ibi, K.;

Ishihara, H.; Murai, T.; Kato, S. *Bull. Chem. Soc. Jpn.* **1993**, *66*, 990. (q) Duddeck, H.; Wagner, P.; Müller, D.; Jaszberenyi, J. C. *Magn. Reson. Chem.* **1990**, *28*, 549. (r) Valeev, R. B.; Kalabin, G. A.; Kushnarev, D. F. *Zh. Org. Khim.* **1980**, *16*, 2482. (s) Kushnarev, D. F.; Kalabin, G. A.; Kyandzhetsian, R. A.; Magdesieva, N. N. *Zh. Org. Khim.* **1982**, *18*, 119. (t) Poleschner, H.; Radeglia, R.; Kuprat, M.; Richter, A. M.; Fandhänel, E. *J. Organomet. Chem.* **1987**, *327*, 7. (u) Gombler, W. *Z. Naturforsch., B: Chem. Sci.* **1981**, *36B*, 535. (v) Kuhn, N.; Henkel, G.; Schumann, H.; Froelich, R. *Z. Naturforsch., B: Chem. Sci.* **1990**, *45B*, 1010. (w) McFarlane, W.; Rycroft, D. S. *J. Chem. Soc., Chem. Commun.* **1972**, 902. (x) McFarlane, W.; Rycroft, D. S. *J. Chem. Soc., Dalton Trans.* **1973**, 2162. (y) Socol, S. M.; Verkade, J. G. *Inorg. Chem.* **1986**, *26*, 2658. (z) Arnold, D. E. J.; Dryburgh, J. S.; Ebsworth, E. A. V.; Rankin, D. W. H. *J. Chem. Soc., Dalton Trans.* **1972**, 2518. (aa) Koch, W.; Lutz, O.; Nolle, A. *Z. Naturforsch., A: Phys. Sci.* **1978**, *33A*, 1025. (bb) Snatzke, G. *The Chemistry of Organic Selenium and Tellurium Compounds*; Patai, S., Rappoport, Z., Eds.; Wiley: Chichester, 1986; Vol. 1, p 667. (cc) Pan, W. H.; Fackler, J. P.; Kargol, J. A.; Burmeister, J. L. *Inorg. Chim. Acta* **1980**, *44*, L95. (dd) Lamoureux, M.; Milne, J. B. *Polyhedron* **1990**, *9*, 589. (ee) Milne, J. B.; Williams, A. J. *Inorg. Chem.* **1992**, *31*, 4534. (ff) Milne, J. B.; Williams, A. J. *Can. J. Chem.* **1992**, *70*, 693. (gg) Collins, M. J.; Ratcliffe, C. I.; Ripmeester, J. A. *J. Magn. Reson.* **1986**, *68*, 172. (hh) Damerius, R.; Huppmann, P.; Lentz, D.; Seppelt, K. *J. Chem. Soc., Dalton Trans.* **1984**, 2821. (ii) Milne, J. *J. Chem. Soc., Chem. Commun.* **1991**, 1048.

(35) (a) Brown, I. D.; Crump, D. B.; Gillespie, R. J. *Inorg. Chem.* **1971**, *10*, 2319. (b) Schrobilgen, G. J.; Burns, R. C.; Granger, P. *J. Chem. Soc., Chem. Commun.* **1978**, 957. (c) Collins, M. J.; Gillespie, R. J.; Sawyer, J. F.; Schrobilgen, G. J. *Inorg. Chem.* **1986**, *25*, 2053.

(36) Tuononen, H. M.; Suontamo, R.; Valkonen, J.; Laitinen, R. S. *J. Phys. Chem. A* **2004**, *108*, 5670.

(37) The gas-phase value is used for H_2Se. The selenide dianion is excluded from the plot.

(38) (a) Wiberg, K. B. *J. Comput. Chem.* **1999**, *20*, 1299. (b) Sefzik, T. H.; Turco, D.; Iuliucci, R. J.; Facelli, J. C. *J. Phys. Chem. A* **2005**, *109*, 1180.

(39) Reference deleted in press.

(40) Trend, J. E. Dissertation, University of Wisconsin, 1976.

(41) Cullen, E. R.; Guziec, F. S.; Murphy, C. J.; Wong, T. C.; Andersen, K. K. *J. Am. Chem. Soc.* **1981**, *103*, 7055.

(42) Dowd, D.; Gettins, P. *Magn. Reson. Chem.* **1988**, *26*, 1.

(43) Paulmier, C.; Lerouge, P.; Outurquin, F.; Chapelle, S.; Granger, P. *Magn. Reson. Chem.* **1987**, *25*, 955.

(44) (a) Nakanishi, W.; Matsomoto, S.; Ikeda, Y.; Sugawara, T.; Kawada, Y.; Iwamura, H. *Chem. Lett.* **1981**, 1353. (b) Nakanishi, W.; Ikeda, Y.; Iwamura, H. *Org. Magn. Reson.* **1982**, *20*, 117.

CT050136T

JCTC Journal of Chemical Theory and Computation

Distributed Multipole Analysis: Stability for Large Basis Sets

Anthony J. Stone*

University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Received August 2, 2005

Abstract: The distributed multipole analysis procedure, for describing a molecular charge distribution in terms of multipole moments on the individual atoms (or other sites) of the molecule, is not stable with respect to a change of basis set, and indeed, the calculated moments change substantially and unpredictably when the basis set is improved, even though the resulting electrostatic potential changes very little. A revised procedure is proposed, which uses grid-based quadrature for partitioning the contributions to the charge density from diffuse basis functions. The resulting procedure is very stable, and the calculated multipole moments converge rapidly to stable values as the size of the basis is increased.

I. Introduction

In many calculations involving the interactions between molecules, it is necessary to calculate the electrostatic interaction between molecular charge distributions. The exact expression, however, involves an integral over both charge distributions and is much too time consuming for most applications. Instead, it is usual to use a multipole approximation, and it is well-understood that, for all but the very smallest molecules, it is necessary to use some form of distributed multipole model, in which multipole moments are attached to several sites in each molecule, usually to each atom but sometimes to groups of atoms and sometimes to additional sites as well as the atoms. Often, only charges are used, but it is recognized that, for accurate work, atomic dipoles and perhaps higher moments are needed.

Distributed multipole analysis (DMA)[1−3] is a simple, fast procedure that assigns multipole moments to each atom or other specified site in a molecule. It is exact, in the sense that the overall multipole moments constructed from the distributed moments exactly reproduce the overall multipole moments of the charge distribution. However, it has become clear, in the 25 years since the method was first introduced, that it is not stable with respect to changes of the basis set. Very large basis sets with diffuse functions are commonly used in accurate work nowadays, and these are particularly troublesome, yielding distributed multipoles that may not correspond to physical expectations. Moreover, there is no convergence as the basis set is increased; on the contrary, the distributed multipoles vary more and more wildly.

It should be emphasized that the overall multipole moments remain exact, but this is not helpful for short-range interactions, where the expansion in terms of overall multipoles diverges, and the interaction is determined by the multipole moments for a few sites on each molecule.

The DMA procedure works as follows. A Gaussian basis function centered at A is the product of a Gaussian function $\exp[-\zeta_A(r - A)^2]$ with a low-degree polynomial (degree 0, i.e., constant, for s functions, 1 for p functions, and so on). Boys showed[4] that the product of two such functions, one at A and the other at B, is another Gaussian of the same form. The degree of its polynomial is the sum of the degrees of the original Gaussians, and the Gaussian part is $\exp[-(\zeta_A + \zeta_B)(r - P)^2]$, where the "overlap center" $P = (\zeta_A A + \zeta_B B)/(\zeta_A + \zeta_B)$. The electron density is a sum of such products, with coefficients determined from the density matrix. Any individual product can be described exactly in terms of a sum of multipole moments of ranks up to the degree of its polynomial; so, the product of two s functions can be expressed as a pure charge, the product of an s with a p as a charge plus dipole, and so on. Distributed multipole analysis evaluates these exact representations and ap-

* Author to whom correspondence should be addressed. Phone: +44 1223 336375. Fax: +44 1223 336362. E-mail: ajs1@cam.ac.uk.

10.1021/ct050190+ CCC: $30.25
Published on Web 09/13/2005

proximates each of them by a multipole expansion (nonterminating, in general) about the nearest of a set of expansion sites, which comprises some or all of the atomic nuclei, and perhaps some additional sites.

The nature of the problem with this method can be easily understood if we contemplate a simple molecule such as carbon dioxide. In a large and diffuse basis, we may have a diffuse p_z primitive on the C atom (we take z along the molecular axis), which is very similar to an out-of-phase combination of diffuse s primitives on the O atoms. The densities associated with these two functions are $(p_z^C)^2$ and $^1/_2(s^A - s^B)^2 = {}^1/_2(s^A)^2 - s^A s^B + {}^1/_2(s^B)^2$, respectively, where A and B denote the two oxygen atoms and C the carbon. In the distributed-multipole procedure, each product of primitives is represented by multipoles on the site nearest to its overlap center. If we assume that both functions are normalized, the density $(p_z^C)^2$ becomes a charge -1 plus a pure quadrupole Θ on the C atom. The density $^1/_2(s^A)^2 - s^A s^B + {}^1/_2(s^B)^2$ becomes a set of charges, $-{}^1/_2 q$ on each O atom and $+q - 1$ on the C.

Now if we change the basis, the exponents of the diffuse functions will be different. In one basis, the C p_z function may be variationally preferred as a description of this particular aspect of the charge distribution, while in another, the combination of s functions may be preferred. The actual charge distribution may change very little, especially if both basis sets are large, but the distributed-multipole description may be very different. As the size of the basis set is increased, the opportunities for this kind of ambiguity increase, so the problem gets worse. Far from approaching a converged description, the distributed multipoles vary ever more wildly as the basis size is increased.

At short distances (for example, in a close end-on contact with one of the oxygen atoms), it may be supposed, and often is assumed, that the electrostatic interaction obtained from the distributed multipoles could also vary significantly. In fact, we shall see that this is not the case, but the lack of convergence in the description is a very unsatisfactory feature.

II. A New Version of Distributed Multipole Analysis

The difficulty with standard DMA arises because the partitioning of the density between atoms is carried out in basis-function space. It has been recognized for some time that a partition in real space is more appropriate. Probably the best method of this type that is currently available is the atoms-in-molecules procedure, originally due to Bader[5] (see also references therein) and more recently developed as a method for distributed multipoles by Popelier.[6] This method works well, but it involves a very time-consuming procedure for determining the surfaces that separate one atom basin from another. Another recent method[7] uses the Hirshfeld partitioning method,[8,9] apportioning electron density between atoms in proportion to the free atom densities at the same distances from the nuclei.

The purpose of this paper is to propose a version of distributed multipole analysis that overcomes the strong basis-set dependence of the original version but is simpler to implement than other real-space partitioning schemes. It uses numerical quadrature for the diffuse functions, while retaining the original method for the more compact ones. This avoids any difficulties associated with the numerical quadrature of strongly peaked functions. The method used for the numerical quadrature is that of Becke,[10,11] with some minor differences. A grid of integration points is constructed around each atom, and the molecular grid is the union of these atom grids. Becke associates a set of atom weights with any point in space; the sum of the weights is unity, the nearest atom having a weight close to 1 and the others having weights close to zero. At the boundary between two atoms, the weight of one atom falls smoothly to zero as the weight of the other increases to 1. This is achieved by a "smoothing function", and the steepness of this function is a parameter of the quadrature algorithm. The density is partitioned in this way into overlapping regions, each assigned to one atom. In density-functional calculations, the parameter is usually chosen so that the smoothing function is quite steep; this means that many of the more distant points on each atom grid have weights close to zero and can be dropped, making the integration more efficient. For the distributed multipole analysis, however, it is advantageous to make it less steep, so that the regions associated with each atom interpenetrate and are more nearly spherical. The density assigned to each atom is integrated over the grid for that atom to yield atomic multipole moments referred to the atom nucleus as the origin.

For calculations of multipole moments, the quadrature grid needs to extend to large distances, because the multipole moment functions (the regular spherical harmonics) include radial factors of r^k for rank k, and the products of basis functions that contribute to moments of rank k themselves include radial factors r^n with $n \geq k$. Moreover, the regular spherical harmonics have a strong angular dependence at high rank, as do the high-angular-momentum basis functions, so the angular integral also needs to be more accurate than is usually necessary for Kohn−Sham calculations, where the integral involves the total density.

These considerations have prompted the development of a new program for distributed multipole analysis. It is called GDMA2, where the G stands for Gaussian and reflects the fact that the program is intended for use with the Gaussian suite of quantum chemistry programs, and it supersedes the older GDMA program, which uses the original distributed multipole analysis.

III. Examples and Discussion

III.1. Carbon Monoxide. The features that led to dissatisfaction with the original distributed multipole analysis are illustrated in Figure 1, where the multipole moments are plotted against the basis sets. It is clear that the distributed multipoles vary substantially with basis set and, in particular, that there is a systematic difference between the augmented and nonaugmented sets. Moreover, the values seem to be diverging rather than converging as the basis set is improved.

In contrast, the behavior of the total moments is just what one would expect. The results are poor for the unaugmented double-ζ basis, but the remaining values rapidly settle down to consistent values. The hexadecapole values are not quite

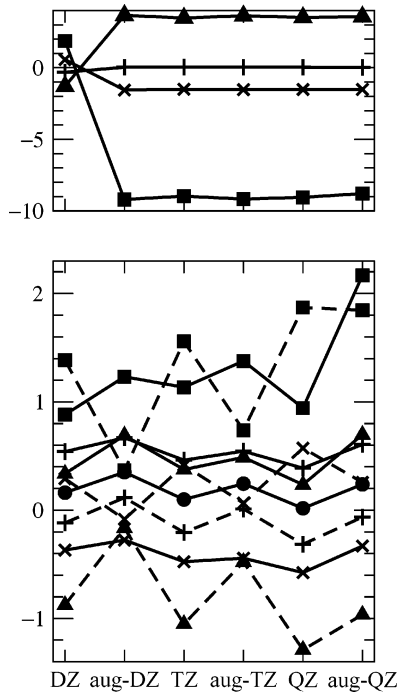


Figure 1. Distributed multipoles (bottom) and total multipole moments (top) for CO, in atomic units, using the original DMA algorithm. ● denotes charge, + denotes dipoles, × denotes quadrupoles, ▲ denotes octopoles, and ■ denotes hexadecapoles. In the lower diagram, solid lines refer to C and dashed lines to O. "*n*Z", *n* = D, T, Q, are abbreviations for "cc-pV*n* Z".

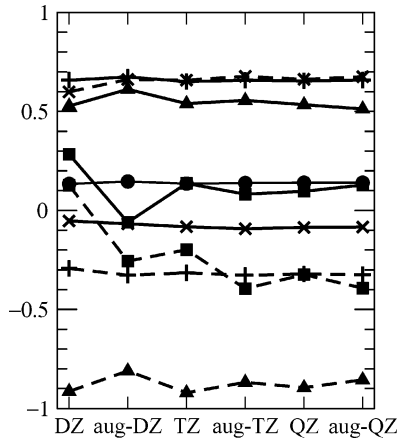


Figure 2. Distributed multipoles for CO, in atomic units, using the new DMA algorithm. ● denotes charge, + denotes dipoles, × denotes quadrupoles, ▲ denotes octopoles, and ■ denotes hexadecapoles. Solid lines refer to C and dashed lines to O. "*n*Z", *n* = D, T, Q, are abbreviations for "cc-pV*n* Z".

converged with the QZ basis, but they are affected by the high angular-momentum (*g*) basis functions that first appear in the QZ basis.

Figure 2 shows the distributed multipoles for the revised method. It is evident that the calculated values are converging steadily; the behavior as the basis set is improved is much more satisfactory.

At the same time, it is important to realize that the difference is a matter more of perception than reality. Although the distributed multipole moments may look very different for the two methods, the resulting electrostatic potentials are almost identical. In Table 1, the aug-cc-pVQZ multipole moments are shown for the cases illustrated in Figures 1 and 2, together with the differences. The atom charges differ by 0.35 *e* and are different in sign between old and new, and the higher multipoles also differ substantially. However, the electrostatic potential on the vdW × 2 surface (i.e., the surface of the solid that is the union of spheres of twice the van der Waals radius around each atom) differs by less than 5×10^{-5} au, or about 1.2 mV, everywhere. Note particularly that the values shown at the bottom of the last column are maximum and minimum potentials, on the vdW × 2 surface, arising from the set of multipole differences shown in the last column.

Table 1. Multipole Moments on the C and O Atoms of Carbon Monoxide Using the Original Calculation Method (Switch = 0) and the New Method, with Grid-Based Quadrature for Products of Primitives with a Total Exponent Less than 4[a]

multipole	switch = 0	switch = 4	difference
C Q_0	0.244 75	−0.108 77	−0.353 52
C Q_1	0.544 70	0.367 55	−0.177 15
C Q_2	−0.443 94	−0.203 48	0.240 46
C Q_3	0.487 55	0.870 63	0.383 08
C Q_4	1.376 77	0.800 72	−0.576 05
O Q_0	−0.244 75	0.108 77	0.353 52
O Q_1	0.011 59	−0.564 95	−0.576 54
O Q_2	0.064 39	0.688 67	0.624 28
O Q_3	−0.478 34	−0.432 66	0.045 68
O Q_4	0.738 07	−0.857 87	−1.595 94
V_{max}/V	0.073 34	0.074 23	0.001 23
V_{min}/V	−0.146 99	−0.148 23	−0.001 06

[a] The last two lines of the table show the maximum and minimum electrostatic potential on the vdW × 2 surface and the maximum and minimum difference.

This corresponds to energy differences, for a unit charge at the vdW × 2 surface, of about 50 microHartrees. Although the *multipole moments* differ considerably, the *electrostatic potentials* hardly differ at all. The wide variation in multipole descriptions provided by different basis sets is not as troublesome as it appears at first sight—any set of multipoles will serve, provided that they are calculated with a reasonably good basis set.

Nevertheless, the more stable description provided by the new method has the advantage that it converges more satisfactorily as the basis set is improved, so that it becomes possible to assess when the basis set is of adequate quality, and multipole moments of related systems are more comparable and transferable.

III.2. Formamide. As a second example, we consider formamide. Here, there are many more multipole moments, too many to tabulate or even to display diagrammatically in detail. Figure 3 shows the behavior of the atom moments as the basis set is improved. The solid lines denote the atom charges; the dashed lines the magnitudes of each atom dipole, $(Q_{10}^2 + Q_{11c}^2 + Q_{11s}^2)^{1/2}$; and the dotted lines the magnitudes of the atom quadrupoles, $(\sum_m Q_{2m}^2)^{1/2}$. The C, N, and O charges are identified, but the other moments have not been identified individually, as the object is just to contrast the

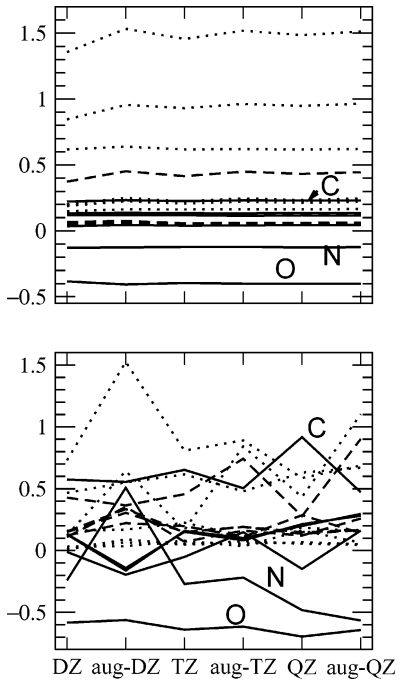


Figure 3. Distributed multipoles for formamide, in atomic units, using the original DMA algorithm (bottom) and the new (top). Solid lines denote atom charges, dashed lines denote atom dipoles, and dotted lines denote quadrupoles. "*n*Z", *n* = D, T, Q, are abbreviations for "cc-pV*n*Z".

Table 2. Maximum and Minimum Electrostatic Potentials (V) on the vdW × 2 Surface of Formamide, for Original and New Methods of Calculation and for Several Basis Sets, and the Maximum Differences between Old and New Potentials for Each Basis

	switch = 0		switch = 4		difference
basis	min	max	min	max	max
cc-pVDZ	−0.744	0.714	−0.745	0.712	0.0034
aug-cc-pVDZ	−0.800	0.748	−0.800	0.749	0.0073
cc-pVTZ	−0.782	0.744	−0.782	0.743	0.0041
aug-cc-pVTZ	−0.796	0.745	−0.782	0.743	0.0072
cc-pVQZ	−0.794	0.749	−0.794	0.748	0.0047
aug-cc-pVQZ	−0.797	0.750	−0.797	0.754	0.0069

behaviors of the two methods. It is clear that the original DMA shows no sign of settling down to a converged value as the basis set is improved; on the contrary, some values, especially the C and N charges, appear to be diverging. The behavior of the new method is considerably better.

As in the case of CO, the electrostatic potential is well-described by most of the basis sets, whichever method is used. In this case, the original DMA leads to very large charges on the atoms when the cc-pVQZ basis is used. Nevertheless, the differences between the old and new methods in the electrostatic potential on the vdW × 2 surface are still less than 1% of the potential itself, as Table 2 shows. The corresponding energy differences, for a unit charge at the vdW × 2 surface, are a few meV, or less than 0.5 kJ mol^{-1}, and are significantly smaller than the differences between the better basis sets when the same method of calculation is used for both.

Nevertheless, the much greater stability of the multipole description with respect to changes of basis set, as illustrated in Figure 3, is sufficient reason for preferring the new method.

IV. The Program

The GDMA program is written in Fortran90 and is self-contained. It requires as input a Gaussian03[12] (or earlier) formatted check-point file, together with a small controlling data file. It can currently handle basis functions up to *g*; an extension to handle *h* basis sets is planned. The program can, in principle, determine distributed multipole moments up to rank 10, but the higher moments are likely to be inaccurate unless very large angular integration grids are used. (The original method gives results to the full accuracy of the input density matrix and may be preferred if very high ranks are needed.) Various parameters can be controlled in the data file, though the default values will usually be adequate. The default number of radial grid points is 80; more may be needed if moments higher than hexadecapole are required. The angular integration is usually carried out on a 590-point Lebedev grid;[13] as noted above, a large angular grid is needed if high-rank moments are required, because of the strong angle-dependence of the multipole operators. The number of grid points may be specified; any number may be requested, and the program will use the next-larger size of grid. Inadequacies in the angular grid are usually evident in small spurious nonzero values for multipole moments that should be zero by symmetry. The program can also use Gauss−Legendre integration over the angular coordinates if preferred.

The radial integration is carried out using Euler−Maclaurin quadrature, as formulated by Murray et al.[11] Following normal practice, the radial grid is scaled by reference to Bragg−Slater covalent radii,[14] but with the hydrogen radius taken to be 0.5 Å, twice the Bragg−Slater value. For the cutoff between atoms, we use the scheme proposed by Becke.[10] The softness or sharpness of the boundary between atoms is controlled by Becke's cutoff parameter *k*. The effect of various values is illustrated in Figure 1 of Becke's paper. Becke recommends a value of 3, which gives a fairly soft cutoff; sharper cutoff functions have been used,[11] but the value of 3 seems to be satisfactory for the present purpose and is the default for the GDMA program. Other values can be specified if required. The position of the cutoff is controlled, again as proposed by Becke,[10] by assigning radii to each site and transforming the cutoff function by reference to the ratio of radii. It is customary to use the Bragg−Slater radii in the Becke partitioning, but this leads to rather implausible multipole moments—rather large and not in keeping with chemical intuition—though, like the multipoles from the original method, they lead to accurate electrostatic potentials. The default procedure in the GDMA program is to set all the radii equal for this purpose, so that the boundary between neighboring atoms comes halfway between them. Any additional multipole sites are also assigned the same radius. However, any of the radii may be changed if required, in the data file controlling the program, and a somewhat smaller radius is probably appropriate for hydrogen atoms.

The results given above for formamide were obtained with the hydrogen radius set to 0.35 Å and the rest to 0.65 Å.

The other main controlling parameter determines the switch between old and new integration methods. If the sum $\zeta_A + \zeta_B$ of exponents in a product of primitives is larger than the specified switch value, the product function is regarded as compact and can be efficiently and accurately integrated using Gauss−Hermite quadrature. This is the method used in the original DMA and GDMA programs. If the sum of exponents is smaller than the switch value, the product function is relatively diffuse, and it is evaluated on the quadrature grid to accumulate the diffuse contribution to the total density. Finally, the multipole moments are evaluated from the diffuse density, the integration being carried out for each atom grid separately and the moments assigned to the corresponding atom. A value of 4 has been found satisfactory for the switch between methods. A value of zero causes the old method to be used throughout.

Integration over a grid is a slow process, and treating every product of primitives individually makes it even slower, so the new method is considerably slower than the old—by a factor of 2 orders of magnitude or more—but the computation times are still short enough not to be a concern. For formamide, with a cc-pVQZ basis set (255 basis functions) and a large integration grid, the time needed is a few minutes on a Pentium 4 workstation. The CPU time scales roughly as the square of the number of diffuse functions in the basis.

The program can be downloaded from my web site, www.stone.ch.cam.ac.uk, and full documentation is provided.

V. Conclusions

Although the distributed multipole analysis procedure, as originally defined, was entirely satisfactory 25 years ago, when small basis sets were the norm and few calculations included diffuse functions, it becomes less satisfactory for modern large basis sets with many diffuse functions. As the basis set size is increased, the distributed multipole moments do not approach converged values but oscillate in a divergent fashion. The overall molecular moments, and the electrostatic potential derived from the distributed multipoles, do converge satisfactorily with increasing basis set size, but the lack of converged distributed multipoles is, nevertheless, an unsatisfactory feature.

The revised method described in this paper deals with this problem by integrating the diffuse contributions to the electron density by grid-based quadrature, while the more compact functions are handled by exact Gauss−Hermite quadrature as in the original procedure. The resulting method is very stable with respect to increasing basis set size and, indeed, reaches convergence for relatively modest basis sets.

References

(1) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233−239.

(2) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047−1064.

(3) Stone, A. J. *The Theory of Intermolecular Forces*, International Series of Monographs in Chemistry; Clarendon Press: Oxford, U. K., 1996.

(4) Boys, S. F. *Proc. R. Soc. London, Ser. A* **1950**, *200*, 542−554.

(5) Bader, R. *Atoms in Molecules*; Clarendon Press: Oxford, U. K., 1990.

(6) Popelier, P. L. A. *Atoms in Molecules: An Introduction*; Prentice-Hall: Harlow, U. K., 2000.

(7) Harrison, J. F. *J. Chem. Phys.* **2003**, *119*, 8763−8764.

(8) Hirshfeld, F. L. *J. Mol. Struct.* **1985**, *130*, 125−141.

(9) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129−138.

(10) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 2547−2553.

(11) Murray, C. W.; Handy, N. C.; Laming, G. J. *Mol. Phys.* **1993**, *78*, 997−1014.

(12) 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*; Gaussian, Inc.: Pittsburgh, PA, 2003. For a catalogue of basis sets, see http://www.gaussian.com/g_ur/m_basis_sets.htm.

(13) Lebedev, V. I.; Laikov, D. N. *Dokl. Math.* **1999**, *59*, 477−481.

(14) Slater, J. C. *J. Chem. Phys.* **1964**, *41*, 3199−3204.

CT050190+

JCTC Journal of Chemical Theory and Computation

SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute−Water Clusters

Casey P. Kelly, Christopher J. Cramer,* and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, 207 Pleasant St. SE, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received June 28, 2005

Abstract: A new charge model, called Charge Model 4 (CM4), and a new continuum solvent model, called Solvation Model 6 (SM6), are presented. Using a database of aqueous solvation free energies for 273 neutrals, 112 ions, and 31 ion−water clusters, parameter sets for the mPW0 hybrid density functional of Adamo and Barone (Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664−675) were optimized for use with the following four basis sets: MIDI!6D, 6-31G(d), 6-31+G(d), and 6-31+G(d,p). SM6 separates the observable aqueous solvation free energy into two different components: one arising from long-range bulk electrostatic effects and a second from short-range interactions between the solute and solvent molecules in the first solvation shell. This partition of the observable solvation free energy allows SM6 to effectively model a wide range of solutes. For the 273 neutral solutes in the test set, SM6 achieves an average error of ∼0.50 kcal/mol in the aqueous solvation free energies. For solutes, especially ions, that have highly concentrated regions of charge density, adding an explicit water molecule to the calculation significantly improves the performance of SM6 for predicting solvation free energies. The performance of SM6 was tested against several other continuum models, including SM5.43R and several different implementations of the Polarizable Continuum Model (PCM). For both neutral and ionic solutes, SM6 outperforms all of the models against which it was tested. Also, SM6 is the only model (except for one with an average error 3.4 times larger) that improves when an explicit solvent molecule is added to solutes with concentrated charge densities. Thus, in SM6, unlike the other continuum models tested here, adding one or more explicit solvent molecules to the calculation is an effective strategy for improving the prediction of the aqueous solvation free energies of solutes with strong local solute−solvent interactions. This is important, because local solute−solvent interactions are not specifically accounted for by bulk electrostatics, but modeling these interactions correctly is important for predicting the aqueous solvation free energies of certain solutes. Finally, SM6 retains its accuracy when used in conjunction with the B3LYP and B3PW91 functionals, and in fact the solvation parameters obtained with a given basis set may be used with any good density functional or fraction of Hartree−Fock exchange.

1. Introduction

Continuum solvent models are an attractive alternative to explicit solvent approaches, because they require less computational effort, making them applicable to larger solutes, extensive conformational analysis, and large libraries of compounds. Continuum solvent models have advanced to a point where aqueous solvation free energies of typical neutral organic solutes can usually be predicted accurately to better than 1 kcal/mol. However, the development of methods for accurately predicting aqueous solvation free energies of ionic solutes has been much less successful. In part, this is due to

* Corresponding author e-mail: cramer@chem.umn.edu (C.J.C.) and truhlar@chem.umn.edu (D.G.T).

10.1021/ct050164b CCC: $30.25
Published on Web 10/22/2005

the limited availability of experimental solvation free energies for ionic solutes. Unlike neutral solutes, for which aqueous solvation free energies can be obtained directly from partition coefficients of solutes between the gas phase and dilute aqueous solution, aqueous solvation free energies of ionic solutes must be determined from other experimental observables. Because of this, there is a certain degree of uncertainty associated with making direct comparisons between calculated and experimental aqueous solvation free energies for ions, making the development of a model that is able to treat neutral and ionic solutes at the same level of accuracy a very challenging task. In addition, because of strong electrostatic effects arising from localized solute−solvent interactions, the magnitudes of solvation free energies are much greater for ions than for neutrals, requiring smaller percentage errors for the same absolute accuracy. Because the differential solvation free energy between a given acid/base pair can be used in various thermodynamic cycles to determine pK_a, developing a model that can be used to predict these free energies accurately is a high priority.

A number of different strategies have been used to account for short-range interactions within the framework of continuum solvation theory.[1,2] For example, the SM*x* series of models developed by our co-workers[3−17] and us augments the electrostatic portion of the calculated solvation free energy with an empirical term that accounts for, among other things, deviations of short-range interactions, primarily those in the first solvation shell, from the bulk electrostatic limit. Although this approach has been very successful in predicting solvation free energies of neutral solutes, it remains unclear whether this type of correction to the solvation free energy can be applied to ionic solutes with the same success.

A key issue in predicting the large electrostatic effects involved in solvation of ions is determining the shape of the cavity that is used to define the boundary between the electronic distribution of the solute and the continuum solvent. In all of our recent SM*x* models[7−16] (including the one presented in this article), a single set of radii that are dependent only on the atomic number of the given atom are used to build up the molecular cavity. More elaborate methods for assigning atomic radii depend on the local chemical environment of the atom.[18−23] One such prescription, called the united atom for Hartree−Fock (UAHF) method,[23] assigns radii to atoms based on their hybridization states, what other atoms are bonded to them, and their formal charge. These radii are often used in conjunction with the popular polarizable continuum models (PCMs)[24−34] to predict aqueous solvation free energies. Other methods have been proposed in which the atomic radii depend on partial atomic charge, and several groups have had some success using charge-dependent atomic radii in continuum solvation calculations,[35−41] although the work has been limited to a small number of solutes. Other methods define the solute cavity as the contour on which the solute electronic density is equal to some constant value.[42] For example, Chipman has recently shown[43] that using a value of 0.001 e/a_0^3 for the isodensity contour along with a continuum model[44] results in accurate aqueous solvation free energies for a series of protonated amines. However, Chipman also showed in his work[43] that for a series of oxygen-containing anions, no single common contour value could be used to accurately predict their absolute aqueous solvation free energies. Chipman attributed this finding to the inability of a continuum model alone to account for strong anion−water interactions in the first solvation shell and suggested that better results might be obtained by augmenting continuum solvent calculations with other complementary methods that are especially designed to account for specific short-range interactions.

Adding explicit solvent molecules has been a popular strategy for trying to incorporate the effects of specific solute−solvent interactions into continuum solvent calculations. Often, this involves treating enough solvent molecules classically or quantum mechanically to account for at least the entire first solvation shell around the given solute.[45] Depending on the solute, this may require a large number of explicit solvent molecules, which can lead to a significant increase in the amount of computational effort expended. In addition to this problem, there are several other potential problems associated with treating solvent molecules explicitly. First, for many solutes, there is no easy way to determine the number or orientation of explicit water molecules in the first solvation shell. For example, X-ray diffraction experiments[46] and various theoretical calculations[47−51] lead to average coordination numbers ranging from 6 to 9.3 for the Ca^{2+} ion, suggesting that several different solvation structures exist. Second, even for solutes for which the first solvation shell is well defined, to properly treat even a few solvent molecules explicitly will most likely involve the need to sample over a large number of conformations that are local minima. Finally, introducing explicit solvent molecules will not yield more accurate solvation free energies if the level of theory used to treat the system is not high enough. Because properly treating nonbonded interactions usually requires treatment of electron correlation and the use of fairly large basis sets,[52] any realistic attempt at using solute−water clusters to calculate aqueous solvation free energies requires an accurate treatment of the entire solute−solvent system, which is practical only for small numbers of solvent molecules. Despite these problems, hybrid approaches combining quantal and classical treatments of solvent molecules have had some success in predicting aqueous solvation free energies of ions. For example, Pliego and Riveros showed[53] that for a test set of 17 ions, including several explicit water molecules in the continuum calculation significantly improved the performance of the model. To determine the number of explicit solvent molecules required in the calculation, these workers developed an approach in which the aqueous solvation free energy of the bare solute is minimized with respect to the number of coordinating waters.[54] Besides predicting solvation free energies of ions, this approach has also been used to predict solvatochromic shifts,[55,56] where explicit solute−solvent effects between the electronically excited solute and surrounding solvent molecules can have large effects on both the magnitude and direction of the shift.

In the present paper, we will present a new continuum solvent model called Solvation Model 6 (SM6). This model is similar to our most recently developed previous continuum model, called SM5.43R,[15,16] but improves on it in a number of significant ways. In both of these models, SM6 and

SM5.43R, the aqueous solvation free energy is calculated as a sum of free energies arising from long-range bulk electrostatic effects, which are calculated by a self-consistent reaction field (SCRF) calculation,[12,57,58] and those from nonbulk electrostatic effects, which are calculated using the solvent accessible surface area (SASA)[59,60] of the solute and a set of atomic surface tensions that depend on a set of empirical parameters and the geometry of the solute. SM6 differs from SM5.43R in two important ways. First, SM6 uses an improved charge model, called Charge Model 4 (CM4), for assigning partial atomic charges. CM4 is a new charge model developed as part of the present effort, and it is presented later in the text and in the Supporting Information. Second, SM6 is parametrized with a training set of aqueous solvation free energies that has been improved in three ways: (1) the neutral portion of the training set has been extended to include molecules containing certain functionalities that were not present in the SM5.43R training set, (2) we use a larger and improved set of data for ionic solutes, and (3) aqueous solvation data for various ion−water clusters and the water dimer have been added, and the entire philosophy of the parametrization of Coulomb radii is changed to reflect the use of cluster data in place of bare-ion data for cases where continuum solvent models are expected to be inadequate for bare ions.

SM6 calculations may use any reasonable gas-phase or liquid-phase geometry of the solute to calculate its aqueous solvation free energy. In addition, geometry optimizations in the liquid phase using analytical free-energy gradients can be efficiently carried out.[13] We previously denoted the case for which aqueous solvation free energies were calculated using gas-phase geometries with the suffix "R" and those in which they were calculated using liquid-phase geometries by dropping the "R" suffix (which stands for "rigid"); here we will drop the "R" suffix in all cases and use the standard Pople notation. For example, a single-point SM6 calculation at the MPW25/6-31+G(d,p) level using a gas-phase geometry optimized at the MPW25/MIDI! level of theory would be written as SM6/MPW25/6-31+G(d,p)//MPW25/MIDI!, whereas if the consistent liquid-phase geometry were used, this calculation would be written as SM6/MPW25/6-31+G(d,p). A solvation free energy calculated by SM6/MPW25/6-31+G(d,p) at a gas-phase geometry computed by the same electronic structure level (i.e., MPW25/6-31+G(d,p)) can be denoted SM6/MPW25/6-31+G(d,p)//MPW25/6-31+G(d,p) or for short SM6/MPW25/6-31+G(d,p)//*g*, where //*g* denotes a gas-phase geometry at the same level.

Four new parametrizations of SM6 for the MPW*X* hybrid density functional will be presented, where each parametrization uses a particular basis set. These four basis sets are MIDI!6D[61,62] and Pople's[63] popular 6-31G(d), 6-31+G(d), and 6-31+G(d,p) basis sets. The MPW*X* functional uses Barone and Adamo's[64] modified version of Perdew and Wang's exchange functional,[65] Perdew and Wang's PW91 correlation functional, and a percentage *X* of Hartree−Fock exchange.[16] The parameters presented here can be used with any value of *X*, which is a stability feature pointed out in a previous paper.[16] This is particularly useful, because depending on the problem, it may be advantageous to optimize *X* in the gas phase or in solution. For example, $X = 42.8$ has been optimized for kinetics (resulting in the MPW1K functional[66]), $X = 40.6$ for $Y^- + RY$ nucleophilic substitution reactions (Y = F, Cl; the MPW1N functional[67]), $X = 6$ for conformations of sugars,[68] and $X = 25$ has been suggested for predicting heats of formation[64] (this is the mPW1PW91 functional of Barone and Adamo,[64] which they also call mPW0 and which we also refer to as MPW25, or MPW*X*, with $X = 25$). We chose to base the present parametrizations on the MPW*X* hybrid density functional for two reasons. First, as mentioned above, methods for predicting various gas-phase properties that employ different values of *X* with this functional have already been developed, and it is useful to have a set of solvation parameters that can also be used with any *X*. Second, the MPW*X* functional has been shown to be more accurate than the popular B3LYP[69] and HF[70] methods for predicting energies of reaction and barrier heights.[66,71] Furthermore, two of the parametrizations presented here are for basis sets containing diffuse functions. This is important because diffuse functions are often required for accurate calculations of conformational energies and barrier heights.[72] Thus, the parametrizations based on the 6-31+G(d) and 6-31G+(d,p) basis sets are of special interest, because they can be applied in cases where one wants to use the same level of theory for calculating relative energies in the gas phase and in the aqueous phase.

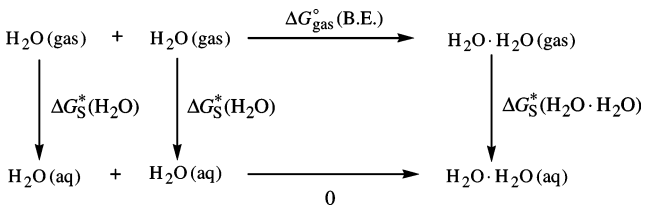
In addition to parametrizing a new charge model and a new aqueous solvent model, the present article has a third goal, namely, to ascertain what effect adding explicit solvent molecules has on the accuracy of continuum solvent models for predicting aqueous solvation free energies. For this, we added a *single* explicit water molecule to some of the solutes in our training set, and we used the resulting solute−solvent cluster to calculate the aqueous solvation free energy. Since the effort in this approach is modest because we are limiting the number of explicit water molecules to one and because the solute−solvent system will be modeled as a single, rigid conformation, the approach is very practical and does not have most of the problems associated with adding several explicit solvent molecules that were outlined above. Furthermore, the comparison between aqueous solvation free energies calculated using bare solutes to those calculated using solute−solvent clusters provides insight into whether continuum solvent models are appropriate for calculating aqueous solvation free energies of solute−water clusters as well as whether the performance of these models can be improved in cases where specific localized solute−solvent interactions are expected to play a large role in determining an aqueous solvation free energy.

Section 2 presents the experimental data used to train and test the new model, which is itself presented in section 3. Section 4 is concerned with parametrization, and section 5 gives the results. Sections 6 and 7 present discussion and conclusions, respectively.

2. Experimental Data

2.1. Standard States. All data and calculations are for 298 K. All experimental and calculated gas-phase free energies are tabulated using an ideal gas at 1 atm as the reference state. Free energies that employ this standard state definition will be denoted by the superscript "°". All experimental and

Scheme 1



calculated solvation free energies are tabulated for an ideal gas at a gas-phase concentration of 1 mol/L dissolving as an ideal solution at a liquid-phase concentration of 1 mol/L. Free energies that employ this standard state definition will be denoted by the superscript "*". The relationship between these two standard states is

$$\Delta G^*_{gas} = \Delta G^\circ_{gas} + \Delta G^{\circ\rightarrow *} \quad (1)$$

and

$$\Delta G^*_S = \Delta G^\circ_S - \Delta G^{\circ\rightarrow *} \quad (2)$$

where[73]

$$\Delta G^{\circ\rightarrow *} = RT\ln(24.46) \quad (3)$$

At 298 K $\Delta G^{\circ\rightarrow *}$ equals 1.89 kcal/mol.

2.2. Neutral Solutes. For neutral solutes, we start with all of the experimental aqueous solvation free energy data from the previously described SM5.43R training set,[15] which includes 257 aqueous solvation free energies for solutes containing at most H, C, N, O, F, P, S, Cl, and/or Br. To these data, we added additional aqueous solvation data for various reasons. Aqueous solvation free energies were added for methylhydrazine and 1,1-dimethylhydrazine in order to test the performance of SM6 without the surface tension previously[7−17,74,75] applied to solutes containing hydrogen atoms in the vicinity of two nitrogen atoms (e.g., hydrazines and hydrazones with a hydrogen attached to one of the individual nitrogen atoms). Data for hydrogen peroxide, methyl peroxide, and ethyl peroxide were added because the SM5.43R training set does not contain any data for peroxides. The SM5.43R training set contains an aqueous solvation free energy for aniline but no other aniline analogues, so we added data for *ortho*-, *meta*-, and *para*-methylaniline as well as *N*-methyl-, *N*-ethyl-, and *N*,*N*-dimethylaniline. We also added 1,2-ethanediamine and 3-aminoaniline because the SM5.43R training set does not contain any data for solutes with more than one amino group in the same solute. Finally, we added urea and benzamide because the SM5.43R training set contains only one solute with urea functionality (1-dimethyl-3-phenylurea) and three solutes with amide functionality (acetamide, Z-*N*-methylacetamide, and *E*-*N*-methylacetamide).

We also examined the accuracy of several experimental solvation free energies that were used in earlier versions of our training sets. In our previous training sets that include hydrazine,[7−17,74,75] we used a value of −9.30 kcal/mol for the aqueous solvation free energy. Here, this value has been replaced by a value of −6.26 kcal/mol, which was obtained using experimental values for the vapor pressure and aqueous solubility.[76,77] For methyl benzoate, the value of −2.22 kcal/mol that was used in our previous training sets[15−17] has been replaced by a value of −3.91 kcal/mol that was also obtained using experimental values for the vapor pressure and aqueous solubility.[76,77]

We also added the aqueous solvation free energy for the water dimer. This free energy can be determined using the free energy cycle shown in Scheme 1, according to

$$\Delta G^*_S(H_2O{\cdot}H_2O) = 2\Delta G^*_S(H_2O) - \Delta G^\circ_{gas}(B.E.) + \Delta G^{\circ\rightarrow *} \quad (4)$$

where $\Delta G^*_S(H_2O)$ is the aqueous solvation free energy of water, and $\Delta G^\circ_{gas}(B.E.)$ is the gas-phase binding free energy, which equals $G^\circ_{gas}(H_2O{\cdot}H_2O) - 2G^\circ_{gas}(H_2O)$. Substituting experimental values of −6.31 kcal/mol for $\Delta G^*_S(H_2O)$ and 3.34 kcal/mol[78] for $\Delta G^\circ_{gas}(B.E.)$ into eq 4 gives −14.06 kcal/mol for the aqueous solvation free energy of the water dimer (at 298 K).

Adding the new data described to the SM5.43R training set and correcting the two values mentioned in the previous paragraph results in a new data set with a total of 273 aqueous solvation free energies for neutral solutes containing at most H, C, N, O, F, P, S, Cl, and/or Br. This will be called the SM6 neutral-solute aqueous free-energy-of-solvation data set. These solvation free energies are listed in Table S1 of the Supporting Information.

2.3. Ionic Solutes. For our previous models that included data for ionic soutes,[3−17,74,75] aqueous solvation free energies were taken from Florián and Warshel[79] and Pearson[80] and then updated based on changes in the accepted absolute aqueous solvation energy of the proton. Based on a careful analysis of the ionic data in the SM5.43R training set (which contains aqueous solvation free energies for 47 ionic solutes), we decided to develop two entirely new data sets of experimental solvation free energies for ionic solutes. The first new data set, which is listed in Tables 1 and 2, contains aqueous solvation free energies for 112 ionic solutes (60 anions and 52 cations). This will be called the SM6 unclustered-ion data set, and it is described in the next two paragraphs and further discussed in the two paragraphs after that.

For the aqueous solvation free energy of the proton, $\Delta G^*_S(H^+)$, we used Zhan and Dixon's value of −264 kcal/mol.[81] For the remaining cations, we used the thermodynamic cycle shown in Scheme 2 along with eq 5

$$\Delta G^*_{aq} = 2.303RT\mathrm{p}K_a \quad (5)$$

where pK_a is the negative common logarithm of the aqueous acid dissociation constant of AH and ΔG^*_{gas} is the same as ΔG^*_{aq} except for the gas phase. Using Scheme 2 then yields the standard-state aqueous solvation free energy of AH^+ as

$$\Delta G^*_S(AH^+) = \Delta G^\circ_{gas}(A) + \Delta G^{\circ\rightarrow *} + \Delta G^*_S(A) + \Delta G^*_S(H^+) - 2.303RT\mathrm{p}K_a(AH^+) \quad (6)$$

where $\Delta G^*_S(A)$ is the aqueous solvation free energy of the neutral species A, $\Delta G^*_S(H^+)$ is the aqueous solvation free energy of the proton, and $\Delta G^\circ_{gas}(A)$ is the gas-phase basicity of A, equal to $G^\circ_{gas}(A) + G^\circ_{gas}(H^+) - G^\circ_{gas}(AH^+)$. The

Table 1. Aqueous Solvation Free Energies (kcal/mol) of Bare Anions[a]

A^-	AH	ΔG°_{gas} (AH)[b]	pK_a (AH)[c]	ΔG^*_S(AH)[d]	$\Delta G^*_S(A^-)$
HC_2^-	acetylene	370.0 ± 1.8	21.7[e]	0.0	−74 ± 3
HCO_2^-	formic acid	338.3 ± 1.5	3.8	−7.0[f]	−78 ± 2
$CH_3CO_2^-$	acetic acid	341.4 ± 2.0	4.8	−6.7	−80 ± 3
$CH_3CH_2CO_2^-$	propanoic acid	340.4 ± 2.0	4.9	−6.5	−78 ± 3
$CH_3(CH_2)_4CO_2^-$	hexanoic acid	339.0 ± 2.1	4.9	−6.2	−76 ± 3
$H_2C{=}CHCO_2^-$	acrylic acid	337.2 ± 2.8	4.3	−6.6[f]	−76 ± 3
$CH_3COCO_2^-$	pyruvic acid	326.5 ± 2.8	2.5	−9.4[f]	−70 ± 3
$C_6H_5CO_2^-$	benzoic acid	333.0 ± 2.0	4.2	−7.9[f]	−73 ± 3
CH_3O^-	methanol	375.0 ± 0.6	15.5	−5.1	−97 ± 2
$C_2H_5O^-$	ethanol	371.3 ± 1.1	15.9	−5.0	−93 ± 2
$CH_3CH_2CH_2O^-$	1-propanol	369.4 ± 1.4	16.1	−4.8	−90 ± 2
$(CH_3)_2CHO^-$	2-propanol	368.8 ± 1.1	17.1	−4.8	−88 ± 2
$CH_3CH_2CHOCH_3^-$	2-butanol	367.5 ± 2.0	17.6	−4.7[f]	−86 ± 3
$C(CH_3)_3O^-$	*t*-butanol	367.9 ± 1.1	19.2	−4.5	−84 ± 2
$H_2C{=}CHCH_2O^-$	allyl alcohol	366.6 ± 2.8	15.5	−5.1	−88 ± 3
$C_6H_5CH_2O^-$	benzyl alcohol	363.4 ± 2.0	15.4	−6.6[f]	−87 ± 3
$CH_3OCH_2CH_2O^-$	2-methoxyethanol	366.8 ± 2.0	14.8	−6.8	−91 ± 3
$C_6H_5O^-$	phenol	342.9 ± 1.3	10.0	−6.6	−74 ± 2
o-$CH_3C_6H_4O^-$	2-methylphenol	342.4 ± 2.0	10.3	−5.9	−72 ± 3
m-$CH_3C_6H_4O^-$	3-methylphenol	343.3 ± 2.0	10.1	−5.5	−73 ± 3
p-$CH_3C_6H_4O^-$	4-methylphenol	343.8 ± 2.0	10.3	−6.1	−74 ± 3
$CH_2OHCH_2O^-$	1,2-ethanediol	360.9 ± 2.0	15.4	−9.3	−87 ± 3
m-$HOC_6H_4O^-$	3-hydroxyphenol	339.1 ± 2.0	9.3[g]	−11.4[f]	−76 ± 3
p-$HOC_6H_4O^-$	4-hydroxyphenol	343.1 ± 2.0	9.9[g]	−11.9[f]	−80 ± 3
CH_3OO^-	methyl hydroperoxide	367.6 ± 0.7	11.5	−5.3[h]	−95 ± 2
$CH_3CH_2OO^-$	ethyl hydroperoxide	363.9 ± 2.0	11.8	−5.3[h]	−91 ± 3
$CH_2(O)CH^-$	acetaldehyde	359.4 ± 2.0	16.5	−3.5[f]	−78 ± 3
$CH_3C(O)CH_2^-$	acetone	362.2 ± 2.0	19.0	−3.9	−78 ± 3
$CH_3CH_2C(O)CHCH_3^-$	3-pentanone	361.4 ± 2.0	19.9	−3.3[f]	−76 ± 3
CH_2CN^-	acetonitrile	366.0 ± 2.0	25.0	−3.9	−74 ± 3
$NCNH^-$	cyanamide	344.0 ± 2.0	10.3[g]	−6.2[f]	−74 ± 3
$C_6H_5NH^-$	aniline	359.1 ± 2.0	27.7	−5.5	−65 ± 3
$(C_6H_5)_2N^-$	diphenylamine	343.8 ± 2.0	22.4	−5.3[f]	−56 ± 3
CN^-	hydrogen cyanide	343.7 ± 0.3	9.2[i]	−3.1[f]	−72 ± 2
o-$NO_2C_6H_4O^-$	2-nitrophenol	329.5 ± 2.0	7.2	−4.5[f]	−62 ± 3
m-$NO_2C_6H_4O^-$	3-nitrophenol	327.6 ± 2.0	8.4	−9.6[f]	−64 ± 3
p-$NO_2C_6H_4O^-$	4-nitrophenol	320.9 ± 2.0	7.1	−10.6[f]	−60 ± 3
$CH_2NO_2^-$	nitromethane	350.4 ± 2.0	10.2	−4.0[f]	−78 ± 3
p-$NO_2C_6H_5NH^-$	4-nitroaniline	336.2 ± 2.0	18.2	−9.9[f]	−59 ± 3
CH_3CONH^-	acetamide	355.0 ± 2.0	15.1	−9.7	−82 ± 3
CH_3S^-	methanethiol	350.6 ± 2.0	10.3	−1.2	−76 ± 3
$CH_3CH_2S^-$	ethanethiol	348.9 ± 2.0	10.6	−1.3	−74 ± 3
$C_3H_7S^-$	1-propanethiol	347.9 ± 2.0	10.7	−1.1	−72 ± 3
$C_6H_5S^-$	thiophenol	333.8 ± 2.0	6.6	−2.6	−65 ± 3
$CH_3S(O)CH_2^-$	dimethyl sulfoxide	366.8 ± 2.0	33.0	−9.8[f]	−70 ± 3
CCl_3^-	chloroform	349.7 ± 2.0	24.0	−1.1	−56 ± 3
$CF_3CO_2^-$	trifluoroacetic acid	316.7 ± 2.0	0.5	−7.3[f]	−61 ± 3
$CH_2ClCO_2^-$	chloroacetic acid	328.9 ± 2.0	2.9	−8.7[f]	−72 ± 3
$CHCl_2CO_2^-$	dichloroacetic acid	321.5 ± 2.0	1.4	−6.6[f]	−64 ± 3
$CF_3CH_2O^-$	2,2,2-trifluoroethanol	354.1 ± 2.0	12.4	−4.3	−80 ± 3
$CH(CF_3)_2O^-$	1,1,1,3,3,3-hexafluoropropan-2-ol	338.4 ± 2.0	9.3	−3.8	−67 ± 3
$ClC_6H_4O^-$	2-chlorophenol	337.1 ± 2.0	8.5	−4.5[f]	−68 ± 3
$ClC_6H_4O^-$	4-chlorophenol	336.5 ± 2.0	9.4	−6.2[f]	−68 ± 3
HO^-	water	383.7 ± 0.2	15.7	−6.3	−107 ± 2
HO_2^-	hydrogen peroxide	368.6 ± 0.6	11.7	−8.6[h]	−99 ± 2
O_2^-	hydroperoxyl radical	346.7 ± 0.8	4.7	−7.0[j]	−85 ± 2
HS^-	hydrogen sulfide	344.9 ± 1.2	7.0	−0.7	−74 ± 2
F^-	hydrofluoric acid				−102 ± 2[k]
Cl^-	hydrochloric acid				−73 ± 2[k]
Br^-	hydrobromic acid				−66 ± 2[k]

[a] Aqueous solvation free energies are for a temperature of 298 K. [b] Gas-phase basicities taken from ref 83. [c] From ref 87, unless otherwise noted. [d] From the current data set unless otherwise noted. [e] Reference 89. [f] Reference 76. [g] Reference 91. [h] Reference 86. [i] Reference 92. [j] Reference 85. [k] Reference 82.

Table 2. Aqueous Solvation Free Energies of Bare Cations[a]

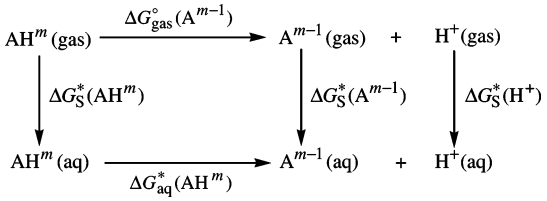
AH^+	AH	ΔG°_{gas} (AH^+)[b]	pK_a (AH^+) [c]	ΔG^*_S(AH)[d]	ΔG^*_S(AH^+)
$CH_3OH_2^+$	methanol	173.2 ± 2.0	−2.1	−5.1	−91 ± 3
$CH_3CH_2OH_2^+$	ethanol	178.0 ± 2.0	−1.9	−5.0	−86 ± 3
$(CH_3)_2OH^+$	dimethyl ether	182.7 ± 2.0	−2.5	−1.8	−78 ± 3
$(C_2H_5)_2OH^+$	diethyl ether	191.0 ± 2.0	−2.4	−1.8	−70 ± 3
$CH_3C(OH)CH_3^+$	acetone	186.9 ± 2.0	−2.9	−3.9	−75 ± 3
$CH_3COHC_6H_5^+$	acetophenone	198.2 ± 2.0	−4.3	−4.6	−63 ± 3
$CH_3NH_3^+$	methylamine	206.6 ± 2.0	10.6	−4.6	−74 ± 3
$CH_3(CH_2)_2NH_3^+$	*n*-propylamine	211.3 ± 2.0	10.6	−4.4	−70 ± 3
$(CH_3)_2CHNH_3^+$	isopropylamine	212.5 ± 2.0	10.6	−3.7[e]	−68 ± 3
$C(CH_3)_3NH_3^+$	*t*-butylamine	215.1 ± 2.0	10.7	−3.9[e]	−65 ± 3
c-$C_6H_{11}NH_3^+$	cyclohexanamine	215.0 ± 2.0	10.7	−5.1[e]	−67 ± 3
$H_2C{=}CHCH_2NH_3^+$	allylamine	209.2 ± 2.0	9.5	−4.3[e]	−70 ± 3
$(CH_3)_2NH_2^+$	dimethylamine	214.3 ± 2.0	10.7	−4.3	−67 ± 3
$(C_2H_5)_2NH_2^+$	diethylamine	219.7 ± 2.0	11.0	−4.1	−62 ± 3
(n-$C_3H_7)_2NH_2^+$	di-*n*-propylamine	222.1 ± 2.0	11.0	−3.7	−59 ± 3
$(H_2C{=}CHCH_2)_2NH_2^+$	diallylamine	219.0 ± 2.0	9.3	−4.0[e]	−60 ± 3
$(CH_3)_3NH^+$	trimethylamine	219.4 ± 2.0	9.8	−3.2	−59 ± 3
$(C_2H_5)_3NH^+$	triethylamine	227.0 ± 2.0	10.8	−3.0[e]	−53 ± 3
(n-$C_3H_7)_3NH^+$	tri-*n*-propylamine	229.5 ± 2.0	10.3	−2.5[e]	−49 ± 3
$C_6H_5NH_3^+$	aniline	203.3 ± 2.0	4.6	−5.5	−70 ± 3
o-$CH_3C_6H_4NH_3^+$	2-methylaniline	205.3 ± 2.0	4.5	−5.6[e]	−68 ± 3
m-$CH_3C_6H_4NH_3^+$	3-methylaniline	206.5 ± 2.0	4.7	−5.7[e]	−68 ± 3
p-$CH_3C_6H_4NH_3^+$	4-methylaniline	206.7 ± 2.0	5.1	−5.6[e]	−68 ± 3
m-$NH_2C_6H_4NH_3^+$	3-aminoaniline	214.9 ± 2.0	5.0	−9.9[e]	−64 ± 3
$C_6H_5NH_2CH_3^+$	*N*-methylaniline	212.7 ± 2.0	4.9[f]	−4.7[e]	−61 ± 3
$C_6H_5NH_2CH_2CH_3^+$	*N*-ethylaniline	213.4 ± 2.0	5.1[f]	−4.6[e]	−60 ± 3
$C_6H_5NH(CH_3)_2^+$	*N*,*N*-dimethylaniline	217.3 ± 2.0	5.1	−3.6[e]	−55 ± 3
p-$CH_3C_6H_4NH(CH_3)_2^+$	4-methyl-*N*,*N*-dimethylaniline	219.4 ± 2.0	5.6	−3.7[e]	−54 ± 3
$C_6H_5NH(CH_2CH_3)_2^+$	*N*,*N*-diethylaniline	221.8 ± 2.0	6.6	−2.9[e]	−52 ± 3
$C_{10}H_7NH_3^+$	1-aminonaphthalene	209.2 ± 2.0	3.9	−7.3[e]	−66 ± 3
$C_2H_4NH_2^+$	aziridine	208.5 ± 2.0	8.0	−4.5[e]	−69 ± 3
$C_3H_6NH_2^+$	azetidine	217.2 ± 2.0	11.3	−5.6	−66 ± 3
$C_4H_8NH_2^+$	pyrrolidine	218.8 ± 2.0	11.3	−5.5	−64 ± 3
$C_5H_{10}NH_2^+$	piperidine	220.0 ± 2.0	11.1	−5.1	−62 ± 3
$C_6H_{12}NH_2^+$	azacycloheptane	220.7 ± 2.0	11.1	−4.9[e]	−61 ± 3
$C_4H_5NH^+$	pyrrole	201.7 ± 2.0	−3.8	−4.3[e]	−60 ± 3
$C_5H_5NH^+$	pyridine	214.7 ± 2.0	5.2	−4.7	−59 ± 3
$C_9H_7NH^+$	quinoline	220.2 ± 2.0	4.8	−5.7[e]	−54 ± 3
$C_4H_8NHNH_2^+$	piperazine	218.6 ± 2.0	9.7	−7.4	−64 ± 3
CH_3CNH^+	acetonitrile	179.0 ± 2.0	−10.0[g]	−3.9	−73 ± 3
p-$CH_3OC_6H_4NH_3^+$	4-methoxyaniline	207.6 ± 2.0	5.3	−7.6[e]	−69 ± 3
p-$NO_2C_6H_4NH_3^+$	4-nitroaniline	199.4 ± 2.0	1.0	−9.9[e]	−74 ± 3
$C_4H_8ONH_2^+$	morpholine	213.0 ± 2.0	8.4	−7.2	−68 ± 3
$CH_3COHNH_2^+$	acetamide	199.0 ± 2.0	−0.6	−9.7	−72 ± 3
$C_6H_5COHNH_2^+$	benzamide	205.8 ± 2.0	−1.4	−10.9[e]	−65 ± 3
$(CH_3)_2SH^+$	dimethyl sulfide	191.5 ± 2.0	−7.0	−1.5	−63 ± 3
$(CH_3)_2SOH^+$	dimethyl sulfoxide	204.0 ± 2.0	−1.5	−9.8[e]	−66 ± 3
m-$ClC_6H_4NH_3^+$	3-chloroaniline	199.9 ± 2.0	3.5	−5.8[e]	−73 ± 3
p-$ClC_6H_4NH_3^+$	4-chloroaniline	201.2 ± 2.0	4.0	−5.9[e]	−72 ± 3
NH_4^+	ammonia	195.7 ± 2.0	9.3[h]	−4.3	−83 ± 3
$HNNH_2^+$	hydrazine	196.6 ± 2.0	8.1	−6.3[e]	−83 ± 3
H_3O^+	water	157.7 ± 0.7	−1.7	−6.3	−108 ± 2

[a] Aqueous solvation free energies are for a temperature of 298 K. [b] Gas-phase basicities from ref 84. [c] From ref 87, unless otherwise noted. [d] From the current data set, unless otherwise noted. [e] Reference 76. [f] Reference 91. [g] Reference 88. [h] Reference 92.

experimental aqueous solvation free energies of F^-, Cl^-, and Br^- were taken from Tissandier et al.[82] and adjusted for a change in the standard state and the value used here for $\Delta G^*_S(H^+)$. For the remaining anions, we used the free energy cycle shown in Scheme 2 and eq 5 to determine the aqueous solvation free energy according to

$$\Delta G^*_S(A^-) = -\Delta G^\circ_{gas}(A^-) - \Delta G^{\circ\rightarrow *} + \Delta G^*_S(AH) - \Delta G^*_S(H^+) + 2.303RT\mathrm{p}K_a(AH) \quad (7)$$

Scheme 2



where $\Delta G^{\circ}_{gas}(A^-)$ is the gas-phase basicity of A^-, equal to $G^{\circ}_{gas}(A^-) + G^{\circ}_{gas}(H^+) - G^{\circ}_{gas}(AH)$.

Experimental gas-phase basicities of anions and acidities of neutral species were taken from the National Institute of Standards and Technology (NIST) database;[83] experimental gas-phase basicities of neutral species were taken from the most recent compilation of Lias et al.[84] For neutral species, experimental aqueous solvation free energies were taken from the data set described in section 2.2 and from several additional sources.[76,85,86] A large part of the experimental aqueous pK_a data used here was taken from the compilation of Stewart;[87] pK_a data not available in this compilation were taken from several additional sources.[88−92]

We note that gas-phase free energies and aqueous solvation energies for neutral solutes can be calculated fairly accurately, and several other compilations of aqueous solvation free energies of ions[79,80,93] use theoretical data for these quantities. Although we did not use theoretical values in deriving any of the free energies shown in Tables 1 and 2, future extensions of the present database could include aqueous solvation energies determined from calculated data.

The aqueous solvation free energies reported here can be compared to the recent compilation of Pliego et al.,[93] who used the same thermodynamic cycle, along with Tissandier et al.'s[82] value of −266 kcal/mol for $\Delta G^*_S(H^+)$, in determining the aqueous solvation free energies for 56 ions. (Note that the above value of −266 kcal/mol for $\Delta G^*_S(H^+)$ reflects the standard-state correction required in order to adjust Tissandier et al.'s reported value of −264 kcal/mol,[82] which is for a gas-phase standard state of 1 atm combined with an aqueous phase standard state of 1 mol/L, to a standard state that uses a concentration of 1 mol/L in both the gas and aqueous phases. Also note that Tissandier's reported value of −264 kcal/mol has sometimes been misinterpreted as corresponding to a standard state of 1 mol/L in both the gas and aqueous phases.[15,16,94−96]) Making an adjustment to account for the difference between the value used for $\Delta G^*_S(H^+)$ here and the value used in Pliego and Riveros' work brings the two sets of data into very good agreement with one another.

2.4. Water−Solute Clusters. We also compiled another data set, to be called the selectively clustered-ion data set, in which 31 bare ions in the unclustered-ion data set were replaced by the ion−water clusters that are listed (along with the water dimer) in Table 3. Using the free energy cycle shown in Scheme 3, the aqueous solvation free energies of these 31 solute−water clusters were determined according to

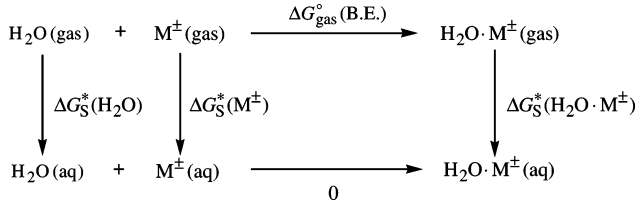
$$\Delta G^*_S(H_2O{\cdot}M^{\pm}) = \Delta G^*_S(H_2O) + \Delta G^*_S(M^{\pm}) - \Delta G^{\circ}_{gas}(\text{B.E.}) + \Delta G^{\circ\rightarrow *} \quad (8)$$

Table 3. Aqueous Solvation Free Energies for Solute−Water Clusters[a]

A·H$_2$O[b]	ΔG°_{gas}(B.E.)[c]	ΔG^*_S(A)[d]	ΔG^*_S(A·H$_2$O)
$H_2O(H_2O)$	3.34 ± 0.50[e]	−6.31 ± 0.20	−14.06 ± 0.57
$CH_3OH_2(H_2O)^+$	−18.5 [f]	−91.1 ± 2.8	−77 ± 3
$CH_3CH_2OH_2(H_2O)^+$	−16.8[f]	−86.5 ± 2.8	−74 ± 3
$(CH_3)_2OH(H_2O)^+$	−15.4[f]	−77.8 ± 2.8	−67 ± 3
$(C_2H_5)_2OH(H_2O)^+$	−11.4 ± 2.0[g]	−69.6 ± 2.8	−62 ± 3
$CH_3C(OH)CH_3(H_2O)^+$	−12.8[f]	−75.1 ± 2.8	−67 ± 3
$CH_3COHC_6H_5(H_2O)^+$	−10.8[f]	−62.6 ± 2.8	−56 ± 3
$NH_4(H_2O)^+$	−12.6[f]	−83.3 ± 2.8	−75 ± 3
$H_3O(H_2O)^+$	−27.0 ± 2.0[g]	−108.4 ± 2.0	−86 ± 3
$C_2H(H_2O)^-$	−10.6 ± 1.0[f]	−78.4 ± 2.6	−72 ± 3
$CN(H_2O)^-$	−8.3 ± 0.7[f]	−72.2 ± 1.9	−68 ± 2
$CH_3O(H_2O)^-$	−17.00 ± 0.30[f]	−96.9 ± 2.0	−84 ± 2
$C_2H_5O(H_2O)^-$	−14.2 ± 2.0[g]	−92.6 ± 2.2	−83 ± 3
$CH_3CH_2CH_2O(H_2O)^-$	−14.6 ± 2.0[g]	−90.2 ± 2.4	−80 ± 3
$(CH_3)_2CHO(H_2O)^-$	−12.3 ± 2.0[g]	−88.2 ± 2.2	−80 ± 3
$CH_3CH_2CHOCH_3(H_2O)^-$	−9.9 ± 2.0[g]	−86.1 ± 2.8	−81 ± 3
$C(CH_3)_3O(H_2O)^-$	−12.2 ± 2.0[g]	−84.2 ± 2.2	−76 ± 3
$H_2C{=}CHCH_2O(H_2O)^-$	−13.5 ± 2.0[g]	−88.5 ± 3.4	−79 ± 4
$C_6H_5CH_2O(H_2O)^-$	−11.6 ± 2.0[g]	−87.0 ± 2.8	−80 ± 3
$CH_3OCH_2CH_2O(H_2O)^-$	−13.5 ± 2.0[g]	−91.4 ± 2.8	−82 ± 3
$CH_2OHCH_2O(H_2O)^-$	−14.0 ± 2.0[g]	−87.2 ± 2.8	−78 ± 3
$CF_3CH_2O(H_2O)^-$	−11.6 ± 2.0[g]	−79.5 ± 2.8	−72 ± 3
$CH(CF_3)_2O(H_2O)^-$	−6.0 ± 2.0[g]	−67.4 ± 2.8	−66 ± 3
$CH_3OO(H_2O)^-$	−14.6 ± 2.0[g]	−95.2 ± 2.0	−85 ± 3
$CH_3CH_2OO(H_2O)^-$	−14.1 ± 2.0[g]	−91.1 ± 2.8	−91 ± 3
$HO(H_2O)^-$	−19.8 ± 1.4[f]	−106.6 ± 1.9	−91 ± 2
$HO_2(H_2O)^-$	−17.0 ± 2.0[g]	−99.2 ± 2.0	−87 ± 3
$O_2(H_2O)^-$	−12.1 ± 2.0[f]	−85.2 ± 2.1	−78 ± 3
$HS(H_2O)^-$	−8.6 ± 2.0[f]	−74.0 ± 2.3	−70 ± 3
$F(H_2O)^-$	−12.5 ± 1.6[f]	−102.5 ± 1.9	−94 ± 3
$Cl(H_2O)^-$	−9.0 ± 4.0[f]	−72.7 ± 1.9	−68 ± 4
$Br(H_2O)^-$	−7.1 ± 2.0[f]	−66.3 ± 1.9	−64 ± 3

[a] Aqueous solvation free energies are for a temperature of 298 K. [b] B97-1/MG3S optimized geometries. [c] Water−solute binding free energies. [d] Aqueous solvation free energy of the bare solute. [e] Experimental value, taken from ref 78. [f] Experimental value, taken from ref 100. [g] Theoretical value, calculated at the B97-1/MG3S level of theory.

Scheme 3



In the above equation, aqueous solvation free energies of the unclustered ions $\Delta G^*_S(M^{\pm})$ were taken from Tables 1 or 2. When available, experimental values for the gas-phase binding energies were used in the above equation. When experimental values were not available, we calculated them at the B97-1[97]/MG3S[98] level of theory, which has recently been shown[52] to perform well for nonbonded interactions in the gas phase.

2.5. Uncertainty of Experimental Data. We have previously estimated an average uncertainty of 0.2 kcal/mol for aqueous solvation free energies of neutral solutes in our data sets.[15] The uncertainties in the aqueous solvation free energies for ionic solutes are significantly greater due in part to their

large magnitudes but also due to uncertainties associated with each of the experimental quantities (except pK_a, as discussed below) used to determine them. The uncertainties in aqueous solvation free energies for ionic solutes are reported here as the root-sum-of-squares combination of each of the uncertainties associated with the individual experimental measurements used to determine them.

All of the gas-phase basicities of anions were taken from the NIST tables.[83] In several cases, more than one experimental measurement was available for a single molecule, in which case the average value was used. For molecules where more than one experimental measurement was available, we calculated the standard deviation from the mean and compared this value to the smallest value for the uncertainty associated with any of the individual measurements. The uncertainty reported here is the larger of these two values. For the gas-phase basicities of neutrals, the only molecule used in this work for which an absolute experimental value has been reported is water; this measurement has an experimental uncertainty of 0.7 kcal/mol. For the remaining molecules considered here, the experimental gas-phase basicities are relative values that have been obtained from bracketing experiments. Following Hunter and Lias[84] we have assigned an uncertainty of 2.0 kcal/mol to data obtained by bracketing.

An additional contribution to the overall uncertainty arises from the value used for the aqueous solvation free energy of the proton $\Delta G^*_S(H^+)$. We use Zhan and Dixon's value of -264 kcal/mol[81] for $\Delta G^*_S(H^+)$, to which we assign an uncertainty of 2 kcal/mol.

Experimental pK_a values that fall in the range $0-14$ can be measured quite accurately, and thus their uncertainty is not included in calculating the overall uncertainties in the aqueous solvation free energies of ions. Experimental pK_a values that fall outside the range $0-14$ are somewhat more uncertain, especially those pK_a values that are 2 pK_a units or more outside this range.[99] Despite this, we feel that the relative uncertainties associated with the experimental pK_a values considered in this work that do fall outside this range are small in comparison to the uncertainties associated with the other experimental data, and thus we will not take them into account.

Finally, for the solute−water clusters, an additional source of uncertainty from the value used for the gas-phase binding free energy ΔG°_{gas}(B.E.) must be considered. For the anion−water clusters, all experimental values for ΔG°_{gas}(B.E.) were taken from the NIST tables,[100] in which the uncertainty of each measurement is reported. The uncertainty associated with the experimental ΔG°_{gas}(B.E.) values of the cations, which were also taken the NIST tables, is negligible. For theoretical ΔG°_{gas}(B.E.) values calculated at the B97-1/MG3S level of theory, we have estimated an uncertainty of 2.0 kcal/mol.

Based on the error analysis presented above, we estimate that the uncertainty in the aqueous solvation free energy of a typical ionic solute is approximately 3 kcal/mol, which is lower than our previously estimated[15] uncertainty of 4−5 kcal/mol.

3. Definition of SM6

In the SM*x* models, the solvation free energy ΔG^*_S is partitioned according to

$$\Delta G^*_S = \Delta E_{elec} + \Delta E_{relax} + \Delta G^*_{conc} + G_P + G_{CDS} \quad (9)$$

where ΔE_{elec} is the change in the solute's internal electronic energy in moving from the gas phase to the liquid phase at the same geometry, ΔE_{relax} is the change in the solute's internal energy due to changes in the geometry accompanying the solvation process, and ΔG^*_{conc} accounts for the concentration change between the gas-phase and the liquid-phase standard states. Following the notation used in our previous models, we will refer to the sums $\Delta E_{elec} + G_P$ and $\Delta E_{relax} + \Delta E_{elec} + G_P$ as ΔG_{EP} and ΔG_{ENP}, respectively. Since we use the same concentrations (1 mol/L) in both phases, ΔG^*_{conc} is zero.[73,101] Also, all calculations reported here are based on gas-phase geometries (although the present model can be used to optimize geometries in the liquid phase[13]), so ΔE_{relax} is assumed to be zero.

The ΔG_{EP} contribution to the total solvation free energy is calculated from a self-consistent molecular orbital calculation,[12,57,58] where the generalized Born approximation[102−106] is used to calculate the polarization contribution to the total free energy according to

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{k,k'} q_k \gamma_{kk'} q_{k'} \quad (10)$$

In the above equation, the summation goes over all atoms k in the solute, ϵ is the dielectric constant of the solvent, q_k is the partial atomic charge of atom k, and $\gamma_{kk'}$ is a Coulomb integral involving atoms k and k'. For water, we use $\epsilon = 78.3$.[107]

The partial atomic charges are obtained from Charge Model 4 (CM4). This new charge model is similar in most ways (except one that is described below) to our most recent previous charge model, CM3.[108−112] In particular, CM4 empirically maps atomic charges obtained from a Löwdin population analysis (LPA)[113−116] or a redistributed Löwdin population analysis (RLPA).[117] The Supporting Information provides a detailed description of CM4, although we note here that an important difference between CM3 and CM4 is the performance of these two models for hydrocarbons and molecules containing aliphatic functional groups. For CM3, the parameter that is used to map Löwdin or redistributed Löwdin C−H bond dipoles was optimized[108,109,112,118] by requiring the average CM3 charge on H in benzene and ethylene to be 0.11, a value that had been justified in a previous paper.[118] More recently, careful analysis of partial atomic charges calculated using CM3 revealed that in some cases, CM3 yielded C−H bonds that were too polar, and this, in turn, had a negative impact on the performance of our solvation models for some solutes as well as the performance of other methods that use CM3 partial atomic charges. Because of this deficiency, we developed a different procedure for optimizing the C−H parameter. In particular, this parameter was optimized by minimizing the error between calculated partial atomic charges and those partial atomic charges used in Jorgenson et al.'s OPLS force field[119]

for a series of 19 hydrocarbons. Once this parameter was optimized it was fixed, and the remaining parameters were optimized in a fashion completely analogous to that used for CM3,[108,109,112,118] resulting in a new charge model called CM4. Details of the CM4 parametrization are given in the Supporting Information. It is important to note that CM4 retains all of the qualities of CM3, except that the new model gives more reasonable partial atomic charges for hydrocarbons and molecules containing aliphatic functional groups, which is important for accurately modeling hydrophobic effects.

The Coulomb integrals $\gamma_{kk'}$ are calculated according to

$$\gamma_{kk'} = [R_{kk'}^2 + \alpha_k\alpha_{k'} \exp(-R_{kk'}^2/d\alpha_k\alpha_{k'})]^{-1/2} \qquad (11)$$

where $R_{kk'}$ is the distance between atoms k and k' and α_k is the effective Born radius of atom k, which is described below. When $k = k'$, eqs 10 and 11 lead to Born's equation[120] for the polarization free energy of a monatomic ion, whereas at large $R_{kk'}$ Coulomb's law for the interaction energy of two point charges in a dielectric continuum is recovered. A d value of 4 was originally proposed by Still et al.[105] because for intermediate values of $R_{kk'}$, it gives polarization free energies that are close to those predicted using the classical equation for a dipolar sphere embedded in a dielectric continuum. Because modeling most solutes as a dipolar sphere is itself a drastic approximation, and because Jayaram et al.[121] demonstrated for a test set of dicarboxylic acids that using different values of d in eq 11 led to improvements in calculated pK_a shifts, we will treat d as a parameter that can be adjusted.

The following equation[122] is used to calculate the Born radius

$$\alpha_k = \left(\frac{1}{R'} + \int_{\rho_{Z_{k'}}}^{R'} \frac{A_k(\mathbf{R},r,\{\rho_{Z'}\})}{4\pi r^4}\mathrm{d}r\right)^{-1} \qquad (12)$$

where R' is the radius of the sphere centered on atom k that completely engulfs all other spheres centered on the other atoms of the solute, and $A_k(R,r,\{\rho_Z\})$ is the exposed area[122] of a sphere of radius r that is centered on atom k. This area depends on the geometry of the solute, **R**, and the radii of the spheres centered on all the other atoms in the solute. The radii of these spheres are given by a set of intrinsic Coulomb radii ρ_{Z_k} that depend on the atomic number Z_k of the atom k. Other methods that have recently been proposed for determining the Born radius include one by Onufriev et al.[123] that takes into account interior regions of the solute inaccessible by solvent molecules (which can have an effect on the calculated solvation free energies of macromolecules[124,125]) and one by Zhang et al.[126] that assigns Born radii based on atom types. We recently tested[16] Onufriev et al.'s method for calculating Born radii during the development of SM5.43R and found that it significantly worsened the performance of that model for predicting the aqueous solvation free energies of ionic solutes. Therefore, we will not test this method here.

The G_{CDS} term is calculated according to

$$G_{CDS} = \sum_k \sigma_k A_k(\mathbf{R},\{R_{Z_k} + r_s\}) \qquad (13)$$

where σ_k is the atomic surface tension of atom k, A_k is the solvent-accessible surface area (SASA)[59,60] of atom k, which depends on the geometry, **R**, atomic radius R_{Z_k}, and solvent radius r_s, which is added to each of the atomic radii. Adding a nonzero value for solvent radius to the atomic radii defines the SASA of a given solute. Note that the atomic radii in eq 13 are not constrained to be equal to the intrinsic Coulomb radii ρ_{Z_k} of eq 12. Although the same values for the atomic radii might have been used in eqs 12 and 13, our experience has shown that the overall performance of our models is relatively insensitive to the values used for the atomic radii in eq 13 (but not eq 12). Therefore, in eq 13 we use Bondi's van der Waals radii.[127] For the solvent radius, we use a value of 0.4 Å, which has been justified in earlier work. [15] The atomic surface tensions σ_k are given by

$$\sigma_k = \sum_k \tilde{\sigma}_{Z_k} + \sum_{k,k'} \tilde{\sigma}_{Z_kZ_{k'}} T_k(\{Z_{k'},R_{kk'}\}) \qquad (14)$$

where $\tilde{\sigma}_Z$ is an atomic-number-specific parameter, $\tilde{\sigma}_{ZZ'}$ is a parameter that depends on the atomic numbers of atoms k and k', and $T_k(\{Z_{k'},R_{k,k'}\})$ is a geometry-dependent switching function called a cutoff tanh or COT. The general form for this function is

$$T(R_{kk'}|\bar{R}_{ZZ'},\Delta R) = \begin{cases} \exp\left[-\left(\dfrac{\Delta R}{\Delta R - R_{kk'} + \bar{R}_{ZZ'}}\right)\right] & R_{kk'} \le \bar{R}_{ZZ'} + \Delta R \\ 0 & \text{otherwise} \end{cases} \qquad (15)$$

where $\bar{R}_{ZZ'}$ is the midpoint of the switch, and ΔR determines the range over which the function switches. These values, which are listed in Table 4, have been assigned based on an analysis of distances between certain atoms for molecules in our database. The COT function has the property that it vanishes identically for all $R_{kk'}$ greater than $\bar{R} + \Delta R$ but is continuous and has an infinite number of continuous derivatives for all $R_{kk'}$. These properties are important because they allow the model to optimize geometries efficiently in the liquid phase.[13] The functional forms that we use for the

Table 4. $\bar{R}_{ZZ'}$ and ΔR Values (Å)

Z	Z'	$\bar{R}_{ZZ'}$
H	C, N, or O	1.55
H	S	2.14
C	C, N	1.84
C	C	1.27[a]
C	O	1.33
C	F or P	2.20
C	Cl	2.10
C	Br	2.30
N	N	1.85
N	O	1.50
O	O	1.80
O	P	2.10
S	S	2.75
S	P	2.50
W		0.30
$W_{CC(2)}$		0.07
W_{CO}		0.10
W_{NC}		0.065

[a] $\bar{R}_{CC}^{(2)}$.

atomic surface tensions are given below, where we write Z for Z_k to simplify the notation:

$$\sigma_Z|_{Z=1} = \tilde{\sigma}_Z + \sum_{\substack{k' \\ Z_{k'}=6,7,8,16}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] \quad (16)$$

$$\sigma_Z|_{Z=6} = \tilde{\sigma}_Z + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}^{(2)}_{ZZ'}[T(R_{kk'}|\bar{R}^{(2)}_{ZZ'},W_{CC(2)})] + \sum_{\substack{k' \neq k \\ Z_{k'}=7}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)]^2 \quad (17)$$

$$\sigma_Z|_{Z=7} = \tilde{\sigma}_Z + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}_{ZZ'}\{[T(R_{kk'}|\bar{R}_{ZZ'},W)] \sum_{\substack{k'' \neq k \\ k'' \neq k'}} [T(R_{kk'}|\bar{R}_{ZZ'},W)]^2\}^{1.3} + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}^{(2)}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] \sum_{\substack{k'' \\ Z_{k''}=8}} [T(R_{k'k''}|\bar{R}_{Z'Z''},W)] + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}^{(3)}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W_{NC})] \quad (18)$$

$$\sigma_Z|_{Z=8} = \tilde{\sigma}_Z + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W_{OC})] + \sum_{\substack{k' \neq k \\ Z_{k'}=7,8,15}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] \quad (19)$$

$$\sigma_Z|_{Z=9} = \tilde{\sigma}_Z \quad (20)$$

$$\sigma_Z|_{Z=15} = \tilde{\sigma}_Z \quad (21)$$

$$\sigma_Z|_{Z=16} = \tilde{\sigma}_Z + \sum_{\substack{k' \neq k \\ Z_{k'}=15,16}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] \quad (22)$$

$$\sigma_Z|_{Z=17} = \tilde{\sigma}_Z \quad (23)$$

$$\sigma_Z|_{Z=35} = \tilde{\sigma}_Z \quad (24)$$

The functional forms shown above are a convenient way to treat different types of chemical environments that a particular atom in a solute might encounter. Unlike models that require the user to assign types to atoms (e.g., molecular mechanics force fields), these functional forms do not require the user to make assignments. This feature means that the user is never in doubt about which parameter to use. Furthermore, because the above equations are smooth functions of the solute geometry, the present model can be applied to systems containing nonbonded or partially bonded pairs of atoms, such as transition states and solute−solvent clusters.

It is important to point out that although separating the free energy of solvation into the above components allows us to effectively model a wide variety of solutes, only the total free energy is a state function, so there is a certain degree of ambiguity associated with separating the aqueous solvation free energy, which is an experimental observable, into several contributions that cannot be measured independently. Thus, there is some flexibility in how to interpret these various contributions.

We interpret the contribution to the free energy arising from G_P as accounting for electrostatic interactions between the charge distribution of the solute (which is modeled as a collection of point charges distributed over atomic spheres) and the bulk electric field of the solvent when it is assumed to begin at a boundary defined by the effective Born radii. For ionic solutes, G_P makes the largest contribution to the overall solvation free energy. Because G_P is calculated under the assumption that the solvent responds linearly to the electronic distribution of the solute (hence[58] the factor $^1/_2$ appearing in eq 10), nonlinear solvent effects, such as changes in the dielectric constant of the solvent in the vicinity of the solute, and strong solute−solvent hydrogen bonds, which prevent the solvent from fully responding to the solute charge distribution, are not explicitly accounted for by this term. One strategy that we have used to partly account for these nonlinear effects has been to empirically adjust the values that we use for the intrinsic Coulomb radii. For neutrals, previous experience has shown that the overall performance of our models is relatively insensitive to the values used for these atomic radii, whereas for ions the performance of our models is quite sensitive to the choice of these radii. By making adjustments to the atomic radii, our previous models have achieved an accuracy of ~5 kcal/mol in aqueous solvation free energies for the majority of the ions used to train these models. However, the present data set is much larger and more diverse than our previous data sets for ions. Thus, one of the goals of this work will be to see if the deficiencies described above can adequately be accounted for by making empirical adjustments to the intrinsic Coulomb radii.

A second strategy to account for deficiencies of a bulk electrostatic model is to include G_{CDS}. In the past, we have interpreted the G_{CDS} term as formally accounting for cavitation (i.e. the free energy cost associated with creating a cavity in the solvent to accommodate the solute), dispersion interactions between the solute and solvent, and specific solute effects on the solvent structure (e.g. the loss of orientational freedom of water molecules around a nonpolar solute). However, because the G_{CDS} term is empirical in nature, it can be more accurately interpreted as accounting for *any* contribution to the total solvation free energy of a given solute that is not explicitly accounted for by the bulk electrostatic (G_P) term. Such effects include, but are not limited to, the nonlinear solvent effects described above, deviations of the true solute−solvent interface defined by the atomic radii, short-range exchange and repulsion forces between the solute and solvent, neglect of charge transfer between the solute and solvent, and any systematic errors that may arise from the GB approximation, the ability of partial atomic charges to represent the true solute charge distribution, or the level of theory used to calculate the electronic wave function of the solute. In addition, several other effects are implicitly accounted for by G_{CDS} that could, in principle, be explicitly calculated, such as the change in the solute's translational, vibrational, or rotational free energy in moving from the gas phase to solution. By using atomic surface tensions, our previous models have been quite

successful at predicting aqueous solvation free energies for a wide variety of neutral solutes (an average error of ~0.5 kcal/mol), including many hydrogen bonding solutes. It is important to point out that because (in the present work and in most, but not all, of our previous work) we optimize the parameters contained in the G_{CDS} term using only neutral solutes, applying this scheme also to ions involves the assumption that first solvation shell interactions are similar for a given pair of solutes containing the same functionality but different formal charge; this may be a major contributor to the residual error. We could of course eliminate this problem by using different surface tension parameters for ions, but, as stated above, unlike some of the other methods in the literature, we avoid using molecular mechanics types.

4. Parameters To Be Optimized

Three different types of parameters will be optimized as part of this work: (1) the atomic radii in eq 12, (2) the value for d in eq 11, and (3) the atomic surface tension parameters $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ in eqs 14−24. The above parameters will be optimized using gas-phase geometries. In all cases, the solutes in our database were represented by a single, lowest-energy conformation. For some of the solutes, in particular the solute−water clusters, this involved performing a conformational analysis to identify the global minimum on the potential energy surface. For the acetamide cation, we used the geometry corresponding to the oxygen-protonated species, which is 9.7 kcal/mol lower in free energy in the gas phase than the nitrogen-protonated species (MPW25/MIDI! level of theory). For the acetamide anion, we used the deprotonated imidate form, which is lower in free energy in the gas phase than the enolate by 15.7 kcal/mol (MPW25/MIDI! level of theory).

For all of the unclustered solutes used in this work, we used geometries optimized at the MPW25/MIDI! level of theory. The MIDI! basis set[61,62,128] is an especially economical basis set for calculations on large organic systems, but it was designed to give particularly accurate geometries. All of the solute−water clusters were optimized at the B97-1/MG3S level of theory.

5. Results

5.1. Partial Atomic Charges. Although there is formally no "correct" method for assigning partial atomic charges because partial atomic charge it is not a quantum mechanical observable,[129] several qualities make CM4 partial atomic charges more suitable for use in the present model than charges obtained from other methods. First, dipole moments derived from CM4 point-charges are generally more accurate than point-charge-derived dipole moments obtained using other charge partitioning schemes. This is demonstrated by the data in Table 5, which lists the mean unsigned errors between experimental dipole moments for 397 neutral molecules and those dipole moments calculated using CM4 point charges and those obtained from a redistributed Löwdin population analysis (RLPA).[117] Also listed are the average errors for 107 unclustered ionic molecules (experimental dipole moments are not available for ionic solutes, so we used density dipole moments calculated at the MPW25/MG3S//MPW25/MIDI! level of theory for comparison). The data in Table 5 show that for most of the molecules tested here, CM4-point-charge-derived dipole moments are more accurate than those calculated using RLPA partial atomic charges (for nondiffuse basis sets, RLPA partial atomic charges are equivalent to those obtained from a Löwdin population analysis[113−116]). One notable exception is for the anions tested here, where at the MPW25/MIDI!6D level of theory, RLPA-point-charge-derived dipole moments are more accurate than the CM4-point-charge-derived dipole moments (for other levels of theory, the CM4-point-charge derived dipole moments are more accurate).

Table 5. Mean Unsigned Errors (Debyes) in Dipole Moments Calculated with Partial Atomic Charges Obtained from CM4 and RLPA at the MPW25 Level of Theory[a]

	CM4			RLPA[b]		
basis set	neutrals	cations	anions	neutrals	cations	anions
MIDI!6D	0.19	0.20	0.61	0.37	0.24	0.35
6-31G(d)	0.23	0.24	0.36	0.62	0.21	0.46
6-31+G(d)	0.27	0.32	0.44	0.76	0.43	0.55
6-31+G(d,p)	0.26	0.33	0.40	0.81	0.67	0.67

[a] For the neutral solutes, point-charge-derived dipole moments were compared to dipole moments taken from the CM3 training set, which is described in ref 112 (397 total dipole moments). For the dipolar ions in Tables 1 and 2, point-charged derived dipole moments were compared to density dipole moments calculated at the MPW25/MG3S level of theory (107 total dipole moments). [b] For nondiffuse basis sets, RLPA partial atomic charges are equivalent to Löwdin partial atomic charges.

A second reason we prefer using CM4 partial atomic charges is because they are less sensitive to changes in the basis set than partial atomic charges obtained from other models. This becomes especially true when diffuse functions are added. Table 5 shows that as polarization and diffuse functions are added to the basis set, the quality of RLPA charges progressively worsens, whereas a much smaller dependency on basis set is observed for the CM4 charges.

Finally, it has recently been shown[112] that CM3 (a model quite similar to CM4) delivers more accurate charges for interior or buried atoms than partial atomic charges calculated from a fit to the electrostatic potential calculated around the molecule of interest (e.g. the ChElPG charge model[130]). In this same paper, it was also shown that CM3 charges are much less sensitive to small conformational changes and to the level of treatment of electron correlation than are ChElPG charges.

5.2. Aqueous Solvation Free Energies Calculated Using Previously Defined Atomic Radii. First, we tested several previously defined sets of atomic radii for predicting aqueous solvation free energies. For this, we developed three intermediate models using three different sets of atomic radii in eq 12, in particular the following: Bondi's atomic radii (which we also use to calculate the solvent-accessible-surface area in eq 13) and the radii used by our SM5.42R[10−14] and SM5.43R[15,16] models. These sets of atomic radii are listed in Table 6. For each intermediate method, we calculated ΔG_{EP} values at the MPW25/6-31+G(d,p) level of theory for all of the solutes in our data set, with d fixed at 4. With calculated values for ΔG_{EP} in hand, we then optimized a set of surface tension coefficients $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ for each intermediate model by minimizing the errors between the experimental

Table 6. Atomic Radii Used by Various Models

atom	Bondi	SM5.42R	SM5.43R	SM6
H	1.20	0.91	0.79	1.02
C	1.70	1.78	1.81	1.57
N	1.55	1.92	1.66	1.61
O	1.52	1.60	1.63	1.52[c]
F	1.47	1.50	1.58	1.47[c]
P	1.80	2.27	2.01	1.80[c]
S	1.80	1.98	2.22	2.12
Cl	1.75	2.13	2.28	2.02
Br	1.85	2.31	2.38	2.60

[a] Not optimized, held fixed at Bondi's value.

Table 7. Mean Unsigned Errors in Aqueous Solvation Free Energies (kcal/mol) Obtained Using Various Sets of Atomic Radii[a]

		ions	
atomic radii[b]	neutrals	unclustered[c]	selectively clustered[d]
Bondi	0.56	6.87	5.55
SM5.42R	0.52	5.64	4.73
SM5.43R	0.52	6.06	5.32

[a] For each set of radii, a different set of atomic surface tensions was optimized. All *d* values were fixed at 4 for the calculations in this table, and (in the whole article) we always use Bondi's radii in eq 13. The calculations in this table were carried out using MPW25/6-31+G(d,p). [b] Intrinsic Coulomb radii. [c] This data set contains all 112 ions listed in Tables 1 and 2. [d] This data set contains 81 of the ions listed in Tables 1 and 2 (those that do not appear in clustered form in Table 3) plus the 31 clustered ions listed in Table 3.

and calculated solvation free energies for all 273 of the neutral solutes. This step was accomplished using a NAG Fortran 90 routine, in particular the linear least squares solver routine.[131] The performance of the three intermediate models is summarized in Table 7.

The data in Table 7 show that all three models lead to similar errors in the solvation free energies of neutral solutes. Although not shown explicitly in Table 7, all three of the intermediate models predict an aqueous solvation free energy for water that is ~4 kcal/mol too negative. We encountered a similar error during the development of earlier models and removed it by including a surface tension that identified oxygen atoms in the vicinity of two hydrogen atoms. We did not include this surface tension here, because doing so would have a negative impact on the performance of the model for hydronium, protonated alcohols, and solute−water clusters. For the water *dimer*, all three of the intermediate models predict its solvation free energy correctly to within 1.2 kcal/mol without using the special surface tension, which is a significant improvement compared to the performance of these models for the bare water solute.

For ions, the errors shown in Table 7 are broken down into two different subsets: the unclustered-ion data set and the selectively clustered-ion data set. The unclustered-ion data set contains all of the experimental aqueous solvation free energies listed in Tables 1 and 2 but none of the solute−water clusters in Table 3 (112 ionic solutes). The selectively clustered-ion data set contains all of the experimental aqueous solvation free energies in Table 3 plus those aqueous solvation free energies that are in Tables 1 or 2 but not Table 3 (i.e. solutes that are included in the selectively clustered-ion data set as solute−water clusters are not included as their analogous bare ions). The criteria we used to decide to which of the solutes in our training set to add an explicit water molecule (i.e. which bare solutes would be deleted from the selectively clustered-ion data set and replaced by their analogous water−solute cluster) is as follows. First, we added an explicit water molecule to any ionic solute containing three or fewer atoms. Second, we added an explicit water molecule to any ionic solute with one or more oxygen atoms bearing a more negative partial atomic charge than bare water solute (as judged by comparison of CM4 gas-phase charges computed at the MPW25/6-31+G(d,p) level of theory). Finally, we added an explicit water molecule to ammonium and to all of the oxonium ions. We singled out solutes that satisfied one or more of these three criteria because we felt these solutes were likely to form strong solute−solvent hydrogen bonds with water and therefore would serve as a useful indicator as to whether including a single explicit water molecule in the calculation is an effective strategy for accounting for strong solute−solvent hydrogen bonding effects.

The data in Table 7 are consistent with the fact that the solvation free energies of ions are more sensitive to the choice of atomic radii than the neutrals. This is not surprising, since the ΔG_{EP} term is the major contributor to the total solvation free energy for ions and because the surface tensions are optimized for a given set of values of ΔG_{EP}. The data in Table 7 also show that all three intermediate models give significantly lower errors for the selectively clustered-ion data set than for the unclustered-ion data set. This indicates that for the above intermediate models including a *single* explicit water molecule in the calculation is at least partly effective in accounting for strong specific solute−solvent hydrogen bonding interactions.

5.3. Aqueous Solvation Free Energies Calculated Using Optimized Atomic Radii and *d* Parameter. Of the three intermediate models presented above, the one based on SM5.42R radii performs the best for ionic solutes, giving a mean unsigned error of 5.64 kcal/mol for the ions in the unclustered-ion data set and 4.73 kcal/mol for the ions in the selectively clustered-ion data set. Next, we examined whether optimizing a new set of radii would lead to more accurate solvation free energies. For this, we again used MPW25/6-31+G(d,p). Throughout the parameter optimization process, for each set of intrinsic Coulomb radii, we optimized a different set of surface tension coefficients for each set of radii by minimizing the overall error between the calculated and experimental aqueous solvation free energies for all of the neutral solutes. This two-step procedure (where first a set of intrinsic Coulomb radii are chosen, and then surface tension coefficients are optimized for that set of atomic radii) was implemented into a genetic algorithm,[132] and the average error between the calculated and experimental aqueous solvation free energies for all of the ionic solutes in the selectively clustered-ion data set (the reason we chose to use only the data from the selectively clustered-ion data set is discussed in section 6.2) was minimized. To ensure that a physical parametrization was achieved, we constrained the optimization in the following ways. First, any set of intrinsic Coulomb radii that yielded positive ΔG_{EP}

values for any of the solutes in our data sets, or for *n*-hexadecane, was disqualified. Second, any set of intrinsic Coulomb radii that yielded a ΔG_{EP} value more negative than −0.40 kcal/mol for any of the *n*-alkanes was disqualified. This second constraint was added because several initial optimizations led to unusually small values for intrinsic Coulomb radii of hydrogen and carbon atoms. A value of −0.40 kcal/mol was chosen as the cutoff because 0.40 kcal/mol is twice the value that we have previously estimated[15] for the average uncertainty in the free energy of solvation of a typical neutral solute.

Throughout the optimization, we found that small changes in the value used for *d* led to some improvement in the model. Therefore, we decided to optimize the *d* parameter simultaneously along with the intrinsic Coulomb radii. For this, we constrained the value of *d* between 3.5 and 4.5. We chose not to vary *d* more widely because it appears in the exponential term of eq 11.

With the above constraints, we found an optimum of 3.7 for *d*, which is close to the value of 4 originally suggested by Still et al.[105] The intrinsic Coulomb radii resulting from the above optimization are listed in the last column of Table 6. For oxygen and fluorine, we found that using Bondi's atomic radii[127] instead of the optimized radii had little effect on the overall performance of the model, so we used Bondi's values instead. For phosphorus, we held the value fixed at Bondi's value throughout the optimization because the current data set does not contain any phosphorus-containing ions.

5.4. Atomic Surface Tension Coefficients. With the intrinsic Coulomb radii and *d* parameter fixed at the values obtained above, again using all 273 of the neutral solutes in our data set, we optimized four different sets of atomic surface tension coefficients for the following levels of theory: MPW25/MIDI!6D, MPW25/6-31G(d), MPW25/6-31+G(d), and MPW25/6-31+G(d,p). The ΔG_{EP} values calculated at the MPW25/6-31+G(d) and MPW25/6-31+G(d,p) levels of theory for the solute *O*-ethyl *O*′-4-bromo-2-chlorophenyl *S*-propyl phosphorothioate are large outliers, so we omitted this solute from the determination of the atomic surface tension coefficients for these two levels of theory (although we did use their calculated aqueous solvation free energies to determine the errors shown in Table 9). The optimized surface tension coefficients for each of the above levels of theory are listed in Table 8.

Shown in the first column of Table 9 are the various classes of solutes in our data set. The next four columns list the average errors in aqueous solvation free energies for each solute class by level of theory (subsequent columns of this table are discussed in the last paragraph of this section). For neutral solutes, the overall performance of SM6 is relatively insensitive to changes in basis set. Closer inspection of the individual solute classes in Table 9 does, however, reveal that for the neutrals there are a few systematic differences between the aqueous solvation free energies calculated with and without diffuse functions. For example, using diffuse functions leads to calculated solvation free energies for the carboxylic acids that are on average 0.2−0.3 kcal/mol less accurate than those calculated without diffuse functions. Conversely, using diffuse functions reduces the average error by a factor of around 3 for the two halogenated sulfur compounds in the data set.

Table 8. SM6 Surface Tension Coefficients (cal/Å^2), Optimized for MPW*X* with Various Basis Sets

Z,*Z*′	MIDI!6D	6-31G(d)	6-31+G(d)	6-31+G(d,p)
H	60.3	60.7	55.2	55.2
C	96.8	92.5	114.6	108.0
N	49.5	14.9	7.1	9.8
O	−120.4	−126.9	−142.8	−152.5
H,C	−79.5	−80.7	−75.2	−73.7
C,C	−77.9	−71.8	−82.4	−80.2
C,C(2)	−21.5	−19.8	−29.5	−30.3
H,O	−116.7	−80.3	−59.4	−55.1
O,C	208.4	209.4	208.6	227.3
O,O	95.3	109.6	127.9	139.5
H,N	−135.9	−110.1	−89.9	−91.8
C,N	−11.6	20.8	17.4	7.5
N,C	−46.1	−50.7	−57.1	−58.0
N,C(2)	−226.1	−158.8	−257.5	−267.4
N,C(3)	15.0	75.9	44.7	54.6
O,N	228.2	236.8	254.0	267.4
F	38.4	32.7	26.3	26.3
Cl	−2.0	−1.6	0.4	0.7
Br	−19.2	−19.2	−22.2	−21.8
S	−59.9	−52.7	−75.6	−74.9
H,S	18.4	6.6	71.0	72.3
S,S	21.6	12.3	51.0	55.3
P	−35.0	12.1	21.7	14.9
O,P	153.9	135.9	196.3	202.7
S,P	114.3	104.9	104.8	122.5

Including diffuse basis functions has a much more significant impact on the calculated aqueous solvation free energies for the anionic solutes. Depending on whether diffuse basis functions are used, the average errors in the selectively clustered-ion data set range from 3.56 kcal/mol to 5.56 kcal/mol. For anions, the MPW25/MIDI!6D level of theory performs the poorest of all the levels of theory, which is not surprising since it was also shown to give the least accurate point-charge-derived dipole moments for anions. For the cations, including diffuse functions has almost no effect on the calculated aqueous solvation free energies—the average errors for the cationic portion of the selectively clustered-ion data set range from only 2.72 kcal/mol to 2.82 kcal/mol, and they do not follow the same trend as for the anions. Although, based on the above results, it is tempting to suggest that diffuse basis functions are a necessary requirement for calculating aqueous solvation free energies of anions, the role diffuse functions play in the liquid phase is not as well understood as their role in the gas phase. Therefore, we will avoid making any general statements regarding the relationship between using diffuse functions and the accuracy of various calculated liquid-phase properties.

Earlier work revealed[16] that for a given basis set and hybrid density functional using the same set of atomic surface tensions for any fraction *X* of Hartree−Fock exchange had little effect on the overall performance of our models. We also found this to be true here, so the surface tension coefficients in Table 8 can be used for any fraction of

Table 9. Mean Unsigned Errors (kcal/mol) in Aqueous Solvation Free Energies Calculated Using SM6, by Solute Class

		MPW25					
solute class	*N*	MIDI!6D	6-31G(d)	6-31+G(d)	6-31+G(d,p)	B3LYP/ 6-31+G(d,p)[a]	B3PW91/ 6-31+G(d,p)[a]
inorganic compounds	8	0.96	0.99	1.11	1.08	1.09	1.09
n-alkanes	8	0.53	0.58	0.62	0.59	0.53	0.57
branched alkanes	5	0.40	0.44	0.39	0.38	0.34	0.38
cycloalkanes	5	0.51	0.48	0.52	0.50	0.43	0.49
alkenes	9	0.30	0.29	0.28	0.26	0.24	0.26
alkynes	5	0.22	0.18	0.28	0.27	0.35	0.29
arenes	8	0.26	0.19	0.30	0.29	0.30	0.26
all hydrocarbons	40	0.37	0.36	0.40	0.38	0.36	0.37
alcohols	12	0.62	0.53	0.46	0.43	0.48	0.44
phenols	4	0.69	0.63	0.67	0.65	0.95	0.70
ethers	12	0.49	0.47	0.51	0.49	0.49	0.48
aldehydes	6	0.42	0.28	0.34	0.39	0.34	0.37
ketones	12	0.31	0.30	0.55	0.57	0.54	0.55
carboxylic acids	5	0.57	0.65	0.90	0.83	0.93	0.84
esters	13	0.33	0.40	0.52	0.46	0.41	0.44
bifunctional H,C,O compounds	5	0.70	0.53	0.47	0.46	0.50	0.46
peroxides[b]	3	0.21	0.20	0.27	0.26	0.22	0.26
all H,C,O compounds[c]	115	0.42	0.40	0.46	0.44	0.45	0.44
aliphatic amines	15	0.67	0.63	0.60	0.61	0.59	0.60
anilines	7	0.48	0.55	0.72	0.79	1.12	0.84
aromatic nitrogen heterocycles (1N in ring)	10	0.21	0.22	0.32	0.35	0.53	0.37
aromatic nitrogen heterocycles (2Ns in ring)	2	0.71	0.56	0.53	0.49	0.51	0.50
hydrazines[d]	3	0.97	0.86	0.91	0.91	0.84	0.90
nitriles	4	0.41	0.34	0.63	0.62	0.69	0.61
bifunctional H,C,N compounds	3	0.33	0.32	0.39	0.39	0.51	0.39
all H,C,N compounds[c]	85	0.45	0.44	0.50	0.50	0.54	0.50
amides	4	1.07	0.89	1.02	0.97	1.10	0.99
ureas	2	0.59	0.31	0.61	0.45	0.25	0.40
nitrohydrocarbons	7	0.58	0.28	0.24	0.30	0.35	0.29
bifunctional H,C,N,O compounds	4	0.67	0.58	0.40	0.46	0.34	0.46
H,C,N,O compounds[c]	177	0.50	0.46	0.53	0.52	0.55	0.52
fluorinated hydrocarbons	6	0.51	0.46	0.41	0.42	0.43	0.42
chlorinated hydrocarbons	27	0.43	0.34	0.37	0.40	0.40	0.39
brominated hydrocarbons	14	0.24	0.20	0.34	0.33	0.22	0.31
multihalogen hydrocarbons	12	0.31	0.30	0.43	0.42	0.42	0.43
halogenated bifunctional compounds	9	1.00	0.95	1.24	1.21	1.23	1.20
thiols	4	0.29	0.32	0.30	0.28	0.29	0.27
sulfides[e]	5	0.65	0.78	0.48	0.46	0.40	0.44
disulfides	2	0.21	0.16	0.37	0.34	0.38	0.34
sulfur heterocycles	1	0.40	0.11	0.86	0.84	0.60	0.80
halogenated sulfur compounds	2	1.70	1.90	0.47	0.29	0.36	0.25
all non-phosphorus sulfur compounds	14	0.61	0.67	0.44	0.39	0.38	0.37
all phosphorus compounds [f]	14	0.64	0.72	1.30	1.18	1.10	1.22
neutrals	273	0.50	0.47	0.55	0.54	0.55	0.54
H,C,N,O ions[c,g]	91	4.26	3.83	3.39	3.30	3.46	3.34
F,Cl,Br,S ions[g,h]	21	4.17	3.63	2.82	2.83	3.01	2.85
anions[g]	60	5.56	4.66	3.68	3.56	3.76	3.62
cations[g]	52	2.72	2.80	2.82	2.81	2.93	2.82
ions[g]	112	4.24	3.80	3.28	3.21	3.37	3.25

[a] Surface tension coefficients optimized for the MPW25/6-31+G(d,p) level of theory were used for these calculations. [b] The inorganic solute hydrogen peroxide is included in this solute class as well as the inorganic compound solute class. [c] Solutes containing at most the listed elements. [d] The inorganic solute hydrazine is included in this solute class as well as the inorganic compound solute class. [e] The inorganic solute hydrogen sulfide is included in this solute class as well as the inorganic compound solute class. [f] The inorganic compound phosphine is included in this solute class as well as the inorganic compound solute class. [g] The ion data in this table are all taken from the selectively clustered-ion data set. [h] Solutes containing at least one of the listed elements plus, in most cases, elements from the previous row.

Hartree−Fock exchange. Furthermore, we propose that they can be used with any good density functionals. To test the accuracy of the surface tensions in Table 8 when used with density functionals other than MPW*X*, we calculated aqueous solvation free energies at the SM6/B3LYP[69]/6-31+G(d,p)//MPW25/MIDI! and SM6/B3PW91[133]/6-31+G(d,p) //MPW25/MIDI! levels of theory (both of these functionals have $X = 20\%$) using the atomic surface tension coefficients in Table

8. The resulting errors are shown in the final two columns of Table 9. In almost all cases, the aqueous solvation free energies calculated at these levels of theory are very close to those calculated at the SM6/MPW25/6-31+G(d,p)//MPW25/MIDI! level of theory. This invariance is due to the ability of these density functionals to deliver accurate electronic wave functions, and the above results are encouraging because they show that the above parameters can be applied to a wide variety of different density functionals, assuming that the given density functional is able to provide a reasonably accurate electronic wave function for the solute of interest.

6. Discussion

6.1. Optimizing Solvation Parameters Based on Gas-Phase Geometries. The first point worth some discussion is the fact that the method is parametrized using gas-phase geometries. We optimize the parameters based on gas-phase geometries here as well as in recent previous work[7,10−12,14−17,74,75] for two reasons. First, for many solutes, less expensive (e.g., semiempirical or molecular mechanics methods) can yield accurate gas-phase geometries. Second, our experience is that optimizing solvation parameters based on gas-phase geometries does not cause a problem because for all or almost all of the molecules in the parametrization set (with the possible exception of those containing an explicit solvent molecule), the difference in solvation free energy between using a gas-phase geometry and using an aqueous geometry is smaller than the mean error of the model. Having obtained the parameters with such a training set, they can be used more broadly. Thus, when the geometry does change significantly in solution, the molecule should be optimized in the aqueous phase, and such optimization would be expected to give more accurate results in such a case. The ASA algorithm[122] that we use for the solvation calculations has excellent analytic gradients that allow for efficient and stable geometry optimization in solution.

6.2. Optimizing Atomic Radii Against the Unclustered-Ion Data Set. The results presented above suggest that for some of the ions considered here as well as water adding an explicit water molecule to the calculation is one way to increase the accuracy of the model for these solutes. However, since the solute−water clusters in Table 3 were included in the training set used to obtain the parameters contained in the present model, an interesting question is whether better results might be obtained if only the unclustered ions were considered. One strategy that could be used to improve the performance of the model for the unclustered-ion data set would simply be to repeat the above optimization using the unclustered instead of the selectively clustered-ion data set. Indeed, it was shown in section 5.3 that the solvation free energies of the ions are quite sensitive to the choice of radii. Based on a careful examination of the solvation free energies for all of the ions in our data set though, we determined that some of the bare ions required a much different set of parameters than the majority of the other solutes in our data set. Therefore, fitting the unclustered-ion data set with a single set of atomic radii gives uneven results, and that strategy was abandoned.

Another strategy that could be used to try to fit all of the data in the unclustered-ion data set would be to make the atomic radius of a given atom not only a function of its atomic number but also a function of its partial atomic charge. In fact, our earliest models (SM1-SM3[3−6]) used charge-dependent intrinsic Coulomb radii. We experimented with several different functional forms for charge-dependent radii, including (as just one example) making intrinsic Coulomb radii a quadratic function of partial atomic charge, and we found that while making the intrinsic Coulomb radii a function of partial atomic charge did improve the performance of the model for many of the ions it did so at the cost of having a deleterious effect on some of the other ions. We eventually abandoned the idea of using charge-dependent radii because of this finding and for two additional reasons. First, a model that uses charge-dependent radii would be more likely to be highly basis-set dependent than a model that uses atomic-number-dependent radii.[35,36] Second, charge-dependent atomic radii might lead to highly questionable results in cases where this dependence has not been carefully examined (e.g. our data set does not contain any zwitterions or large biomolecules) or in cases where the atomic radius might not be a smooth function of its partial atomic charge (e.g. transition states that involve the displacement of a charged or partially charged leaving group).

Several additional strategies that were not considered here, but that have been used by others, include using atom-typed radii[18−23] or using different sets of atomic radii for neutrals and ions.[134] An example of this former type of approach is the united atom for Hartree−Fock (UAHF)[23] method of Barone and co-workers, in which the atom's radius depends on its hybridization state, connectivity, and formal charge (so, in a sense, this method falls under both of the above categories). Here, we did not consider using atom-typed radii because it is not completely clear whether they can be used for modeling chemical reactions or predicting activation free energies (which require modeling a transition state) in the liquid phase.

We feel that adding one or more explicit water molecules to the calculation is the most reasonable approach to modeling some ionic solutes with a continuum solvent model. Ideally we would give a definite prescription for when the user should add a specific water molecule in using this model. However, it is not possible to do this in a way that covers the great diversity of possible cases that occur in applications, especially if one includes reaction paths and enzymes. One prescription would be to add an explicit water whenever one wants to improve the accuracy, since adding an explicit water should almost always improve the accuracy when the effect is large, but it is relatively safe because it cannot make the accuracy much worse when the effect is small. In considering this question, we should note that although we obtain better results when we add an explicit water in cases where the given solute satisfies one or more of the three criteria explained in section 5.2, we also obtain reasonably good results in most cases even without the explicit water (see Table 10).

Although not tested as part of this work, the above strategy of adding one or more explicit solvent molecules should also improve the accuracy for predicting solvation free energies of some solutes in nonaqueous solvents, in particular those where specific solute−solvent hydrogen bonds are expected

Table 10. Mean Unsigned Errors (kcal/mol) in Aqueous Solvation Free Energies Calculated Using Different Continuum Solvent Models

	MPW25/6-31+G(d,p)		HF/6-31G(d)					MPW25/6-31+G(d,p)
	SM6	SM5.43R	DPCM/98	DPCM/03	CPCM/98	CPCM/03	IEF-PCM/03	IEF-PCM/03
neutrals[a]	0.54	0.62	1.02	1.40	1.06	1.11	1.10	1.21
unclustered ions[b]	4.19	6.12	4.40	14.95	5.02	6.32	6.37	7.78
clustered ions[b]	3.21	6.16	5.86	14.32	6.33	7.58	7.63	8.98
all ions[c]	4.38	6.92	5.83	15.60	6.40	7.47	7.53	8.88
all data[d]	1.86	2.79	2.69	6.28	2.91	3.30	3.31	3.85

[a] 273 molecules. [b] 112 ions. [c] 143 ions. [d] 416 data.

to occur. Of course, testing this strategy for solvents other than water would require a careful comparison between experimental and calculated solvation free energies of solutes in nonaqueous solvents, which is beyond the scope of the present work.

6.3. Performance of Other Continuum Models. Using our database of aqueous solvation free energies, along with gas-phase geometries optimized at the MPW25/MIDI! level of theory, we tested the performance of the SM5.43R and PCM continuum models for predicting aqueous solvation free energies. For all PCM calculations, we used the UAHF method for assigning atomic radii,[23] which is the recommended method for predicting aqueous solvation free energies with PCM according to the *Gaussian 03* manual.[135] Because the parameters contained in the UAHF method were originally optimized for use with the HF/6-31G(d) level of theory,[23] we used this level of theory to calculate aqueous solvation free energies for all of the PCM methods tested here. (Thus the PCM methods are tested in a way that should allow them to perform at their best.) However, we also wanted to see what effect changing the level of theory had on the accuracy of PCM, so we tested one of the PCM methods described below at the PCM/MPW25/6-31+G(d,p)//MPW25/MIDI! level of theory also.

There are several different varieties of PCM, and most of these are implemented differently in *Gaussian 98* [136] and *Gaussian 03*.[135] Here, we tested the dielectric version[26−28] of PCM (DPCM) as implemented in both *Gaussian 98* and *Gaussian 03*.[29] These two models will be referred to as DPCM/98 and DPCM/03, respectively. We also tested CPCM/98[25,137,138] and CPCM/03[24,137,138] as well as the default PCM method in *Gaussian 03*, IEF-PCM/03.[29−32] The IEF-PCM/03 model is particularly interesting because Chipman[139] found that it includes charge penetration effects "extremely well for all solutes". The results of these calculations are summarized in Table 10.

The data in Table 10 show that SM6 outperforms all of the other models tested above for both neutral and ionic solutes. For PCM, the most accurate solvation free energies are obtained using the older DPCM/98 implementation. The data in Table 10 also show that changing the level of theory has a negative effect on the performance of IEF-PCM/03, which is not surprising, since the UAHF method for assigning atomic radii was optimized[23] using DPCM/98 at the HF/6-31G(d) level of theory. Of course, the performance of SM6 for anions is dependent on the basis set used (see the data in Table 9), although its performance improves as the basis set size in increased (for neutrals and cations, there is very little basis-set dependence in aqueous solvation free energies calculated with SM6).

Comparing the overall errors for the unclustered-ion data set to those in the selectively clustered-ion data set shows that only the performance of SM6 improves significantly when a single explicit solvent molecule is added to the calculation (the overall error decreases from 4.19 kcal/mol for the ions in the unclustered-ion data set to 3.21 kcal/mol for the ions in the selectively clustered-ion data set). Again, this suggests that including a small number of explicit water molecules in SM6 calculations may be an effective strategy for predicting the aqueous solvation free energies of some ions in cases where strong solute−solvent hydrogen bonds are expected to play an important role in the aqueous phase. Furthermore, we feel that this is a much more reasonable strategy than trying to use drastically scaled values for the atomic radii or atom-typed or charge-dependent radii. The excellent performance of the SM6 model as compared to all the models in the popular *Gaussian* packages is especially remarkable when one remembers that the atomic radii in SM6 are functions of only atomic number (and the radii for O and F are not even optimized), whereas the recommended radii used in *Gaussian* depend on connectivity, hybridization state, and formal charge.

7. Concluding Remarks

We have presented a new database of experimental aqueous solvation free energies that contains 273 neutral and 143 ionic solutes, including 31 ion−water clusters. Using these data, we developed a new continuum solvent model called SM6. This model can be used to calculate aqueous solvation free energies and, although not demonstrated here, liquid-phase geometries in aqueous solution. SM6 uses partial atomic charges obtained from a new charge model, Charge Model 4 (CM4), which has been shown to give accurate partial charges for both neutral and ionic solutes. In addition, we have shown that the partial atomic charges obtained from CM4 are much less dependent upon changes in the basis set than partial atomic charges obtained from a Löwdin or Redistributed Löwdin population analysis of the wave function.

For some of the ions in our data set, we showed that the addition of a single explicit water molecule to the calculation (i.e., modeling the solute as a solute−water cluster) improved the performance of SM6 for predicting aqueous solvation free energies, indicating that large numbers of solvent molecules are not necessarily required for improving the treatment of some strong solute−solvent hydrogen bonds in

the first solvation shell. This is encouraging, because treating large numbers of solvent molecules explicitly often presents many problems. Furthermore, we feel that this strategy is more reasonable than using unphysical values for the atomic radii or using atom-typed or charge-dependent radii.

We also used our new database of aqueous solvation free energies to test the performance of several other continuum solvent models, namely SM5.43R and several different implementations of PCM. For both neutral and ionic solutes, SM6 outperforms all of the methods against which it was tested for predicting aqueous solvation free energies. Furthermore, we found that SM6 is the only model of those tested here (except for one model with a mean error 3.4 times larger) that improves upon the addition of a single explicit water molecule to the calculation. Thus, unlike the other models tested here, adding a single explicit water molecule to SM6 calculations in cases where strong solute−solvent hydrogen bonds are expected to occur in the aqueous phase appears to be both practical and effective for improving the accuracy of the present model for these types of solutes.

Finally, it was shown that SM6 retains its accuracy when used in conjunction with the B3LYP and B3PW91 functionals. Based on this analysis, we proposed that the charge and solvation parameters obtained with a given basis set (charge and solvation parameters have been optimized for the MIDI!6D, 6-31G(d), 6-31+G(d), and 6-31+G(d,p) basis sets) may be used with any good density functional or fraction of Hartree−Fock exchange.

Availability of SM6. All of the SM6 parametrizations presented in this article are available in the SMXGAUSS[140] program. This program can read a *Gaussian* output file corresponding to a gas-phase calculation of a given solute and carry out a single-point calculation with SM6. In addition, the above program allows liquid-phase geometry optimizations and Hessian calculations to be carried out with SM6. Although SMXGAUSS requires only a *Gaussian* output file to perform SM6 calculations, users that have a *Gaussian 03* executable can use SMXGAUSS in conjunction with the powerful geometry optimizers available in *Gaussian*. For non-*Gaussian* users, the CM4 and SM6 parametrizations are also available in the GAMESSPLUS[141] and HONDOPLUS[142,143] software programs. All three of these programs are available free of charge and can be downloaded from our Web site, http://comp.chem.umn.edu/software.

Acknowledgment. We thank Adam Moser for his help during preparation of the ionic portion of our data set. This work was supported by the NIH training grant for Neurophysical-computational Sciences, by the U.S. Army Research Office under Multidisciplinary Research Program of the University Research Initiative (MURI) through grant number DAAD19-02-1-0176, by the Minnesota Partnership for Biotechnology and Medical Genomics, and by the National Science Foundation (CHR-230446).

Supporting Information Available: Further details regarding CM4, all 273 experimental aqueous solvation free energies for the neutral solutes, and a breakdown of errors by data set for all of the SM6 methods tested in this paper. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.

(2) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(3) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305−8311.

(4) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *6*, 629−666.

(5) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213−217.

(6) Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *13*, 1089−1097.

(7) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257−3271.

(8) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385−16398.

(9) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.

(10) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293−298.

(11) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc*. **1999**, *103*, 9−63.

(12) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117−9133.

(13) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 5503−5513.

(14) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 2178−2182.

(15) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532−6542.

(16) Thompson, D. J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 107−131.

(17) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem*. **2000**, *21*, 340−366.

(18) Banavali, N. K.; Roux, B. *J. Phys. Chem. B* **2002**, *106*, 11026−11035.

(19) Curutchet, C.; Bidon-Chanal, A.; Soteras, I.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2005**, *109*, 3565−3574.

(20) Marten, B.; Kim, K.; Cortis, C.; Friesner, R.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775−11778.

(21) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.

(22) Swanson, J. M. J.; Adcock, S. A.; McCammon, J. A. *J. Chem. Theory. Comput.* **2005**, *1*, 484−493.

(23) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210−3221.

(24) Cossi, M.; Rega, N.; Scalamani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669−681.

(25) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(26) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327−335.

(27) Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* **1981**, *55*, 117−129.

(28) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239−245.

(29) Cossi, M.; Scalamani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43−54.

(30) Cancés, M. T.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032−3041.

(31) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253−260.

(32) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151−5158.

(33) Mennucci, B.; Cancés, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506−10517.

(34) Tomasi, J.; Mennucci, B.; Cancés, E. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 211−226.

(35) Aguilar, M. A.; Martin, M. A.; Tolosa, S.; Olivares del Valle, F. J. *J. Mol. Struct. (THEOCHEM)* **1988**, *166*, 313−318.

(36) Aguilar, M. A.; Olivares del Valle, F. J. *Chem. Phys.* **1989**, *129*, 439−450.

(37) Gonçalves, P. F. B.; Livotto, P. R. *Chem. Phys. Lett.* **1999**, *304*, 438−444.

(38) Olivares del Valle, F. J.; Aguilar, M. A.; Contador, J. C. *Chem. Phys.* **1993**, *170*, 161−165.

(39) Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1994**, *116*, 3949−3954.

(40) Takahashi, O.; Sawahata, H.; Ogawa, Y.; Kikuchi, O. *J. Mol. Struct. (THEOCHEM)* **1997**, *393*, 141−150.

(41) Camaioni, D. M.; Dupuis, M.; Bentley, J. *J. Phys. Chem. A* **2003**, *107*, 5778−5788.

(42) Rinaldi, D.; Rivail, J.-L.; Rguini, N. *J. Comput. Chem.* **1992**, *13*, 675−680.

(43) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413−7422.

(44) Chipman, D. M. *J. Chem. Phys.* **2002**, *116*, 10129−10138.

(45) Rode, B. M.; Schwenk, C. F.; Tongraar, A. *J. Mol. Liq.* **2004**, *110*, 105−112.

(46) Licheri, G.; Piccaluga, G.; Pinna, G. *J. Chem. Phys.* **1976**, *64*, 2437−2441.

(47) Kalko, S. G.; Sesé, G.; Padró, J. A. *J. Chem. Phys.* **1996**, *104*, 9578−9585.

(48) Obst, S.; Bradaczek, H. *J. Phys. Chem.* **1996**, *100*, 15677−15687.

(49) Probst, M. M.; Radnai, T.; Heinzinger, K.; Bopp, P.; Rode, B. M. *J. Phys. Chem.* **1985**, *89*, 753−759.

(50) Tongraar, A.; Liedl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1997**, *101*, 6299−6309.

(51) Kerdcharoen, T.; Morokuma, K. *J. Chem. Phys.* **2003**, *118*, 8856−8862.

(52) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory. Comput.* **2005**, *1*, 415−432.

(53) Pliego Jr., J. R.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434−7439.

(54) Pliego Jr., J. R.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241−7247.

(55) Fox, T.; Rösch, N. *Chem. Phys. Lett.* **1992**, *191*, 33−37.

(56) Karelson, M.; Zerner, M. C. *J. Am. Chem. Soc.* **1990**, *112*, 9405−9406.

(57) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Viellard, A., Eds.; Reidel: Dordrecht, 1980; pp 25ff.

(58) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Dordrecht, 1996; pp 1−80.

(59) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754−2759.

(60) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, 379−400.

(61) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281−301.

(62) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1998**, *99*, 192−196.

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

(64) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664−675.

(65) Perdew, J. P. *Electronic Structure of Solids '91*; Zieesche, P., Eshrig, H., Eds.; Akademie: Berlin, 1991.

(66) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811−4815.

(67) Kormos, B. L.; Cramer, C. J. *J. Phys. Org. Chem.* **2002**, *15*, 712−720.

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

(69) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(70) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69−89.

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

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

(73) Cramer, C. J.; Truhlar, D. G. In *Free Energy Calculations in Rational Drug Design*; Reddy, M. R., Erion, M. D., Eds.; Kluwer/Plenum: New York, 2001; pp 63−95.

(74) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 7147−7157.

(75) Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1998**, *63*, 4305−4313.

(76) *Physical/Chemical Property Database (PHYSPROP);* SRC Environmental Science Center: Syracuse, NY, 1994.

(77) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 1661−1670.

(78) Curtiss, L. A.; Frurip, D. L.; Blander, M. *J. Chem. Phys.* **1979**, *71*, 2703−2711.

(79) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583−5595.

(80) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109−6114.

(81) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105*, 11534−11540.

(82) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. J.; Earhart, A. D.; Coe, J. V. *J. Phys. Chem. A* **1998**, *102*, 7787−7794.

(83) Lias, S. G.; Bartness, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. Ion Energetics Data. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, March 2003.

(84) Hunter, E. P. L.; Lias, S. G. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413−656.

(85) Régimbal, J. M.; Mozurkewich, M. *J. Phys. Chem. A* **1997**, *101*, 8822−8829.

(86) O'Sullivan, D. W.; Lee, M.; Noone, B. C.; Heikes, B. G. *J. Phys. Chem.* **1996**, *100*, 3241−3247.

(87) Stewart, R. *The Proton: Applications to Organic Chemistry*; Academic: New York, 1985.

(88) Arnett, E. M. *Prog. Phys. Org. Chem.* **1963**, *1*, 223−403.

(89) Kresge, A. J.; Pruszynski, P.; Stang, P. T.; Williamson, B. J. *J. Org. Chem.* **1991**, *56*, 4808−4811.

(90) Perrin, D. D. *Ionisation Constants of Inorganic Acids and Bases in Aqueous Solution*; Pergamon: New York, 1982.

(91) Albert, A.; Serjeant, E. P. *The Determination of Ionization Constants: A Laboratory Manual*; Chapman and Hall: New York, 1984.

(92) *Ionisation Constants of Organic Acids in Aqueous Solution*; Serjeant, E. P., Dempsey, B., Eds.; Pergamon: New York, 1979.

(93) Pliego Jr., J. R.; Riveros, J. M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 1622−1627.

(94) Lewis, A.; Bumpus, J. A.; Truhlar, D. G.; Cramer, C. J. *J. Chem. Educ.* **2004**, *81*, 596−604.

(95) Winget, P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2004**, *112*, 217−227.

(96) Palascak, M. W.; Shields, G. C. *J. Phys. Chem. A* **2004**, *108*, 3692−3694.

(97) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264−6271.

(98) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407−410.

(99) Lowry, T. H.; Schueller-Richardson, K. *Mechanism and Theory in Organic Chemistry*, 3rd ed.; Harper Collins: New York, 1987.

(100) Meot-Ner, M. M.; Lias, S. G. Binding Energies Between Ions and Molecules, and the Thermochemistry of Cluster Ions. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, March 2003.

(101) Ben-Naim, A. *Solvation Thermodynamics*; Plenum: New York, 1987.

(102) Hoijtink, G. J.; Boer, E. D.; Meij, P. H. v. D.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas.* **1956**, *75*, 487−503.

(103) Peradejori, F. *Cah. Phys.* **1963**, *17*, 393−447.

(104) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164−170.

(105) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(106) Cramer, C. J.; Truhlar, D. G. In *Rev. Comp. Chem.*; Boyd, D. B., Lipkowitz, K. B., Eds.; VCH Publishers: New York, 1995; Vol. 6; pp 1−72.

(107) Winget, P.; Dolney, D. M.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *Minnesota solvent* descriptor database [http://comp.chem.umn.edu/solvation/mnsddb.pdf], 1999.

(108) Thompson, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291−1304.

(109) Winget, P.; Thompson, D. J.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 10707−10717.

(110) Brom, J. M.; Schmitz, B. J.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 6483−6488.

(111) Kalinowski, J. A.; Lesyng, B.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2545−2549.

(112) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133−151.

(113) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365−375.

(114) Golebiewski, A.; Rzescowska, E. *Acta Phys. Pol.* **1974**, *45*, 563−568.

(115) Baker, J. *Theor. Chim. Acta* **1985**, *68*, 221−229.

(116) Kar, T.; Sannigrahi, A. B.; Mukherjee, D. C. *J. Mol. Struct. (THEOCHEM)* **1987**, *153*, 93−101.

(117) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. *PhysChemComm* **2002**, *5*, 117−134.

(118) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.

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

(120) Born, M. *Z. Physik.* **1920**, *1*, 45−48.

(121) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465−1471.

(122) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422−440.

(123) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712−3720.

(124) Qui, D.; Shenkin, P.; Hollinger, F.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.

(125) Srinvasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426−434.

(126) Zhang, W.; Hou, T.; Xu, X. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 88−93.

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

(128) Thompson, J. D.; Winget, P.; Truhlar, D. G. *PhysChemComm* **2001**, *4*, 72−77.

(129) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.

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

(131) *NAG Fortran 90 Library*, 4th ed.; The Numerical Algorithms Group, Inc.: Oxford, 2000.

(132) Carroll, D. L. *GA Driver, Version 1.7a*; CU Aerospace, University of Illinois: Urbana, IL, 2001.

(133) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(134) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588−5593.

(135) 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.; Scalamani, G.; Rega, N.; Petersson, G. A.; Nagatsuji, 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 C.01*; Pittsburgh, PA, 2003.

(136) 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. W.; 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*; Pittsburgh, PA, 1998.

(137) Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995**, *103*, 9312−9320.

(138) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.

(139) Chipman, D. M. *Theor. Chem. Acc.* **2002**, *107*, 80−89.

(140) Kelly, C. P.; Thompson, D. J.; Lynch, B. J.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Volobuev, Y.; Dupuis, M.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. SMXGAUSS-version 3.0; University of Minnesota: Minneapolis, MN 55455, 2005.

(141) Kelly, C. P.; Pu, J.; Thompson, D. J.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Lynch, B. J.; Liotard, D. A.; Rinaldi, D.; Gao, J.; Cramer, C. J.; Truhlar, D. G. GAMESSPLUS-version 4.7; University of Minnesota, Minneapolis, 2005, based on the General Atomic and Molecular Electronic Structure System (GAMESS) as described in Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(142) Dupuis, M.; Marquez, A.; Davidson, E. R. *HONDO 99.6*; based on HONDO 95.3, Dupuis, M.; Marquez, A.; Davidson, E. R., Quantum Chemistry Program Exchange (QCPE); Indiana University: Bloomington, IN, 1999.

(143) Kelly, C. P.; Nakamura, H.; Xidos, J. D.; Thompson, D. J.; Li, J.; Hawkins, G. D.; Zhu, T.; Lynch, B. J.; Volobuev, Y.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. HONDOPLUS-version 4.7; University of Minnesota, Minneapolis, 2005, based on HONDO v. 99.6.

CT050164B

JCTC Journal of Chemical Theory and Computation

Electron-Impact Ionization Cross Sections of Molecules Containing Heavy Elements ($Z > 10$)

Gregory E. Scott

Department of Chemistry, Davidson College, Davidson, North Carolina 28035-7120

Karl K. Irikura*

Computational Chemistry Group, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8380

Received March 25, 2005

Abstract: The binary-encounter-Bethe (BEB) theory has been successful for computing electron-impact ionization cross sections of many molecules. For molecules that contain heavy atoms (defined here as atoms with valence principal quantum number $n > 2$), there are two alternative BEB procedures in the literature. The first involves a kinetic-energy correction for molecular orbitals that are dominated by atomic orbitals with $n > 2$. The second alternative is to use effective core potentials (ECPs), which were developed for other purposes but yield valence pseudo-orbitals with reduced kinetic energies. In the present study, the results of these two approaches are compared with experimental cross sections for several molecules containing heavy elements. Although both procedures perform well, the ECP results agree somewhat better with experimental measurements. Cross sections are presented for C_2Cl_6, C_2HCl_5, C_2Cl_4, both isomers of $C_2H_2Cl_4$, CCl_4, $TiCl_4$, CBr_4, $CHBr_3$, CH_2Br_2, P_2, P_4, As_2, As_4, GaCl, CS_2, H_2S, CH_3I, $Al(CH_3)_3$, $Ga(CH_3)_3$, hexamethyldisiloxane, and $Zn(C_2H_5)_2$. Incorrect BEB calculations have been reported in the literature for several of these molecules. As an ancillary result, the dipole polarizability of $Zn(C_2H_5)_2$ is predicted to be 12.1 Å^3.

Introduction

Electron impact ionization cross sections are essential quantities for modeling plasma chemistry, which is important in a variety of practical processes. For example, low-temperature plasmas are important in semiconductor processing, in the destruction of volatile organic compounds, for modifying the mechanical properties of surfaces, and in wall-chemistry in nuclear fusion reactors. Absolute ionization cross sections are also necessary to obtain quantitative gas densities from mass-spectrometric measurements, as in flame sampling and Knudsen-cell thermochemistry.

Absolute ionization cross sections are difficult to measure precisely. Even for convenient, stable molecules, experimental groups often disagree on the values of the cross sections.[1] Furthermore, many of the interesting cross sections are for molecules that are especially difficult to measure, such as free radicals and molecular ions. Thus, reliable theoretical predictions are valuable.

The binary-encounter-Bethe (BEB) model[2,3] has been shown to produce reliable total cross sections for a wide variety of molecules.[4] For molecules containing heavy atoms, here meaning atoms with atomic number $Z > 10$, the ab initio computations required by the BEB model can be performed in two different ways: with all electrons explicit or by using core pseudopotentials, also known as effective core potentials (ECPs). Experimental cross sections are available for several such molecules,[5−7] warranting a comparison of the two computational approaches to BEB cross sections. Note that partial ionization cross sections, i.e., cracking patterns, cannot yet be predicted theoretically.

* Corresponding author phone: (301)975-2510; e-mail: karl.irikura@nist.gov.

10.1021/ct050077j CCC: $30.25
Published on Web 09/30/2005

Theoretical Procedures

In BEB theory, the total electron-impact ionization cross section is expressed as a sum of cross sections for each molecular orbital (MO)

$$\sigma_{\mathrm{MO}}(T;n) = \frac{S}{t + (u+1)/n}\left[\frac{\ln t}{2}\left(1 - \frac{1}{t^2}\right) + 1 - \frac{1}{t} - \frac{\ln t}{t+1}\right] \tag{1}$$

where T is the energy of the incident electron, $t = T/B$, $u = U/B$, $S = 4\pi a_0^2 N(R/B)^2$, a_0 is the Bohr radius, and R is the Rydberg energy. For each molecular orbital, B is the binding energy (i.e., the vertical ionization energy), U is the kinetic energy, and N is the number of electrons in the orbital (i.e., the occupation number). When $T < B$, the MO cannot be ionized, so $\sigma_{\mathrm{MO}} = 0$. The constant n is a kinetic-energy scaling factor whose role will be discussed below as it pertains to the two different methods of calculation. When B exceeds the second ionization energy (IE_2), the molecule is presumed to acquire a second charge through an Auger process. In that case, the contribution of σ_{MO} is doubled to correspond to experimental measurements of ion current, that is, the gross ionization cross section. For a molecule composed only of light atoms, eq 1 is used with $n = 1$ for all orbitals. When a molecule contains heavy atoms, there are two computational alternatives for applying the BEB model.

In the first method,[8] all electrons are included explicitly in the ab initio calculations. The scaling factor n in eq 1 is unity except for orbitals that are dominated by atomic orbitals with principal quantum number > 2, as judged by a Mulliken population greater than some threshold. When this is the case, n is set equal to the principal quantum number of the dominant atomic orbitals. We have typically used a threshold of 50%; choosing a different value generally changes the peak cross section by less than 10%.[1] However, changing the threshold affects the predicted cross sections systematically. As suggested by a referee, we consider the threshold as an adjustable parameter in this study.

In the second method,[9] the core electrons of the heavy atoms are replaced by effective potentials (ECPs). Since there are no core orbitals to which they must be orthogonal, the resulting valence pseudo-orbitals lack the radial nodes of normal orbitals, making their kinetic energies much lower than normal. [Nodes indicate oscillatory behavior, which corresponds to large gradients and curvatures in the orbital function. Eliminating radial nodes decreases the kinetic energy, especially in the core, because the principal radial dependence of the kinetic energy is proportional to $[\partial^2/\partial r^2 + (2/r)(\partial/\partial r)]$.[10]] Thus, eq 1 is used with $n = 1$ for all valence molecular pseudo-orbitals. No Mulliken populations or arbitrary thresholds are needed. However, since many core orbitals are missing in an ECP calculation, their contribution to the ionization cross section can only be obtained from a separate, all-electron calculation.

An intermediate situation sometimes arises in which certain orbitals do not lose any nodes upon introduction of an ECP. In the present study, this occurred for $Zn(C_2H_5)_2$. The 3*d* orbitals on Zn are considered to be valence orbitals and are treated explicitly even when an ECP is in use. However, 3*d* orbitals lack radial nodes, so their kinetic energy is nearly unaffected by using the ECP. Consequently, one should use $n = 3$ for the 3*d* orbitals even when an ECP is used. A similar situation would occur in the case of valence 4*f* orbitals.

Computational Methods[11]

All ab initio calculations in support of the BEB model were performed with the Gaussian 03 program suite using basis sets as implemented therein.[12] Basis sets listed in square brackets were used on atoms with atomic number $Z > 36$ (that is, beyond Kr). Molecular geometries were computed using the B3LYP hybrid density functional[13,14] with the 6-31G(d) [3-21G(d)] basis sets, with all electrons explicit. Vibrational frequencies were computed to verify that all structures are energy minima. These geometries were used for all subsequent calculations.

Orbital binding energies, B, kinetic energies, U [Gaussian 03 keyword iop(6/81=3)], and Mulliken populations were computed at the Hartree−Fock (HF) level using the 6-311G(d,p) [3-21G(d)] basis sets. For pseudopotential calculations, the Stuttgart[15,16] ECPs and corresponding basis sets were used on heavy atoms according to the defaults in the Gaussian 03 software. For the elements relevant to this study, the ECPs are MWB10 (Al−Cl, Zn), MDF10 (Ti), MWB28 (Ga−Br), and MWB46 (I), where "MWB" and "MDF" denote quasi-relativistic and relativistic ECPs, respectively, and the numerical suffix indicates the number of core electrons replaced by the effective potential. A set of *d* polarization functions, taken from the corresponding 6-311G(d,p) [3-21G(d)] all-electron basis, was added to each heavy center. The combination of ECP, basis, and polarization set is labeled ECP(d) here.

The BEB ionization cross section is sensitive to the vertical ionization energy of the molecule, that is, the ionization threshold. Koopmans (i.e., Hartree−Fock) binding energies are too approximate for this purpose. More accurate values of B for the valence orbitals were computed using the outer-valence Green's function (OVGF) method[17,18] with the 6-311+G(d,p) [3-21G(d)] basis sets. For the chlorofluoromethanes, the binding energies from this procedure are lower than corresponding experimental values by only 0.3 eV.[1] OVGF results were rejected for pole strengths < 0.75. When available, experimental vertical ionization energies were used for the outermost valence orbitals. Second ionization energies, IE_2, were taken either from experiment or from B3LYP/6-311G(d,p) [ECP(d)] calculations at the geometry of the neutral molecule, considering both singlet and triplet dications. Mulliken populations for each MO were generated by using the program MullPop,[19] which was kindly modified by its author to accommodate Gaussian 03 output files.

At low impact energies, the core orbitals make only a small contribution to the ionization cross section and can be neglected. Thus, if one is interested only in low energies, ECP calculations alone are adequate for computing ionization cross sections. For such calculations, geometries were optimized at the B3LYP/6-31G(d,p) [ECP(d)] level.

Table 1. Experimental Ionization Energies Used in BEB Calculations[e]

molecule	IE_v[a] (eV)	IE_2[b] (eV)
C_2Cl_6	$2a_{2g} = 11.22$, $9e_g = 11.37$, $2a_{1u} = 11.79$;[48] (11.02, 11.17, 11.52)	28.8 ± 0.5[49] (26.3)
C_2HCl_5	$20a'' = 11.28$, $19a'' = 29a' = 11.56$, $28a' = 18a'' = 12.09$;[48] (10.98, 11.28, 11.45, 11.79, 11.74)	28.7 ± 0.5[49] (26.4)
C_2Cl_4	$3b_{3u} = 9.51$, $7b_{3g} = 11.37$, $2a_u = 7b_{2u} = 12.19$, $2b_{1g} = 8b_{1u} = 9a_g = 12.58$;[48] (9.59, 11.13, 11.94, 12.11, 12.26, 12.43, 12.73)	
1,1,1,2-$C_2H_2Cl_4$	$14a'' = 11.45$, $27a' = 13a'' = 11.67$, $12a'' = 26a' = 11.91$, $25a' = 12.37$, $24a' = 11a'' = 12.83$;[48] (11.13, 11.30, 11.32, 11.57, 11.67, 12.04, 12.50, 12.55)	28.2 ± 0.5[49] (26.8)
1,1,2,2-$C_2H_2Cl_4$	$9b_g = 11.17$, $12a_g = 9a_u = 11.62$;[48] (10.84, 11.34, 11.56)	28.7 ± 0.5[49] (26.8)
CCl_4	$2t_1 = 11.69$, $7t_2 = 12.62$, $2e = 13.44$;[48] (11.39, 12.31, 13.03)	29.1 ± 0.1[50] (28.0)
$TiCl_4$	$2t_1 = 11.69$, $9t_2 = 12.67$, $2e = 13.17$, $8t_2 = 13.46$;[51] (11.69, 12.94, 13.41, 14.12)	
CBr_4	$5t_1 = 10.39$, $13t_2 = 11.07$, $5e = 12.11$;[52] (10.27, 11.12, 11.82)	
$CHBr_3$	$5a_2 = 10.5$, $18e = 10.8$, $15a_1 = 11.3$, $17e = 11.7$;[52] (10.18, 10.65, 10.83, 11.38)	27.9 ± 0.2[53] (26.2)
CH_2Br_2	$13b_2 = 10.6$, $6b_1 = 10.8$, $5a_2 = 15a_1 = 11.3$;[52] (10.28, 10.47, 10.76, 10.90)	27.4 ± 0.2[53] (27.0)
P_2	$2\pi_u = 10.62$, $5\sigma_g = 10.81$;[54] (10.49, 10.71)	
P_4	$2e = 9.46, 9.92$;[c] $6t_2 = 10.36, 10.53, 10.72$;[c] $5a_1 = 11.85$;[55] (9.53, 10.29, 11.70)	
As_2	$4\pi_u = 9.82, 9.96$;[d] $8\sigma_g = 10.22$;[56] (9.75, 10.13)	
As_4	$5e = 8.75, 9.16$;[c] $12t_2 = 9.76, 9.97, 10.11$;[c] $8a_1 = 11.06$;[55] (8.84, 9.76, 10.94)	
GaCl	$12\sigma_g = 10.07$, $5\pi = 11.38$;[36] (9.55, 10.93)	
CS_2	$2\pi_g = 10.06$, $2\pi_u = 12.83$;[48] (9.79, 13.12)	27.22 ± 0.08[57] (27.0)
H_2S	$2b_1 = 10.48$, $5a_1 = 13.25$;[48] (9.87, 12.95)	32.8[58] (31.9)
CH_3I	$9e = 9.54, 10.16$;[d] $13a_1 = 12.50$;[48] (9.15, 12.07)	
$Al(CH_3)_3$	$5e' = 9.85$, $6a' = 12.6$;[59] (9.71, 12.84)	
$Ga(CH_3)_3$	$7e' = 9.76$, $8a' = 13.5$;[60,61] (9.54, 13.16)	
TMS_2O	$16a_1 = 8b_1 = 9.88$, $14b_2 = 7a_2 = 10.73$, $15a_1 = 7b_1 = 12.5$;[62] (9.77, 9.79, 10.59, 10.59, 12.31, 12.31)	
$Zn(C_2H_5)_2$	$15b = 8.6$, $17a = 10.5$;[63] (8.53, 10.28)	

[a] Vertical ionization energy. [b] Double-ionization energy. [c] Jahn−Teller split. [d] Spin−orbit split. [e] Calculated OVGF values listed parenthetically, for comparison. All values in eV.

The series of computations were performed automatically by means of a Perl-language script. The output of this script is a summary of the MO data in the tab-delimited format used in the NIST database.[4] This file serves as input to another Perl script, which computes energy-dependent cross sections.

Electric dipole polarizabilities, which are not part of the BEB procedure but are relevant in the Results section, were computed using the Gaussian 03 and GAMESS programs as convenient.[12,20]

Results

We selected several molecules for which total ionization cross sections have been measured experimentally and which contain as many heavy atoms as possible yet which do not contain atoms so heavy that all-electron basis sets are unavailable. This set of molecules was intended to accentuate differences between the all-electron and ECP predictions. For computational convenience, only closed-shell molecules were considered. Along with literature references for the experimental cross sections, the molecules are as follows: C_2Cl_6, C_2HCl_5, C_2Cl_4, 1,1,1,2-$C_2H_2Cl_4$, and 1,1,2,2-$C_2H_2Cl_4$;[7] CCl_4;[7,21,22] $TiCl_4$;[23] CBr_4, $CHBr_3$, and CH_2Br_2;[6] P_2, P_4, As_2, and As_4;[24] GaCl;[25] CS_2;[26−28] H_2S;[27,29,30] CH_3I;[5,31] $Al(CH_3)_3$ and $Ga(CH_3)_3$;[32] hexamethyldisiloxane (TMS_2O);[33] $Zn(C_2H_5)_2$.[34]

Tables of MO data used to generate the BEB cross sections are available as Supporting Information. As mentioned above, some vertical ionization energies and double-ionization energies were taken from experimental measurements. These experimental values are listed in Table 1 with their sources. The corresponding calculated values are listed parenthetically for comparison. In some cases, the theoretical results led us to reassign the experimental photoelectron spectra.

Ionization cross sections are used for a variety of applications, suggesting a variety of criteria for comparing theoretical predictions with experimental measurements. We consider four criteria in this study:

(1) Peak cross section: The difference between theory and experiment may be expressed in absolute terms ($Å^2$) or as a percentage of the experimental value.

(2) Cross section at energy of maximum difference between AE and ECP predictions: This is the energy at which the two models are most easily distinguished. The difference between theory and experiment may be expressed in absolute terms ($Å^2$) or as a percentage of the experimental value.

(3) Shape of the cross-section curve, plotted as a function of incident electron energy: Predicting the shape of the curve is most useful when reliable experimental measurements are only available at a few impact energies.[35]

(4) Initial slope: The energy dependence of the cross section at low energies is important for plasma modeling. The difference between theory and experiment may be expressed in absolute terms ($Å^2$/eV) or as a percentage of the experimental value.

Peak Cross Section. The simplest comparison between theoretical and experimental cross sections is simply the peak value. Table 2 lists the peak cross sections from experimental measurements and from BEB calculations using both all-electron (AE) and ECP approaches. The effect of changing the Mulliken population threshold is addressed below; the

Table 2. Experimental and Calculated Peak Total Ionization Cross Sections in Å^2 (1 Å^2 = 10^{-20} m^2)[c]

molecule	expt	BEB-AE	BEB-ECP
C_2Cl_6	26.61 ± 1.06[7]	25.9	24.7
C_2HCl_5	23.61 ± 0.94[7]	22.7	21.6
C_2Cl_4	21.74 ± 0.87[7]	18.0	17.8
1,1,1,2-$C_2H_2Cl_4$	21.24 ± 0.85[7]	19.1	18.4
1,1,2,2-$C_2H_2Cl_4$	19.66 ± 0.79[7]	19.6	18.8
CCl_4	15.45 ± 0.62[7]	16.4	15.6
	15.15 ± 0.76[21]		
	14.9 ± 1.5[22]		
$TiCl_4$	16.45 ± 2.47[23]	19.9	17.5
CBr_4	19.0 ± 0.76[6]	20.6	19.7
$CHBr_3$	13.75 ± 0.55[6]	15.4	14.9
CH_2Br_2	11.67 ± 0.47[6]	11.5	11.4
P_2	9.0 ± 1.4[24]	8.5	7.6
P_4	21.0 ± 3.4[24]	17.0	15.1
As_2	13.2 ± 2.1[24]	9.7	8.6
As_4	27.2 ± 4.3[24]	19.4	17.1
GaCl	9.25 ± 0.93[25]	8.6	8.3
CS_2	9.03 ± 0.54[27]	9.7	9.4
	11.70 ± 0.47[28]		
	8.85 ± 1.33[26]		
H_2S	3.93 ± 0.51[30]	4.9	4.6
	5.53 ± 0.33[27]		
	6.28[29]		
CH_3I	10.3 ± 0.3[5]	8.9	8.7
	9.64 ± 0.58[31]		
$Al(CH_3)_3$	13 ± 1[32] [a]	14.0[a]	14.4[a]
$Ga(CH_3)_3$	12 ± 1[32] [a]	14.1[a]	14.9[a]
TMS_2O	26.41 ± 3.96[33] [b]	28.0[b]	28.6[b]
$Zn(C_2H_5)_2$	7.9 ± 1.4[34]	17.1	17.6

[a] At 70 eV, not necessarily the peak. [b] At 100 eV, not necessarily the peak. [c] A Mulliken population threshold of 75% was used in the AE calculations.

AE values in Table 2 were computed using a threshold of 75%, which we denote AE75. The differences between the calculated and experimental peak cross sections are plotted in Figure 1. Anomalously large discrepancies are apparent for $Zn(C_2H_5)_2$ and for As_4. These anomalies and questions about these experiments, described below, lead us to exclude from primary consideration the five molecules [P_2, P_4, As_2, As_4, $Zn(C_2H_5)_2$] measured in these two experiments.

From Table 2, the average reported experimental uncertainty is 1.2 Å^2. Excluding the five molecules mentioned above, the average reported experimental uncertainty is 0.9 Å^2, and the mean deviations from experimental values are 0.1 and −0.3 Å^2 for the AE75 and ECP approaches, respectively. The standard deviations of the discrepancies are 1.6 Å^2 for both AE75 and ECP methods. On average, the peak cross sections from AE75 calculations are 0.5 Å^2 larger than from ECP calculations (standard deviation = 0.6 Å^2).

The discrepancies in Table 2 may also be considered on a fractional basis. The mean reported experimental uncertainty is then 8%. When the same five molecules are excluded as above, the mean experimental uncertainty is 7%, and the mean deviations from the experimental values are 1% and −2% for the AE75 and ECP methods, respectively, both with standard deviation of 13%, all as percentages of the experimental values. On average, the peak cross sections from AE75 calculations are 4% larger than from ECP calculations (standard deviation = 4%). As expected, the fractional discrepancy between the AE and ECP predictions is greatest at low electron energy, where the cross section has a small value. For example, the AE75 prediction exceeds the ECP prediction for C_2Cl_6 by 31% at 15 eV, 5% at 70 eV, and 1% at 200 eV.

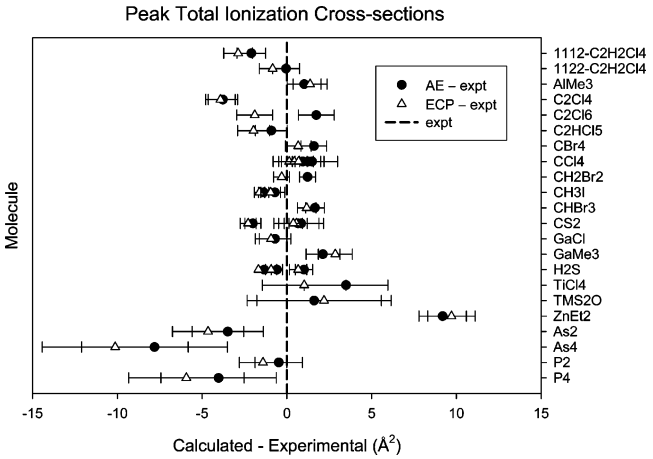


Figure 1. Differences between theoretical peak cross sections (both all-electron and effective-core-potential calculations) and experimental peak cross sections for 22 molecules. A Mulliken population threshold of 75% was used in the all-electron calculations. Error bars indicate the reported experimental uncertainties. A difference of zero corresponds to perfect agreement between theory and experiment.

Cross Section at Energy of Maximum Difference. The absolute difference between the AE and ECP predictions is greatest somewhat below the peak in the cross section. For example, the AE75 prediction exceeds the ECP prediction for C_2Cl_6 by 1.0 Å^2 at 15 eV, 2.1 Å^2 at 40 eV, 1.2 Å^2 at 70 eV, and 0.2 Å^2 at 200 eV. This suggests that the AE and ECP predictions are best compared not at the peak in the cross section, as in Table 2, but at some lower energy, where their contrast is greatest. Table 3 lists cross sections at energies where the difference between the AE75 and ECP predictions is greatest. Experimental values are included when available. Excluding the same five molecules as before, the mean deviations from experimental values are 0.8 and −0.1 Å^2 for the AE75 and ECP approaches, respectively. As fractions of the experimental values, the respective mean deviations are 10% and −1% for the AE75 and ECP methods.

Shape of the Cross-Section Curve. There is no standard procedure for comparing experimental and theoretical curve shapes. Here we use the minimized root-mean-square difference (rmsd) given by eq 2, where the integral is over the range of the experimental data and the scaling parameter s is chosen to minimize rmsd for each data set individually.

$$\mathrm{rmsd} = [(T_{\max} - T_{\min})^{-1} \int_{T_{\min}}^{T_{\max}} (s\sigma_{\mathrm{BEB}} - \sigma_{\mathrm{expt}})^2 \mathrm{d}T]^{1/2} \quad (2)$$

The values of rmsd (and s) for each experimental data set are listed in Table 4. Excluding the same five molecules as before, the average values of rmsd are 0.8 and 0.6 Å^2 for

Table 3. Experimental and Calculated Total Ionization Cross Sections at Energies Where AE and ECP Predictions Differ Most[a]

molecule	energy	expt	AE75	ECP
C_2Cl_6	28	15.3 ± 0.6[7]	17.7	15.4
C_2HCl_5	26	13.0 ± 0.5[7]	14.2	12.3
C_2Cl_4	24	11.6 ± 0.5[7]	9.9	9.1
1,1,1,2-$C_2H_2Cl_4$	26	12.4 ± 0.5[7]	11.8	10.3
1,1,2,2-$C_2H_2Cl_4$	26	11.6 ± 0.5[7]	12.4	10.8
CCl_4	28	9.6 ± 0.4[7]	10.9	9.5
		9.0 ± 0.4[21]		
		11.4 ± 1.1[22]		
$TiCl_4$	38	14.9 ± 2.2[23]	17.6	14.5
CBr_4	23.5		13.4	11.5
$CHBr_3$	23	6.5 ± 0.3[6]	9.9	8.6
CH_2Br_2	21.5	5.2 ± 0.2[6]	6.6	5.7
P_2	30	5.3 ± 0.8[24]	7.6	6.4
P_4	30	15.7 ± 2.5[24]	14.8	12.5
As_2	28	3.4 ± 0.5[24]	8.7	7.2
As_4	28	14.2 ± 2.3[24]	17.1	14.2
GaCl	23	7.9 ± 0.8[25]	5.3	4.6
CS_2	30	6.9 ± 0.4[27]	7.8	7.1
		6.4 ± 0.3[28]		
		8.7 ± 1.3[26]		
H_2S	30	3.3 ± 0.4[30]	4.0	3.5
		5.1 ± 0.3[27]		
		5.1[29]		
CH_3I	19.5	3.4 ± 0.1[31]	4.8	4.3
$Al(CH_3)_3$	34	8.9 ± 0.7[32]	11.2	11.6
$Ga(CH_3)_3$	38	9.7 ± 1.0[32]	11.2	12.1
TMS_2O	40	22.3 ± 3.3[33]	24.5	25.4
$Zn(C_2H_5)_2$	50	7.3 ± 1.3[34]	16.0	16.6

[a] Energies in eV, cross sections in Å^2. A Mulliken population threshold of 75% was used in the AE calculations.

the AE75 and ECP approaches, respectively. On average, the ECP value is 15% smaller than the AE75 value.

Initial Slope. Cross-section curves are not linear. The initial slope is defined here as the slope of the least-squares line through the experimental set of low-energy data points, i.e., from the threshold up to twice the threshold energy. These same low energies are used to compute theoretical cross sections, which are then fitted with a line to determine the corresponding theoretical initial slope. Results are collected in Table 5. Excluding the same five molecules as before, the average discrepancies with experimental values are -0.1 and -0.5 Å^2/eV for the AE75 and ECP methods, respectively. As fractions of the experimental values, the respective mean deviations are 2% and -6% for the AE75 and ECP methods. For the molecule GaCl, the experimental ionization threshold (6 eV) is much lower than the molecular ionization energy (10.07 eV).[36] The authors attributed this to ion pair formation, $Ga^+ + Cl^-$ (although the threshold is 1.4 eV too low),[25] which is not included in the BEB theory. Another possibility is unwanted, atomic Ga (IE = 5.9993 eV)[37] formed during the ion neutralization process. Thus, GaCl data points below 10 eV were omitted from our analysis.

Statistics for all the comparisons are collected in Table 6. For the AE calculations, results are tabulated for Mulliken population thresholds of 25%, 50%, 75%, and 85% as relevant to the choice of n in eq 1. The upper part of Table 6 excludes the five questionable molecules. The lower part of Table 6 includes all molecules. Differences between theory and experiment are listed in both absolute and fractional (percentage) terms. Since a small mean difference can conceal large discrepancies, root-mean-square (rms) differences are also listed. For all entries in Table 6, a value of zero corresponds to ideal agreement between theory and experiment.

Table 4. Comparison of the Shapes of the Experimental and Calculated Cross-Section Curves[a]

	rmsd (Å^2)		fitting parameter s	
molecule	AE75	ECP	AE75	ECP
C_2Cl_6[7]	2.06	1.04	0.97	1.09
C_2HCl_5[7]	1.50	0.65	1.01	1.11
C_2Cl_4[7]	1.03	0.30	1.10	1.23
1,1,1,2-$C_2H_2Cl_4$[7]	0.97	0.38	1.07	1.17
1,1,2,2-$C_2H_2Cl_4$[7]	1.00	0.39	0.97	1.06
CCl_4[7,21,22]	0.51	0.21	0.88	0.98
	0.99	0.55	0.97	1.05
	0.79	1.22	0.87	1.01
$TiCl_4$[23]	0.37	0.66	0.83	0.90
$CHBr_3$[6]	1.19	0.68	0.85	0.93
CH_2Br_2[6]	0.75	0.31	0.93	1.02
P_2[24]	1.85	1.51	1.15	1.29
P_4[24]	3.64	3.06	1.45	1.61
As_2[24]	4.19	3.72	1.26	1.42
As_4[24]	5.41	4.23	1.55	1.72
GaCl[25]	0.85	1.12	1.00	1.07
CS_2[26−28]	0.80,	0.60	0.95	1.03
	0.44	0.20	0.92	1.00
	1.73	1.22	1.13	1.28
H_2S[27,29,30]	0.16,	0.17	1.18	1.23
	0.15	0.12	1.27	1.37
	0.73	0.68	1.04	1.08
CH_3I[31]	0.53	0.28	1.10	1.16
$Al(CH_3)_3$[32]	0.47	0.51	0.84	0.81
$Ga(CH_3)_3$[32]	0.80	0.66	0.86	0.81
TMS_2O[33]	1.64	1.57	0.91	0.89
$Zn(C_2H_5)_2$[34]	0.23	0.25	0.45	0.44

[a] A better match yields a smaller value of rmsd from eq 2. A Mulliken population threshold of 75% was used in the AE calculations.

Comments on Experimental Measurements for P_2, P_4, As_2, and As_4. The experimental results for As_2, As_4, P_4, and P_2 were reported in the same paper.[24] All were measured using a Knudsen cell. Different precursors were required for the tetramers and the dimers. The number densities of the neutral gases were inferred from the rate of mass flow and the equations describing molecular flow. A multiplicative instrumental correction factor was derived from room-temperature measurements of N_2 and Ar. We suggest that the number densities estimated by this procedure were somewhat too low.

Comments on Experimental Measurements for $Zn(C_2H_5)_2$. In the $Zn(C_2H_5)_2$ experiment,[34] it was assumed that the Ar:$Zn(C_2H_5)_2$ pressure ratio in the gas reservoir was retained in the mass spectrometer. This neglects the expected mass discrimination in the leak valve, which depends on the molecular masses as $(M_{Zn(C_2H_5)_2}/M_{Ar})^{1/2} \approx 1.76$.[38] Applying this correction raises the peak experimental cross section to 13.9 ± 2.5 Å^2, much closer to the theoretical values.

Table 5. Experimental and Calculated Initial Slopes (Å²/eV) of Total Ionization Cross-Section Curves, as Defined in the Text[b]

molecule	expt	AE75	ECP
C_2Cl_6	0.95[7]	0.91	0.85
C_2HCl_5	0.86[7]	0.78	0.74
C_2Cl_4	0.81[7]	0.61	0.60
1,1,1,2-$C_2H_2Cl_4$	0.86[7]	0.77	0.71
1,1,2,2-$C_2H_2Cl_4$	0.77[7]	0.68	0.65
CCl_4	0.50,[7] 0.57[21], 0.77[22]	0.76	0.65
$TiCl_4$	0.76[23]	1.05	0.81
$CHBr_3$	0.50[6]	0.78	0.70
CH_2Br_2	0.44[6]	0.58	0.53
P_2	0.37[24]	0.32	0.28
P_4	1.09[24]	0.59	0.52
As_2	0.24[24]	0.33	0.30
As_4	0.99[24]	0.62	0.56
GaCl[a]	0.63[25]	0.48	0.42
CS_2	0.48,[27] 0.34,[28] 0.61[26]	0.51	0.46
H_2S	0.17,[30] 0.20,[27] 0.32[29]	0.19	0.18
CH_3I	0.38[31]	0.40	0.39
$Al(CH_3)_3$	0.60[32]	0.61	0.64
$Ga(CH_3)_3$	0.66[32]	0.50	0.57
TMS_2O	1.36[33]	1.06	1.12
$Zn(C_2H_5)_2$	0.23[34]	0.69	0.71

[a] Points deleted below thermodynamic threshold. [b] A Mulliken population threshold of 75% was used in the AE calculations.

Mixing between the heavy and light gases in the reservoir may also have been incomplete, which would reduce the proportion of the zinc compound if gas were withdrawn from the upper part of the reservoir, further underestimating the ionization cross section.

Mass discrimination is often problematic in Fourier transform mass spectrometry (FTMS).[39] Most seriously, FTMS suffers from a low-mass cutoff; it fails to detect ions with cyclotron frequencies too high (that is, masses too low) for the analog-to-digital converter. Although remedies have been developed,[40−42] they have not become popular. No ions lighter than C_2H^+ were reported in either of the FTMS studies cited here.[32,34] This suggests the possibility that the total ion yields were underestimated.

Reaction rate constants provide another way to determine pressure inside an FTMS cell. The authors made the reasonable assumption that charge transfer with Ar^+ occurs at the collision rate. The second-order collision rate constant can be computed from ion−molecule collision theory and combined with the pseudo-first-order rate constant for charge transfer to deduce the pressure of $Zn(C_2H_5)_2$. For this purpose, we calculate the isotropic polarizability of $Zn(C_2H_5)_2$ to be $\alpha = 11.38$ Å^3 (and dipole moment $\mu_D = 5.2 \times 10^{-31}$ Cm = 0.16 D) at the MP2/6-311++G(3d2f,3p2d) level, using a B3LYP/6-31G(d) geometry. Analogous calculations for Zn atom, C_2H_6, and CH_4 yield $\alpha = 5.64$, 4.20, and 2.43 Å^3, respectively, which are lower than the corresponding experimental values,[43] 5.75, 4.47, and 2.593 Å^3, by approximately 0.14 Å^3 per non-hydrogen atom. After applying this correction, our predicted polarizability for $Zn(C_2H_5)_2$ is 12.1 Å^3, which corresponds to a collision rate with Ar^+ of 1.5×10^{-9} cm^3 molecule^{-1} s^{-1}, according to the average dipole orientation[44] (ADO) theory. Unfortunately, since the pseudo-first-order rate constant was not reported, we cannot infer the pressure of $Zn(C_2H_5)_2$ in the FTMS experiment.

Table 6. Differences between Theory and Experiment: Summary Statistics and Effects of Varying the Mulliken Population Threshold[a]

comparison with expt	ECP	AE25	AE50	AE75	AE85
Five Molecules [P_2, P_4, As_2, As_4, $Zn(C_2H_5)_2$] Excluded					
peak: mean (Å²)	−0.3	1.2	1.1	*0.1*	−0.2
peak: mean %	−2%	8%	8%	*1%*	−3%
peak: rms (Å²)	**1.6**	1.9	1.9	*1.6*	1.8
peak: rms %	**12%**	15%	15%	*12%*	14%
max|AE-ECP|: AE mean (Å²)		1.7	1.6	0.8	*0.4*
max|AE-ECP|: ECP mean (Å²)		−0.2	−0.2	−0.1	**0.0**
max|AE-ECP|: AE mean %		14%	14%	10%	*6%*
max|AE-ECP|: ECP mean %		−2%	−2%	−1%	**1%**
max|AE-ECP|: AE rms (Å²)		2.3	2.3	*1.7*	1.8
max|AE-ECP|: ECP rms (Å²)		1.7	1.7	**1.7**	1.6
max|AE-ECP|: AE rms %		*22%*	*22%*	23%	23%
max|AE-ECP|: ECP rms %		**17%**	**17%**	20%	18%
shape [eq 2]: mean (Å²)	**0.61**	0.89	0.88	0.75	*0.72*
initial slope: mean (Å²/eV)	−0.05	0.05	0.05	*−0.01*	−0.03
initial slope: mean %	−6%	11%	11%	*2%*	−3%
initial slope: rms (Å²/eV)	**0.12**	0.16	0.16	*0.15*	0.16
initial slope: rms %	**19%**	29%	28%	*23%*	26%
All Molecules Included					
peak: mean (Å²)	−0.7	0.8	0.7	*−0.1*	−0.4
peak: mean %	−2%	8%	8%	2%	*−1%*
peak: rms (Å²)	3.3	3.0	3.0	*2.9*	*2.9*
peak: rms %	28%	27%	27%	*26%*	*26%*
max|AE-ECP|: AE mean (Å²)		2.0	1.9	1.3	*1.0*
max|AE-ECP|: ECP mean (Å²)		0.2	0.2	0.3	**0.4**
max|AE-ECP|: AE mean %		24%	24%	20%	*17%*
max|AE-ECP|: ECP mean %		8%	8%	8%	**10%**
max|AE-ECP|: AE rms (Å²)		2.8	2.8	*2.6*	2.7
max|AE-ECP|: ECP rms (Å²)		2.4	2.4	**2.5**	2.5
max|AE-ECP|: AE rms %		44%	44%	44%	44%
max|AE-ECP|: ECP rms %		**37%**	**37%**	**38%**	**37%**
shape [eq 2]: mean (Å²)	**0.97**	1.29	1.29	1.18	*1.16*
initial slope: mean (Å²/eV)	−0.06	0.03	*0.02*	*−0.02*	−0.04
initial slope: mean %	**−1%**	14%	14%	7%	*3%*
initial slope: rms (Å²/eV)	**0.20**	0.21	0.21	*0.20*	0.21
initial slope rms %	**46%**	48%	48%	*46%*	47%

[a] E.g., AE85 denotes AE calculations using a threshold of 85%. See the Discussion section for the meaning of the italic and bold fonts.

Discussion

Some published BEB calculations have incorrectly computed the parameter n from eq 1. As discussed elsewhere,[45] this has led to inaccurately unfavorable characterizations of the reliability of BEB theory for molecules containing the heavier elements ($Z > 10$). The present calculations are compared with experimental measurements for CCl_4, C_2Cl_6, and CS_2 in Figures 2-4. These figures exemplify the performance of BEB theory for molecules containing 3p elements such as chlorine and sulfur.

Although the reported experimental uncertainties are small, the disparity among measurements suggests that the total experimental uncertainty, including bias, may exceed 20%.[1] For example, multiple experimental measurements are available for CCl_4 (three measurements), CS_2 (three), H_2S (three),

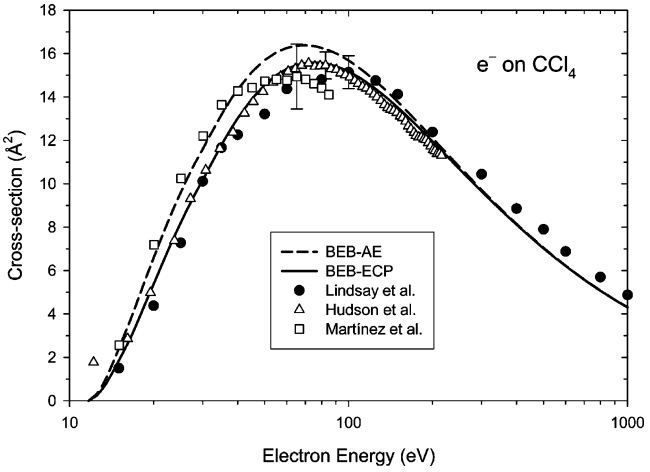


Figure 2. Total ionization cross section for CCl_4. Theoretical values are indicated by the solid (effective core-potential) and dashed (all-electron) curves. Experimental values are indicated by the symbols.[7,21,22]

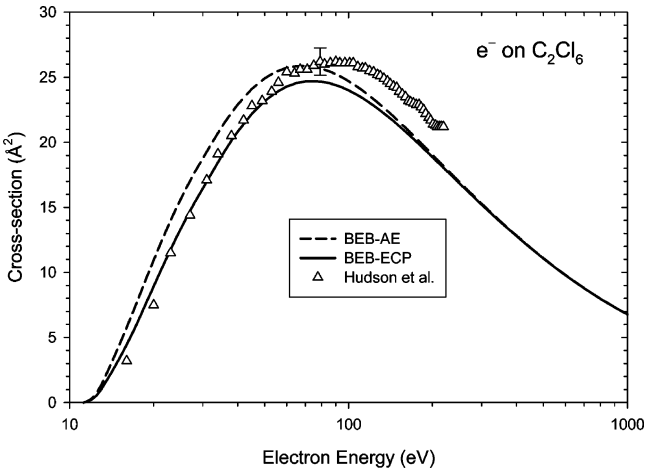


Figure 3. Total ionization cross section for C_2Cl_6. Theoretical values are indicated by the solid (effective core-potential) and dashed (all-electron) curves. Experimental values are indicated by the symbols.[7]

and CH_3I (two) (Table 2), with peak cross sections spanning 4%, 29%, 45%, and 7% of the respective mean values. Thus, the cross sections from both the all-electron (AE) and effective-core-potential (ECP) calculations agree with experimental measurements as well as different experimental measurements agree with each other. As judged from Tables 2−6 and the figures, differences between the AE the ECP predictions are generally minor.

Table 6 summarizes quantitatively the agreement with experiment that is achieved by both the AE and ECP implementations of BEB theory. We exclude the five problematic molecules from this discussion (top half of Table 6). The effects of including them, shown in the bottom half of Table 6, are minor.

Mulliken Population Threshold in AE Calculations. For each row in Table 6, values in italics indicate which value of the Mulliken population threshold (in the AE calculations) agrees best with the available experimental data. The first four rows refer to the peak value of the ionization cross section. All four agree best when the Mulliken population threshold is 75%.

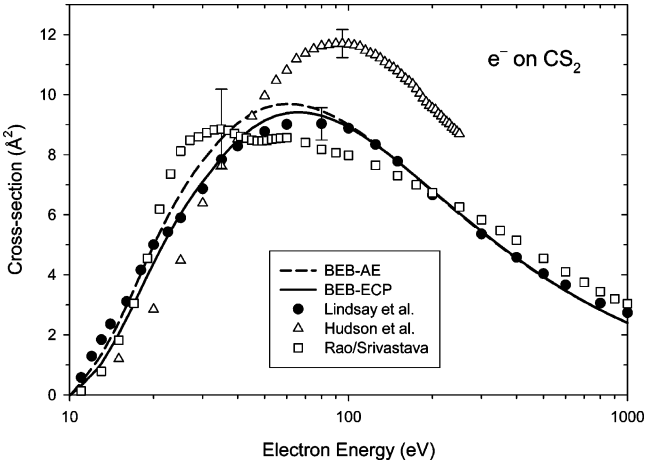


Figure 4. Total ionization cross section for CS_2. Theoretical values are indicated by the solid (effective core-potential) and dashed (all-electron) curves. Experimental values are indicated by the symbols.[26−28]

The next eight rows refer to the impact energy that maximizes the absolute difference between the AE and ECP predictions. This energy is generally different for each molecule. Considering only the mean discrepancies with experiment at these energies, the best Mulliken population threshold appears to be 85%. However, small mean values can mask large positive and negative discrepancies that happen to counterbalance each other. If absolute root-mean-square (rms) discrepancies with experimental values are considered instead, the best threshold appears to be 75%. Relative rms discrepancies weakly favor a lower threshold of 50% or 25%.

The non-negative shape comparison [eq 2] favors a Mulliken population threshold of 85%. The last four rows of the top half of Table 6, all dealing with the initial slope of the cross-section curve, all favor a threshold of 75%. Considering all rows of the top half of Table 6, the best choice of Mulliken population threshold appears to be 75%.

ECP vs AE Calculations. In Table 6, results of ECP calculations are shown in boldface when they agree with experiment at least as well as the best AE results. For peak cross sections, the AE mean value is slightly better than the ECP mean value. The ECP and AE peak values are equally good on an rms basis.

At the energies where the ECP and AE predictions differ most, the ECP mean value agrees with experiment better than the AE value does. The ECP results are slightly better than the AE results on an rms basis.

The shapes of the ECP curves agree better with experimental data than do the AE results, as measured using eq 2. On an averaged basis, the AE initial slopes agree with experiment better than the ECP slopes do. However, the order is reversed when the slopes are compared on an rms basis. Considering all the data of Table 6, the ECP predictions agree better with experimental measurement by some metrics and the AE predictions agree better by other metrics. Overall, the ECP predictions agree better with experiment, although the AE calculations agree nearly as well.

The ECP approach has operational advantages over the AE method. The ECP procedure is more clearly defined. In the AE procedure, one must use Mulliken populations to identify valence orbitals that are dominated by atomic orbitals with principal quantum number n greater than 2. The Mulliken population threshold for identifying such molecular orbitals, although recommended here as 75%, is rather arbitrary. There is further ambiguity when multiple heavy atoms are involved. For example, in the molecule MoO_2Cl_2 (not included in the Tables), orbital $10b_1$ has 38% Mo 4d character and 24% Cl 3p character; the appropriate value of n in eq 1 is debatable. In contrast, no Mulliken populations are needed in the ECP approach.

The ECP method is more easily extensible to molecules containing very heavy atoms, for which good all-electron basis sets are often unavailable. As a bonus, many ECPs also recover scalar relativistic effects, thus providing more accurate energies. We find that the results are independent of the choice of effective potential. For example, the Stuttgart,[15,16] NIST,[46] and Los Alamos[47] ECPs, with their corresponding valence basis sets, produce nearly superimposable cross-section curves for C_2Cl_4 (differences less than 0.1 Å^2).

The ECP procedure still requires performing all-electron calculations to obtain cross sections for the core orbitals. This is cumbersome for molecules containing very heavy atoms that lack good all-electron basis sets. Alternatively, binding and kinetic energies for the core orbitals can be obtained from atomic Dirac−Fock calculations,[9] independent of the particular molecule at hand. For the molecules examined here, completely ignoring the contribution of the core by using ECPs for all calculations reduces the cross section by only a small percentage, particularly at energies below 100 eV. Thus, the core can be neglected entirely if the energies of interest are not too high. This is the situation in applications such as mass spectrometry or low-energy plasmas.

Conclusions

Both the all-electron (AE) and effective-core-potential (ECP) implementations of the BEB theory produce total ionization cross sections that agree with experimental measurements about as well as different experiments agree with each other. The ECP predictions show slightly better agreement with experimental measurements, as evaluated using a variety of metrics. The ECP method is also more convenient, since it avoids Mulliken population thresholds, is applicable to larger molecules and heavier atoms, carries a lower computational cost, and includes scalar relativistic effects. We recommend the ECP approach for calculating BEB cross sections for molecules that contain elements heavier than neon.

Acknowledgment. We are grateful to Prof. Reinaldo Pis Diez (Universidad Nacional de La Plata) for providing and modifying the MullPop program and to Dr. Yong-Ki Kim (NIST) for general encouragement and for pointing out the importance of the shape of the cross-sections curves. We also thank the referees for their comments, especially for suggesting that the Mulliken population threshold be considered as an adjustable parameter and that the initial slope is an interesting quantity. G.E.S. is a NIST Summer Undergraduate Research Fellow, 2003.

Supporting Information Available: Molecular orbital data for the 22 molecules listed in the tables, as obtained by using the AE25, AE50, AE75, AE85, and ECP methods (110 ASCII files). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Irikura, K. K.; Ali, M. A.; Kim, Y.-K. *Int. J. Mass Spectrom.* **2003**, *222*, 189−200.

(2) Kim, Y.-K.; Rudd, M. E. *Phys. Rev. A* **1994**, *50*, 3954−3967.

(3) Hwang, W.; Kim, Y.-K.; Rudd, M. E. *J. Chem. Phys.* **1996**, *104*, 2956−2966.

(4) Kim, Y.-K.; Irikura, K. K.; Rudd, M. E.; Zucker, D. S.; Zucker, M. A.; Coursey, J. S.; Olsen, K. J.; Wiersma, G. G. *Electron-Impact Ionization Cross Section Database* (vers. 2.1); National Institute of Standards and Technology, Gaithersburg, MD: http://physics.nist.gov/ionxsec.

(5) Vallance, C.; Harris, S. A.; Hudson, J. E.; Harland, P. W. *J. Phys. B: At., Mol. Opt. Phys.* **1997**, *30*, 2465−2475.

(6) Bart, M.; Harland, P. W.; Hudson, J. E.; Vallance, C. *Phys. Chem. Chem. Phys.* **2001**, *3*, 800−806.

(7) Hudson, J. E.; Vallance, C.; Bart, M.; Harland, P. W. *J. Phys. B: At., Mol. Opt. Phys.* **2001**, *34*, 3025−3039.

(8) Kim, Y.-K.; Hwang, W.; Weinberger, N. M.; Ali, M. A.; Rudd, M. E. *J. Chem. Phys.* **1997**, *106*, 1026−1033.

(9) Huo, W. M.; Kim, Y.-K. *Chem. Phys. Lett.* **2000**, *319*, 576−586.

(10) Berry, R. S.; Rice, S. A.; Ross, J. *Physical Chemistry*; John Wiley & Sons: New York, 1980.

(11) Certain commercial materials and equipment are identified in this paper in order to specify procedures completely. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

(12) 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 (vers. B.05)*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(13) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(14) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(15) Andrae, D.; Häussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123−141.

(16) Bergner, A.; Dolg, M.; Küchle, W.; Stoll, H.; Preuss, H. *Mol. Phys.* **1993**, *80*, 1431−1441.

(17) von Niessen, W.; Schirmer, J.; Cederbaum, L. S. *Comput. Phys. Rep.* **1984**, *1*, 57−125.

(18) Zakrzewski, V. G.; Ortiz, J. V.; Nichols, J. A.; Heryadi, D.; Yeager, D. L.; Golab, J. T. *Int. J. Quantum Chem*. **1996**, *60*, 29−36.

(19) Pis Diez, R. *MullPop*; National University of La Plata: La Plata, 2003.

(20) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem*. **1993**, *14*, 1347−1363.

(21) Lindsay, B. G.; McDonald, K. F.; Yu, W. S.; Stebbings, R. F. *J. Chem. Phys.* **2004**, *121*, 1350−1356.

(22) Martínez, R.; Sierra, B.; Redondo, C.; Sánchez Rayo, M. N.; Castaño, F. *J. Chem. Phys.* **2004**, *121*, 11653−11660.

(23) Basner, R.; Schmidt, M.; Becker, K.; Tarnovsky, V.; Deutsch, H. *Thin Solid Films* **2000**, *374*, 291−297.

(24) Monnom, G.; Gaucherel, P.; Paparoditis, C. *J. Phys. (Paris)* **1984**, *45*, 77−84.

(25) Shul, R. J.; Freund, R. S.; Wetzel, R. C. *Phys. Rev. A* **1990**, *41*, 5856−5860.

(26) Rao, M. V. V. S.; Srivastava, S. K. *J. Geophys. Res.* **1991**, *96*, 17563−17567.

(27) Lindsay, B. G.; Rejoub, R.; Stebbings, R. F. *J. Chem. Phys.* **2003**, *118*, 5894−5900.

(28) Hudson, J. E.; Vallance, C.; Harland, P. W. *J. Phys. B: At., Mol. Opt. Phys.* **2004**, *37*, 445−455.

(29) Belić, D. S.; Kurepa, M. V. *Fizika* **1985**, *17*, 117−127.

(30) Rao, M. V. V. S.; Srivastava, S. K. *J. Geophys. Res., [Planets]* **1993**, *98*, 13137−13145.

(31) Rejoub, R.; Lindsay, B. G.; Stebbings, R. F. *J. Chem. Phys.* **2002**, *117*, 6450−6454.

(32) Jiao, C. Q.; DeJoseph, C. A., Jr.; Haaland, P.; Garscadden, A. *Int. J. Mass Spectrom.* **2000**, *202*, 345−349.

(33) Basner, R.; Foest, R.; Schmidt, M.; Becker, K.; Deutsch, H. *Int. J. Mass Spectrom.* **1998**, *176*, 245−252.

(34) Jiao, C. Q.; DeJoseph, C. A., Jr.; Garscadden, A. *Int. J. Mass Spectrom.* **2004**, *235*, 83−89.

(35) Beran, J. A.; Kevan, L. *J. Phys. Chem.* **1969**, *73*, 3866−3876.

(36) Grabandt, O.; Mooyman, R.; de Lange, C. A. *Chem. Phys.* **1990**, *143*, 227−238.

(37) Martin, W. C.; Musgrove, A.; Kotochigova, S.; Sansonetti, J. E. *Ground Levels and Ionization Energies for the Neutral Atoms*; Sept. 2003; National Institute of Standards and Technology: http://physics.nist.gov/PhysRefData/IonEnergy/ionEnergy.html.

(38) Atkins, P. W. *Physical Chemistry*; W. H. Freeman and Co.: San Francisco, 1978.

(39) Gordon, E. F.; Muddiman, D. C. *J. Mass Spectrom.* **2001**, *36*, 195−203.

(40) Cody, R. B.; Kinsinger, J. A.; Goodman, S. D. *Anal. Chem.* **1987**, *59*, 2567−2569.

(41) Verdun, F. R.; Ricca, T. L.; Marshall, A. G. *Appl. Spectrosc.* **1988**, *42*, 199−203.

(42) Schweikhard, L.; Alber, G. M.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 177−181.

(43) *CRC Handbook of Chemistry and Physics*, 84th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 2004.

(44) Su, T.; Bowers, M. T. In *Gas-Phase Ion Chemistry*; Bowers, M. T., Ed.; Academic: New York, 1979; Vol. 1; pp 83−118.

(45) Scott, G. E.; Irikura, K. K. *Surf. Interface Anal.* **2005**, in press.

(46) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612−630.

(47) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284−298.

(48) Kimura, K.; Katsumata, S.; Achiba, Y.; Yamazaki, T.; Iwata, S. *Handbook of HeI Photoelectron Spectra of Fundamental Organic Molecules*; Japan Scientific Societies: Tokyo, 1981.

(49) Griffiths, W. J.; Harris, F. M.; Andrews, S. R.; Parry, D. E. *Int. J. Mass Spectrom. Ion Proc.* **1992**, *112*, 45−61.

(50) Grant, R. P.; Harris, F. M.; Andrews, S. R.; Parry, D. E. *Int. J. Mass Spectrom. Ion Proc.* **1995**, *142*, 117−124.

(51) Bancroft, G. M.; Pellach, E.; Tse, J. S. *Inorg. Chem.* **1982**, *21*, 2950−2955.

(52) von Niessen, W.; Åsbrink, L.; Bieri, G. *J. Electron Spectrosc. Relat. Phenom.* **1982**, *26*, 173−201.

(53) Grant, R. P.; Andrews, S. R.; Parry, D. E.; Harris, F. M. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 382−388.

(54) Bulgin, D. K.; Dyke, J. M.; Morris, A. *J. Chem. Soc., Faraday Trans. 2* **1976**, *72*, 2225−2232.

(55) Wang, L.-S.; Niu, B.; Lee, Y. T.; Shirley, D. A.; Ghelichkhani, E.; Grant, E. R. *J. Chem. Phys.* **1990**, *93*, 6327−6333.

(56) Dyke, J. M.; Elbel, S.; Morris, A.; Stevens, J. C. H. *J. Chem. Soc., Faraday Trans. 2* **1986**, *82*, 637−645.

(57) Mazumdar, S.; Marathe, V. R.; Kumar, S. V. K.; Mathur, D. *Int. J. Mass Spectrom. Ion Proc.* **1988**, *86*, 351−355.

(58) Cesar, A.; Ågren, H.; Naves de Brito, A.; Svensson, S.; Karlsson, L.; Keane, M. P.; Wannberg, B.; Baltzer, P.; Fournier, P. G.; Fournier, J. *J. Chem. Phys.* **1990**, *93*, 918−931.

(59) Vass, G.; Tarczay, G.; Magyarfalvi, G.; Bödi, A.; Szepes, L. *Organometallics* **2002**, *21*, 2751−2757.

(60) Barker, G. K.; Lappert, M. F.; Pedley, J. B.; Sharp, G. J.; Westwood, N. P. C. *J. Chem. Soc. Dalton Trans*. **1975**, 1765−1771.

(61) Ibuki, T.; Hiraya, A.; Shobatake, K.; Matsumi, Y.; Kawasaki, M. *Chem. Phys. Lett.* **1989**, *160*, 152−156.

(62) Starzewski, K. A. O.; Dieck, H. T.; Bock, H. *J. Organomet. Chem.* **1974**, *65*, 311−325.

(63) Creber, D. K.; Bancroft, G. M. *Inorg. Chem.* **1980**, *19*, 643−648.

CT050077J

JCTC Journal of Chemical Theory and Computation

Ab Initio Study of Spin-Vibronic Dynamics in the Ground $\tilde{X}^2E$ and Excited $\tilde{A}^2A_1$ Electronic States of $CH_3S^\bullet$

Aleksandr V. Marenich and James E. Boggs*

Institute for Theoretical Chemistry, Department of Chemistry and Biochemistry, The University of Texas at Austin, 1 University Station A5300, Austin, Texas 78712

Received June 6, 2005

Abstract: A spin-vibronic Hamiltonian including the linear, quadratic, cubic, and quartic Jahn−Teller terms with account for all important anharmonic effects was applied to study electronic and nuclear dynamics in the ground $\tilde{X}^2E$ and first excited $\tilde{A}^2A_1$ electronic states of the CH_3S methylthio radical (C_{3v}). The $E\otimes(3a_1+3e)$ problem of spin-vibronic eigenvalues and eigenfunctions was solved in a basis set of products of electronic, electron spin, and vibrational functions. The Jahn−Teller distortions in $\tilde{X}^2E$ CH_3S are totally quenched by the strong spin−orbit coupling. However, Jahn−Teller interaction terms in the spin-vibronic Hamiltonian cannot be neglected for the high precision evaluation of energy levels of CH_3S. The results of calculations show the importance of inclusion of at least quadratic vibronic terms into variational treatment. The nonadiabatic (pseudo-Jahn−Teller) coupling of the $\tilde{X}^2E$ and $\tilde{A}^2A_1$ electronic states was found small and safely removable from the spin-vibronic Hamiltonian of CH_3S.

Introduction

This paper is intended to extend our previous study[1] of the Jahn−Teller (JT) effect and spin−orbit coupling (SOC) in the ground electronic state $\tilde{X}^2E$ of the methylthio (CH_3S) radical by means of a variational solution of the E⊗e problem using ab initio methods to parametrize the Hamiltonian. The formalism of our calculations is described in detail in refs 2 and 3. The ground $\tilde{X}^2E$ electronic state of CH_3S was studied here with the simultaneous treatment of spin−orbit coupling, all linear and quadratic JT interactions including multimode couplings, the cubic and most important quartic vibronic terms, and all significant anharmonic effects. Eigenenergies and eigenfunctions of a model spin-vibronic Hamiltonian were calculated in a basis set of products of electronic, electron spin, and vibrational functions transformed according to irreducible representations ($E_{3/2}$ and $E_{1/2}$) of the double C_{3v} symmetry group.[4,5]

A study of the nuclear dynamics in the first excited $\tilde{A}^2A_1$ state of CH_3S within the adiabatic approximation[4] was also performed. The smallness of nonadiabatic coupling between the ground $\tilde{X}^2E$ and excited $\tilde{A}^2A_1$ electronic states was proven with calculation of linear vibronic constants responsible for the pseudo-Jahn−Teller (PJT) effect[6] in CH_3S. The equilibrium geometry and force constants, vibrational energy levels in $\tilde{A}^2A_1$ CH_3S, and energies (T_e and T_o) of the $\tilde{A}^2A_1 \leftrightarrow \tilde{X}^2E$ electronic transition were found.

Details of Computations

Calculations of the Spin-Vibronic Dynamics in $\tilde{X}^2E$ CH_3S. The model Hamiltonian describing the dynamics of electrons and nuclei in $\tilde{X}^2E$ CH_3S (refs 2 and 3) is obtained with use of refs 7 and 8 in terms of the spin-vibronic matrix that is formed in the basis set of four spin-electronic functions $X = |\Lambda\rangle\ |\Sigma\rangle$: $|-1\rangle|-^1/_2\rangle$, $|+1\rangle|-^1/_2\rangle$, $|-1\rangle|+^1/_2\rangle$, and $|+1\rangle|+^1/_2\rangle$, where the label $\Lambda = \pm 1$ is intended to distinguish the two components of the E-term, lowest (−1) and upper (+1), and the value of $\Sigma = \pm^1/_2$ is a projection of the electron spin ($S = {}^1/_2$) on the symmetry axis of the molecule

$$\begin{pmatrix} (\hat{T} + V_E^{\circ} \mp \Delta_{SO}) - E & V_{-+} \\ V_{+-} & (\hat{T} + V_E^{\circ} \pm \Delta_{SO}) - E \end{pmatrix} \begin{pmatrix} X_{\Lambda=-1\Sigma=\mp(1/2)} \\ X_{\Lambda=+1\Sigma=\mp(1/2)} \end{pmatrix} = 0 \quad (1)$$

where $V_E^{\circ} = \langle \pm 1|\hat{V}(r,Q)|\pm 1\rangle$ and $2\Delta_{SO}$ is the value of spin−orbit splitting ($-A_{SO}\zeta_e$) at the electronic degeneracy point

* Corresponding author e-mail: james.boggs@mail.utexas.edu.

10.1021/ct0501452 CCC: $30.25
Published on Web 09/28/2005

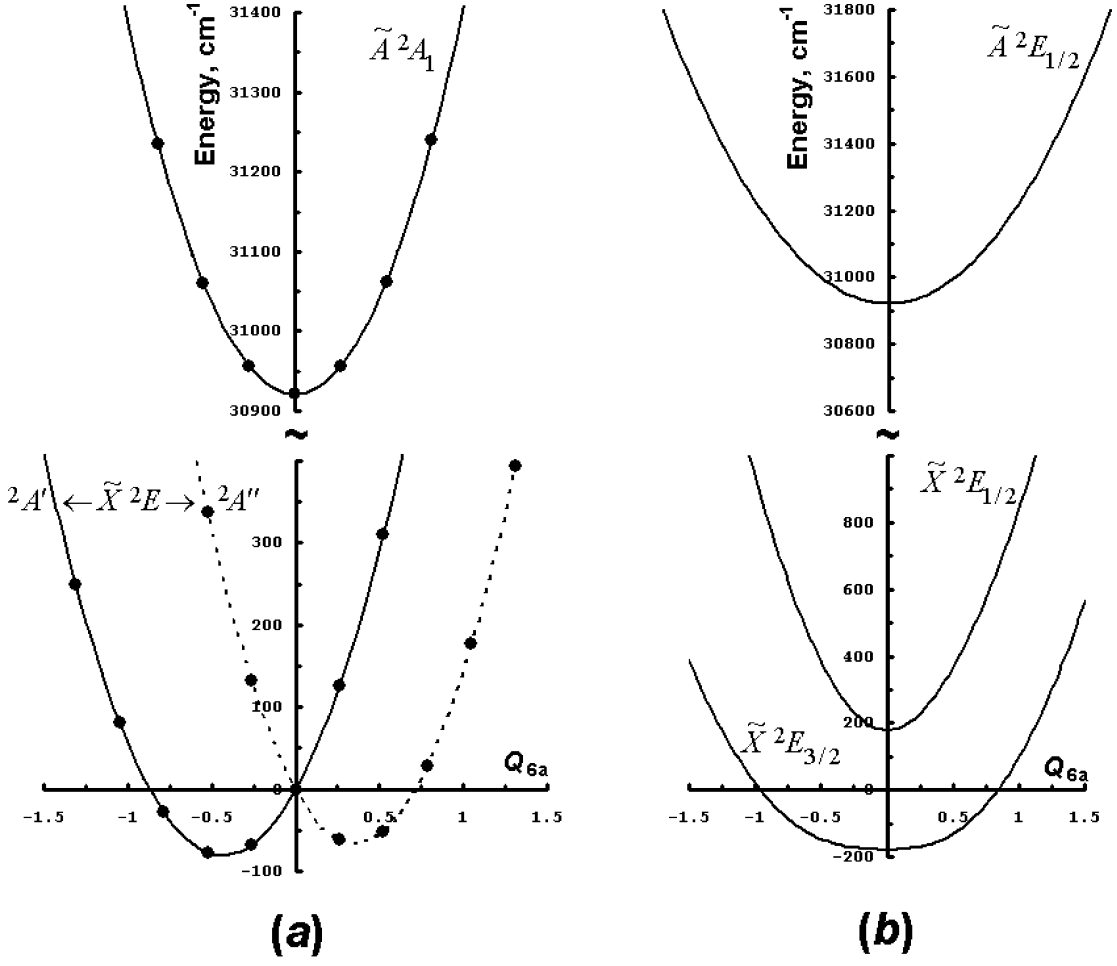


Figure 1. Vertical energies for $\tilde{X}^2E$ and $\tilde{A}^2A_1$ CH_3S without spin−orbit coupling (a) and with SOC (b) projected on the normal mode Q_6(E). Ab initio single point energies calculated by EOMIP/aug-cc-pCVTZ are depicted with solid circles. The a component of Q_6 is dimensionless. Other normal coordinates **Q** (**S** = **LQ**) were equal to zero. The **L**-matrix was calculated at the C_{3v} symmetry nuclear configuration optimized for the $\tilde{X}^2E$ state.

(Q_0), r and Q are electronic and nuclear coordinates, respectively, $\hat{V}(r,Q)$ is the electrostatic potential involving electrons and nuclei, $\hat{T}(Q)$ is the nuclear kinetic energy operator. In case of $\Delta_{SO} = 0$, diagonalization of eq 1 at different nuclear configurations gives two adiabatic potential surfaces (A′ and A″) crossing at Q_0 (Figure 1, part a). With inclusion of the spin−orbit coupling ($\Delta_{SO} \neq 0$), the two surfaces corresponding to the two spin−orbit states of $E_{3/2}$ (lower) and $E_{1/2}$ (upper) symmetry avoid crossing (Figure 1, part b).

There are six normal coordinates (Q) determined for the C_{3v} symmetry nuclear configuration of CH_3S: symmetric CH_3-stretching $Q_1(A_1)$, symmetric CH_3-deformation $Q_2(A_1)$ ("umbrella"), CS-stretching $Q_3(A_1)$, asymmetric CH_3-stretching Q_4(E), and asymmetric HCH− ("scissors") and HCS-deformations, Q_5(E) and Q_6(E), respectively. The electronic matrix (V) elements in eq 1 are expressed via the complex $Q_\pm = Q_a \pm (-1)^{1/2}Q_b$ components of doubly degenerate normal modes Q(E) as

$$V_E^{\,o} = \sum_{i=1}^{3}\frac{1}{2}\omega_i Q_i^{\,2} + \sum_{i=4}^{6}\frac{1}{2}\omega_i Q_{i+}Q_{i-} + \Delta V_{anh} \qquad (2)$$

and

$$V_{\pm\mp} = \sum_{i=4}^{6} k_i Q_{i\mp} + \sum_{i=4}^{6}\frac{1}{2} g_{ii} Q_{i\pm}^{\,2} + \sum_{i=4}^{6}\sum_{j=i+1}^{6} g_{ij} Q_{i\pm} Q_{j\pm} + \sum_{i=1}^{3}\sum_{j=4}^{6} b_{ij} Q_i Q_{j\mp} + V_{\pm\mp}(\mathrm{III}) + V_{\pm\mp}(\mathrm{IV}) \quad (3)$$

The quantities ω_i in eq 2 refer to harmonic frequencies calculated for the three nondegenerate and three 2-fold degenerate vibrational modes. The potential term ΔV_{anh} is responsible for anharmonic effects.

The vibronic term $V_{\pm\mp}$ in eq 3 includes Jahn−Teller parameters of up to the fourth order. The functions $V_{\pm\mp}$(III) and $V_{\pm\mp}$(IV) worked out in detail[3] refer to cubic and quartic vibronic couplings, respectively. The linear (k_i), quadratic (g_{ij}), cubic (g_{ijk}), and quartic (g_{ijkl}) vibronic constants (see also eqs 6−8 in ref 3) mix electronic states $\Lambda = \pm 1$ via single JT active vibrations or via their combinations. Other quadratic, cubic, and quartic constants of type "b" are responsible for the vibronic coupling between nondegenerate ($i = 1-3$) and JT active ($j = 4-6$) vibrational modes in

CH_3S. All the constants of *i*, *ij*, *ijk*, and *iijj* type were considered here.

The term ΔV_{anh} in eq 2 was expressed via the symmetrized Taylor series:

$$\Delta V_{anh} = \sum_{i=1}^{3}\frac{1}{3!}f_{iii}Q_i^3 + \sum_{i=1}^{3}\sum_{j=i+1}^{3}\frac{1}{2!}f_{iij}Q_i^2Q_j + \sum_{i=1}^{3}\sum_{j=i+1}^{3}\frac{1}{2!}f_{ijj}Q_iQ_j^2 + \sum_{i=1}^{3}\sum_{j=i+1}^{3}\sum_{k=j+1}^{3}f_{ijk}Q_iQ_jQ_k + \sum_{i=4}^{6}\frac{1}{3!2}f_{iii}(Q_{i+}^3 + Q_{i-}^3) + \sum_{i=4}^{6}\sum_{j=i+1}^{6}\frac{1}{2!2}f_{iij}(Q_{i+}^2Q_{j+} + Q_{i-}^2Q_{j-}) + \sum_{i=4}^{6}\sum_{j=i+1}^{6}\frac{1}{2!2}f_{ijj}(Q_{i+}Q_{j+}^2 + Q_{i-}Q_{j-}^2) + \sum_{i=4}^{6}\sum_{j=i+1}^{6}\sum_{k=j+1}^{6}\frac{1}{2}f_{ijk}(Q_{i+}Q_{j+}Q_{k+} + Q_{i-}Q_{j-}Q_{k-}) + \sum_{i=1}^{3}\sum_{j=4}^{6}\frac{1}{2!}f_{ijj}Q_iQ_{j+}Q_{j-} + \sum_{i=1}^{3}\sum_{j=4}^{6}\sum_{k=j+1}^{6}\frac{1}{2}f_{ijk}Q_i(Q_{j+}Q_{k-} + Q_{j-}Q_{k+}) + \sum_{i=1}^{3}\frac{1}{4!}f_{iiii}Q_i^4 + \sum_{i=1}^{3}\sum_{j=i+1}^{3}\frac{1}{2!2!}f_{iijj}Q_i^2Q_j^2 + \sum_{i=4}^{6}\frac{1}{4!}f_{iiii}Q_{i+}^2Q_{i-}^2 + \sum_{i=4}^{6}\sum_{j=i+1}^{6}\frac{1}{2!2!}f_{iijj}Q_{i+}Q_{j+}Q_{i-}Q_{j-} + \sum_{i=1}^{3}\sum_{j=4}^{6}\frac{1}{2!2!}f_{iijj}Q_i^2Q_{j+}Q_{j-} \quad (4)$$

where the quantities f_{iii}, f_{ijj}, and f_{ijk} are cubic force constants (the integers *i* and *j* run over all vibrations). The coefficients f_{iiii} and f_{iijj} refer to quartic constants. The force constants of higher order are found to be small and can be omitted.

The general variational solution of eq 1 was obtained in a basis set of products of electronic, electron spin, and vibrational functions transformed according to irreducible representations ($E_{3/2}$ and $E_{1/2}$) of the double C_{3v} symmetry group[4,5] (relativistic symmetry). In the absence of SOC, the vibronic eigenfunctions obviously comply to irreducible representations (A_1, A_2, and E) of the C_{3v} group (nonrelativistic symmetry). The vibrational portion of the basis set consists of the orthonormalized harmonic oscillator eigenfunctions expressed via Hermite polynomials.[4]

Calculations of the Nuclear Dynamics in $\tilde{A}^2A_1$ CH_3S with Use of the Adiabatic Approximation. The nuclear dynamics in the first excited electronic state $\tilde{A}^2A_1$ was studied with the model vibrational Hamiltonian written in terms of the one-dimensional harmonic oscillator for symmetric normal modes $Q_1(A_1)$, $Q_2(A_1)$, and $Q_3(A_1)$ and the two-dimensional (isotropic) harmonic oscillator for asymmetric modes $Q_4(E)$, $Q_5(E)$, and $Q_6(E)$[4]

$$\hat{H} = \sum_{i=1}^{3}\frac{1}{2}\omega_i\left(-\frac{\partial^2}{\partial Q_i^2} + Q_i^2\right) + \sum_{i=4}^{6}\frac{1}{2}\omega_i\left(-\frac{\partial^2}{\partial Q_{i+}\partial Q_{i-}} + Q_{i+}Q_{i-}\right) + \Delta V_{anh} \quad (5)$$

where the potential term ΔV_{anh} expressed in eq 4 accounts for anharmonic effects in $\tilde{A}^2A_1$ CH_3S. The dimensionless normal coordinates were used in eqs 1−5.

Eigenenergies and eigenfunctions of eq 5 are calculated in a basis set of products of the one- or two-dimensional harmonic oscillator eigenfunctions[4] transformed according to irreducible representations (A_1, A_2, and E) of the C_{3v} point symmetry group.[4,5]

Miscellaneous Details on Calculation of Model Parameters. Optimizations of geometry parameters and harmonic analyses of the C_{3v} symmetry nuclear configurations in the $\tilde{X}^2E$ and $\tilde{A}^2A_1$ electronic states of CH_3S were performed numerically using the symmetry-adapted internal coordinates **S** related to the normal coordinates as $\mathbf{S} = \mathbf{L}\,\mathbf{Q}$ (see ref 1) and the program ANOCOR[9] for generation of distorted geometries and harmonic analysis. The unperturbed (diabatic) harmonic frequencies $\omega_4(E)$, $\omega_5(E)$, and $\omega_6(E)$ in $\tilde{X}^2E$ CH_3S were calculated by differentiation of the function $V_E^\circ(Q) = {}^1/_2(U_+ + U_-)$ averaged over total energies of the upper (U_+) and lower (U_-) adiabatic electronic states arising from the degenerate $\tilde{X}^2E$ term.

The vibrational and vibronic parameters from eqs 3 and 4 were found through the least-squares fit of model potential energy values to "exact" ab initio energies for the two components of $\tilde{X}^2E$ and for the $\tilde{A}^2A_1$ state over 1153 nuclear configurations distorted on the normal coordinates Q_i and Q_{ja} ($i = 1-3$, $j = 4-6$) (see details in refs 2 and 3). For the ab initio calculations of single point energies, we used the equation-of-motion coupled cluster method[10] with the coupled cluster singles and doubles (CCSD) reference wave function for the CH_3O^- -anion ground state (further abbreviated to EOMIP) and with the augmented correlation consistent polarized valence basis sets of Dunning et al.[11] of triple-ζ quality including extra functions[12] for C and S to account for core−core and core-valence correlation (further abbreviated to aug-cc-pCVTZ). The EOMIP calculations were performed with a local version of the ACES II program package.[13]

The value of spin−orbit splitting ($-A_{SO}\xi_e = 358\ cm^{-1}$) at the electronic degeneracy point (C_{3v}) in $\tilde{X}^2E$ CH_3S has been found in our previous study[1] with the use of multiconfiguration quasi-degenerate second-order perturbation theory[14] followed by an SOC perturbative calculation within the full Breit-Pauli spin−orbit operator.[15] This spin−orbit calculation was performed with the GAMESS (US) package.[16]

Evaluation of the Nonadiabatic Coupling of the Ground $\tilde{X}^2E$ and First Excited $\tilde{A}^2A_1$ Electronic States. To study the role of the pseudo-Jahn−Teller interaction between the electronic states of interest, we constructed a diabatic electronic matrix relevant to the $(E+A_1)\otimes e$ problem to introduce the nonadiabatic coupling of three states by means of generalization of findings in refs 17−19

$$V = \begin{pmatrix} V_E^{\,o} \mp \Delta_{SO} & V_{-+}^{\,JT} & \frac{1}{\sqrt{2}}V_{+-}^{\,PJT} \\ V_{+-}^{\,JT} & V_E^{\,o} \pm \Delta_{SO} & -\frac{1}{\sqrt{2}}V_{-+}^{\,PJT} \\ \frac{1}{\sqrt{2}}V_{-+}^{\,PJT} & -\frac{1}{\sqrt{2}}V_{+-}^{\,PJT} & V_{A_1}^{\,o} \end{pmatrix} \quad (6)$$

where the diabatic potentials $V_E^\circ(Q)$ and $V_{A_1}^\circ(Q)$ are

expressed with use of eqs 2 and 4 in terms of normal coordinates Q_i ($i = 1-3$) and $Q_{i\pm} = Q_{ia} \pm (-1)^{1/2}Q_{ib}$ ($i = 4-6$) corresponding to the ground $\tilde{X}^2E$ state of CH_3S. The value of $T_v = V_{A_1}^{\circ}(0) - V_E^{\circ}(0)$ is the vertical energy of the $\tilde{A}^2A_1 \leftrightarrow \tilde{X}^2E$ transition. The Jahn−Teller and pseudo-Jahn−Teller coupling constants $V_{\pm\mp}(Q)$ are expressed with use of eq 3. As in eq 1, the electronic term V_E° in eq 6 may undergo spin−orbit splitting of $2\Delta_{SO}$. Diagonalization of eq 6 at $\Delta_{SO} = 0$ and $Q_b = 0$ leads to the three adiabatic potentials $U_{E(A')}$, $U_{E(A)}$, and U_{A_1} representing ab initio total energies (Figure 1, part a):

$$U_{E(A'')} = V_E^{\circ} + A_{JT} \tag{7}$$

$$U_{E(A')} = \frac{1}{2}(V_{A_1}^{\circ} + V_E^{\circ} - A_{JT} - \sqrt{4A_{PJT}^2 + (V_{A_1}^{\circ} - V_E^{\circ} + A_{JT})^2}) \tag{8}$$

$$U_{A_1} = \frac{1}{2}(V_{A_1}^{\circ} + V_E^{\circ} - A_{JT} + \sqrt{4A_{PJT}^2 + (V_{A_1}^{\circ} - V_E^{\circ} + A_{JT})^2}) \tag{9}$$

The functions of A_{JT} and A_{PJT} are defined as $V_{+-}^{(P)JT}(Q_a) = V_{-+}^{(P)JT}(Q_a)$. Taking into account only linear vibronic (JT and PJT) coupling, one can readily represent the expressions for these potentials given by Woywod et al.[17] in the study of the Jahn−Teller and pseudo-Jahn−Teller interactions in the ammonia cation (cf. eqs 2.17 in ref 17 following the phase convention for the choice of a sign on λ).

In the case of well-separated adiabatic states with $4A_{PJT}^2 \ll (V_{A_1}^{\circ} - V_E^{\circ} + A_{JT})^2$, eqs 8 and 9 can be reduced to

$$U_{E(A')} = V_E^{\circ} - A_{JT} - \frac{A_{PJT}^2}{T_v} \tag{10}$$

$$U_{A_1} = V_{A_1}^{\circ} + \frac{A_{PJT}^2}{T_v} \tag{11}$$

with the further simplification $V_{A_1}^{\circ} - V_E^{\circ} + A_{JT} \approx T_v = V_{A_1}^{\circ}(0) - V_E^{\circ}(0)$. Thus, even linear pseudo-Jahn−Teller effect ($A_{PJT} \sim \lambda Q_a$) giving a quadratic contribution to the adiabatic potentials $U_{E(A')}$ and U_{A_1} may cause an additional instability of the high-symmetry nuclear configuration with reduction of the curvature of the low sheet.[20] According to eqs 10 and 11, a pseudo-Jahn−Teller coupling may be heavily quenched with a large value of the vertical energy T_v (a case of well-separated states). If $(A_{PJT}^2/T_v) \approx 0$, the expressions for the adiabatic potentials (eqs 10 and 11) are reduced to the formulas for uncoupled electronic states ($E = A' + A''$ and A_1) which can be treated individually by use of eqs 1 and 5.

The linear pseudo-Jahn−Teller constants λ_4, λ_5, and λ_6 which couple these electronic states via the terms $V_{\pm\mp}^{PJT} = \sum_{i=4}^{6}\lambda_i Q_{i\mp}$ in eq 6 were evaluated numerically by means of EOMIP/aug-cc-pCVTZ single point calculations with use of eqs 7−9 and $A_{PJT} = \sum_{i=4}^{6}\lambda_i Q_{ia}$. A variational solution of the spin-vibronic Hamiltonian with inclusion of the PJT coupling of the $\tilde{X}^2E$ and $\tilde{A}^2A_1$ states was beyond the scope of the present study.

Results of Calculations and Discussion

The Ground $\tilde{X}^2E$ Electronic State of CH_3S. As predicted by the Jahn−Teller theorem,[6] the C_{3v} symmetry nuclear configuration in the $\tilde{X}^2E$ ground electronic state of CH_3S is unstable as shown in Figure 1(a) (no spin−orbit coupling included). There are two specific points on the adiabatic potential energy surface (PES) of $\tilde{X}^2E$ CH_3S corresponding to distorted (C_s symmetry) nuclear configurations (a minimum and a saddle point) with electron wave functions of A′ or A″ symmetries. The Jahn−Teller stabilization energy $E_{JT} = 92$ cm^{-1} is determined as $E_{JT} = U(C_{3v}, E) - U(C_s, A')$ where $U(C_{3v}, E)$ and $U(C_s, A')$ are total energies corresponding to the $C_{3v}(E)$ and $C_s(A')$ equilibrium geometries, respectively. The barrier to pseudorotation or the quadratic Jahn−Teller stabilization energy is $\Delta_{JT} = U(C_s, A'') - U(C_s, A')$, $\Delta_{JT} = 15$ cm^{-1}.

After account for the spin−orbit coupling, the shape of adiabatic potential surfaces of $\tilde{X}^2E$ CH_3S is changed dramatically as shown in Figure 1(b). The value of spin−orbit splitting in $\tilde{X}^2E$ CH_3S ($-A_{SO}\zeta_e = 358$ cm^{-1}) is significantly larger than the Jahn−Teller stabilization energy $E_{JT} = 92$ cm^{-1}. As shown by our previous calculations,[1] the strong spin−orbit coupling totally quenches the Jahn−Teller distortions in the ground electronic state of CH_3S (Figure 1, part b). Instead of the three equivalent minima corresponding to the Jahn−Teller distorted geometries on the lower potential energy surface of $\tilde{X}^2E$ CH_3S (the Mexican hat[6]) as depicted in Figure 2(a), there is a minimum at the C_{3v} symmetry nuclear configuration (Figure 2, part b). The reduction of the Jahn−Teller effect by spin−orbit coupling does *not* mean though that the vibronic effects may be completely neglected.[7] The shape of potential surfaces for the two final spin−orbit states is determined by the relation between the $A_{SO}\zeta_e$ quantity and the values of E_{JT} and Δ_{JT}. Hence, the thorough evaluation of vibronic parameters and force constants is still important for the most accurate theoretical predictions of spin-vibronic levels of the CH_3S radical.

The EOMIP/aug-cc-pCVTZ values of molecular parameters with inclusion of all significant force constants of higher order and linear, quadratic, cubic, and quartic vibronic constants of $\tilde{X}^2E$ CH_3S are presented in Table 1. The optimal geometry parameters (R_e, α_e), unperturbed harmonic frequencies (ω), and Jahn−Teller stabilization energies (E_{JT}, Δ_{JT}) of $\tilde{X}^2E$ CH_3S (Table 1) are close to results of our previous EOMIP calculations[1] with electronic basis sets of quadruple-ζ quality: $R_e(C-S) = 1.7940$ Å, $R_e(C-H) = 1.0872$ Å, $\alpha_e(HCS) = 109.80°$; $\omega_i = 3066(A_1)$, $1367(A_1)$, $758(A_1)$, $3164(E)$, $1483(E)$, and $938(E)$ cm^{-1} ($i = 1-6$); $E_{JT} = 93.2$ and $\Delta_{JT} = 14.9$ cm^{-1}.[1] Values of linear and quadratic Jahn−Teller constants calculated in ref 1 ($k_4 = -90$, $k_5 = -166$, $k_6 = -372$, $g_{44} = 0.25$, $g_{55} = -24$, and $g_{66} = 93$ cm^{-1}) are also similar to relevant numbers presented in Table 1. Hence, further extension of the basis set would be expected to improve our results only a little. The large magnitudes of force constants corresponding to the CH_3-stretchings $Q_1(A_1)$ and $Q_4(E)$ indicate that these vibrations exhibit the most pronounced anharmonicity. Calculations of diagonal vibronic constants with the use of multiconfiguration quasi-degenerate second-order perturbation theory result in the values ($k_4 =$

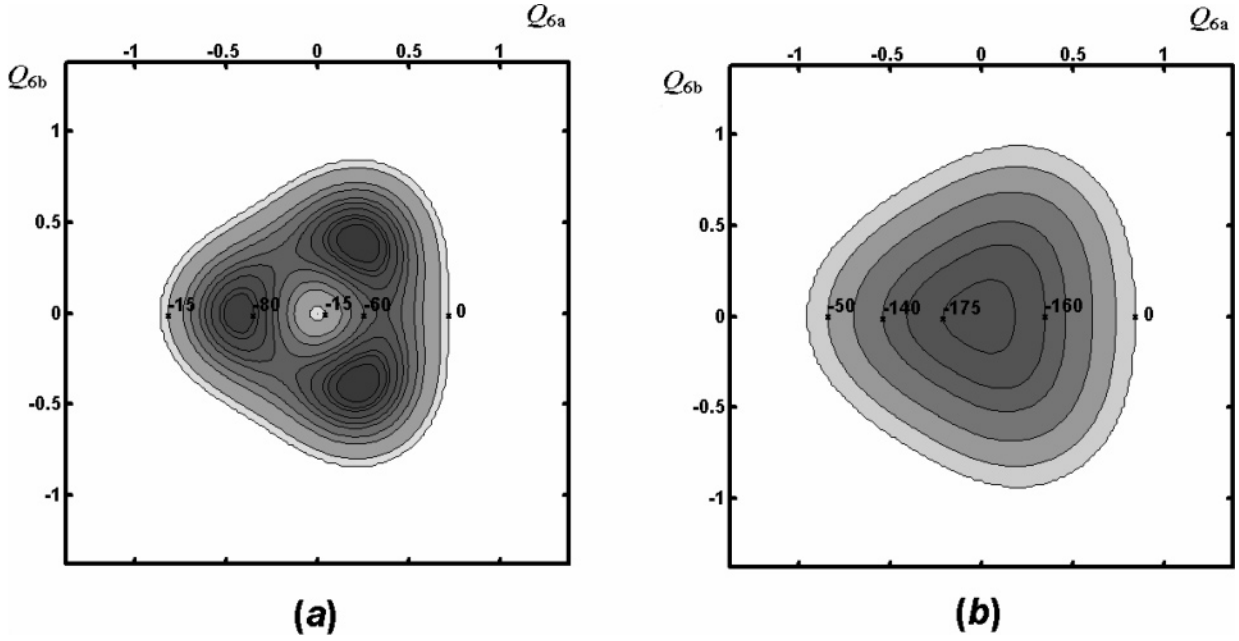


Figure 2. The map of the potential energy surface of $\tilde{X}^2E$ CH_3S (the low sheet) without spin−orbit coupling (a) and with SOC (b) for the normal mode Q_6(E). Energies are given in cm^{-1}. The a and b components of Q_6 are dimensionless.

Table 1. Structure and Spin-Vibronic Parameters of CH_3S in the Ground $\tilde{X}^2E$ and First Excited $\tilde{A}^2A_1$ Electronic States[a]

			$\tilde{X}^2E$ CH_3S		
$E_{JT} = 92$	$\Delta_{JT} = 15$	$-A_{SO}\zeta_e = 358$	R_e(C−S) = 1.7944	R_e(C−H) = 1.0880	α_e(HCS) = 109.79
$\omega_1(A_1) = 3062$	$\omega_2(A_1) = 1373$	$\omega_3(A_1) = 752$	$\omega_4(E) = 3157$	$\omega_5(E) = 1490$	$\omega_6(E) = 943$
$f_{111} = -1067$	$f_{222} = 136$	$f_{333} = -251$	$f_{444} = 823$	$f_{555} = 77$	$f_{666} = -32$
$f_{1111} = 330$	$f_{2222} = -20$	$f_{3333} = 73$	$f_{4444} = 578$	$f_{5555} = -25$	$f_{6666} = 35$
$k_4 = -90$	$k_5 = -168$	$k_6 = -369$	$g_{44} = -7$	$g_{55} = -28$	$g_{66} = 98$
$g_{444} = -1$	$g_{555} = -25$	$g_{666} = -25$	$g_{4444} = 2$	$g_{5555} = -6$	$g_{6666} = -0.02$
$g_{45} = -5$	$g_{46} = 34$	$g_{56} = -43$	$b_{14} = 3$	$b_{15} = 11$	$b_{16} = -25$
$b_{24} = 24$	$b_{25} = -38$	$b_{26} = 52$	$b_{34} = 8$	$b_{35} = 25$	$b_{36} = 64$
			$\tilde{A}^2A_1$ CH_3S		
$T_e = 27308$			R_e(C−S) = 2.0643	R_e(C−H) = 1.0794	α_e(HCS) = 99.75
$\omega_1(A_1) = 3124$	$\omega_2(A_1) = 1203$	$\omega_3(A_1) = 465$	$\omega_4(E) = 3299$	$\omega_5(E) = 1454$	$\omega_6(E) = 718$
$f_{111} = -1070$	$f_{222} = 217$	$f_{333} = -161$	$f_{444} = 833$	$f_{555} = 170$	$f_{666} = -23$
$f_{1111} = 312$	$f_{2222} = -18$	$f_{3333} = 66$	$f_{4444} = 545$	$f_{5555} = -14$	$f_{6666} = 33$

[a] Values and dimensions: distances R (Å), valence angles α (deg), all remaining quantities are in cm^{-1}. The values of geometric parameters (R and α), unperturbed harmonic frequencies (ω), force (f) and vibronic (k, g, b) constants were found numerically by means of EOMIP/aug-cc-pCVTZ total energy calculations over 1153 single points. The off-diagonal force (f) and vibronic (g and b) constants of *ijj*, *ijk*, and *iijj* type were small (with the exception of $f_{144} = -1131$, $f_{1144} = 358$ cm^{-1} for $\tilde{X}^2E$ and $f_{144} = -1137$, $f_{1144} = 335$ cm^{-1} for $\tilde{A}^2A_1$) and were omitted here (but still included in our dynamical calculations). The magnitude of $-A_{SO}\zeta_e$ shows the spin−orbit splitting of the $\tilde{X}^2E$ state.[1] The Jahn−Teller stabilization energy and barrier to pseudorotation are $E_{JT} = U(C_{3v}, E) - U(C_s, A')$ and $\Delta_{JT} = U(C_s, A'') - U(C_s, A')$, respectively. The value of T_e is an adiabatic energy of the $\tilde{A}^2A_1 \leftrightarrow \tilde{X}^2E_{3/2}$ transition.

-91, $k_5 = -154$, $k_6 = -385$, $g_{44} = -10$, $g_{55} = -29$, and $g_{66} = 102$ cm^{-1}) that are close to the ones obtained by EOMIP (Table 1). The values of Jahn−Teller stabilization energies E_{JT} and Δ_{JT} were calculated considering Jahn−Teller coupling of different orders. The inclusion of only linear constants reproduces the ab initio value of E_{JT} relatively poorly (82 cm^{-1}), whereas the addition of quadratic constants improves the quality of the model ($E_{JT} = 91$ and $\Delta_{JT} = 14$ cm^{-1}). Accounting for the cubic terms reproduces the ab initio values of $E_{JT} = 92$ and $\Delta_{JT} = 15$ cm^{-1} exactly, as shown in Table 1, and further addition of the quartic JT constants produces no further change.

To study the role of Jahn−Teller coupling of different orders, we performed a solution of the E⊗2e eigenvalue problem for $\tilde{X}^2E$ CH_3S (Table 2) with inclusion of two vibrational coordinates only: asymmetric HCH- and HCS-deformation normal modes, Q_5(E) and Q_6(E), respectively. The Q_5(E) and Q_6(E) modes, strongly coupled together, are almost solely responsible for the Jahn−Teller effect in $\tilde{X}^2E$ CH_3S, whereas the linear and quadratic vibronic couplings predicted for the asymmetric CH_3-stretching Q_4(E) are smaller than for Q_5(E) and Q_6(E) (compare k_4 and g_{44} with k_5 and g_{55} or with k_6 and g_{66} in Table 1). The spin-vibronic eigenstates were approximately assigned in terms of single normal oscillators. The eigenvalues of the model spin-vibronic Hamiltonian for $\tilde{X}^2E$ CH_3S calculated within the fourth-order vibronic coupling (eq 3, Table 2) are expected to be the most accurate energies in Table 2. The exclusion of all the quartic Jahn−Teller constants (the third-order vibronic coupling) does not lead to a noticeable loss of the

Table 2. Low Spin-Vibronic Energy (in cm^{-1}) Levels in $\tilde{X}^2E$ CH_3S Assigned to the Normal Modes Q_5 and Q_6 (E⊗2e Problem)[a]

assignment		first order	second order	third order	fourth order
0_0(e)	$e_{3/2}^-$	0	0	0	0
0_0(e*)	$e_{1/2}^+$	262	261	255	255
$6_1(a_1)$	$e_{1/2}^-$	848	818	804	804
$6_1(a_2)$	$e_{1/2}^+$	1066	1107	1088	1088
6_1(e)	$e_{1/2}^-$	1109	1102	1103	1104
6_1(e*)	$e_{3/2}^+$	1266	1257	1240	1241
$5_1(a_1)$	$e_{1/2}^-$	1466	1457	1451	1452
5_1(e)	$e_{1/2}^-$	1479	1487	1483	1484
6_2(e)	$e_{1/2}^-$	1730	1692	1670	1671
5_1(e*)	$e_{3/2}^+$	1704	1710	1688	1691
$5_1(a_2)$	$e_{1/2}^+$	1724	1713	1698	1699
6_2(e*)	$e_{3/2}^+$	1938	1861	1846	1846
6_2(e)	$e_{3/2}^-$	1986	2060	2045	2045
6_2(e*)	$e_{1/2}^+$	2071	2111	2082	2083
$6_2(a_1)$	$e_{1/2}^-$	2169	2146	2133	2136
$6_2(a_2)$	$e_{1/2}^+$	2250	2246	2226	2227

[a] Calculations were performed with inclusion of all linear (vibronic coupling of first order), all quadratic (second order), all cubic (third order), and quartic vibronic (fourth order) constants with the use of $-A_{SO}\zeta_e = 2\Delta_{SO} = 358$ cm^{-1} for eq 1. Occupation quantum numbers for the rest of the vibrational modes were equal to zero. The eigenstates were approximately correlated to vibronic states. The symmetry of spin-vibronic levels ($e_{3/2}$ or $e_{1/2}$) complies to the following rules: $E_{1/2}⊗e = e_{3/2} + e_{1/2}$ and $E_{1/2}⊗a_1 = E_{1/2}⊗a_2 = e_{1/2}$, where the symmetry of an electronic spin function ($S = 1/2$) is $E_{1/2}$ and the symmetries of vibronic levels are a_1, a_2, e (e* denotes another component of a doubly degenerate level). The sign −(+) indicates an assignment of a spin-vibronic level to the lower $\tilde{X}^2E_{3/2}$ (upper $\tilde{X}^2E_{1/2}$) spin−orbit state.

accuracy of eigenvalues at least for the lowest ones. The inaccuracy of eigenenergies calculated after exclusion of all the cubic and quartic vibronic terms from eq 3 varies in the range of 2−28 cm^{-1} for values of Table 2 (the second-order vibronic coupling). The error caused by omission of all the vibronic parameters except linear (the first-order coupling) is 5−92 cm^{-1}. Therefore, the quartic vibronic terms in the spin−orbit Hamiltonian for $\tilde{X}^2E$ CH_3S are quite small (Table 1) and can be neglected even for calculations of higher energy levels without a significant loss of accuracy. The inclusion of cubic vibronic constants is still desirable for a benchmark calculation of the spin-vibronic spectrum of CH_3S. On the other hand, the evaluation of thermochemical properties of CH_3S can be performed with eigenvalues of lower accuracy (the second-order or even first-order vibronic coupling, Table 2). This conclusion is in agreement with our previous result obtained for $\tilde{X}^2E$ CH_3O.[3] The spin−orbit splitting in $\tilde{X}^2E$ CH_3S ($-A_{SO}\zeta_e = 358$ cm^{-1}) is much larger than in $\tilde{X}^2E$ CH_3O (134 cm^{-1}).[2,3] The Jahn−Teller effect in $\tilde{X}^2E$ CH_3S ($E_{JT} = 92$ cm^{-1}) is rather weaker than in $\tilde{X}^2E$ CH_3O ($E_{JT} = 270$ cm^{-1}), and the Jahn−Teller distortions in $\tilde{X}^2E$ CH_3S are totally quenched by a strong spin−orbit coupling (unlike the case[2,3] of $\tilde{X}^2E$ CH_3O). Hence, the spin-vibronic dynamics in CH_3S can be described by means of the vibronic terms of lower order (linear and quadratic for instance), whereas the inclusion of cubic terms in the case of $\tilde{X}^2E$ CH_3O is more vital to a better accuracy.

The large magnitudes of some vibronic constants in Table 1 show the significant Jahn−Teller coupling between different vibrational modes: the constants of g_{ij} type are responsible for vibronic coupling between asymmetric vibrations Q_4(E), Q_5(E), and Q_6(E), whereas the constants b indicate a coupling between nondegenerate ($i = 1-3$) and JT active ($j = 4-6$) vibrations $Q_i(A_1)$ and Q_j(E).

We performed a variational solution of the full E⊗(3e + $3a_1$) eigenvalue problem for $\tilde{X}^2E$ CH_3S with account for all possible vibrational and vibronic couplings of up to the fourth order. Table 3 shows results of calculations of the low spin-vibronic energy levels assigned in terms of single normal oscillators. The calculated energy levels were correlated to one of the two spin−orbit states $^2E_{3/2}$ and $^2E_{1/2}$ arising from the electronic E-term split by SOC. Due to a strong coupling between different normal modes, all the assignments in Table 3 are approximate. For instance, two lowest spin-vibronic eigenstates at 0 and 252 cm^{-1} are definitely correlated to the two spin−orbit components of the lowest vibronic eigenstate. They are made up of the following combinations of basis functions

$$0.93|-\rangle|0_0\rangle + 0.27|+\rangle|6_1, -1\rangle \text{ at } 0\ (e_{3/2}^-) \text{ and}$$
$$0.84|+\rangle|0_0\rangle + 0.45|-\rangle|6_1, 1\rangle \text{ at } 252\ (e_{1/2}^+)\ cm^{-1}$$

where the notation $|-\rangle$ defines the lower spin−orbit state $^2E_{3/2}$ and $|+\rangle$ corresponds to the upper state $^2E_{1/2}$, a function $|6_v, l\rangle$ defines the two-dimensional (isotropic) harmonic oscillator with the principal vibrational quantum number v and the vibrational angular momentum quantum number l ($l = v, v-2, v-4, ...$) for a normal mode Q_6(E),[4] the indexes for other vibrational modes are omitted.

The higher energy spin-vibronic levels are more strongly mixed than the lower ones, and their assignments in terms of harmonic oscillators may be more difficult or meaningless. For instance, the spin-vibronic eigenstate at 1826 cm^{-1} (Table 3) is made up of the following combinations only slightly dominated by the basis function $|+\rangle|6_2, 2\rangle$:

$$0.54|+\rangle|6_2,2\rangle + 0.47|+\rangle|5_1,-1\rangle + 0.34|-\rangle|6_3,3\rangle$$
$$-0.29|-\rangle|6_2,0\rangle + 0.28|+\rangle|6_1,-1\rangle + 0.24|-\rangle|5_1,-1\rangle|6_1,1\rangle$$

Hence, the state at 1826 $(e_{3/2}^+)$ cm^{-1} can be approximately assigned to the upper spin−orbit component of the doubly degenerate vibronic level 6_2(e) arising from a two-quantum vibrational excitation of the normal oscillator corresponding to the asymmetric HCS-deformation Q_6(E).

The role of inclusion of the vibronic coupling between symmetric and Jahn−Teller active normal modes into the spin-vibronic Hamiltonian becomes clear after comparison of results of solutions of the E⊗2e eigenvalue problem (the right column of Table 2) to relevant eigenenergies from Table 3: the difference of these wavenumbers varies from 3 to 24 cm^{-1}. The spin−orbit splitting of the vibronic levels 1_v(e), 2_v(e), and 3_v(e) is a nearly constant value equal to the spin−orbit splitting $-A_{SO}\zeta_e d = 252$ cm^{-1} in the zero vibronic level 0_0(e) because these eigenstates are mostly dominated by a single harmonic oscillator basis function related to a vibrational excitation $1_v(A_1)$, $2_v(A_1)$, or $3_v(A_1)$, respectively, with only a little contribution from 4_v(E), 5_v(E), or 6_v(E). The strong spin−orbit coupling almost totally reduces the vibronic splitting for the eigenstates at 3022($e_{1/2}^-$), 3026($e_{1/2}^-$),

Table 3. Theoretical Spin-Vibronic Energy (in cm^{-1}) Levels in $\tilde{X}^2E$ CH_3S ($E\otimes(3a_1 + 3e)$ Problem) versus Experimental Values[a,b]

assignment	theory	experiment[c]	assignment	theory	experiment[c]
$0_0(e)$	0 ($e_{3/2}^-$)	0	$6_2(e^*)$	1826 ($e_{3/2}^+$)	
$0_0(e^*)$	252 ($e_{1/2}^+$)	256−259,[d] 264 [e]	$3_16_1(e)$	1844 ($e_{1/2}^-$)	
$3_1(e)$	742 ($e_{3/2}^-$)	738−745,[f] 736[e]	$3_16_1(e^*)$	1971 ($e_{3/2}^+$)	
$6_1(a_1)$	793 ($e_{1/2}^-$)	586[d]	$6_2(e)$	2037 ($e_{3/2}^-$)	
$3_1(e^*)$	993 ($e_{1/2}^+$)	980−986,[d] 998[e]	$6_2(e^*)$	2070 ($e_{1/2}^+$)	
$6_1(a_2)$	1078 ($e_{1/2}^+$)		$2_13_1(e)$	2095 ($e_{3/2}^-$)	2045−2057,[f] 2036[e]
$6_1(e)$	1105 ($e_{1/2}^-$)		$6_2(a_1)$	2112 ($e_{1/2}^-$)	
$6_1(e^*)$	1231 ($e_{3/2}^+$)		$2_16_1(a_1)$	2159 ($e_{1/2}^-$)	
$2_1(e)$	1358 ($e_{3/2}^-$)	1325−1332,[f] 1312[e]	$3_15_1(a_1)$	2176 ($e_{1/2}^-$)	
$5_1(a_1)$	1436 ($e_{1/2}^-$)		$3_15_1(e)$	2201 ($e_{1/2}^-$)	2204−2212[d]
$5_1(e)$	1463 ($e_{1/2}^-$)	1489−1501[d]	$6_2(a_2)$	2215 ($e_{1/2}^+$)	
$3_2(e)$	1487 ($e_{3/2}^-$)	1463−1464,[f] 1456[e]	$3_3(e)$	2229 ($e_{3/2}^-$)	2177−2179,[f] 2159[e]
$3_16_1(a_1)$	1540 ($e_{1/2}^-$)		$5_16_1(e)$	2230 ($e_{1/2}^-$)	
$2_1(e^*)$	1613 ($e_{1/2}^+$)	1583−1595,[f] 1559[e]	$5_16_1(e)$	2245 ($e_{3/2}^-$)	
$6_2(e)$	1659 ($e_{1/2}^-$)		$3_26_1(a_1)$	2286 ($e_{1/2}^-$)	
$5_1(e^*)$	1674 ($e_{3/2}^+$)		$2_13_1(e^*)$	2348 ($e_{1/2}^+$)	2310−2319,[f] 2283[e]
$5_1(a_2)$	1680 ($e_{1/2}^+$)		$3_16_2(e)$	2402 ($e_{1/2}^-$)	
$3_2(e^*)$	1734 ($e_{1/2}^+$)	1718−1723,[f] 1713[e]	$3_15_1(e^*)$	2410 ($e_{3/2}^+$)	
$3_16_1(a_2)$	1823 ($e_{1/2}^+$)		$3_15_1(a_2)$	2421 ($e_{1/2}^+$)	

[a] See footnotes to Table 2. [b] All energies are related to the ground level of the lower spin−orbit state $\tilde{X}^2E_{3/2}$. The intervals (in cm^{-1}) calculated for the CH_3-stretchings are 2938 for $1_1(e, e_{3/2}^-)$, 3022 for $4_1(a_1, e_{1/2}^-)$, 3026 for $4_1(e, e_{1/2}^-)$, 3194 for $1_1(e^*, e_{1/2}^+)$, 3275 for $4_1(a_2, e_{1/2}^+)$, and 3282 for $4_1(e^*, e_{3/2}^+)$. See comparison of these wavenumbers to observed intervals in the text. [c] The intervals from the $\tilde{A}^2A_1 \rightarrow \tilde{X}^2E_{3/2}$ and $\tilde{A}^2A_1 \rightarrow \tilde{X}^2E_{1/2}$ laser-induced fluorescence (LIF) spectra of jet-cooled CH_3S observed[21,22] for several pumped bands (0_0^0, 3_0^1, 3_0^2, or 2_0^1) are written here with a hyphen. Tentative assignments are italicized. [d] LIF by Chiang and Lee.[21] [e] LIF by Suzuki, Inoue, and Akimoto.[23] [f] LIF by Misra, Zhu, and Bryant.[22]

3275($e_{1/2}^+$), and 3282($e_{3/2}^+$) cm^{-1} approximately assigned to a single excitation $4_1(e)$ of the asymmetric CH_3-stretching normal mode $Q_4(E)$: a remaining vibronic splitting in the lower spin−orbit component ($E_{3/2}\otimes E = 2e_{1/2}$) and in the upper component ($E_{1/2}\otimes E = e_{3/2} + e_{1/2}$) is 4 and 7 cm^{-1}, respectively, where $E_{3/2}$ ($E_{1/2}$) are relativistic symmetry of a lower (upper) spin−orbit state (see Figure 1, part b).

Theory versus Experiment: the Ground $\tilde{X}^2E$ Electronic State. The CH_3S radical has been studied during the last three decades with use of different spectroscopic techniques (see refs 7 and 21−36 and therein).

In agreement with theory, experiment demonstrates a relatively large spin−orbit splitting of the ground electronic state, in excess of 200 cm^{-1}, and the Jahn−Teller distortions vanish.[7] The theoretical value of spin−orbit splitting in the ground vibronic state of $\tilde{X}^2E$ CH_3S ($-A_{SO}\zeta_e d = 252$ cm^{-1}, Table 3) is close to the observed ones (in cm^{-1}): 220.3[29] (microwave spectroscopy); 259.1,[21] 266,[22] 280 ± 20,[23] and 255.5[30] (laser-induced fluorescence); 280 ± 50[27] and 265 ± 15[34] (photoelectron spectroscopy). The Ham reduction factor d ($d \le 1$) in $-A_{SO}\zeta_e d$ indicates how much the spin−orbit coupling ($-A_{SO}\zeta_e$) is reduced by the Jahn−Teller effect.[7]

Chiang and Lee,[21] Misra et al.,[22] and Suzuki et al.[23] observed the CH_3S $\tilde{X}^2E - \tilde{A}^2A_1$ systems by laser-induced fluorescence (LIF) spectroscopy. The main ($\tilde{X}^2E_{3/2} \rightarrow \tilde{A}^2A_1$) and weak ($\tilde{X}^2E_{1/2} \rightarrow \tilde{A}^2A_1$) progressions detected in the excitation spectrum of CH_3S were assigned to the transitions from the $^2E_{3/2}$ and $^2E_{1/2}$ spin−orbit components of the ground electronic state $\tilde{X}^2E$ CH_3S to different CS-stretching vibrational levels (3^v) of the $^2A_1$ state. Less intense progressions assigned to the $\tilde{X}^2E_{3/2} \rightarrow 2_0^1 3_0^v \tilde{A}^2A_1$ transitions were also identified.[21]

Some intervals for $\tilde{X}^2E_{3/2}$ and $\tilde{X}^2E_{1/2}$ CH_3S observed in the dispersed fluorescence spectra[21−23] are shown in Table 3. The agreement between theory and experiment is good for all the intervals assigned to the symmetric vibrations $Q_2(A_1)$ and $Q_3(A_1)$: deviations vary between 4 and 80 cm^{-1} with the growth for higher energy overtones (Table 3). The interval observed[21] at 1496 ± 6 cm^{-1} (1463 cm^{-1} by theory) in the main progressions $3_0^v\tilde{A}^2A_1 \rightarrow \tilde{X}^2E_{3/2}$, $2_0^1\tilde{A}^2A_1 \rightarrow \tilde{X}^2E_{3/2}$ is assigned to the lower spin−orbit component of the vibronic level $5_1(e)$ (Table 3), but the transition to the upper spin−orbit component (1674 cm^{-1} by theory) was not detected in ref 21 apparently because of the low intensity of relevant lines in the LIF spectrum. Due to a relatively small Jahn−Teller coupling in CH_3S, those lines in the spectrum, which may be assigned to vibrations along the Jahn−Teller active modes, $Q_j(E)$, $j = 4-6$, are likely to be weak,[7] and they may not be detected or identified properly within the experimental accuracy. Some weak lines in the LIF spectrum of CH_3S were *tentatively* assigned by the authors of ref 21 to the symmetric $Q_1(A_1)$ and asymmetric $Q_4(E)$ CH_3-stretchings and the asymmetric HCS-deformation $Q_6(E)$, with $\nu_1(A_1) = 2776$, $\nu_4(E) = 2706$, and $\nu_6(E) = 586$ cm^{-1} being suggested.[21] In light of the present study and the results of our previous calculations[2,3] on $\tilde{X}^2E$ CH_3O (the methoxy radical CH_3O is an analogue of CH_3S but more extensively studied), some (or all) of these assignments may be incorrect. The value of $\nu_6(E) = 586$ cm^{-1} (ref 21) is significantly underestimated in comparison with the theoretical predictions for the unperturbed harmonic frequency $\omega_6(E) = 943$ cm^{-1} (Table 1) and the lowest 6_1 spin-vibronic level (at 793 cm^{-1}, Table 3). The intervals calculated for the CH_3-stretchings in $\tilde{X}^2E_{3/2}$ CH_3S (in cm^{-1}, see the symmetry notations in Table

2) are 2938 for $1_1(e, e_{3/2}^-)$, 3022 for $4_1(a_1, e_{1/2}^-)$, 3026 for $4_1(e, e_{1/2}^-)$, 3194 for $1_1(e^*, e_{1/2}^+)$, 3275 for $4_1(a_2, e_{1/2}^+)$, and 3282 for $4_1(e^*, e_{3/2}^+)$. They lie much higher than suggested in ref 21. Moreover, the value of $\nu_1(A_1) = 2776$ cm^{-1} (ref 21) is in disagreement with the interval observed in the CH_3S^- anion photoelectron spectrum[34] at 2960 ± 30 cm^{-1} and assigned to the lower spin−orbit component of $1_1(e)$ (cf. 2938 cm^{-1} by theory). The interval at 3225 ± 30 cm^{-1} in ref 34 was assigned to the upper spin−orbit component of $1_1(e^*)$ (3194 cm^{-1} by theory).

The First Excited $\tilde{A}^2A_1$ Electronic State of CH_3S. The EOMIP/aug-cc-pCVTZ values of geometric parameters, harmonic frequencies, and all significant force constants of higher order for $\tilde{A}^2A_1$ CH_3S (Figure 1) are presented in Table 1. The adiabatic energy of the $\tilde{A}^2A_1 \leftrightarrow \tilde{X}^2E_{3/2}$ electronic transition is $T_e = 27\,308$ cm^{-1}. The value of $T_o = 27\,117$ cm^{-1} calculated as a difference of the lowest level energies in the electronic states $\tilde{X}^2E_{3/2}$ and $\tilde{A}^2A_1$ CH_3S is ∼2% overestimated about a wavenumber of the $0_0^0\tilde{A}^2A_1 \rightarrow \tilde{X}^2E_{3/2}$ transition observed in the laser-induced excitation spectra of CH_3S: 26 526.3,[21] 26 529,[22] and 26 531[23] cm^{-1}. In respect to the double C_{3v} symmetry group, the $\tilde{A}^2A_1$ state is correlated to $\tilde{A}^2E_{1/2}$ (Figure 1, part b). There is no spin−orbit coupling between $\tilde{X}^2E$ and $\tilde{A}^2A_1$ CH_3S.

The equilibrium C_{3v} geometries of the ground $\tilde{X}^2E$ and first excited $\tilde{A}^2A_1$ electronic states differ significantly: the value of R_e(C−S) for $\tilde{A}^2A_1$ is larger by 0.27 Å and the value of α_e(HCS) is smaller by 10°. The reason for this is the excitation of an electron from the doubly occupied C−S σ molecular orbital (A_1) to a singly occupied MO of E symmetry interpreted as one of the 3pπ lone pairs localized on the sulfur atom.[1,32] In accordance with the increase of R_e(C−S) in $\tilde{A}^2A_1$ CH_3S, the normal vibration frequency correlated to the CS-stretching decreases: cf. $\omega_3(A_1) = 752$ cm^{-1} for $\tilde{X}^2E$ and $\omega_3(A_1) = 465$ cm^{-1} for $\tilde{A}^2A_1$. A vertical energy of the $\tilde{A}^2A_1 \leftrightarrow \tilde{A}^2E_{3/2}$ transition calculated at the C_{3v} nuclear configuration optimized for $\tilde{X}^2E$ is 31100 cm^{-1} (Figure 1, part b), much higher than $T_e = 27\,308$ cm^{-1}.

To study a possible role of nonadiabatic coupling of the ground $\tilde{X}^2E$ and first excited $\tilde{A}^2A_1$ electronic states of CH_3S, the linear pseudo-Jahn−Teller constants λ_4, λ_5, and λ_6 involved in the terms $V_{\pm\mp}^{PJT} = \sum_{i=4}^{6}\lambda_i Q_{i\mp}$ of the diabatic electronic matrix in eq 6 were calculated numerically by means of least-squares fitting of eqs 7−9 to ab initio values of total energies of three states ($E = A' + A''$ and A_1). Setting the vibrational parameters of ω- and f-type for $\tilde{X}^2E$ CH_3S fixed at values shown in Table 1, we varied all other magnitudes: vibrational parameters ω and f for the $\tilde{A}^2A_1$ state, linear PJT parameters λ, and all the JT constants k and g. The final values of λ_i are $|\lambda_i| \leq 10$ cm^{-1} that gives an extremely small (below the computational accuracy) contribution (A_{PJT}^2/T_v) ≤ 0.003 cm^{-1} to the adiabatic potentials $U_{E(A')}$ and U_{A_1} in eqs 10 and 11 with $T_v = 31\,100$ cm^{-1}. According to the results of CASSCF calculations with the state-averaged density matrix, the adiabatic electronic wave functions $\Psi_{\tilde{X}^2E}$ for the two components of the E-term are dominated by Slater determinants $\Phi_1[(a_1)^2(e_a)^2(e_b)^1(a_1)^0]$ and $\Phi_2[(a_1)^2(e_a)^1(e_b)^2(a_1)^0]$, whereas the $\tilde{A}^2A_1$ state is assigned in terms of a single electron excitation $\Phi_3[(a_1)^1(e_a)^2(e_b)^2(a_1)^0]$. Analysis of these wave functions calculated at the $C_s(A')$ equilibrium geometries corresponding to the Jahn−Teller minimum (Figure 1, part a) shows that the ground $\tilde{X}^2E$ and first excited $\tilde{A}^2A_1$ electronic states are coupled only very weakly by the vibrational distortions:

$$\Psi_{\tilde{X}^2E(A')} = 0.413\Phi_1 + 0.903\Phi_2 + 0.018\Phi_3$$

$$\Psi_{\tilde{X}^2E(A)} = 0.903\Phi_1 - 0.413\Phi_2$$

$$\Psi_{\tilde{A}^2A_1} = -0.007\Phi_1 - 0.016\Phi_2 + 0.993\Phi_3$$

Thus, the terms $V_{\pm\mp}^{PJT} = \sum_{i=4}^{6}\lambda_i Q_{i\mp}$ can be safely removed from the model Hamiltonian and the ground $\tilde{X}^2E$ and first excited $\tilde{A}^2A_1$ electronic states of CH_3S can be treated individually with use of eqs 1 and 5. Couplings of the ground $\tilde{X}^2E$ state to other (higher than $\tilde{A}^2A_1$) excited states are expected to be still smaller. For comparison, vertical energies of the lowest electronic states of the same multiplicity were calculated with the use of multiconfiguration quasi-degenerate second-order perturbation theory: 0 ($\tilde{X}^2E$), 30206 ($\tilde{A}^2A_1$), 46108 ($^2A_2$), and 54 378 (^2E) cm^{-1}.

As known from the literature,[32,33,35,37] the $\tilde{A}^2A_1$ adiabatic surface of CH_3S, like any radical of the methoxy family (CH_3X, CF_3X, X = O, S), correlates asymptotically to $CH_3(\tilde{X}^2A_2'') + S(^1D)$, being crossed by three repulsive surfaces, which belong to the CH_3S states ^4E, $^2A_2$, and $^4A_2$ and correlate to $CH_3(\tilde{X}^2A_2'') + S(^3P)$. According to an ab initio study of the photodissociation of CH_3S by Cui and Morokuma,[37] the energy of intersection between the bonding $\tilde{A}^2A_1$ and first repulsive $^4A_2$ states is only 0.13 eV (1049 cm^{-1}) above the $\tilde{A}^2A_1$ state minimum. The threshold energy for the nonadiabatic predissociative photofragmentation of CH_3S is expected to be lower than for the dissociation CH_3S ($\tilde{A}^2A_1$) → CH_3 ($\tilde{X}^2A_2''$) + S (^1D). The energy of the latter reaction is 0.73 eV.[37] For comparison, the stability of the ground state $\tilde{X}^2E$ in respect to the channel CH_3S ($\tilde{X}^2E$) → CH_3 ($\tilde{X}^2A_2''$) + S (^3P) is 2.93 eV.[37] Experimental evidence of predissociation in $\tilde{A}^2A_1$ CH_3S includes the sharp decrease of radiative lifetimes of the vibrational states for the 3_0^v(v = 1 and 2) transitions relative to 0_0^0 and the absence of the band signal for $v > 2$ in the excitation spectrum.[23] Transitions to the vibrational levels of $\tilde{A}^2A_1$ CH_3S lying above the predissociation threshold (≥800 cm^{-1})[21,23,33] were resolved in the CH_3S photofragment yield spectrum.[33]

The energies of low vibrational levels in $\tilde{A}^2A_1$ CH_3S calculated in the present study are compared to observed intervals in Table 4. Agreement between theory and experiment is somewhat worse in the case of $\tilde{A}^2A_1$ (Table 4) than for $\tilde{X}^2E$ CH_3S (Table 3). A more accurate dynamical calculation should account for the dissociative (predissociative) nature of the vibrational modes in $\tilde{A}^2A_1$ CH_3S. The remaining disagreement between theory and experiment (Tables 3 and 4) is likely to be caused by computational errors in our study (mainly omission of triple- and higher-order excitations in the EOMIP−CCSD method[10] and inaccuracy of the ab initio calculation of spin−orbit splitting in the ground state) and also by uncertainties in experimental measurements and errors in band assignments. With respect to the latter, experiments on CH_3S have not yet provided

Table 4. Theoretical and Observed Vibrational Energy Intervals (in cm^{-1}) in the First Excited Electronic State $\tilde{A}^2A_1$ CH_3S

assignment	theory	experiment[a]
$0^0(A_1)$	0	0
$3^1(A_1)$	452	401 ± 2,[b,c] 403 ± 1,[d] 402[e]
$6^1(E)$	689	635 ± 10[b]
$3^2(A_1)$	905	795,[b,c] 799,[d] 794[e]
$3^16^1(E)$	1134	
$2^1(A_1)$	1165	1098 ± 2[b,c,e]
$6^2(E)$	1342	
$6^2(A_1)$	1367	
$3^3(A_1)$	1369	1181,[b] 1180[c,e]
$5^1(E)$	1446	
$3^26^1(E)$	1585	
$2^13^1(A_1)$	1613	1490,[b] 1489[c]
$3^16^2(E)$	1787	
$3^16^2(A_1)$	1808	
$1^1(A_1)$	3001	2966[c]
$4^1(E)$	3165	

[a] The intervals from the $\tilde{X}^2E_{3/2} \rightarrow \tilde{A}^2A_1$ laser-induced excitation spectra of CH_3S were calculated from the wavelengths of excited bands (3_0^1, 3_0^2, 3_0^3, $2_0^13_0^1$, 6_0^1, and 1_0^1 in refs 21−23 and 36) relative to the origin 0_0^0. [b] Reference 21. [c] Reference 36. [d] Reference 23. [e] Reference 22.

complete and totally reliable data about the structure and transition frequencies of CH_3S, especially for the fundamentals $\nu_1(A_1)$, $\nu_4(E)$, $\nu_5(E)$, $\nu_6(E)$ and relevant overtones. Since the Jahn−Teller coupling in $\tilde{X}^2E$ CH_3S is small, the laser-induced excitation and dispersed fluorescence $\tilde{A}^2A_1 \leftrightarrow \tilde{X}^2E$ transitions (for instance, M_0^1 and M_1^0, respectively) involving excitations of the Jahn−Teller active vibrations ($M = 4-6$) may exhibit very weak intensities unresolved or undetected within the accuracy of measurements. The absence of spectral transitions correlated to the symmetric CH_3-stretching $Q_1(A_1)$ is likely due to unfavorable Franck−Condon factors.[35]

Conclusions

A variational ab initio study of the spin-vibronic dynamics in the ground $\tilde{X}^2E$ and first excited $\tilde{A}^2A_1$ electronic states of CH_3S has been performed for the first time with inclusion of all the important anharmonic effects and higher-order Jahn−Teller couplings and with account for the vibrational and vibronic interactions between different normal modes. The electronic problem has been solved with use of the equation-of-motion coupled cluster method and augmented basis sets of triple-ζ quality. A nonadiabatic coupling of the ground $\tilde{X}^2E$ and first excited $\tilde{A}^2A_1$ electronic states is shown to be small, which validates the adiabatic separation of these two states for studying the nuclear motion. However, unlike the case of $\tilde{A}^2A_1$, the nuclear dynamics in $\tilde{X}^2E$ CH_3S simultaneously affected by two components of the degenerate electronic state has been studied here beyond the adiabatic Born−Oppenheimer approximation that is broken down by the Jahn−Teller effect. The Jahn−Teller stabilization energy and barrier to pseudorotation are $E_{JT} = 92$ cm^{-1} and $\Delta_{JT} = 15$ cm^{-1}, respectively. The Jahn−Teller distortions of the CH_3S high symmetry nuclear configuration are totally quenched by a relatively strong spin−orbit coupling ($-A_{SO}\xi_e = 358$ cm^{-1}). Nevertheless, the remaining spin-vibronic coupling still makes a heavy impact on the spectroscopic properties of the CH_3S radical.

Despite the growth of computational capacities during the last two decades, there is a distinct lack of explicit ab initio calculations of vibronic or spin-vibronic spectra (see refs 7 and 8 for calculations on the methoxy radical $\tilde{X}^2E$ CH_3O and the methoxy family of radicals, see also references in refs 6 and 19 on other types of vibronic systems). The methylthio radical CH_3S is an important intermediate in environmental chemistry.[21] Hence, an accurate prediction of its molecular properties is critical for practical applications. The agreement between theory and available experimental data is considered as good, especially for the theoretical model using no empirical adjustments. Results of calculations within the complete set of vibrational modes predict spin-vibronic intervals never observed in experiment and significantly supplement the spectroscopic data presently available in the literature.[38] Theoretical predictions presented here can be used for interpretations of results of future CH_3S spectroscopy studies.

Acknowledgment. We are grateful to John Stanton for an opportunity to use his expanded version of the ACESII program and wish to thank Pablo Garcia Fernandez and Xinzheng Yang (Austin, Texas) for discussing the manuscript. Acknowledgments are made to the Donors of the ACS Petroleum Research Fund and to the Welch Foundation (Grant F-100) for support of this research. It is also a part of Project No. 2000-013-1-100 of the IUPAC entitled "Selected Free Radicals and Critical Intermediates: Thermodynamic Properties from Theory and Experiment" (inquiries to T. Berces, berces@chemres.hu).

References

(1) Marenich, A. V.; Boggs, J. E. *J. Phys. Chem. A* **2004**, *108*, 10594−10601.

(2) Marenich, A. V.; Boggs, J. E. *J. Chem. Phys.* **2005**, *122*, 024308/1−024308/11.

(3) Marenich, A. V.; Boggs, J. E. *Chem. Phys. Lett.* **2005**, *404*, 351−355.

(4) Bunker, P. R.; Jensen, P. *Molecular Symmetry and Spectroscopy*, 2nd ed.; NRC Research: Ottawa, Canada, 1998.

(5) Flurry, R. L., Jr. *Symmetry Groups*; Prentice-Hall: Englewood Cliffs, NJ, 1980.

(6) Bersuker, I. B. *Chem. Rev.* **2001**, *101*, 1067−1114.

(7) Barckholtz, T. A.; Miller, T. A. *Int. Rev. Phys. Chem.* **1998**, *17*, 435−524.

(8) Höper, U.; Botschwina, P.; Köppel, H. *J. Chem. Phys.* **2000**, *112*, 4132−4142.

(9) Sliznev, V. V. Private communication. Solomonik, V. G. Doctor in Chemistry Thesis, Moscow State University, 1993.

(10) Stanton, J. F.; Gauss, J. *J. Chem. Phys.* **1994**, *101*, 8938−8944.

(11) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023. Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806. Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358−1371. Basis sets were obtained from the Extensible Computational Chemistry

Environment Basis Set Database, Version 02/25/04, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352, U.S.A., and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multiprogram laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RLO 1830. Contact Karen Schuchardt for further information.

(12) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1995**, *103*, 4572−4585. Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **2002**, *117*, 10548−10560.

(13) Stanton, J. F.; Gauss, J.; Watts, J. D.; Lauderdale, W. J.; Bartlett, R. J. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1992**, *26*, 879−894.

(14) Nakano, H. *J. Chem. Phys.* **1993**, *99*, 7983−7992.

(15) Fedorov, D. G.; Koseki, S.; Schmidt, M. W.; Gordon, M. S. *Int. Rev. Phys. Chem.* **2003**, *22*, 551−592.

(16) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(17) Woywod, C.; Scharfe, S.; Krawczyk, R.; Domcke, W.; Köppel, H. *J. Chem. Phys.* **2003**, *118*, 5880−5893.

(18) Mahapatra, S.; Vallet, V.; Woywod, C.; Köppel, H.; Domcke, W. *Chem. Phys.* **2004**, *304*, 17−34.

(19) Köppel, H. *Adv. Ser. Phys. Chem.* **2004**, *15*, 429−472.

(20) Haller, E.; Köppel, H.; Cederbaum, L. S.; von Niessen, W.; Bieri, G. *J. Chem. Phys.* **1983**, *78*, 1359−1370.

(21) Chiang, S.-Y.; Lee, Y.-P. *J. Chem. Phys.* **1991**, *95*, 66−72.

(22) Misra, P.; Zhu, X.; Bryant, H. L., Jr. *Pure Appl. Opt.* **1995**, *4*, 587−598.

(23) Suzuki, M.; Inoue, G.; Akimoto, H. *J. Chem. Phys.* **1984**, *81*, 5405−5412.

(24) Callear, A. B.; Dickson, D. R. *Trans. Faraday Soc.* **1970**, *66*, 1987−1995.

(25) Ohbayashi, K.; Akimoto, H.; Tanaka, I. *Chem. Phys. Lett.* **1977**, *52*, 47−49.

(26) Engelking, P. C.; Ellison, G. B.; Lineberger, W. C. *J. Chem. Phys.* **1978**, *69*, 1826−1832.

(27) Janousek, B. K.; Brauman, J. I. *J. Chem. Phys.* **1980**, *72*, 694−700.

(28) Black, G.; Jusinski, L. E. *J. Chem. Soc., Faraday Trans. 2* **1986**, *82*, 2143−2151.

(29) Endo, Y.; Saito, S.; Hirota, E. *J. Chem. Phys.* **1986**, *85*, 1770−1777.

(30) Hsu, Y.-C.; Liu, X.; Miller, T. A. *J. Chem. Phys.* **1989**, *90*, 6852−6857.

(31) Anastasi, C.; Broomfield, M.; Nielsen, O. J.; Pagsberg, P. *Chem. Phys. Lett.* **1991**, *182*, 643−648.

(32) Hsu, C.-W.; Liao, C.-L.; Ma, Z.-X.; Tjossem, P. J. H.; Ng, C. Y. *J. Chem. Phys.* **1992**, *97*, 6283−6290.

(33) Bise, R. T.; Choi, H.; Pedersen, H. B.; Mordaunt, D. H.; Neumark, D. M. *J. Chem. Phys.* **1999**, *110*, 805−816.

(34) Schwartz, R. L.; Davico, G. E.; Lineberger, W. C. *J. Electron. Spectrosc. Relat. Phenom.* **2000**, *108*, 163−168.

(35) Pushkarsky, M. B.; Applegate, B. E.; Miller, T. A. *J. Chem. Phys.* **2000**, *113*, 9649−9657.

(36) Liu, C.-P.; Matsuda, Y.; Lee, Y.-P. *J. Chem. Phys.* **2003**, *119*, 12335−12341.

(37) Cui, Q.; Morokuma, K. *Chem. Phys. Lett.* **1996**, *263*, 54−62.

(38) Jacox, M. E. *J. Phys. Chem. Ref. Data* **2003**, *32*, 1−441.

CT0501452

JCTC Journal of Chemical Theory and Computation

Metal−Polyhydride Molecules Are Compact Inside a Fullerene Cage

Laura Gagliardi*

Dipartimento di Chimica Fisica "F. Accascina", Università di Palermo, Viale delle Scienze Parco d'Orleans, 90128 Palermo, Italy

Received July 29, 2005

Abstract: Quantum chemical calculations show that metal−hydride molecules are more compact when they are placed inside a fullerene cage than when they are isolated molecules. The metal−hydrogen bond distance in ZrH_4 becomes 0.15 Å shorter when it is placed inside a C_{60} cage. Metal−polyhydride molecules with a large number of H atoms such as ScH_{15} and ZrH_{16}, which are not bound as isolated molecules, are predicted to be bound inside a fullerene cage. It is also shown that two TiH_{16} clusters are bound inside a bicapped (9,0) carbon nanotube. Possible ways to make metal−hydrides inside C_{60} and nanotubes are suggested.

Introduction

Metal−hydrides are of considerable interest as potential hydrogen storage systems. A design target of 6.5 weight (wt)% has been regarded as adequate.[1] Species of MH_n type, with $n = 9$, as for example ReH_9^{2-}, are known experimentally.[2,3] We have recently predicted the existence of some MH_{12} molecules, where M is a group 6 atom, Cr, Mo, and W.[4] The isoelectronic ions VH_{12}^-, TiH_{12}^{2-}, and MnH_{12}^+ also turned out to be stable. In a hypothetical solid $Li(VH_{12})$, the hydrogen weight percentage would be 17. This study suggested that the number of hydrogen atoms, bound to a metal atom, could be increased from the previously known $n = 9$ to $n = 12$. The CrH_{12} species has six H_2 moieties, while MoH_{12} and WH_{12} have four H_2 moieties and four M−H σ bonds. The H−H distance in the H_2 moieties is larger than 0.8 Å in all species, while a calculation on isolated H_2 gives an H−H bond distance of 0.77 Å. These values for the H−H bond distances can be considered a fingerprint of the fact that the H_2 moieties are effectively bound to the central metal.

The possible existence of MH_n species with n greater than 12 has also been investigated. Examples are TiH_{14}, ZrH_{14}, ScH_{15}, and ZrH_{16}. Calculations indicate that such species have long metal−hydrogen bonds (greater than 2.0 Å). It should be noted that TiH_{14}, ZrH_{14}, and ScH_{15} fulfill the 18-electron rule, while ZrH_{16} would be a 20-electron species. ZrH_{16} turns out not to be stable and, instead, dissociates to ZrH_{12} and $2H_2$. Moreover, TiH_{14}, ZrH_{14}, ScH_{15}, and ZrH_{16} are not more stable than $TiH_4 + 5H_2$, $ZrH_4 + 5H_2$, $ScH_3 + 6H_2$, and $ZrH_4 + 6H_2$, respectively. It thus seems unlikely that MH_n species with n larger than 12 will exist as isolated molecules.

The following question was then posed: is there an alternative way to store hydrogen in its molecular form. A conceptually new route to stabilize hydrogen-rich metal−polyhydride clusters would consist in placing a metal−hydride inside a C_{60} fullerene cage. The inside sphere of a fullerene is large enough to enclose atoms and small molecules. A variety of endohedral complexes, having metal fragments, noble gases, and atoms inside the cage, have been synthesized and characterized.[5−10] Recently, encapsulation of molecular hydrogen into an open-cage fullerene having a 16-membered ring orifice has been investigated.[11]

Would the C_{60} cage have the effect of stabilizing the MH_n cluster or would dissociation or reaction occur? We performed quantum chemical calculations on the supermolecular systems MH_n+C_{60}. To our knowledge, this is the first study on metal−hydrides inside a fullerene.

A recent paper by Zhao et al.[12] proposes to use transition-metal atoms bound to fullerenes as adsorbents for storage of hydrogen. Compounds such as $C_{60}[ScH_2(H_2)_4]_{12}$ have the Sc atoms bound externally to the fullerene cage. Here we propose a completely different approach, consisting of placing both the metal and the hydrogen inside the fullerene cage. We report results of a computational study on the species $ZrH_4@C_{60}$, $ScH_{15}@C_{60}$, and $ZrH_{16}@C_{60}$. While ZrH_4 is experimentally known as a single molecule[13] and has been

* Corresponding author e-mail: laura.gagliardi@unipa.it.

10.1021/ct0501856 CCC: $30.25
Published on Web 09/30/2005

Table 1. Typical Bond Distances (Å) for the MH_n Species[a]

	M−H(σ)	M−H(H_2)	H−H(H_2)	H−C
ZrH_4	1.859			
$ZrH_4@C_{60}$	1.706			2.095−2.633
ScH_{15}	1.865	2.018−2.232	0.806−0.861	
$ScH_{15}@C_{60}$	1.739	1.775−1.891	0.802−0.875	1.848−2.643
ZrH_{16}	1.871−1.882	2.124−2.136	0.827−0.829	
$ZrH_{16}@C_{60}$	1.753	1.807−1.853	0.807−0.917	1.841−2.635

[a] M−H(σ) is the distance between M and a nonbound H atom. M−H(H_2) is the distance between M and H in a H_2 molecule. H−H is the distance in the H_2 moiety. See Figures 1−3.

previously predicted by Pykkö and Desclaux,[14] ScH_{15} and ZrH_{16}, on the other hand, are unknown. All these $MH_n@C_{60}$ species turned out to be stable with all real harmonic frequencies. We extended the idea by considering also two TiH_{16} clusters inside a bicapped (9,0) carbon nanotube, C_{114}. This system also turned out to be stable.

Details of the Calculations

The program TURBOMOLE[15−17] was employed. The calculations were performed using density functional theory, DFT, with the Becke-Perdew, BP86, exchange-correlation functional. For the Zr metal, the energy-adjusted Stuttgart ECPs were used in order to take into account relativistic effects.[18] The number of valence electrons is 12. The basis sets accompanying the ECPs 5s3p2d were used.[19] For Sc, Ti, H, and C an all electron, split-valence basis set, contracted to 5s3p2d, 5s3p2d, 2s, 3s2p, respectively, was used. Equilibrium geometries and harmonic frequencies were computed for all species at the BP86/DFT level of theory, using the resolution-of-the-identity (RI) variant available in the TURBOMOLE package[15−17] to make the calculations feasible. The auxiliary basis set of split-valence plus polarization type, available in the TURBOMOLE library, were used for all atoms. Comparative equilibrium geometry calculations on some selected structures were performed at the second order perturbation theory level with the RI variant, RI-MP2, to check if possible weak interactions between the H atoms and the fullerene cage would imply a substantial rearrangement of the structures. Since the RI-MP2 structure of $ZrH_{16}@C_{60}$ turned out to be very similar to the BP86 structure, with a slightly shorter Zr−H bond distance at the RI-MP2 level, only BP86 structures will be discussed below. Frequency calculations, on the other hand, were not performed at the RI-MP2 level of theory.

Results

The calculations show that the MH_n systems have shorter M−H bonds when placed inside the C_{60} cage than as isolated molecules. All supersystems turn out to be local minima, having all harmonic frequencies real. The M−H bonds are significantly shorter when the MH_n species are placed inside the C_{60} cage, and the H−H bonds in the H_2 moieties are much longer than in a H_2 isolated molecule.

The M−H and H−H bond distances are reported in Table 1. In general the M−H bond distances become about 0.15 Å shorter when the MH_n cluster is inside the fullerene than when it is an isolated molecule.

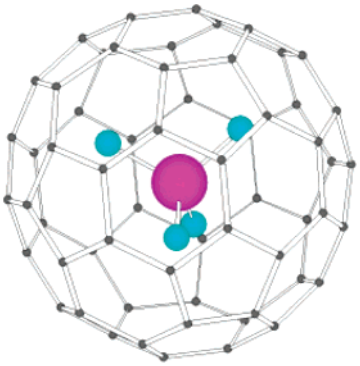


Figure 1. The structure of $ZrH_4@C_{60}$.

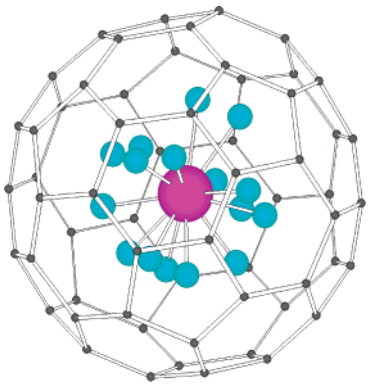


Figure 2. The structure of $ScH_{15}@C_{60}$.

ZrH_4 has a tetrahedral structure both as an isolated molecule and in C_{60}. The Zr−H bond distance is 1.86 Å in an isolated ZrH_4 molecule and 1.70 Å in $ZrH_4@C_{60}$. In $ZrH_4@C_{60}$ (Figure 1) the carbon−hydrogen distances vary between 1.69 and 2.63 Å. The infrared spectrum of ZrH_4 in solid argon has been observed by Chertihin and Andrews.[13] They measured an antisymmetric stretching frequency of 1623.6 cm^{-1}. Our calculation on isolated ZrH_4 gives a value of 1648.4 cm^{-1} in the harmonic approximation. The same mode for $ZrH_4@C_{60}$ is calculated to be 1119.5 cm^{-1}.

Isolated ScH_{15} has one Sc−H σ bond (1.86 Å) and seven Sc−H_2 bonds (2.00−2.22 Å). In C_{60}, ScH_{15} undergoes a significant structural rearrangement: the Sc atom is not at the center of the fullerene cage, surrounded symmetrically by the hydrogen atoms, as one might expect. Instead, the hydrogen atoms are all on one side of the Sc atom, and the Sc−H bonds are in the center of the fullerene cage (Figure 2). The Sc−H σ bond reduces to 1.74 Å and the Sc−H_2 bonds to 1.78−1.89 Å.

As already mentioned, our calculations predict isolated ZrH_{16} not to be stable and, instead, exist as ZrH_{12} and $2H_2$. ZrH_{12} has four Zr−H σ bonds of 1.88 Å and four Zr−H_2 bonds of 2.13 Å. In C_{60}, ZrH_{16} is bound (Figure 3), and it has two Zr−H σ bonds of 1.75 Å and seven Zr−H_2 bonds of 1.83 Å. The H−H bonds are longer than in an isolated H_2 molecule (0.82 Å compared with 0.77 Å).

$ZrH_4@C_{60}$, $ScH_{15}@C_{60}$, and $ZrH_{16}@C_{60}$ have all harmonic frequencies real. The calculated values of the frequencies are available upon request.

In all cases, placing the MH_n cluster inside a fullerene cage leads to a significant shortening of the M−H bond

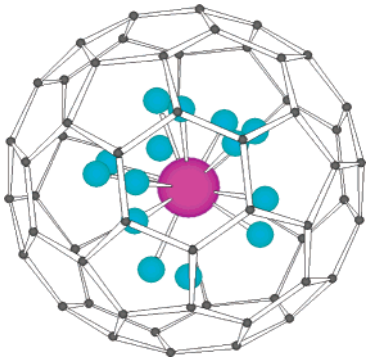


Figure 3. The structure of $ZrH_{16}@C_{60}$.

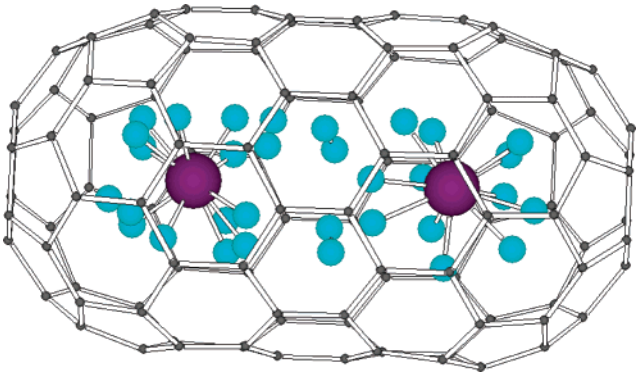


Figure 4. The structure of two TiH_{16} inside C_{114}.

distances. The large clusters (ScH_{15} and ZrH_{16}) change from a M−H nonbonding situation (the M−H bond distances are larger than 2 Å) to a bonding situation (the bond distances are about 1.8 Å). One possible explanation for the structural rearrangement is that the fullerene repels the electrons of the metal−polyhydride cluster and pushes the cluster as far as possible from its surface, i.e., toward the center of the sphere. One may have expected that the metal−hydride cluster would break the fullerene cage, but this does not seem to be the case.

The reason for the polarization of the ScH_{15} cluster in C_{60} is not clear to us. One possibility is the formation of a dipole that can interact favorably with the polarizability of the C_{60} unit. This could be the case if the energy difference between the symmetric and the polarized structure is small. For the isolated clusters the polarized structure (Sc on the side) has a dipole moment of ca. 3.95 D, while the symmetric structure has a dipole moment of 1.87 D.

As previously mentioned, ScH_{15} is an "18-electron" species, while ZrH_{16} is a "20-electron" species. Even if the results are not reported here, structural optimization had also been performed for $TiH_{14}@C_{60}$ and $ZrH_{14}@C_{60}$, as other "18-electron" species, and they also are bound local minima.

Up to now it has been shown that the MH_n systems here considered are more compact when placed inside a C_{60} than as isolated molecules. The next step is trying to understand if these systems are energetically stable inside C_{60}. Thermodynamical stability implies these species to be lower in energy than other isomers of the same molecules. In the $ZrH_4@C_{60}$ case, we considered two other isomers besides the one previously discussed (Figure 1): the first isomer is a Zhao[12]-style $ZrH_4@C_{60}$ species, with ZrH_4 attached to the external C_{60} surface (structure A), and the other isomer has ZrH_4 inside the C_{60}, but close to the C_{60} internal surface in order that one H atom is van der Waals bonded to C_{60} (structure B). Both structures A and B are significantly higher in energy (more than 150 kcal/mol) than the original $ZrH_4@C_{60}$ structure (Figure 1). The optimization of A leads to separated ZrH_4 and C_{60} fragments, while the optimization of B leads to the original $ZrH_4@C_{60}$ structure (Figure 1). These results confirm the original $ZrH_4@C_{60}$ isomer to be a global minimum.

Kinetic stability of these systems, on the other hand, would imply that they are stable with respect to one or more dissociation pathways. Possible ways of forming/dissociating the supermolecular systems were investigated. It seems unfeasible to 'put' an already formed metal−hydride inside a fullerene cage. Simple energetic balances (differences between calculated electronic energies) show that the formation of $ZrH_4@C_{60}$ from ZrH_4 and C_{60}, for example, is highly endothermic and requires about 100 kcal/mol of energy, allowing for both the zero point vibrational energy and the basis set superposition error corrections (less than 2 kcal/mol). An alternative and, probably more realistic route, would consist of inserting the metal and molecular hydrogen into a fullerene. The formation of the MH_n species might then occur directly inside C_{60}. The formation of $ZrH_4@C_{60}$ from a Zr atom, two H_2 molecules, and C_{60}, for example, is about 10 kcal/mol exothermic and thus energetically much more favorable than the above reaction. This energetic balance has been estimated as the difference between the electronic energy of $ZrH_4@C_{60}$ minus the electronic energy of one Zr atom in its triplet state and the energy of two H_2 molecules, allowing for both the zero point vibrational energy and the basis set superposition error corrections.

We also considered the possibility of forming metal−hydride clusters inside a nanotube, and we studied, as a model example, a system formed by two TiH_{16} clusters inside a single wall nanotube (9,0) which was closed with two partial fullerenes at the two extremities, C_{114} (Figure 4). The system turned out to be stable with all real harmonic frequencies. One of the TiH_{16} clusters inside the nanotube evolved to $TiH_{12} + 2H_2$, while the other one remained bound. A third TiH_{16} cluster, on the other hand, does not fit inside C_{114}, and some hydrogen atoms bind to the internal surface of C_{114}.

Conclusions

Quantum chemical calculations have shown that metal−hydride molecules are more compact when they are placed inside a fullerene cage than when they are isolated molecules. Species such as $ZrH_4@C_{60}$, $ScH_{15}@C_{60}$, and $ZrH_{16}@C_{60}$ exist and have very short metal−hydrogen bonding**.** These results pose several questions. First of all it is important to find a synthetic route to make hydrogen−metal clusters inside the fullerene. It might be possible to insert the metal and molecular hydrogen into a fullerene. The formation of the MH_n species might then occur directly inside C_{60}. This study indicates that the $MH_n@C_{60}$ supersystems require a large amount of energy to be made, but once they are synthesized, they may have a chance to be stable.

An alternative route, probably easier from the synthetic point of view, would be to synthesize the clusters inside nanotubes, instead of fullerenes. The calculations have shown that two TiH_{16} clusters are stable inside a (9,0) bicapped nanotube, C_{114}. A general question that we pose is the following: is it possible to make also other species inside a fullerene. Preliminary results indicate that also $CH_4@C_{60}$ is stable. Can we make clusters with a larger number of hydrogen atoms? It seems evident from the present study that metal−polyhydride clusters have indeed a chance to exist inside a fullerene cage and a nanotube, and this approach may represent a conceptually new way of synthesizing, stabilizing, and storing metal−polyhydrides.

Acknowledgment. This work was supported by the Italian Ministero dell'Istruzione dell'Università e della Ricerca, MIUR. The author thanks Professor Pekka Pyykkö, Helsinki, and Professor Reinhart Ahlrichs, Karlsruhe, for useful discussions.

References

(1) Grochala, W.; Edwards, P. P. *Chem. Rev.* **2004**, *104*, 1283.

(2) Kubas, G. J. *Metal Dihydrogen and σ-bond complexes structure, theory and reactivity*; Kluwer: New York, 1999.

(3) Maseras, F.; Lledo's, A.; Clot, E.; Eisenstein, O. *Chem. Rev.* **2000**, *100*, 601.

(4) Gagliardi, L.; Pyykkö, P. *J. Am. Chem. Soc.* **2004**, *126*, 15014.

(5) Kadish, K. M.; Ruoff, R. S. *Fullerenes: Chemistry, Physics and Technology*; John Wiley and Sons: New York, 2000; pp 357−436.

(6) Akasaka, T.; Nagase, S. *Endofullerenes: A New Family of Carbon Clusters*; Kluwer Academic Publisher: Dordrecht, 2000.

(7) Wang, C. R.; Kai, T.; Tomiyama, T.; Yoshida, T.; Kobayashi, Y.; Nishibori, E.; Takata, M.; Sakata, M.; Shinohara, H. *Angew. Chem., Int. Ed.* **2001**, *40*, 397.

(8) Olmstead, M. M.; Lee, H. M.; Duchamp, J. C.; Stevenson, S.; Marciu, D.; Dorn, H. C.; Balch, A. L. *Angew. Chem., Int. Ed.* **2003**, *42*, 900.

(9) Syamala, M.; Cross, R. J.; Saunders: M. *J. Am. Chem. Soc.* **2002**, *124*, 6216.

(10) Kobayashi, K.; Nagase, S.; Dinse, K. P. *Chem. Phys. Lett.* **2003**, *377*, 93.

(11) Iwamatsu, S.; Murata, S.; Andoh, Y.; Minoura, M.; Kobayashi, K.; Mizorogi, N.; Nagase, S. *J. Org. Chem.* **2005**, *70*, 4820.

(12) Zhao, Y.; Kim, Y. H.; Dillon, A. C.; Heben, M. J.; Zhang, S. B. *Phys. Rev. Lett.* **2005**, *94*, 155504.

(13) Chertihin, G. V.; Andrews, L. *J. Am. Chem. Soc.* **1995**, *117*, 6402.

(14) Pyykkö, P.; Desclaux, J. P. *Chem. Phys Lett.* **1977**, *50*, 503.

(15) Ahlrichs, R.; Bär, M.; Häser, M. *Chem. Phys. Lett.* **1989**, *162*, 165.

(16) Häser, M.; Ahlrichs, R. *J. Comput. Chem.* **1989**, *10*, 104.

(17) von Arnim, M.; Ahlrichs, R. *J. Chem. Phys.* **1989**, *90*, 1007.

(18) Andrae, D.; Häussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123.

(19) Institut für Theoretische Chemie, Universität Stuttgart. *ECPs and corresponding valence basis sets*. http://www.theochem.uni-stuttgart.de/.

CT0501856

JCTC Journal of Chemical Theory and Computation

An Efficient Real Space Multigrid QM/MM Electrostatic Coupling

Teodoro Laino,*[,†,‡] Fawzi Mohamed,[‡] Alessandro Laio,[‡] and Michele Parrinello[‡]

Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, I-56125 Pisa, Italy, and Computational Science, DCHAB, ETH Zurich, USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

Received May 3, 2005

Abstract: A popular strategy for simulating large systems where quantum chemical effects are important is the use of mixed quantum mechanical/molecular mechanics methods (QM/MM). While the cost of solving the Schrödinger equation in the QM part is the bottleneck of these calculations, evaluating the Coulomb interaction between the QM and the MM part is surprisingly expensive. In fact it can be just as time-consuming as solving the QM part. We present here a novel real space multigrid approach that handles Coulomb interactions very effectively and implement it in the CP2K code. This novel scheme cuts the cost of this part of the calculation by 2 orders of magnitude. The method does not need very fine-tuning or adjustable parameters, and it is quite accurate, leading to a dynamics with very good energy conservation. We exemplify the validity of our algorithms with simulations of water and of a zwitterionic dipeptide solvated in water.

Introduction

The rapid growth of computer technology has drastically changed the way in which molecular simulations and quantum chemical calculations are used to shed light on the driving forces of chemical reactivity in complex environments. However, the computational treatment of quantum systems made of several hundreds/thousands of atoms is still challenging,[1,2] especially if long simulation times are required. One of the currently followed strategies is the development of linear scaling quantum mechanical (QM) methods.[3−9] However, even with present-day hardware and the most efficient linear scaling method, it is not possible to study with fully ab initio methods many of the complex systems that are of current interest in biology and nanotechnology. An alternative approach is to employ multiscale quantum mechanics/molecular mechanics (QM/MM) schemes. They are particularly useful whenever a chemical reaction involves atoms in a small region, usually labeled QM. The QM part is treated with ab initio methods. The rest of the system, usually labeled MM, is instead treated with a less computationally expensive theory, generally molecular mechanics. Since the pioneering work of Warshel and Levitt,[10] the use and calibration of these techniques applied to the study of chemical reactivity has increased enormously.[11−18] Several quantum mechanics programs have been adapted to perform hybrid QM/MM simulations,[14,17,19−21] using either semiempirical,[20,22] ab initio Hartee-Fock,[19] post Hartree−Fock,[23] or DFT Hamiltonian.[14,17,19,24,25]

The study of chemical reactions in condensed phases is computationally demanding, owing not only to the size of the simulating system but also to the large degree of configurational sampling necessary to characterize a chemical reaction. This places severe demands on the efficiency of the implementation of any QM/MM scheme. Two main bottlenecks can be identified in such calculations: one concerns the evaluation of the QM energy and derivatives, while the other is associated with the evaluation of the electrostatic interaction between the QM and the MM part. In this respect we can identify two classes of codes, those based on Gaussian-type orbitals (GTOs) to represent both the wave function and the charge density[26,27] and those using

* Corresponding author fax: +41919138817; e-mail: teodoro.laino@sns.it. Corresponding author address: Computational Science, DCHAB, ETH Zurich, USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland.

† Scuola Normale Superiore di Pisa.

‡ Computational Science, DCHAB, ETH Zurich, USI Campus.

10.1021/ct050123f CCC: $30.25
Published on Web 10/07/2005

grids in real space to represent the charge density.[1,28,29] The latter encompasses both codes fully based on plane waves (PWs) and the more recent mixed approaches based on Gaussian plane waves (GPWs). It is on this second class of algorithms that our paper is focused.

For localized basis sets (GTOs), the use of an efficient prescreening technique is imperative in order to avoid the quadratic construction of the one-electron QM/MM Hamiltonian matrix. For nonlocal basis sets (PWs), if the interaction is evaluated analytically, the computational price is proportional to the number of grid points times the number of MM atoms. Surprisingly the evaluation of the QM/MM electrostatic interaction, for the latter scheme, requires between 20% and 100% of the time needed by the QM calculation, this despite the use of sophisticated hierarchical multipole (HMP) methods[24] or of clever implementations based on electrostatic cutoffs.[17] Furthermore these techniques require a fine-tuning of parameters to yield optimal performance and lead to a loss of accuracy that makes error control difficult.

The aim of this work is to describe a new implementation of the QM/MM coupling term that avoids the use of any hierarchical method or multipole technique. This novel scheme is based on the use of multigrid techniques in conjunction with the representation of the Coulomb potential through a sum of functions with different cutoffs, derived from the new Gaussian expansion of the electrostatic potential (GEEP for short). The QM/MM driver is based on the quantum mechanical program Quickstep and the molecular mechanics driver FIST (both part of the CP2K package).

The overall speedup is of 1−2 orders of magnitude with respect to other PW-based implementations of the QM/MM coupling Hamiltonian.[14,17] The lack of tuning parameters and electrostatic cutoffs makes this implementation a totally free parameter scheme, once the cutoff of the finest grid level has been specified. Consequently, very stable simulations can be obtained with optimal energy conservation properties.

We test the present implementation by computing the pair correlation function of quantum water in classical water and the pair correlation function of a zwitterionic dipeptide in classical water. Both tests address the correctness of the new coupling scheme.

Wave Function Optimization

We establish our implementation on the use of an additive[11,12,30] QM/MM scheme, where the total energy of the molecular system can be partitioned into three disjoint terms

$$E_{\mathrm{TOT}}(\mathbf{r}_\alpha,\mathbf{r}_\mathrm{a}) = E^{\mathrm{QM}}(\mathbf{r}_\alpha) + E^{\mathrm{MM}}(\mathbf{r}_\mathrm{a}) + E^{\mathrm{QM/MM}}(\mathbf{r}_\alpha,\mathbf{r}_\mathrm{a}) \quad (1)$$

where E^{QM} is the pure quantum energy, E^{MM} is the classical energy, and $E^{\mathrm{QM/MM}}$ represents the mutual interaction energy of the two subsystems. These energy terms depend parametrically on the coordinates of the quantum nuclei ($\mathbf{r}_\alpha$) and classical atoms ($\mathbf{r}_\mathrm{a}$).

The quantum subsystem is described at the density functional theory (DFT) level, exploiting the Quickstep[31] algorithm. This is based on the expansion of the Kohn−Sham orbitals in GTOs and on the use of an auxiliary PW basis set to evaluate the Coulomb interactions[31−33] and is included in the CP2K package.[1] Although the program has not been optimized to yield order N scaling for large systems, it is extremely fast and efficient in comparison with conventional DFT Gaussian or plane wave schemes for medium/large molecules. An extensive review of technical details and performance regarding Quickstep can be found in previous work.[31,34]

The classical subsystem is described through the use of the MM driver called FIST, also included in the CP2K package. This driver allows the use of the most common force fields employed in molecular mechanics simulations.[35,36]

The interaction energy term $E^{\mathrm{QM/MM}}$ contains all nonbonded contributions between the QM and the MM subsystem, and in a DFT framework we express it as

$$E^{\mathrm{QM/MM}}(\mathbf{r}_\alpha,\mathbf{r}_\mathrm{a}) = \sum_{\mathrm{a}\in\mathrm{MM}} q_\mathrm{a} \int \frac{\rho(\mathbf{r},\mathbf{r}_\alpha)}{|\mathbf{r}-\mathbf{r}_\mathrm{a}|} d\mathbf{r} + \sum_{\substack{\mathrm{a}\in\mathrm{MM}\\ \alpha\in\mathrm{QM}}} \upsilon_{\mathrm{VdW}}(\mathbf{r}_\alpha,\mathbf{r}_\mathrm{a}) \quad (2)$$

where $\mathbf{r}_\mathrm{a}$ is the position of the MM atom a with charge q_a, $\rho(\mathbf{r},\mathbf{r}_\alpha)$ is the total (electronic plus nuclear) charge density of the quantum system, and $\upsilon_{\mathrm{VdW}}(\mathbf{r}_\alpha,\mathbf{r}_\mathrm{a})$ is the van der Waals interaction between classical atom a and quantum atom α. The implementation of the electrostatic $E^{\mathrm{QM/MM}}$ term poses serious theoretical and technical problems, related to both its short-range and its long-range behavior. An important issue, connected with the short-range behavior, is related to the so-called *electron spill-out* problem. In classical calculations the atoms are normally represented by a simple point charge. Whenever electrons of a QM atom come very close to the core of a classical atom, the electrons can be trapped into the pointlike classical potential energy source. The use of diffuse basis sets (or of plane waves) can enhance this behavior. A natural solution to this problem is to represent MM atoms with a finite width density of charge in the interaction with QM atoms. This width will be dependent on the atom type and can be expected to be similar to the covalent radius of the atom. The recently proposed implementation[17] describes the short-range interactions accurately, employing a pseudopotential-like approach to remedy the unphysical problem of overpolarization that arises when MM atoms are in close contact with the QM region.

To handle properly the short-range interaction we decided to use, rather than the functional form introduced by Laio et al.,[17] the following expression for the classical potential

$$\upsilon_\mathrm{a}(\mathbf{r},\mathbf{r}_\mathrm{a}) = \frac{\mathrm{Erf}\left(\frac{|\mathbf{r}-\mathbf{r}_\mathrm{a}|}{r_{\mathrm{c,a}}}\right)}{|\mathbf{r}-\mathbf{r}_\mathrm{a}|} \quad (3)$$

where $r_{\mathrm{c,a}}$ is an atomic parameter, close to the covalent radius of the atom a. This function is the exact potential energy function originated by a Gaussian charge distribution $\rho(\mathbf{r},\mathbf{r}_\mathrm{a}) = (1/\sqrt{\pi}*r_{\mathrm{c,a}})^3 \exp(-(|\mathbf{r}-\mathbf{r}_\mathrm{a}|/r_{\mathrm{c,a}})^2)$ and is commonly employed in the Ewald summation techniques. Moreover, the expression in eq 3 has the desired property of tending to $1/r$ at large distances and going smoothly to a constant for

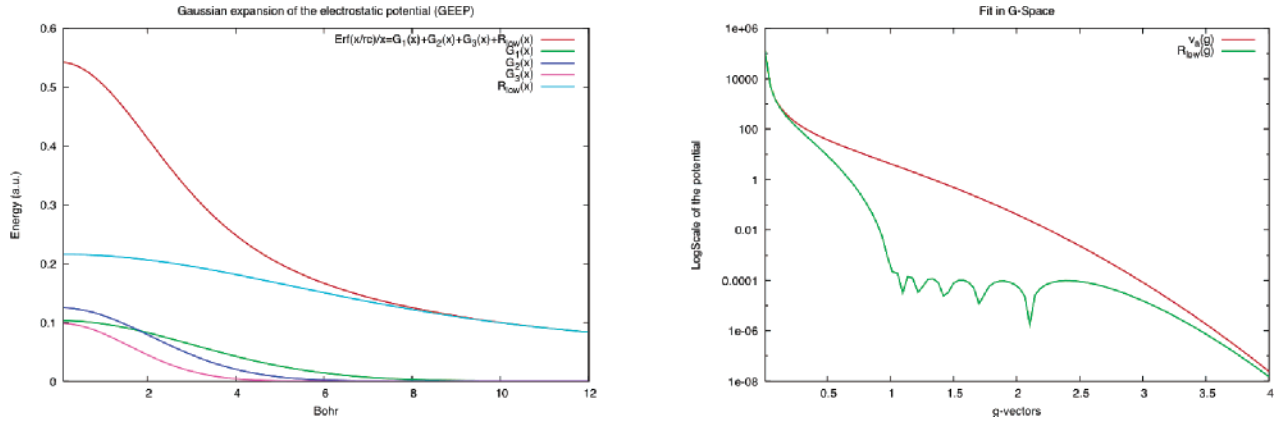


Figure 1. On the left: Gaussian expansion of the electrostatic potential (GEEP). The picture shows the components of the fit for the value $r_{c,a} = 1.1$ Å. On the right: Fourier transform of the potential in eq 3 (in red) and Fourier transform of the residual function R_{low} in eq 5 (in green). For this particular case ($r_{c,a} = 1.1$) we can define for the residual function a $G_{cut} \approx 1.0$.

small r. In the contest of QM/MM calculations,[37,38] the MM charges have been Gaussian smeared as a means to repair the broken covalent bonds at the QM/MM boundaries. In contrast here, we do not address the issue of treating QM/MM regions crossed by a covalent bond, and the smearing of all MM charges is exploited in order to prevent the spill-out problem and to accelerate the calculation of the electrostatic interactions.

Due to the Coulomb long-range behavior, the computational cost of the integral in eq 2 is surprisingly large. When using a localized basis set like GTOs, the most natural way to handle this term is to modify the one-electron Hamiltonian by adding to it the contribution of the MM classical field

$$H^{\mu\nu}_{\mathrm{QM/MM}} = -\int \phi_\mu(\mathbf{r},\mathbf{r}_\alpha) \sum_{a\in \mathrm{MM}} \frac{q_a}{|\mathbf{r}_a - \mathbf{r}|} \phi_\nu(\mathbf{r},\mathbf{r}_\alpha) d\mathbf{r} \qquad (4)$$

ϕ_μ and ϕ_ν being Gaussian basis functions and q_a the atomic charge of classical atom a with coordinates $\mathbf{r}_a$. In this case a suitable prescreening procedure has to be applied for the integral evaluation, to effectively compute only the nonzero terms and thus avoiding the quadratically scaling construction of the core Hamiltonian with respect to the number of elements of the basis set. When using a fully delocalized basis set like PWs, on the other hand, the QM/MM interaction term is evaluated by modifying the external potential and collocating on the grid nodes the contribution coming from the MM potential. Unfortunately the number of operations that a direct evaluation of eq 2 requires is of the order of $N_u N_{MM}$, where N_u is the number of grid points, usually of the order of 10^6 points, and N_{MM} is the number of classical atoms, usually of the order of 10^4 or more in systems of biochemical interest. It is evident that in a real system a brute force computation of the integral in eq 2 is impractical.

GEEP: Gaussian Expansion of the QM/MM Electrostatic Potential

The key to our method is the efficient decomposition of the electrostatic potential in terms of Gaussian functions with different cutoffs. The most general representation of the electrostatic potential eq 3 in terms of Gaussian functions with different cutoffs is

$$v_a(\mathbf{r},\mathbf{r}_a) = \frac{\mathrm{Erf}\left(\frac{|\mathbf{r}-\mathbf{r}_a|}{r_{c,a}}\right)}{|\mathbf{r}-\mathbf{r}_a|} = \sum_{N_g} A_g \exp\left(-\left(\frac{|\mathbf{r}-\mathbf{r}_a|}{G_g}\right)^2\right) + R_{low}(|\mathbf{r}-\mathbf{r}_a|) \qquad (5)$$

where the smoothed Coulomb potential is expressed as a sum of N_g Gaussian functions and of a residual function R_{low}. The A_g are the amplitudes of the Gaussian functions, and G_g are their width. If the parameters A_g and G_g are properly chosen, the residual function R_{low} will be smooth, i.e., its Fourier transform will have a compact domain for very small g vectors, and will be approximately zero for $g \gg G_{cut}$. The G_{cut} parameter is related to the spacing of the grid on which the R_{low} function will be mapped. We performed the fit of eq 5 by a least-squares approach in Fourier space, using the analytical expression of the **g**-representation of the modified electrostatic potential:[39]

$$\tilde{v}_a(\mathbf{g}) = \left[\frac{4\pi}{g^2}\right] \exp\left(-\frac{g^2 r_{c,a}^2}{4}\right) \qquad (6)$$

In Figure 1 we show the result of the fitting procedure in G-space with $r_{c,a} = 1.1$ Å, comparing the Fourier components of the modified Coulomb potential with the Fourier components of the residual function R_{low}. In this case the compact support of R_{low} is truncated at $G_{cut} \approx 1.0$ which should be compared with the value of $G_{cut} \approx 3.0$ needed to achieve the same accuracy when using $v_a(\mathbf{r},\mathbf{r}_a)$. This implies that the residual function can be mapped on a grid with a spacing 1 order of magnitude bigger than the one required to map the v_a function.

In Figure 1 we show the same result of the fit in real space, and in Table 1 we provide coefficients for selected values of $r_{c,a}$.

The advantage of this decomposition scheme is that grids of different spacing can be used to represent the different

Table 1. Amplitudes and Coefficients of the Optimal Gaussian Functions as Derived by the Fit

no. of Gaussians	radius $r_{c,a} = 1.1$ Å		radius $r_{c,a} = 0.44$ Å	
	A_g (au)	G_g (bohr)	A_g (au)	G_g (bohr)
1	0.103103	4.243060	0.230734	1.454390
2	0.125023	2.966300	0.270339	1.094850
3	0.098613	2.254250	0.075855	4.906710
4			0.190667	0.883485
5			0.173730	1.965640
6			0.127689	2.658160
7			0.095104	3.591640

contributions to $v_a(\mathbf{r},\mathbf{r}_a)$. In fact, the evaluation of a function on a grid relies on the assumption that the grid spacing is optimally chosen on the basis of its Fourier transform spectrum. Writing a function as a sum of terms with compact support and with different cutoffs, the mapping of the function is achieved using different grid levels, in principle as many levels as contribution terms, each optimal to describe the corresponding function. In our case, sharp Gaussians require fine grids, while coarser grids are necessary for the smoothest components. In addition the Gaussians can be truncated beyond a certain threshold value, which makes the collocation of the Gaussians on the grid a very efficient process.

The problem of mapping a noncompact function on a fine grid is then converted into the mapping of N_g compact functions on grids with cutoffs lower or at least equal to the fine grid, plus a noncompact very smooth function R_{low} mapped on the coarsest available grid. The sum of the contributions of the several grids, suitably interpolated, will be approximately equal to the function mapped analytically on the fine grid within errors due to the interpolation procedure.

Multigrid Framework

Multigrid methods are well-established tools in the solution of partial differential equations.[40,41] In the present implementation multigrid techniques are employed to combine functions with different cutoffs, i.e., represented on different grid levels.

Let us start by considering two grids, a coarse grid C with N_c points and a fine grid F with N_f points, whose grid-level is k-1 and k, respectively. The *interpolation* operator

$$I_{k-1}^{k}: C \rightarrow F \tag{7}$$

is by definition a transfer operator of a low cutoff function to a grid with a higher cutoff. The extension of the function to more points requires some regularity assumptions on its behavior. Two limiting cases can be identified: C^1 and C^∞, which can be handled with a simple linear interpolation scheme and with a *G-space* interpolation, respectively. If the function is C^∞, as in the case of a Gaussian, it is normally better to use an interpolator that assumes a high regularity. This ceases to be true once a collocation threshold is defined for the mapping of the Gaussians. In fact, the function on the grid becomes less regular, and an interpolation of a lower order might perform better. Another reason to avoid G-space interpolation comes from the fact that periodic boundary conditions with respect to the QM grid cannot be applied to the QM/MM potential. This makes the normal G-space interpolation unsuitable for our purpose. Thus we preferred to use an interpolation based on splines working entirely in real space unlike the scheme exploited by Yarne[16] et al. that relies on the use of splines in G-space. Moreover, the coding and the parallelization in real space is easier due to the use of commensurate grid levels and in our hands more efficient than working in G-space. For simplicity we use a set of commensurate grids, in which all the points of the coarse grid are points of the fine grid. Moreover, the number of points in each direction doubles going from the coarse to the fine grid level immediately above ($N_f = 8N_c$ in 3D). In the case of 1D space, the interpolation operator can be defined as

$$I_{k-1}^{k}(i,j) = \sum_n T(i,n)S^{-1}(n,j) \tag{8}$$

where for the points away from the border $T(i,n) = N_3(n-i/2)$ and $S(i,j) = N_3(j-i)$, with N_3 being the characteristic B-spline function of order 3.[42] The border was treated as a nonuniform B-spline. Higher dimensional spaces can be treated using the direct product of the transformation along the single dimensions. The opposite operation, the *restriction* J_k^{k-1}, is defined through the condition that the integral of the product of a function defined on the coarse grid with a function defined on a fine grid should give the same result both on the fine and on the coarse grid. Thus the restriction is simply the transpose of the interpolation

$$J_k^{k-1}(i,j) = [I_{k-1}^{k}(i,j)]^T = \sum_n S^{-1}(i,n)T(n,j) \tag{9}$$

Using N_{grid} grid levels and choosing the finer and coarser grid levels in order to treat correctly the sharpest and smoothest Gaussian components, respectively, we can achieve good accuracy and performance.

QM/MM Energy

The QM/MM electrostatic energy within DFT can be expressed with the following equation

$$E^{QM/MM}(\mathbf{r}_\alpha,\mathbf{r}_a) = \int d\mathbf{r}\rho(\mathbf{r},\mathbf{r}_\alpha)V^{QM/MM}(\mathbf{r},\mathbf{r}_a) \tag{10}$$

where $V^{QM/MM}$ is the electrostatic QM/MM potential evaluated on the finest grid, the same on which the final QM total density is evaluated. The overall description of the algorithm used to evaluate the QM/MM electrostatic potential on the finest grid can be outlined as follows:

(1) Each MM atom is represented as a continuous Gaussian charge distribution. The electrostatic potential generating from this charge is fitted through a Gaussian expansion using functions with different cutoffs, as shown in section 2.1.

(2) Every Gaussian function is mapped on one of the available grid levels, chosen to be the first grid whose cutoff is equal to or bigger than the cutoff of that particular Gaussian function. Using this collocation criterion, every

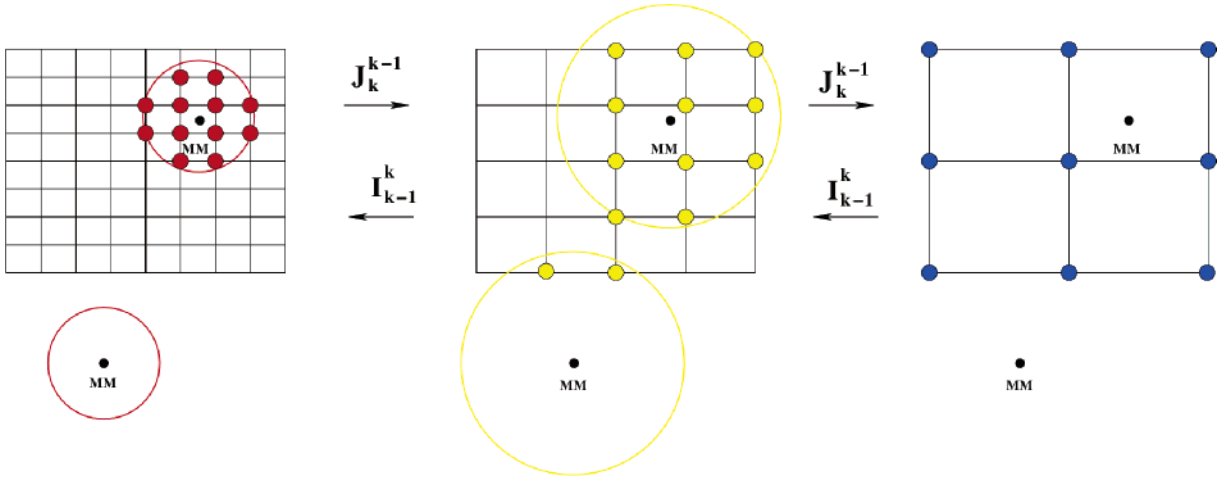


Figure 2. Schematic representation of the collocation procedure. Two MM atoms and three grid levels have been depicted. The circles (in the first and second grid levels) are the collocation regions of the Gaussian centered on the two MM atoms. Atoms whose distance from the QM box is greater than the Gaussian collocation radius do not contribute to the potential on that grid level. However, all atoms contribute to the coarsest grid level through the long-range R_{low} part.

Gaussian will be represented on the same number of grid points irrespective of its width. In practice a sub mesh of size $\approx 25 \times 25 \times 25$ suffices for an optimal Gaussian representation. Moreover, once a collocation threshold is defined, the Gaussian can be considered a compact domain function, i.e., it is zero beyond a certain distance, usually called a Gaussian radius. Thus only MM atoms embedded into the QM box, or close to it, will contribute to the finest grid levels, as shown in Figure 2.

The result of this collocation procedure is a multigrid representation of the QM/MM electrostatic potential $V_i^{QM/MM}(\mathbf{r},\mathbf{r}_a)$, where i labels the grid level, represented by a sum of single atomic contributions $V_i^{QM/MM}(\mathbf{r},\mathbf{r}_a) = \sum_{a\in MM} v_a^i(\mathbf{r},\mathbf{r}_a)$, on that particular grid level. In a realistic system the collocation represents most of the computational time spent in the evaluation of the QM/MM electrostatic potential, that is around 60−80%.

(i) Afterward, the multigrid expansion $V_i^{QM/MM}(\mathbf{r},\mathbf{r}_a)$ is sequentially interpolated starting from the coarsest grid level up to the finest level. The QM/MM electrostatic potential on the finest grid level can then be expressed as

$$V^{QM/MM}(\mathbf{r},\mathbf{r}_a) = \sum_{i=\text{coarse}}^{\text{fine}} \prod_{k=i}^{\text{fine}-1} I_{k-1}^{k}\, V_i^{QM/MM}(\mathbf{r},\mathbf{r}_a) \qquad (11)$$

where $V_i^{QM/MM}(\mathbf{r},\mathbf{r}_a)$ is the electrostatic potential mapped on grid level i, and I_{k-1}^{k} is the interpolation operator in real space. This operation does not depend on the number of MM atoms but only on the number of grid points, i.e., on the cutoff used in the calculation and on the dimensions of the QM box. For realistic systems the computational cost is around 20−40% of the overall cost of the evaluation of the QM/MM electrostatic potential.

Using the real space multigrid technique together with the GEEP expansion, the prefactor in the evaluation of the QM/MM electrostatic potential has been lowered from $N_f{*}N_f{*}N_f$ to $N_c{*}N_c{*}N_c$, where N_f is the number of grid points on the finest grid and N_c is the number of grid points on the coarsest grid. The computational cost of the other operations for evaluating the electrostatic potential, such as the mapping of the Gaussians and the interpolations, becomes negligible in the limit of a large MM system, usually more than 600−800 MM atoms.

Using the fact that grids are commensurate ($N_f/N_c = 2^{3(Ngrid-1)}$), and employing for every calculation 4 grid levels, the speed-up factor is around 512 (2^9); this means that the present implementation is 2 orders of magnitude faster than the direct analytical evaluation of the potential on the grid. The number of grid levels that can be used is limited by two technical factors. The first is that the coarsest grid needs to have at least as many points per dimension as the ones corresponding to the cutoff of the residual function R_{low} in order to perform the interpolation/restriction in an efficient manner. The second limitation is due to the constraint of using commensurate grid levels. The more grid levels are required in the calculation, the more the finest grid level cutoff will increase. This leads to an increment in memory requirements and to an unnecessary precision when handling the higher cutoff grids. Usually it is a combination of cutoff and grid levels that provides maximum efficiency. The two parameters can be chosen by checking that the coarsest grid level has no more than 5−10 grid points per dimension within the specified cutoff for the finest grid. Following the previous rule, the number of operations required for the direct evaluation of eq 2 is of the order of $N{*}100{*}N_{MM}$, where N is an integer between 1 and 10, and N_{MM} is the number of classical atoms.

QM/MM Energy Derivatives

The evaluation of the QM/MM derivatives on classical atoms are obtained by taking the derivative of eq 10 with respect to the classical atomic positions $\mathbf{r}_a$. These are given by

$$\frac{\partial E^{QM/MM}(\mathbf{r}_\alpha,\mathbf{r}_a)}{\partial_a} = \int d\mathbf{r}\rho(\mathbf{r},\mathbf{r}_\alpha)\frac{\partial V^{QM/MM}(\mathbf{r},\mathbf{r}_a)}{\partial \mathbf{r}_a} \qquad (12)$$

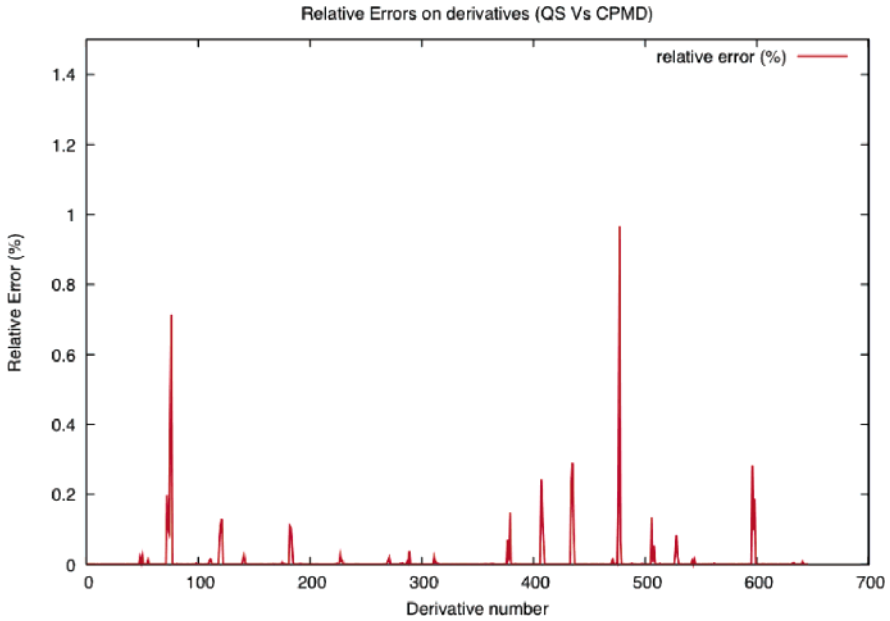


Figure 3. Relative errors on derivatives evaluated with the different functional form of eq 3 implemented in CPMD code and the new scheme implemented in CP2K. The average relative error is 0.01%.

The integral evaluation can be divided into terms deriving from the different grid levels

$$\frac{\partial E^{\mathrm{QM/MM}}(\mathbf{r}_\alpha,\mathbf{r}_a)}{\partial \mathbf{r}_a} = \sum_{i=\mathrm{coarse}}^{\mathrm{fine}} \int d\mathbf{r}\rho(\mathbf{r},\mathbf{r}_\alpha)\frac{\partial V_{\mathrm{fine}}^{i,\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a)}{\partial \mathbf{r}_a} \tag{13}$$

where the $V_{\mathrm{fine}}^{i,\mathrm{QM/MM}}$ labels the potential term on the finest grid level coming from the corresponding grid level i. Using the multigrid expression for terms $V_{\mathrm{fine}}^{i,\mathrm{QM/MM}} = \prod_{k=i}^{\mathrm{fine}} I_{k-1}^{k}\, V_i^{\mathrm{QM/MM}}$, the derivatives on MM atoms can be written as

$$\frac{\partial E^{\mathrm{QM/MM}}(\mathbf{r}_\alpha,\mathbf{r}_a)}{\partial \mathbf{r}_a} = \sum_{i=\mathrm{coarse}}^{\mathrm{fine}} \int d\mathbf{r}\rho(\mathbf{r},\mathbf{r}_\alpha)\prod_{k=i}^{\mathrm{fine}-1} I_{k-1}^{k}\frac{\partial V_i^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a)}{\partial \mathbf{r}_a} = \sum_{i=\mathrm{coarse}}^{\mathrm{fine}} \int d\mathbf{r}[\prod_{k=i+1}^{\mathrm{fine}} J_k^{k-1}]\rho(\mathbf{r},\mathbf{r}_\alpha)\frac{\partial V_i^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a)}{\partial \mathbf{r}_a} \tag{14}$$

In the previous equation the property that the interpolation operator is equal to the transpose of the restriction operator (and vice versa) was used. The MM derivatives are then evaluated applying the restriction operator to the converged QM $\rho(\mathbf{r},\mathbf{r}_\alpha)$. This leads to a multigrid expansion of the density, and each integral is evaluated on the appropriate grid level. The overall derivative is the sum of the contributions of the different grid levels.

We now consider the forces on the QM atoms. If $n_c^\alpha(\mathbf{r})$ is the Gaussian density used to represent the core charge distribution of the αth quantum ions and labeling with $P^{\mu\nu}$ the $\mu\nu$ element of the density matrix in the Gaussian basis set $\{\phi_\mu\}$, the derivatives on quantum ions due to the QM/MM interaction potential are

$$\frac{\partial E^{\mathrm{QM/MM}}(\mathbf{r}_\alpha,\mathbf{r}_a)}{\partial \mathbf{r}_\alpha} = \sum_{\mu\nu}\left(\frac{\partial P^{\mu\nu}}{\partial \mathbf{r}_\alpha}\right)V_{\mu\nu}^{\mathrm{QM/MM}} + 2\sum_{\mu\nu}P^{\mu\nu}\int d\mathbf{r}\left(\frac{\partial \phi_\mu(\mathbf{r},\mathbf{r}_\alpha)}{\partial \mathbf{r}_\alpha}\right)V_{\mu\nu}^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a)\phi_\nu(\mathbf{r},\mathbf{r}_\alpha) + \int d\mathbf{r}\left(\frac{\partial n_c^\alpha(\mathbf{r},\mathbf{r}_\alpha)}{\partial \mathbf{r}_\alpha}\right)V_{\mu\nu}^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a) \tag{15}$$

where $V_{\mu\nu}^{\mathrm{QM/MM}} = \int d\mathbf{r}\phi_\mu(\mathbf{r},\mathbf{r}_\alpha)V^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a)\phi_\nu(\mathbf{r},\mathbf{r}_\alpha)$ is the QM/MM Hamiltonian interaction term in the Gaussian basis set $\{\phi_\mu\}$. The first term is the so-called Pulay term[43] and is present because of the atom position dependent basis set.[31] The evaluation of the gradients on QM atoms is relatively inexpensive compared to a full quantum calculation. All considerations raised in section 3, regarding the scaling of the present scheme in the evaluation of the QM/MM potential, remain valid in the evaluation of the forces on classical atoms.

The calculation of the forces within the present implementation has been compared with the calculation of the forces using the method published elsewhere,[17] which is a QMMM of the CPMD code.[28] Comparison with the CPMD-QMMM code is complicated by the fact that in this last scheme[17] a multipolar expansion is used for the long-range part of the QM/MM electrostatic coupling, leading to inaccuracies. For this reason we compare only forces on atoms of the first MM solvation shell, which are treated exactly also in the CPMD-QMMM code. The realistic problem was made up of 215 classical SPC[44] water molecules and 1 QM water molecule. Although the system size is

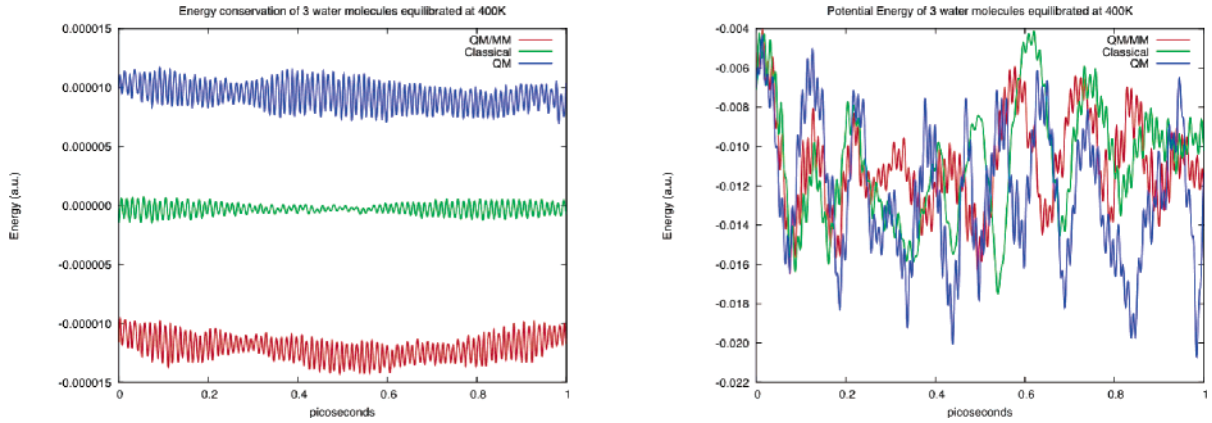


Figure 4. On the left: energy conservation of a system composed of 3 water molecules equilibrated at 400 K during 1 ps of simulation. The red line shows the total energy for the QM/MM run, the green line represents a pure classical run, and the blue line shows a pure quantum run. The total energies have been shifted for better visualization. No drift is observed, and all energy conservations are consistent. On the right: potential energy during the same run. Its variation is 3 orders of magnitude larger than the total energy variation.

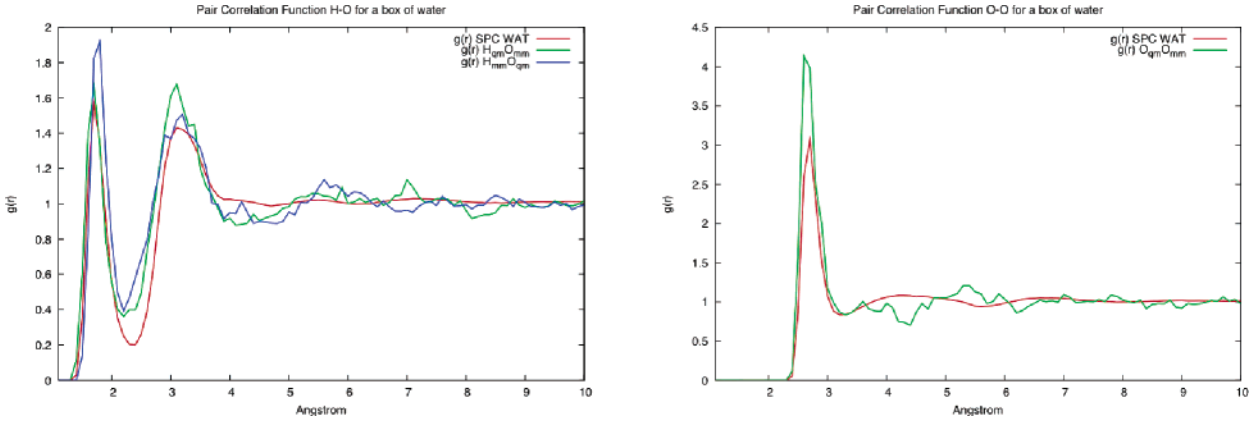


Figure 5. H−O and O−O pair correlation functions for QM water. QM/MM values are compared with the full classical SPC calculation. The QM/MM calculations are performed with $r_{c,a}$ equal to 1.2 Å for oxygen and 0.44 Å for hydrogen. The quantum box and the classical box employed in the simulation have a cube box size of 10.0 Å and 42.0 Å, respectively. The roughness of the QM/MM curve is due to the much shorter simulation time. The bin size for the evaluation of the pair correlation function is of 0.2 Å.

relatively small, the number of molecules present is comparable to the number of molecules normally treated exactly in CPMD-QMMM. In Figure 3 we show the relative error between the previous and the present implementations. The highest relative errors (less than 1.0%) correspond to forces that have small modules ($\leq 10^{-3}$ au). The average relative error is $\approx$0.01% with a speed-up in the energy and derivative evaluation of a factor of 40 wrt the CPMD.

An important benchmark for QM/MM codes that are aimed at molecular dynamics (MD) simulation is their ability to conserve energy. The system studied was composed of 3 water molecules (2 MM and 1 QM for the QM/MM run) previously equilibrated at 400 K. The simulation time was 1 ps, and results are shown in Figure 4. For comparison the energy of the pure classical and the quantum run are shown in the same picture. No drift is observed during 1 ps of simulation. We also show the potential energy during the simulation, whose oscillation is $\approx$3 orders of magnitude bigger than the total energy oscillation.

Results and Discussion

Consistently with checks done in previous work,[17,24] we test the accuracy of our implementation by computing the pair correlation function of a QM system embedded in a classical solvent. As found elsewhere,[17] the smoothing radius plays an important role in determining the bond properties of the system, and the choice of this parameter can have dramatic effects on pair correlation functions. The use of a different functional form (cf. eq 3 with eq 3 of Laio et al.[17]) forced us to reparametrize the $r_{c,a}$.

For the classical water molecules, the cutoff radii $r_{c,a}$ were chosen in order to reproduce the coordination number and the main peaks of the classical SPC water pair correlation function. A system of 2560 water molecules (2559 classical SPC water and 1 quantum water) in a cubic box, subject to periodic boundary conditions, was investigated. The system was previously equilibrated at $\rho = 1$ g/cm^3, $T = 298$ K. One SPC water molecule was then replaced by a QM water

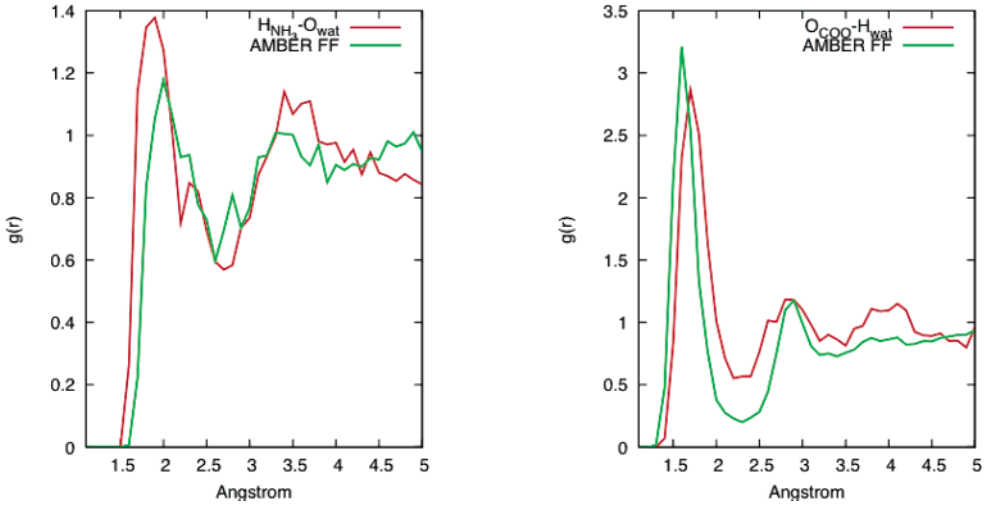


Figure 6. $H_{pept}-O_{wat}$ and $O_{pept}-H_{wat}$ pair correlation functions for the ALA-GLY dipeptide in SPC water. QM/MM values are compared with the full classical SPC calculation. The QM/MM calculations are performed with $r_{c,a}$ equal to 1.2 Å for oxygen and 0.44 Å for hydrogen. The quantum box and the classical box employed in the simulation have a cube box size of 15.0 Å and 50.0 Å, respectively. A plane-wave cutoff of 280 Hartree was used during all the simulations, in conjunction with the GTH pseudopotential and the BLYP exchange-correlation density functional. The roughness of the QM/MM curve is due to the much shorter simulation time. The bin size for the evaluation of the pair correlation function is of 0.2 Å.

molecule. GTH pseudopotentials[45] were used to describe the core charge distribution, and B-LYP exchange-correlation density functional[46,47] was employed in all the calculations, in conjunction with a cell plane-wave cutoff of $E_{cut} = 280$ Hartee. Several runs, with different values of the radius parameter $r_{c,a}$, were performed. The optimized radii are 0.44 Å for hydrogen and 1.20 Å for oxygen and allow the full classical SPC pair correlation function to be reproduced, as shown in Figure 5. Due to the different functional form of eq 3, the optimal values found with our implementation are different from the ones previously published.[17]

To test the transferability of the $r_{c,a}$ parameters determined for water, we also evaluated the pair correlation function of a QM dipeptide (GLY-ALA) zwitterion solvated in 3352 SPC water. We aimed at reproducing the pair correlation function obtained with the AMBER force field.[36] The pair correlation functions obtained with the present QM/MM implementation are indeed extremely close to the full classical results (see Figure 6).

Conclusions

We have presented an algorithm for evaluating the QM/MM coupling term with a fast linear scaling implementation. The main result is the dropping of the prefactor in the linear scaling, with a gain in the number of floating point operations proportional to $2^{3(N\text{grid}-1)}$, where N_{grid} is the number of grid levels used in the multigrid framework. The evaluation of the electrostatic potential on a grid is proportional to the number of MM atoms times the number of grid points. In real systems the linear scaling evaluation of the potential is therefore characterized by a prefactor $\approx 10^6$. In this scheme the prefactor is instead $\approx 10^3$. The number of floating point operations is reduced several orders of magnitude, and the computational time is 10−100 times smaller.

The algorithm is now implemented in the package CP2K, released under GPL license, and freely available on the Internet.[1] The scheme was validated by checking the energy conservation, and for a realistic system numerical accuracy was verified by comparing the forces with the analytical method, with a mean relative error of 0.01%. In addition, we computed the pair correlation function of a QM water molecule in classical water and of a QM zwitterionic dipeptide in classical water. The modified Coulomb interaction and the multigrid approach reproduce correctly the structural properties of a QM water molecule solvated in classical water and the parameters obtained therein can be used effectively to describe the properties of an organic molecule containing both negatively and positively charged moieties, as in the case of the zwitterion.

All tests address the correctness of results. The performance analysis confirms the present algorithm as the state of the art for the evaluation of QM/MM interaction coupling. Moreover, at variance with the majority of present-day QM/MM methods, our scheme does not rely on electrostatic cutoffs and so avoids all related problems. Consequently, the present method offers a fast, easy-to-use code for QM/MM calculations of large biological and inorganic systems.

Acknowledgment. The authors wish to thank Joost VandeVondele and Juerg Hutter for useful discussions and Christopher Mundy and William I-Feng Kuo for the very precious help they provided with the classical module FIST.

References

(1) Freely available at the URL: http://cp2k.berlios.de, released under GPL license.

(2) Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213−13225.

(3) York, D.; Lee, T.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 5011.

(4) Soler, J. M.; Artacho E.; Gale J. D.; Garcia A.; Junquera J.; Ordejon P.; Sanchez-Portal D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745−2779.

(5) Gogonea, V.; Merz, K. M. *J. Phys. Chem. A* **1999**, *103*, 5171.

(6) Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*, 4782.

(7) Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell, A. P. *J. Comput. Chem.* **2000**, *21*, 788.

(8) Khandogin, J.; York, M. D. *J. Phys. Chem. B* **2002**, *106*, 7693.

(9) Krajewski, F.; Parrinello, M. *Phys. Rev. B* **2005**, *71* (23), Art. No. 233105.

(10) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(11) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.

(12) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(13) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170−1179.

(14) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejon, P.; Roitberg, A. E.; Estrin, D. A. *J. Phys. Chem. B* **2003**, *107*, 13728−13736.

(15) Thompson, M. A. *J. Phys. Chem.* **1995**, *99*, 4794−4804.

(16) Yarne, D. A.; Tuckerman, M. E.; Martyna, G. J. *J. Chem. Phys.* **2001**, *115*, 3531−3539.

(17) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941−6947.

(18) Nam, K.; Gao, J.; York, D. M. *J. Chem. Th. Comput.* **2005**, *1*, 2−13.

(19) Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462−3471.

(20) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088−1100.

(21) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French S. A.; Sokol, A. A.; Bromley, S.T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1−28

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

(23) Ferre, N.; Olivucci, M. *J. Am. Chem. Soc.* **2003**, *125*, 6868−6869.

(24) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452−10467.

(25) Loferer, M. J.; Loeffler, H.H.; Liedl, K. R. *J. Comput. Chem.* **2003**, *24*, 1240−1249.

(26) Karlström, G.; Lindh, R.; Malmqvist, P.-A.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222.

(27) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(28) Hutter, J.; Alavi, A.; Deutsch, T.; Ballone, P.; Bernasconi, M.; Focher, P.; Fois, E.; Goedecker, S.; Marx, D.; Parrinello, M.; Tuckerman, M. *CPMD v.9.1*; Technical Report; MPI für Festkörperforschung, Stuttgart and IBM: Zurich Research Laboratory, 2005.

(29) Fusti-Molnar, L.; Pulay, P. *J. Chem. Phys.* **2002**, *116*, 7795−7805.

(30) Sherwood, P. In *Modern Methods and Algorithms of Quantum Chemistry*; John von Neumann Institute for Computing: 2000; Chapter Hybrid quantum mechanics/molecular mechanics approaches, Vol. 1, pp 257−277.

(31) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Comm.* **2005**, *167*, 103−128.

(32) Kohn, W.; Sham, L. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(33) *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

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

(35) Lee, Y.; Hodoscek, M.; Brooks, B.; Kador, P. *Biophys. Chem.* **1998**, *70*, 203−216.

(36) Case, D.; et al. *AMBER v.7.0*; Technical Report; University of California: San Francisco, 2002.

(37) Amara, P.; Field M. J. *Theor. Chem. Acc.* **2003**, *109*, 43−52

(38) Das, D.; Eurenius, P. K.; Billings, E. M.; Sherwood, P.; Chatfield, C. D.; Hodoscek, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534−10547

(39) The Mathematica Notebook used to develop the GEEP technology is part of the CP2K distribution and can be freely downloaded. Released under GPL license.

(40) *Multi-Grid Methods and Applications;* Springer-Verlag: Berlin, 1985; Vol. 4 of *Series in Computational Mathematics.*

(41) *A Multigrid Tutorial*; SIAM Books: Philadelphia, PA, 1987.

(42) Feng, G. *IEEE Trans. Signal Process.* **1998**, *46*, 2790−2796.

(43) Pulay, P. *Mol. Phys.* **1969**, *17*, 197.

(44) *Intermolecular Forces*; Reidel: Dordrecht, The Netherlands, 1981; Chapter Interaction models for water in relation to protein hydration, pp 331−342.

(45) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703−1710.

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

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

CT050123F

JCTC Journal of Chemical Theory and Computation

1-Fluoropropane. Torsional Potential Surface

Lionel Goodman* and Ronald R. Sauers

Department of Chemistry and Chemical Biology, Wright and Rieman Laboratories, Rutgers, the State University of New Jersey, New Brunswick, New Jersey 08903

Received August 15, 2005

Abstract: The systematic deletion of orbital interactions, using natural bond orbital (NBO) theory at the B3LYP/ 6-311++G(3df,2p) level, provides validation for the anti-C−H/C−F* hyperconjugative interaction providing the backbone for the gauche preference of 1-fluoropropane (FP). The FCCC torsional coordinate taking trans FP to gauche FP is predicted to be strongly contaminated by CCC bending with the result that a large part of the trans → gauche stabilization energy stems from mode coupling. The anti-C−H/C−F* hyperconjugative interaction is also found to play a major, if not determining, role in the coupling. The results of Rydberg deletion calculations suggest that Rydberg interactions play a role in NBO analysis, contrary to the usual assumption that interactions involving Rydberg orbitals can be ignored.

I. Introduction

The term "gauche effect"[1] is frequently used for the counterintuitive conformational preference of polar group 1,2-disubstituted ethanes. The equilibrium conformer of 1,2-difluoroethane (DFE), for example, is the gauche geometry shown in Figure 1a, where the two fluorines are adjacent (with a FCCF dihedral angle ∼ 72°) to each other.[2] The gauche conformer is calculated to be approximately 0.8 kcal/mol more stable than the metastable trans conformer, where the fluorines are in anti orientation (Figure 1b). The gauche preference is counterintuitive since dipole repulsions between the polar C−F bonds and steric effects both favor the trans conformation. The accepted explanation for the gauche equilibrium geometry is based on σ hyperconjugation involving charge transfer from C−H electron donor bonds to C−F* acceptor antibonds.[2,3] This type of orbital interaction is maximized when C−H bonds and C−F* antibonds are in anti orientation to each other. The consequence of the hyperconjugation model is that gauche conformational preference will be highly expressed for 1,2-disubstituted ethanes having two strongly electron-accepting substituents, as in 1,2-difluoroethane. This brings up the question of conformational preference in singly fluorine-substituted ethanes. A particularly interesting category is the 2-substituted 1-fluoroethanes, where maximal donor−acceptor interaction will depend on the C_2-X substituent bond electron-donating ability relative to both the C_2−H bond and their relative orientation. Rablen et al.[4] carried out an extensive study of the conformational preference and energetics of several 2-substituted 1-fluoroethanes and, in particular for the basic molecule, propane with a single fluorine substituent, 1-fluoropropane (FP; Figure 2). FP can be considered as a 2-substituted 1-fluoroethane with a methyl group in the 2 position. The conformational preference of FP is weakly gauche (by only 0.4 kcal/mol or less). Although Rablen's computations suggest that electrostatic attraction between fluorine and methyl strongly contributes to the gauche preference, they were not able to unambiguously establish the role of hyperconjugation. Since FP represents one of the simplest molecules exhibiting the gauche effect, its origin deserves detailed study for what it reveals about interactions leading to structural preference. We do this by using a feature

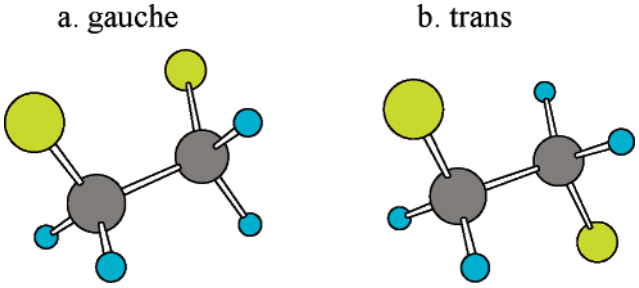


Figure 1. 1,2-Difluoroethane conformers.

* Corresponding author. E-mail: goodman@rutchem.rutgers.edu.

10.1021/ct050204b CCC: $30.25
Published on Web 10/11/2005

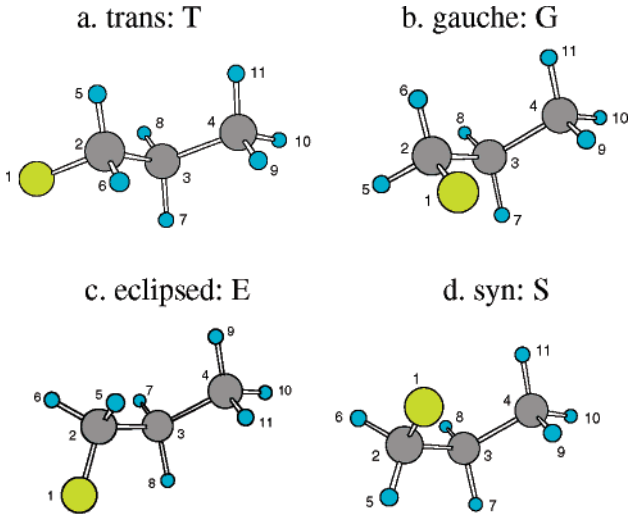


Figure 2. 1-Fluoropropane conformers.

of natural bond orbital (NBO) theory[3,5] that allows the hyperconjugative interactions to be switched off.

Dipole−dipole repulsion is usually invoked as the interaction that forces the FCCF dihedral angle to 72° in DFE.[6] However, a recently proposed alternate rationalization for the >60° angle invokes both anti interactions, which maximize near 60°, and cis interactions, which maximize at a much larger dihedral angle.[2] The absence of dipole−dipole repulsion in FP provides an opportunity to further address conformational determinants in fluorohydrocarbons. The >3 kcal/mol G−T barriers are sufficient to regard FP as a semirigid molecule, but given the presence of several low lying torsional modes, we are persuaded to examine the role of mode coupling as a significant factor in determining the gauche preference energetics.

Microwave and infrared experiments[7] support FP's gauche conformational preference (Figure 2b) in accord with theoretical prediction.[4] The microwave-determined trans−gauche energy difference is 0.37 kcal/mol, larger than either the DFT- or CCSD(T)-calculated difference, however. The experimental barrier height (3.47 kcal/mol at the eclipsed geometry) is in reasonable agreement with Rablen's MP2 (3.89 kcal/mol) and our DFT calculations (given in the next section). The microwave-spectroscopy-determined FCCC dihedral angle is 62.6°, in close agreement with Rablen's calculated 62.2°, and in somewhat less agreement with our slightly larger DFT value (Section II). Despite the reasonable agreement between experiment and the levels of theory used in ref 4 and in this study, interesting questions remain. Why is the energetic preference for gauche over trans so much higher in DFE than in FP, despite the absence of gauche destabilizing dipole−dipole repulsion in FP? Rablen et al. raised the possibility that the gauche conformational preference actually could be largely due to electrostatic attraction between the negatively charged fluorine and the C−H bond of the methyl group in FP.[4] However, a hyperconjugation origin was not thoroughly investigated. In this paper, we will examine this possibility in detail.

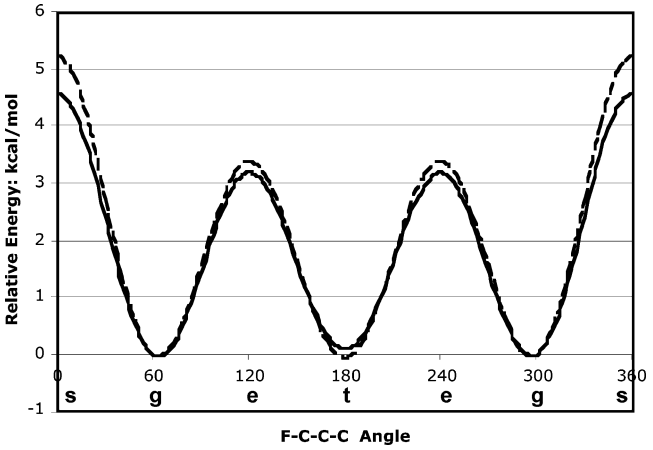


Figure 3. Torsional potential energy curves for 1-fluoropropane showing dependence on the F−C−C−C dihedral angle. Fully relaxed rotation (solid curve); skeletal geometry frozen at trans geometry (dashed curve).

Table 1. Energy of 1-Fluoropropane Conformers at Various Calculation Levels

method	gauche (Hartrees)	trans[a]	eclipsed[a]	syn[a]
HF/6-311++G(3df,2p)	−217.194 033	0.016	3.70	4.99
B3LYP/6-311++G(3df,2p)	−218.460 253	0.13	3.23	4.62
MP4/6-311++G(3df,2p)[b]	−218.037 724	0.19	3.55	4.70
CCSD(T) /6-311++G(2d,p)[b]	−218.035 210	0.19	3.53	4.67
MP2/6-311++G(3df,2p)	−217.968 472	0.21	3.73	4.94
MP2/6-31+G(d)[c]		0.38	3.85	5.26
MP2/6-311++G**(6D)[c,d]		0.28	3.89	5.27

[a] Energy (kcal/mol) of fully optimized geometry relative to gauche conformer unless otherwise noted. [b] Single-point calculations at B3LYP/6-311++G(3df,2p) optimized geometry. [c] Rablen et al.[4] [d] Single-point energies at MP2/6-31+G(d).

II. Torsional Potential Surface Landscape

The coordinate path between various conformers in FP can be usefully probed by FCCC dihedral angle rotation from 0° (for syn) through 60° (gauche) and 120° (eclipsed) to 180° (trans). Figure 3 shows the calculated[8] fully optimized potential surface in terms of the dihedral angle (solid curve). Both geometry optimization and energies were calculated at the B3LYP/ 6-311++G(3df,2p) level.[8] The potential surface landscape shows four features: a trans conformer (Figure 2a) at 180°, 0.13 kcal/mol higher than the equilibrium gauche conformer at 63.2° (Figure 2b), a prominent barrier (3.23 kcal/mol) between the trans and gauche conformers at the eclipsed geometry (119.9°; Figure 2c), and a still larger eclipsed-syn geometry barrier at 0° (Figure 2d), 4.62 kcal/mol above the gauche conformer. These are in qualitative agreement with both Rablen et al.'s MP2/6-311++G**(6D)[4] and Guirgis et al.'s MP2/6-311+G(2d,2p)[7b] single-point calculations. However, the smaller DFT gauche preference compared to that reported by Rablen's[4] 0.28−0.38 kcal/mol leads us to carry out additional calculations at different correlation levels. These are given in Table 1. Our results for single-point CCSD(T) calculations lead to an energy difference between the gauche and trans conformers of 0.19 kcal/mol, somewhat larger than the B3LYP value. In

Table 2. 1-Fluoropropane B3LYP/6-311++G(3df,2p) Optimized Geometries[a]

	gauche (G)	trans (T)	eclipsed (E)	syn (S)
	Skeletal Coordinates[b]			
R(C−F)	1.400	1.398	1.401	1.399
$R(C_2-C_3)$	1.514	1.514	1.531	1.541
$R(C_3-C_4)$	1.528	1.530	1.529	1.531
$\angle F-C_2-C_3$	110.44	110.37	110.60	111.86
$\angle C_2-C_3-C_4$	113.55	111.79	113.63	114.34
$\angle C_2-C_3-H_8$	107.24	108.74	106.96	108.86
$\angle F-C_2-C_3-C_4$	63.18	180.00	120.03	0.02
	C−H Coordinates			
$R(C_2-H_5)$	1.091	1.092	1.091	1.090
$R(C_2-H_6)$	1.093	1.092	1.092	1.090
$R(C_3-H_7)$	1.093	1.092	1.091	1.092
$R(C_3-H_8)$	1.094	1.092	1.092	1.092
$R(C_4-H_9)$	1.089	1.092	1.091	1.089
$R(C_4-H_{10})$	1.090	1.090	1.090	1.090
$R(C_4-H_{11})$	1.092	1.092	1.092	1.089

[a] See Figure 2 for conformer structure. [b] Bond lengths in Å, bond angles in degrees.

summary, these respectable high-level calculations lead to a gauche preference of only 0.1−0.3 kcal/mol.

III. Coupling between Skeletal Motion and FCCC Dihedral Rotation

The nonrotational skeletal relaxations accompanying FCCC dihedral angle rotation are given in Table 2. Optimized geometries for the gauche and trans conformers show that the major skeletal relaxation between these conformers is the 1.8° ∠CCC decrease accompanying gauche → trans rotation. All other bond length and bond angle changes are insignificant (<0.003 Å and <0.1°). However, there are additional skeletal changes aside from the mentioned ones at the barrier maxima, occurring for the syn and eclipsed dihedral geometries; for example, the C_2-C_3 bond undergoes significant (0.02−0.03 Å) lengthening. Insight into the role that the nonrotational coordinates play in the potential surface features, in particular for the trans → gauche stabilization energy and for the G−T barrier at the eclipsed geometry, can be seen from two potential curves in terms of the FCCC dihedral angle given in Figure 3, one with frozen trans skeletal geometry (dashed curve), the other obtained from global optimized geometries (solid curve).

The landscape of both curves closely resemble each other, with three minima for 360° rotation and maxima at the 0° and 120° $F-C_2-C_3-C_4$ angles (see Figure 2). However, there is a significant effect on the T → G stabilization energy, E_G-E_T. Frozen rotation essentially obliterates the weak gauche preference, reversing the trans−gauche stabilization energy from 0.13 kcal/mol, gauche-conformer-preferred (for the fully optimized potential curve), to −0.014 kcal/mol, that is, a slight trans conformer preference (Table 3). On the other hand, the 3.2 kcal/mol fully relaxed trans → gauche barrier is only slightly increased. The largest energetic change is at the syn geometry: from 4.6 kcal/mol (fully relaxed) to 5.3 kcal/mol (rigid rotation). We conclude that skeletal relaxation, that is, the nonrotational phase space of the trans → gauche reaction path, plays an important role in the FP gauche preference. It is also a component of the syn barrier that cannot be ignored.

Table 3. Relaxation Effects on the Conformer Energies (kcal/mol)

	relaxation[a]		
	E_G-E_T	E_E-E_T	E_S-E_T
fully relaxed	−0.13	3.10	4.49
rigid[b]	0.014	3.42	5.28
	partial relaxation[c]		
all bond lengths	−0.003	3.29	4.88
all bond angles	−0.089	3.25	4.77
$R(C_2-C_3)$	−0.013	3.29	4.88
$\angle C_2-C_3-C_4$	−0.094	3.35	4.92
$\angle C_2-C_3-C_4 + R(C_2-C_3)$	−0.094	3.23	4.74
$\angle C_2-C_3-H_8$	−0.016	3.50	5.42

[a] Skeletal relaxation accompanying FCCC dihedral angle rotation. [b] All skeletal coordinates are held at frozen trans geometry. [c] All other skeletal coordinates are held at frozen trans geometry.

To obtain further insight into the energetic consequences of relaxation, we decompose the full skeletal relaxation into two categories: (1) angular and (2) bond length relaxations. The results, given in Table 3, show that it is the angular ones that are primary.

Parsing of the skeletal angular relaxations, where only one angle is allowed to relax from its trans description and the other coordinates are frozen, is also given in Table 3. Partially relaxed rotations allow insight into the roles that individual internal coordinates play in trans−gauche energetics, and Table 2 shows that the paramount angular relaxation is ∠CCC scissor opening. This motion represents the largest amplitude oscillation in the 313 cm^{-1} gauche conformer "scissor" vibrational mode, and Table 3 clearly demonstrates that the coupling of dihedral torsion to this mode provides the major relaxation effect on G−T energetics.

IV. Hyperconjugation Model

The link between hyperconjugation and the gauche effect can be validated by removing all the electron charge transfers from bonds and lone pairs to antibond NBOs and Rydberg orbitals. The results of these single-point calculations at the gauche, trans, eclipsed, and syn geometries are given in Table 4, and the potential curve at 10° FCCC dihedral angle intervals is given in Figure 4. When all hyperconjugative interactions are switched off (dashed curve in Figure 4), the potential curve is inverted from that for the real molecule (solid curve). There still are two energy minima, but these occur at the eclipsed and syn geometries rather than at trans and gauche, which are now maxima. Furthermore, the B3LYP/6-311++G(3df,2p) energy of the gauche conformer is 0.5 kcal/mol higher than that of the trans instead of lower; the most stable structures are now syn and eclipsed, with eclipsed the strongly preferred conformer. The important conclusion: inversion of the potential curve from that of the real molecule when hyperconjugation is deleted clearly establishes that hyperconjugation is an essential part of the machinery driving the FP gauche preference. An interesting ancillary conclusion, not revealed in Table 4, is that the

Table 4. Deletion Effects on Conformer Energies (kcal/mol)

hyperconjugative deletion	T conformer deletion energy change[a]	G−T deletion difference	E−T deletion difference	S−T deletion difference
all hyperconjugation	117.4	0.46	−3.72	−2.10
all vicinal	116.0	−0.96	−2.56	−2.90
all geminal	17.7	−0.64	4.46	4.78
all remote	2.5	−0.53	2.30	4.07

[a] *E*(after deletion) − *E*(total).

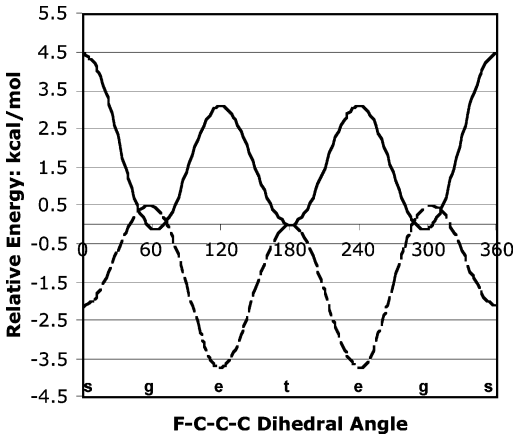


Figure 4. Potential energy curves for 1-fluoropropane showing the effect of orbital interaction. Full Hamilton, full optimization (solid curve); all hyperconjugative interactions deleted (dashed curve). Both curves have the zero-energy reference point at the 180° trans geometry.

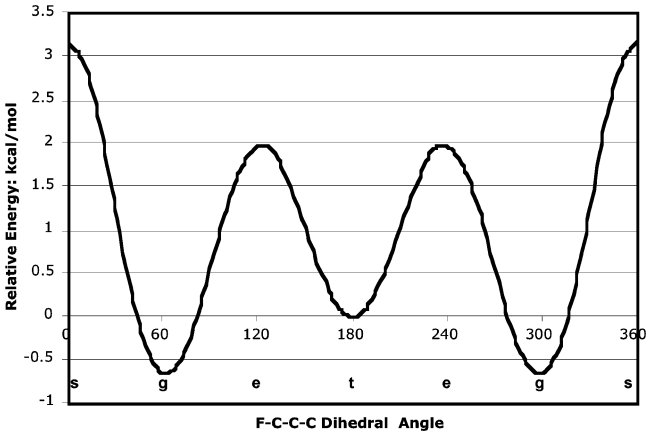


Figure 5. Torsional potential-energy curve for 1-fluoropropane with all geminal hyperconjugative interactions deleted using full Hamiltonian optimized geometry.

energies of both the syn and eclipsed structures are less affected by hyperconjugation deletion than those of the gauche and trans.

The removal of only the geminal hyperconjugations (e.g., C_2-H_5/C_2-F^*) leaves the lowest-energy FP conformer unchanged from gauche (Figure 5 and Table 4). Thus, geminal interactions, which include such bond/antibond charge transfers as $C_2-H_5/C_2-C_3^*$ and $C-F/C_2-H_5^*$, have little influence on the gauche geometry of the preferred conformer, but they do have a strong effect on syn and eclipsed energetics (Table 4).

However, when all vicinal hyperconjugations alone are removed, the landscape of the completely deleted hyperconjugations case is retained (Figure 6); the maxima are again at trans and gauche. There is also a parallel with the no-hyperconjugation calculation of Figure 4 in that the energies of the syn and eclipsed structures are less affected by vicinal hyperconjugation deletion than those of the gauche and trans. The consequence is that the deepest energy minima are (as in the all-hyperconjugation deletion case) at the syn and eclipsed geometries. But, surprisingly, as in the full Hamiltonian calculation, the gauche energy is lower than the trans.

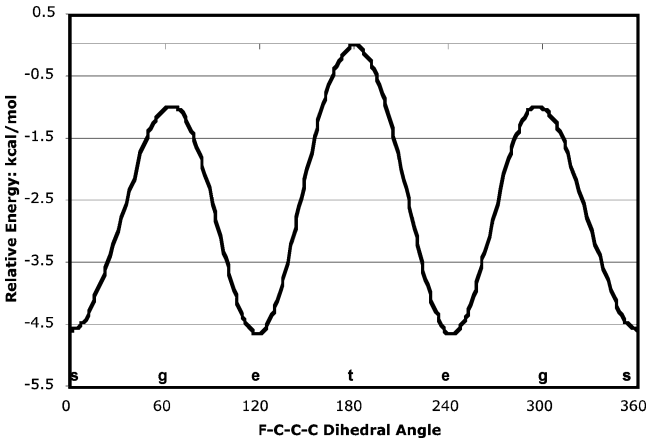


Figure 6. Torsional potential energy curve for 1-fluoropropane with all vicinal hyperconjugative interactions deleted. Single-point calculations using full Hamiltonian optimized geometry.

One additional interaction category remains: remote hyperconjugations between bonds, fluorine lone pairs (lp), and distant antibonds such as $C_4-H_{10}/C-F^*$ and $F(lp)/C_3-H_8^*$. The deletion of all 36 possible remote interactions has only small effects on the conformer energies and does not change the preferred conformation. It remains gauche (Table 4). Thus, systematic deletion calculations establish that only the removal of vicinal hyperconjugative interactions leads to potential-curve inversion. The surprising result of vicinal deletion, that gauche energy is lower than that of trans, is discussed further in Section VI.

The vicinal interactions can be further parsed into anti ($C_3-H_7/C-F^*$) and cis (e.g., $C_3-H_8/C-F^*$) orbital interactions. Their relative importance can be ascertained by separate deletions, retaining all other vicinal interactions. For the gauche conformer, cis interaction deletion raises the energy by only 0.81 kcal/mol, compared to 5.63 kcal/mol for anti deletion. The anti interaction does not occur in the trans conformer; both interactions are now cis. Deletion of the two possible cis interactions in this conformer raises the energy by 1.32 kcal/mol, the magnitude of each cis interaction remaining essentially unchanged from its value in

Table 5. Deletion Energies Generated by Rigid Rotation from trans Conformer (kcal/mol)

hyperconjugative deletion	T conformer deletion energy change[a]	G−T deletion difference	E−T deletion difference	S−T deletion difference
all hyperconjugaton	117.4	0.45	−1.44	1.98
all vicinal	116.0	−1.12	−2.62	−1.53
all geminal	17.6	−0.58	2.66	4.50

[a] *E*(after deletion) − *E*(total).

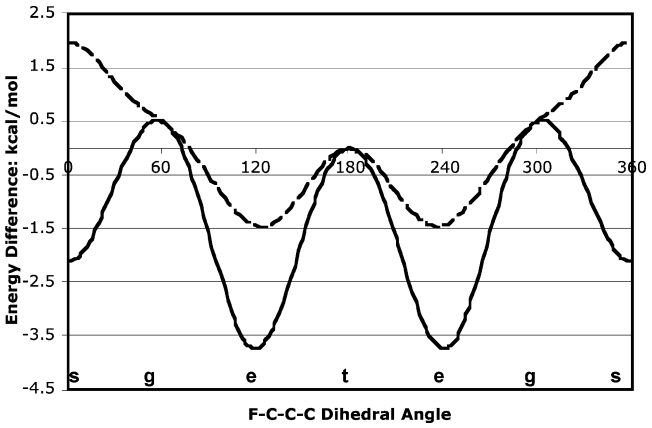


Figure 7. Potential-energy curves with all hyperconjugative interactions deleted. Optimized geometry (solid curve); frozen at trans skeletal geometry (dashed curve). Both curves have the zero-energy reference point at the 180° trans geometry.

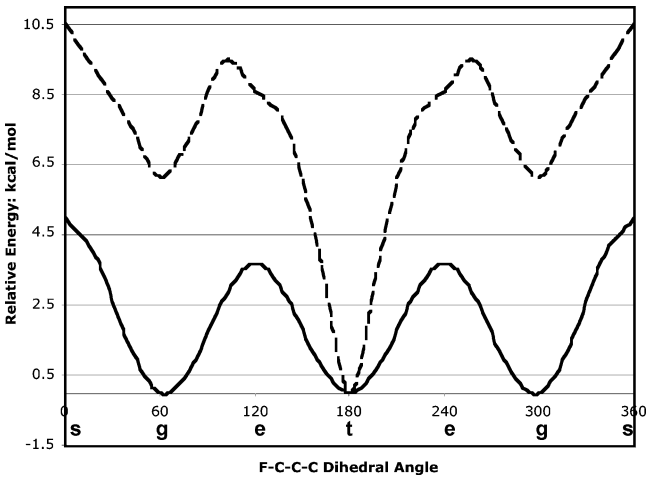


Figure 8. Energy curves for 1-fluoropropane. Full Hamiltonian potential curve (solid line); steric exchange repulsion single-point calculations (dashed line). Both curves utilize HF/6-311++G(3df,2p) wave functions and have the zero-energy reference point at the 180° trans geometry.

gauche. The clear conclusion is that the anti interaction is much greater (nearly 7 times greater) than that of the cis. This disparity is much larger than in 1,2-difluoroethane, where an account of both types of interactions is necessary to understand the gauche−trans potential surface.[2] The overarching conclusion is that the anti orbital hyperconjugative interaction provides the backbone for the gauche effect in FP.

Because of the importance of skeletal relaxation to the T−G stabilization energy (see Section III), it is instructive to obtain the hyperconjugation deletion energies for rigid rotation, thereby eliminating changes in skeletal bond lengths and angles. The results of these calculations referenced to the frozen trans optimized geometry are given in Table 5. The same general pattern for deletions with and without relaxation (i.e., vicinal interactions are more important than geminal ones, as expected from local symmetry considerations) and the inversion of trans and gauche conformer energetics by extinguishing hyperconjugation interactions are obtained. The clear outcome: the determinant controlling the G and T orbital interactions is the rotation itself, even though the torsional coordinate is contaminated by skeletal displacements.

However, Figure 7 shows that the landscape for the rigid rotation potential curve with all hyperconjugated interactions quenched is considerably different from the fully relaxed equivalent (also shown in the same figure); the pronounced gauche geometry maximum in the fully relaxed curve disappears in the rigid rotation curve. Most significant is that the disparities in the two curves largely relate to the eclipsed and syn conformers, where the effect of excluding hyperconjugation interactions in the rigid rotation case reverses the S−T energy difference found for relaxed rotation. Thus, the conclusion that the G and T orbital interaction energies are primarily determined by torsional rotation is diluted for the E and, even more so, the S interaction energies.

V. Steric Exchange Repulsion Model

Steric exchange repulsion is classically defined as the effect of Pauli exchange repulsion, a short-range force, which spatially separates pairs of electrons. The conceptual foundation of NBO steric analysis can be found in several publications by Weinhold[9−11] showing that quantitative appreciation of total steric exchange repulsion for a given molecular geometry can be obtained from the energy difference between orthogonal natural localized molecular orbital (NLMOs) and nonorthogonal preorthogonal natural localized molecular orbital (PNLMOs) descriptions of the molecular wave function. The PNLMOs and NLMOs are closely related to preNBOs (PNBOs) and NBOs with somewhat different localizations designed to improve exchange repulsion quantifications.[11]

The dashed curve in Figure 8 represents the calculated HF/6-311++G(3df,2p) total steric exchange repulsion at 20° intervals using the NBO steric analysis, described above. (A B3LYP repulsion plot is almost identical to the HF one; however, the HF curve is conceptually better founded.) A striking feature of the repulsion plot is the decreased repulsion at the trans geometry. There is also a much less pronounced decrease near the gauche geometry. These features allow the conclusion that lowered steric exchange

Table 6. Rydberg Deletion Effects on Conformer Energies (kcal/mol)

hyperconjugative deletion	T conformer deletion energy change[a]	G−T deletion difference	E−T deletion difference	S−T deletion difference
all Rydberg[b]	49.6	−0.38	0.95	2.09
only vicinal valence[c]	61.8	0.65	−2.03	−0.55

[a] *E*(after deletion) − *E*(total). [b] Valence interactions remain. [c] Only vicinal Rydberg interactions remain.

repulsion plays a major if not determining role in stabilizing the metastable 180° trans conformer. Moreover, the maxima at the syn and eclipsed FCCC dihedral angles strongly suggest that the syn and eclipsed geometry maxima in the potential curve (solid curve in Figure 8) arise at least in part from steric repulsion.

As seen from the comparison of repulsion values at the gauche and trans dihedral angles in Figure 8, the effect of shutting off steric exchange repulsion would be to greatly increase the gauche−trans preference, a consequence of greater exchange repulsion destabilization of the gauche conformer. The clear conclusion is that exchange repulsion does not play a significant role in forming the FP gauche conformational preference over the trans. However, it plays an important role in stabilizing the gauche conformer relative to the eclipsed.

VI. Discussion

An interesting conundrum that arises from the systematic deletion calculations pointed out in Section IV is that, even with all-vicinal interactions deleted, the energy of the gauche conformer remains lower than trans—opposite to that found for complete hyperconjugation removal. The well-respected Brunck−Weinhold[3] rules for gauche conformer preference link vicinal hyperconjugative interactions to gauche preference, and their deletion is predicted to change the preference to trans. To address this problem, we separate the vicinal interactions involving Rydberg orbitals from those involving valence orbitals by restoring the Rydberg interactions into the all-vicinal interactions deletion calculation. The outcome of this new calculation, now deleting only vicinal *valence* interactions, is that the gauche conformer becomes less energetically favorable than the trans (Table 6). If the complete hyperconjugation deletion removal ("no-star") calculation is modified so that only Rydberg orbitals are deleted (i.e., interactions involving the antibond valence orbitals are restored), the gauche conformer becomes more energetically favorable than the trans (Table 6).

In a sense, these calculations are illegal because Rydberg orbitals are brought into the NBO scheme to ensure orbital orthogonality; thus, vicinal deletion excluding Rydberg orbitals is equivalent to deleting nonorthogonal PNBO vicinal antibonds.[10] Deletion of only the Rydberg orbitals has a similar flavor since it is equivalent to a full PNBO deletion calculation. Since a tenet of NBO analysis is that the extra valence natural atomic orbitals in the Rydberg basis play practically no significant role in the theory,[10] these energy inversions provide a challenge. One of the most dramatic and characteristic simplifying features of NBO analysis is that they can effectively be ignored.[10]

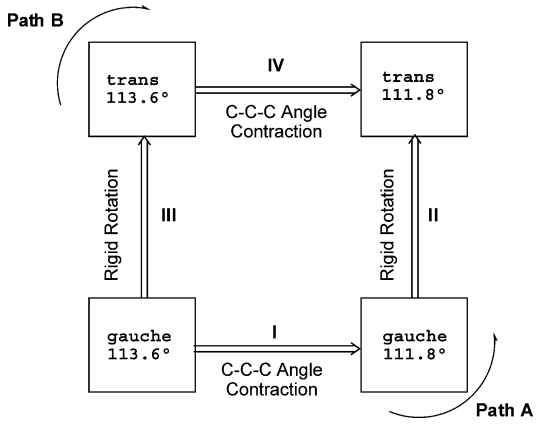


Figure 9. Stepwise analysis of $C_2C_3C_4$ angular contraction energetics for the gauche equilibrium conformation. Step I: angular contraction leading to a prepared state with the trans-$C_2C_3C_4$ angle. Step II: rigid rotation from the prepared state leading to partially relaxed *trans*-1-fluoropropane. Step III: rigid rotation of the equilibrium gauche conformer leading to *trans*-1-fluoropropane with gauche skeletal geometry. Step IV: $C_2C_3C_4$ angular contraction of the prepared trans state leading to partially relaxed *trans*-1-fluoropropane.

The relaxation studies described in Section III demonstrate that the gauche preference is linked to the CCC angle in the propane skeleton of FP, evidenced by expansion of the 111.8° CCC angle in the trans conformer to 113.6° in the gauche (Table 2). A facile explanation for the CCC angle opening in the gauche conformer is that it reduces increased repulsion from the fluorine lone pairs to the C_4−H bonds. Although we show in Sections IV and V that the stereoelectronic origin of the gauche preference comes from hyperconjugative attractive electronic effects, which depend primarily on the FCCC dihedral angle, it remains to obtain insight into the underlying factors affecting the CCC angle as part of the dihedral angle rotation process.

To do this, we focus on separate considerations of Pauli exchange, Lewis energy,[12] and hyperconjugation for two alternate CCC angle-contracting paths for FCCC dihedral rotation from the gauche conformer, shown in Figure 9. One path (path A) first contracts the CCC angle in the optimized gauche conformer (113.6°) to its calculated value in the trans (111.8°), step I. There is no dihedral rotation in step I; that is, the 111.8 °CCC angle (with all other bond lengths and angles frozen to their values in the gauche equilibrium conformer) represents a prepared state of the gauche conformer. Step II is a dihedral rotation step from the prepared gauche state to a partially relaxed trans geometry. Path B starts with frozen dihedral rotation (step III) from the gauche conformer (all geometrical parameters remain those of the gauche conformer), ending in a 113.6° ∠CCC prepared state of the trans conformer. Step IV contracts the

Table 7. Energy Contributions for 1-Fluoropropane Rigid Rotation/Angular Relaxation (kcal/mol)[a]

step	ΔE_{exch}	ΔE_{deloc}	ΔE_{Lewis}	ΔE_{total}
I	1.68	−0.14	0.23	0.089
II	−2.94	−0.48	0.55	0.074
III	−2.07	−0.49	0.73	0.24
IV	0.81	−0.12	0.04	−0.079

[a] ΔE_{exch}, ΔE_{deloc}, and ΔE_{Lewis} refer to exchange repulsion, delocalization, and Lewis energy changes, respectively, for steps I−IV in Figure 9. ΔE_{exch} is calculated as the energy difference between occupied (orthogonal) NLMOs and the associated preorthogonal (Pauli Principle violating) PNLMOs; ΔE_{deloc} represents the energy change attributed to electron transfer from nearly doubly occupied (bonding) orbitals to nearly vacant (antibonding) orbitals; ΔE_{Lewis} refers to the energy of NBOs with ≅2 occupancy and can be associated to the localized electron pairs of the Lewis structure.

CCC angle in the trans prepared state to its value in the fully optimized trans conformer, with all other geometrical parameters remaining frozen. The associated energy changes accompanying paths A and B are shown in Table 7.

An analysis of the total energy changes in Table 7 shows an increase on going to the prepared state of step I and a decrease for step IV, as expected. Our first focus is on the exchange repulsion changes: ΔE_{exch} strongly increases (1.7 kcal/mol) for the CCC angle contraction occurring in step I but shows a much smaller increase (0.8 kcal/mol) for the same angular contraction in step IV. Thus, although exchange repulsion increases upon CCC angular contraction in both the equilibrium gauche and prepared trans states, it increases much less in the prepared trans one. We conclude that the angular expansion that accompanies the T → G rotational process does relieve the increased exchange repulsion occurring in the gauche conformation.

Different behavior is found for ΔE_{deloc}, the delocalization energy change. It is slightly stabilizing (−0.14 kcal/mol) for step I, CCC angular contraction of the gauche conformer, and almost identically so (only −0.12 kcal/mol) in step IV, angular contraction of the prepared trans state. This identity leads to the conclusion that hyperconjugative interactions are not selectively linked to the CCC angular expansion in the gauche conformer.

Additional interest is found in the change in Lewis energy for step I. The Lewis structure of a molecule can be defined in an NBO representation by nearly doubly occupied bonding orbitals and lone pairs, without the delocalization effects of conjugation and hyperconjugation; $E_{Lewis} = E_{total} - E_{deloc}$. E_{Lewis} then is linked to changes in bond strength. The 0.23 kcal/mol increase for step I shows that the Lewis structure of the prepared gauche state is destabilized relative to the equilibrium gauche conformer. In terms of the bond strengths, the contracted angle of the prepared state represents an unfavorable geometry. Thus, in contrast to the behavior of ΔE_{exch}, there does not appear to be a link between ΔE_{Lewis} and the CCC angular contraction on going to the trans geometry.

Even though overall hyperconjugative interaction does not appear to link to the CCC angle change between the G and T conformers, there does appear to be a link to the CCC angle. This is exposed in Figure 10 by comparison of the expunged anti-C−H/C−F* interaction and full Hamiltonian

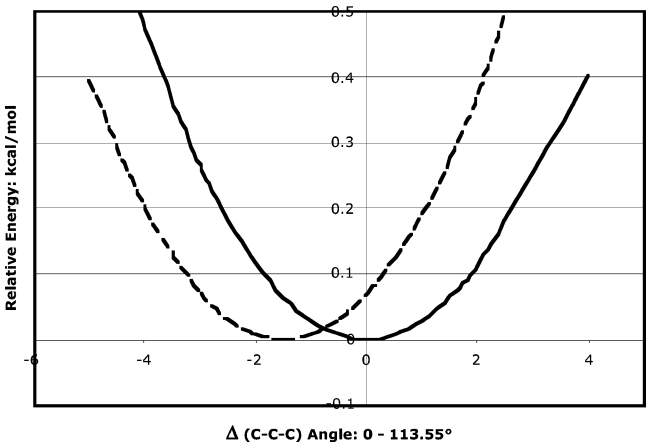


Figure 10. Potential energy curves for C−C−C angular expansion at fixed gauche geometry. Full Hamiltonian (solid curve); deletion of anti-C−H/C−F* interactions (dashed curve).

energy curves versus the CCC angle for the gauche conformer. The point of this comparison is that the potential minima represent the change in gauche conformer geometry when only anti-C−H/C−F* hyperconjugation is absent, thus, giving the needed chemically insightful smoking gun otherwise buried in the full hyperconjugation deletions. The shift in minima demonstrate that the CCC angle is decreased by ∼1.7° when the anti-C−H/C−F* interaction is removed. This provides the needed insight into the 1.8° opening of the CCC angle on going from the trans conformer (where there are no anti-C−H/C−F* orbital interactions) to the gauche one. The effect of this interaction in the gauche conformer is to increase the CCC angle close to the global optimization increase.

Our overall conclusion is that both increased exchange repulsion and increased delocalization accompany CCC angle expansion in the T → G reaction path, with consequent preferential stabilization of the more-open angle gauche conformer. It is difficult to reliably assess which of these two interactions is primary, but on the basis of Figure 10, the attractive factor appears to be more important.

VII. Conclusions

Because FP is an example of an elementary molecule exhibiting a gauche effect (without the complication of dipole−dipole repulsion), it is of considerable importance to study it thoroughly. The DFT/ab initio calculations on 1-fluoropropane conformational preference and torsional potential surface reported here and NBO analysis all provide insight into the conformational preference machinery and provide thought-provoking results on the role of Rydberg orbitals in NBO analysis. We emphasize that the approach taken in this study is the systematic deletion of interaction types (lone-pair/σ^* hyperconjugations, steric repulsion, etc.) and not the comparison of calculated relative magnitudes of these interactions. In this sense, the deletion-caused inversions of the potential-curve landscape represent a kind of truth table and avoid quantitative pitfalls of NBO analysis.[13]

Our deletion calculations provide convincing validation for the anti-C−H/C−F* hyperconjugative interaction providing the backbone for the gauche effect in FP. The cis orbital

interaction is found to be much less important than in 1,2-difluoroethane. Thus, cis interactions, unlike in 1,2-difluoroethane, are not linked to the FP equilibrium conformer 63° FCCC dihedral angle.

Geometry optimization predicts that a ~2° expansion of the CCC angle is the major important skeletal change in the T → G reaction coordinate. A detailed analysis shows that both increased hyperconjugative attraction and decreased exchange repulsion at the larger angle account for the angle opening.

Finally, the results of Rydberg orbital deletion calculations, while not altering any of our overall conclusions concerning stereoelectronic interactions in FP, do not support the supposition that Rydberg interactions can be ignored in NBO analysis. We intend to pursue this aspect of Rydberg orbital participation in a future study.

Acknowledgment. We thank Vojislava Pophristic, Paul R. Rablen, Frank Weinhold, and Kenneth B. Wiberg for many helpful comments on the manuscript.

References

(1) Wolfe, S. *Acc. Chem. Res.* **1972**, *5*, 102−111.

(2) Goodman, L.; Gu, H.; Pophristic, V. *J. Phys. Chem. A* **2005**, *109*, 1223−1229 and references therein.

(3) Brunck, T. K.; Weinhold, F. *J. Am. Chem. Soc.* **1979**, *101*, 1700−1709.

(4) Rablen, P. R.; Hoffmann, R. W.; Hrovat, D. A.; Borden, W. T. *J. Chem. Soc., Perkin Trans. 2* **1999**, 1719−1726 and references therein.

(5) Weinhold, F. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, U. K., 1998; pp 1792−1811.

(6) Trindle, C.; Crum, P.; Douglass, K. *J. Phys. Chem. A* **2003**, *107*, 6236−6242.

(7) (a) Hayashi, M.; Fujitake, M. *J. Mol. Struct.* **1986**, *146*, 9−24. (b) Guirgis, G. A.; Zhu, X.; Durig, J. R. *Struct. Chem.* **1999**, *10*, 445−461.

(8) *Gaussian 98*, revision A.11.2; 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., Rega, N., 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. W., Gill, P. M., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S., Pople, J. A., Eds.; Gaussian, Inc.: Pittsburgh, PA, 2001.

(9) (a) Badenhoop, J. K.; Weinhold, F. *J. Chem. Phys.* **1997**, *107*, 5406−5421. (b) Badenhoop, J. K.; Weinhold, F. *J. Chem. Phys.* **1997**, *107*, 5422−5432.

(10) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO 5.0*; Theoretical Chemical Institute, University of Wisconsin: Madison, WI, 2002.

(11) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736−1740.

(12) Suidan, L.; Badenhoop, J. K.; Glendening, E. D.; Weinhold, F. *J. Chem. Educ.* **1995**, *72*, 583−586.

(13) Salzner, U.; Schleyer, P. v. R. *Chem. Phys. Lett.* **1992**, *190*, 401−406.

CT050204B

JCTC Journal of Chemical Theory and Computation

Slater's Exchange Parameters α for Analytic and Variational Xα Calculations

Rajendra R. Zope*

Department of Chemistry, George Washington University, Washington, D.C., 20052

Brett I. Dunlap

Code 6189, Theoretical Chemistry Section, U.S. Naval Research Laboratory, Washington, D.C. 20375

Received July 7, 2005

Abstract: Recently, we formulated a fully analytical and variational implementation of a subset of density functional theory using Gaussian basis sets to express orbital and the one-body effective potential. The implementation, called the Slater-Roothaan (SR) method, is an extension of Slater's Xα method, which allows arbitrary scaling of the exchange potential around each type of atom in a heteroatomic system. The scaling parameter is Slater's exchange parameter, α, which can be determined for each type of atom by choosing various criteria depending on the nature of problem undertaken. Here, we determine these scaling parameters for atoms H through Cl by constraining some physical quantity obtained from the self-consistent solution of the SR method to be equal to its *exact* value. Thus, the sets of α values that reproduce the *exact* atomic energies have been determined for four different combinations of basis sets. A similar set of α values that is independent of a basis set is obtained from numerical calculations. These sets of α parameters are subsequently used in the SR method to compute atomization energies of the G2 set of molecules. The mean absolute error in atomization energies is about 17 kcal/mol and is smaller than that of the Hartree−Fock theory (78 kcal/mol) and the local density approximation (40 kcal/mol) but larger than that of a typical generalized gradient approximation (∼8 kcal/mol). A second set of α values is determined by matching the highest occupied eigenvalue of the SR method to the negative of the first ionization potential. Finally, the possibility of obtaining α values from the exact atomization energy of homonuclear diatomic molecules is explored. We find that the molecular α values show much larger deviation than what is observed for the atomic α values. The α values obtained for atoms in combination with an analytic SR method allow elemental properties to be extrapolated to heterogeneous molecules. In general, the sets of different α values might be useful for calculations of different properties using the analytic and variational SR method.

The Hohenberg−Kohn−Sham (HKS) formulation of density functional theory is by far the choice for today's electronic structure calculation.[1−4] Prior to the HKS formulation of the density functional theory (DFT), Slater formulated its basis in an attempt to reduce the computational complexity of the Hartree−Fock (HF) method.[3] Unlike the HF method, this Xα method has an exchange potential v_x that is local and proportional to the one-third power of the electron density ρ. It is given by

$$v_x[\rho] = -\alpha\frac{3}{2}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3}(\vec{r})$$

* Corresponding author e-mail: rzope@alchemy.nrl.navy.mil. Corresponding author address: Theoretical Chemistry Section, Naval Research Laboratory, Washington, DC 20375-5345.

10.1021/ct050166w CCC: $30.25
Published on Web 10/04/2005

where α is called Slater's exchange parameter and was originally equal to one. Gáspár[6] and Kohn−Sham[2] obtained a similar expression on other rigorous grounds. They variationally minimized the total energy functional and determined α to be 2/3, in contrast to Slater's value. The difference in the two values is rooted in the averaging process employed in the simplification of the HF exchange potential. Slater, taking a cue from Dirac's earlier work,[5] averaged the exchange potential over the entire Fermi sphere of radius $k_f = (3\pi^2\rho(\vec{r}))^{1/3}$, while in the Gáspár and Kohn−Sham procedure the value of the exchange potential at $k = k_f(\vec{r})$ is used. It was suggested that the α parameter in Slater's method could be treated as an adjustable parameter.[7] The set of α parameters for atoms that make self-consistent Xα energy match the HF energy was determined by Schwarz.[8] The α values he obtained typically range from 0.77 for light atoms to 0.69 for heavy atoms. He also noted that different atomic configurations lead to only slight changes in the α values, suggesting that the Xα method can also be applied to the molecules or solids. Several other ways to determine the α parameters have been put forth.[9−16]

Early implementations of the Xα method for molecules with atom-dependent α parameters employed the muffin tin (MT) approximation.[17−21] In this scheme, atoms or ions are enclosed by the atomic spheres in which the potential is approximated to be spherically symmetric, while, in the interstitial region, the potential is constant. The effective one body potential in the MT implementation is discontinuous at the surface of the MT sphere, and therefore the energy was mathematically undefined. This method, however, had an advantage over all the other quantum-chemical methods which was that the molecules dissociated correctly. At infinite interatomic distances a sum of atomic energies could be reproduced. When the MT model was discarded due to difficulties in geometry optimization, a single value of α (= 0.7) had to be chosen for all types of atoms.[22] There are, however, some attempts to obtain good thermochemistry from the Xα calculations by correcting or improving the uniform α calculations in a secondary calculation.[23−26]

Electronic structure calculations using traditional quantum chemical methods such as the HF theory, and beyond, generally employ basis sets to expand molecular orbitals. In these methods, calculations of matrix elements and other quantities of interests are analytic. On the other hand, almost all implementations of density functional models, until recently, required use of numerical grids to compute contributions to matrix elements from the exchange-correlation terms. Cook and co-workers[27−31] successfully demonstrated fully analytic implementation of a density functional model. His variational implementation used Gaussian basis sets to express molecular orbitals, the one-body effective potential,[27] and the uniform α Xα exchange potential.[28,29] The advantages of analytic calculations are obvious. The calculations are fast and accurate to machine precision (within the limitations of the model and basis set). Round-off error, which grows as the square root of the number of points and other numerical problems are eliminated.[32] Consequently, smooth potential energy surfaces are obtained.[28,29] Unfortunately, modern sophisticated exchange-correlation functionals are too complex to allow a fully analytic solution at this time. Analytic DFT, at this stage, is restricted to Slater's exchange-type functionals. We have recently proposed an algorithm that permits fully analytic solutions and also allows for atom dependent values of Slater's exchange parameter α for heteroatomic systems.[33] This method, called the Slater-Roothaan (SR) method, is based on robust and variational fitting and requires four sets of Gaussian bases. It can have the advantage of MT Xα that molecules dissociate correctly as interatomic distances tend to infinity.[34] Intuitively, it is natural to expect that when molecules dissociate the constituent atoms would have *exact* atomic energies at infinite separation. This important physical requirement is satisfied within the SR method if α parameters that reproduce exact atomic energies are used in molecular calculations, although one might want to use other sets of α's for other molecular properties.

In this work, we report different sets of α parameters that could be used in the SR method. The first such a set is determined by requiring that the self-consistent atomic SR atomic energy be equal to the *exact* atomic energies. The α values in this set will be hereafter referred to as atomic alpha values. We have recently used some of the α values from this set to calculate the *total* energies for the G2 set[35] of molecules and found that these α values give remarkably good total energies.[36] Here, we use them to compute the *atomization* energies for the G2 set of molecules and compare them with other models. Further, we also obtain two additional sets of α based on other criteria for possible use in the SR method. The second set of atomic α values is determined to provide the negative of the eigenvalue of the highest occupied molecular orbital (HOMO), ϵ_{HOMO}, that matches with the *exact* value of the first ionization potential determined.[37−42] The third set of α is determined to provide the *exact* atomization energy for the selected homonuclear diatomic molecules of the first and second rows of the periodic table. All calculations are repeated for four different combinations of analytic Gaussian basis sets.

It is apparent that the analytic Slater-Roothaan method is empirical in nature as it uses adjustable α values. It, however, differs from the semiempirical[43−45] methods which use a minimal basis set, or empirical[46−49] models which do not compute electronic structure, and the minimal-basis tight binding methods[50] reported in the literature. It is a variant of density functional models that is also computationally very efficient because it is fully analytic and requires *no* numerical integration. We have successfully used it for studying the heterogeneous systems such as boron and aluminum nitride nanotubes containing up to 200 atoms.[51,52] Furthermore, it also provides the flexibility of being tuned, through the α parameters, according to the need of the problem. The sets of α parameters reported in this work will be useful starting points in this regard. In the following section we outline the analytic SR method used in this work. In section II, we describe computational details which are followed by results and discussion in section III.

I. Theoretical Method

The total electronic energy in the DFT for an *N*-electron system is a functional of electronic density ρ and is given

by

$$E^{KS}[\rho] = \sum_{i}^{N} < \phi_i | f_1 | \phi_i > + E_{ee} + E_{xc}[\rho_\uparrow, \rho_\downarrow] \quad (1)$$

where the first term contains the kinetic energy operator and the nuclear attractive potential due to the M nuclei

$$f_1 = -\frac{\nabla^2}{2} - \sum_{A}^{M} \frac{Z_A}{|\vec{r} - \vec{R}_A|} \quad (2)$$

The total electron density is expressed in terms of the Kohn−Sham orbitals $\phi_{i,\sigma}$ as

$$\rho(\vec{r}) = \rho_\uparrow + \rho_\downarrow \quad (3)$$

with

$$\rho_\sigma(\vec{r}) = \sum_i n_{i,\sigma} \phi^*_{i,\sigma} \phi_{i,\sigma}(\vec{r}) \quad (4)$$

where $n_{i,\sigma}$ is the occupation number for the $\phi_{i,\sigma}$ orbital. The second term in eq 1 represents the classical Coulombic interaction energy of electrons

$$E_{ee} = \langle \rho | | \rho \rangle = \frac{1}{2} \int\int \frac{\rho(\vec{r}) \rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d^3r \, d^3r' \quad (5)$$

This energy is approximated by expressing the charge density as a fit to a set of Gaussian functions

$$\rho(\vec{r}) \approx \bar{\rho}(\vec{r}) = \sum_i d_i G_i(\vec{r}) \quad (6)$$

where $\bar{\rho}(\vec{r})$ is the fitted density and d_i is the expansion coefficient of the charge density Gaussian basis function G_i. The elimination of the first-order error in the total energy due to the fit leads to[27]

$$E_{ee} = 2\langle \bar{\rho} | | \rho \rangle - \langle \bar{\rho} | | \bar{\rho} \rangle \quad (7)$$

The expansion coefficients $\{d_i\}$ are determined by variationally minimizing this energy with respect to $\{d_i\}$. The last term E_{xc} in eq 1 is the exchange energy

$$E_{xc}[\rho_\uparrow, \rho_\downarrow] = -\frac{9}{8} \alpha \left(\frac{6}{\pi}\right)^{1/3} \int d^3r [\rho_\uparrow^{4/3}(\vec{r}) + \rho_\downarrow^{4/3}(\vec{r})] \quad (8)$$

The form of above functional allows analytic calculations with the Gaussian basis to be performed. For this purpose the one-third and two-third powers of the electron density are also expanded in Gaussian basis sets

$$\rho^{1/3}(\vec{r}) \approx \bar{\rho}^{1/3} = \sum_i e_i E_i(\vec{r}) \quad (9)$$

$$\rho^{2/3}(\vec{r}) \approx \bar{\rho}^{2/3} = \sum_i f_i F_i(\vec{r}) \quad (10)$$

Here, $\{E_i\}$ and $\{F_i\}$ are independent Gaussian basis functions, while e_i and f_i are expansion coefficients. The exchange energy is then given by[28−31,53]

$$E_{xc} = C_\alpha \left[\frac{4}{3} \langle \rho \bar{\rho}^{1/3} \rangle - \frac{2}{3} \langle \bar{\rho}^{1/3} \, \bar{\rho}^{1/3} \, \bar{\rho}^{2/3} \rangle + \frac{1}{3} \langle \bar{\rho}^{2/3} \, \bar{\rho}^{2/3} \rangle \right] \quad (11)$$

where $C_\alpha = -(9/8)\alpha(6/\pi)^{1/3}$. Thus, using the four LCGO basis sets (one for orbital expansion and three fitting basis sets) the total energy is calculated analytically. The LCAO orbital coefficients and the vectors **d**, **e**, and **f** are found by (constrained if desired) variation.

Slater-Roothaan Method. The expression for the total electronic energy in the Slater-Roothaan method has the following form:[33]

$$E^{SR} = \sum_i < \phi_i | f_1 | \phi_i > + 2\langle \rho | | \bar{\rho} \rangle - \langle \bar{\rho} | | \bar{\rho} \rangle + \sum_{\sigma=\uparrow,\downarrow} \left[\frac{4}{3} \langle g_\sigma \, \bar{g}_\sigma^{1/3} \rangle - \frac{2}{3} \langle \bar{g}^{1/3}{}_\sigma \, \bar{g}^{1/3}{}_\sigma \, \bar{g}^{2/3}{}_\sigma \rangle + \frac{1}{3} \langle \bar{g}^{2/3}{}_\sigma \, \bar{g}^{2/3}{}_\sigma \rangle \right] \quad (12)$$

Here, the partitioned 3/4 power of the exchange energy density

$$g_\sigma(\vec{r}) = C_x \sum_{ij} \alpha(i) \, \alpha(j) \, D^\sigma_{ij}(\vec{r}) \quad (13)$$

where $C_x = C_\alpha/\alpha$ and $D^\sigma_{ij}(\vec{r})$ is the diagonal part of the spin density matrix, and the function

$$\alpha(i) = \alpha_i^{3/8} \quad (14)$$

contains the α_i, the α in the Xα, for the atom on which the atomic orbital i is centered.

II. Computational Details

The α values are calculated for the atoms that occur in a standard set of (Becke's)[54] 56 molecules. These atoms are H, Be through F, and Na through Cl. Magnesium and aluminum atoms do not occur in the Becke's set but have been included in the present work. The set of 56 molecules actually consist of 54 molecules with two molecules in two different electronic states. The SR method requires four basis sets, of which one is the orbital basis. The remaining three basis sets are required for fitting of the Kohn−Sham potential (see ref 33 for more details). For the orbital basis sets, our choices are the triple-ζ (TZ) 6-311G** basis[55,56] and the DGauss[57] valence double-ζ (DZ) basis set[58] called DZVP2. The most reliable and thus best *s*-type fitting bases are those that are scaled from the *s*-part of the orbital basis.[27] The scaling factors are two for the density, (2/3) for $\bar{\rho}^{1/3}$ and (4/3) for $\bar{\rho}^{2/3}$. These scaled bases are used for all *s*-type fitting bases. A complete package of basis sets has been optimized[58] for use with DGauss.[57] In addition to the valence double-ζ orbital basis, called DGDZVP2 herein, we use the *pd* part of the (4,3;4,3) for Be−F (A2) charge density fitting basis. Ahlrichs' group has generated a RI-J basis for fitting the charge density of a valence triple-ζ orbital basis set used in the TURBOMOLE program.[59] These fitting bases are used in combination with the 6-311G** and DGDZVP2 basis sets. The α are obtained for the different combination of above basis sets. The molecule geometries were optimized using the Broyden−Fletcher−Goldfarb−Shanno (BFGS) algorithm.[60−64] Forces on ions are rapidly computed nonrecursively using the 4-j generalized Gaunt coefficients.[65,66] The calculations are spin-polarized for open-shell systems.

III. Results and Discussion

The optimal α values that give the exact atomic energies can be obtained by the Newton−Raphson procedure that

Table 1. Optimal α Values for the Different Basis Sets that Yield the *Exact* Atomic Energies in the Highest Symmetry for Which the Solutions Have Integral Occupation Numbers[a]

	basis I	basis II	basis III	basis IV	numerical	*E* (au)
H	0.77739	0.77739	0.78124	0.78124	0.77679	−0.500
Li	0.79169	0.79169	0.79211	0.79211	0.79118	−7.478
Be	0.79574	0.79574	0.79614	0.79614	0.79526	−14.667
B	0.78675	0.78668	0.78684	0.78677	0.78744	−24.654
C	0.77677	0.77672	0.77670	0.77665	0.77657	−37.845
N	0.76747	0.76747	0.76726	0.76726	0.76654	−54.590
O	0.76500	0.76495	0.76454	0.76448	0.76454	−75.067
F	0.76066	0.76067	0.76002	0.76001	0.75954	−99.731
Na	0.75204	0.75204	0.75287	0.75287	0.75110	−162.260
Mg	0.74994	0.74994	0.75120	0.75120	0.74942	−200.060
Al	0.74822	0.74819	0.74872	0.74869	0.74797	−242.370
Si	0.74539	0.74540	0.74600	0.74602	0.74521	−289.370
P	0.74324	0.74324	0.74397	0.74397	0.74309	−341.270
S	0.74262	0.74260	0.74352	0.74350	0.74270	−398.140
Cl	0.74197	0.74196	0.74273	0.74272	0.74183	−460.200

[a] The numerical values are for the spherically symmetric atoms and are obtained by the numerical atomic structure code. The *exact* atomic energies given in the last column are from ref 67. The basis sets are I: 6-311G**/RI-J, II: 6-311G**/A2, III: DGDZVP2/RI-J, and IV: DGDZVP2/A2.

finds zeros of the function $f(\alpha) = E^{SR}(\alpha) - E_{exact} = 0$, where E and E_{exact} are the self-consistent Xα (SR) and the *exact* total energies, respectively. Finding zero of this function requires frequent calculation of this function and its derivative. Determination of the derivative, $f'(\alpha)$, is straightforward. Only the exchange term in the energy functional $E^{SR}(\alpha)$ depends explicitly on the α parameter. The derivative obtained as the exchange-energy divided by the α parameter provides sufficiently accurate approximation to the actual derivative. The LCAO fits depend weakly on α. The Newton−Raphson procedure was implemented using PERL scripts, and the energy and its derivative with respect to α were obtained from the FORTRAN90 SR code. The sets of α values obtained using this procedure are given in Table 1. It is evident from the table that the choice of fitting basis does not significantly affect the α values. The α values do show some dependence on the orbital basis set. However, the changes in alpha values are small with the α values for the 6-311G** basis set being consistently smaller than those for the DGDZVP2 basis set for the same fitting basis set. For the same choice of fitting basis sets, the 6-311G** orbital basis set will give lower energy than the DGDZVP2 basis. Therefore, the α values for the DGDZVP2 basis should be larger than the 6-311G** basis to provide the more negative (or binding) potential required to give the energy equal to the *exact* energy. Except for H, which has no correlation energy, these α values are larger than the reported HF values[8] which is expected as these values are obtained for the exact energy (bounded from above by the HF energy). The HF α values show systematic monotonic decrease with the atomic number.[19] The current α values show overall decrease with the atomic number except for peaks at Li and Be. The peak at Li is caused by the fact that H has no correlation energy, while the peak for Be is caused by electron pairing and correlation in the outer orbital.

Table 2. Optimal α Values for the Different Basis Sets for Which the Eigenvalue of the Highest Occupied Molecular Orbital Is Exactly Equal to the *Exact* Ionization Potential (IP)[a]

	basis I	basis II	basis III	basis IV	IP[68]
H	1.1901	1.1901	1.1877	1.1877	13.60
Li	1.1246	1.1246	1.1152	1.1152	5.39
Be	1.2749	1.2749	1.2581	1.2581	9.32
B	1.1227	1.1220	1.0993	1.0989	8.30
C	1.1006	1.1000	1.0775	1.0770	11.26
N	1.0832	1.0832	1.0629	1.0629	14.52
O	1.2360	1.2345	1.1958	1.1946	13.61
F	1.1696	1.1694	1.1501	1.1498	17.42
Na	1.1181	1.1181	1.1221	1.1221	5.14
Mg	1.2256	1.2256	1.2311	1.2311	7.64
Al	1.1113	1.1112	1.1074	1.1067	5.98
Si	1.0864	1.0836	1.0875	1.0844	8.15
P	1.0501	1.0501	1.0662	1.0662	10.49
S	1.1543	1.1562	1.1759	1.1770	10.35
Cl	1.1211	1.1218	1.1344	1.1347	12.96

[a] The experimental values of the first ionization potentials (in eV) are also included in the last column. The basis sets are I: 6-311G**/RI-J, II: 6-311G**/A2, III: DGDZVP2/RI-J, and IV: DGDZVP2/A2.

The atomic α values can also be determined by different criteria. In the KS DFT it has been argued that the negative of the eigenvalue of the highest occupied orbital $-\epsilon_{max}$ equals the first ionization potential.[37−42] Although such an interpretation of the $-\epsilon_{max}$ is not yet settled, it gives us another way to determine set of α parameters. Such a such set of α's might be useful in future for calculations of polarizability or optical spectra. These α values are given in Table 2 for the four combinations of basis set. All the values presented are larger than those obtained from the total energy matching criteria. This is not surprising as the $-\epsilon_{max}$ for the local density type ($\alpha = 2/3$) functionals typically underestimate the ionization potential by as much as 30−50%. It is known that this occurs due to the incorrect asymptotic behavior of the effective potential. The asymptotic behavior of the effective potential is governed by the exchange potential. In the present Xα method, the exchange potential decays exponentially in the asymptotic region, due to which the valence electrons experience shallower long-range potential than they otherwise should. In the present work, we, however, treat α as purely a fitting parameter to get the right ϵ_{max}. These α values do not decrease left to right across the periodic table and are rather close to Slater's value of α. The removal of self-interaction of an electron also leads to a better asymptotic description.[69] We think that the self-interaction correction can also be implemented in the present analytic SR method. Orbital densities are non-negative.

Finally, we explore the possibility of obtaining exact atomization energy (AE) within the present SR method, for selected diatomic molecules. For this purpose we use the Newton−Raphson procedure to obtain the zero of the following function: $f(\alpha) = E^{SR}_{mol}(\alpha) - 2E^{SR}_{atom}(\alpha) - AE = 0$. Here, E^{SR}_{mol} is the self-consistent total energy of the optimized molecules for a given α, and E^{SR}_{atom} is the self-consistent total atomic energy for a given alpha. The bond length of the molecule is optimized during each Newton−Raphson step. These calculations were performed using

Table 3. Optimal α Values for the Selected Homonuclear Dimers that Yield the *Exact* Atomization Energies for the 6-311G**/RI-J Basis Set[a]

	α	D_0 (kcal/mol)	R_e present	R_e expt[76]
H_2	1.39172	−103.5	0.59	0.74
Li_2	1.45747	−24.0	2.12	2.67
Be_2	0.56596	−2.4	2.82	2.46
C_2	0.81530	−148.7	1.21	1.24
N_2	0.88901	−225.1	1.04	1.10
O_2	0.42790	−118.0	1.36	1.21
F_2	0.29685	−36.9	1.67	1.41
Na_2	1.61278	−16.6	2.27	3.08
Si_2	0.78366	−74.0	2.10	2.24
P_2	1.43361	−116.1	1.58	1.89
S_2	0.64867	−100.7	1.96	1.89
Cl_2	0.66335	−57.2	2.06	1.99

[a] The *exact* bond lengths (in Å) are also included in the last column.

PERL scripts, and the calculated set of α values is presented in Table 3. These values show a large variation as a function of the atomic number. The general trend is that the α values for the dimers of atom on the left of the periodic table are larger than 1.0 and smaller than 1.0 for the atoms on the right. For the exact values of the atomization energies the bond lengths of these dimers show the general trend that the atoms on the left side of the periodic table are larger and those on the right are smaller. The exception are the Be_2 and N_2 dimers, perhaps due to their weak and very strong bonding, respectively. Another noteworthy observation is that these dimers with exception of Be_2 are still bound for vanishingly small value of α. This is interesting as no molecules are bound in the Thomas-Fermi model.[70] It seems the bonding therefore occurs due to the exact treatment of kinetic energy in the present Xα method. Here, it should be noted that the 6-311G** basis sets used in this work may not provide a satisfactory description for very small α values as the electrons then will experience a very shallow potential and will be rather delocalized leading thereby to the artifacts like positive eigenvalues for the outermost electrons. For this reason the α for the F_2 dimer should be used cautiously. Further, these α values when viewed relative to the Kohn−Sham's value of 2/3 provide some hints on how the corrections to the Kohn−Sham exchange should energetically contribute for an accurate determination of atomization energies. The molecules with molecular α values larger than the Kohn−Sham value of 2/3 will be underbound in the exchange only KS scheme and those with α lower will be overbound. Any universal correction to the KS exchange functional should be such that it simultaneously reduces overbinding in some molecules and underbinding in others.

The atomization energies for the set of 54 molecules in 56 of the electronic states are presented in Table 4. These are calculated by the SR (EA) method using the set of α values given in Table 1. The calculations are performed for the two sets of basis sets. These molecules in the dissociation limit will give the *exact* atomized energies. It is evident from Table 4 that the Xα underbinds some molecules, while it overbinds others. This trend is in contrast with the local density approximation and the HF theory. The former

Table 4. Atomization Energies D_0 (kcal/mol) for the 56 Set of Molecules Calculated Using the Two Sets of α Values Given in Table 1 for the Two Basis Sets that Reproduce the *Exact* Atomic Energies[a]

	basis I	basis II	exact
H_2	85.1	87.3	110.0
LiH	38.1	33.8	57.7
BeH	58.2	32.7	49.6
CH	67.0	68.2	83.7
$CH_2(^3B_1)$	195.3	197.0	189.8
$CH_2(^1A_1)$	158.1	161.4	180.5
CH_3	299.2	302.4	306.4
CH_4	407.3	411.8	419.1
NH	67.1	68.8	83.4
NH_2	157.1	161.5	181.5
NH_3	271.8	278.4	297.3
OH	98.6	100.0	106.3
H_2O	228.0	232.8	232.1
HF	144.8	148.4	140.7
Li_2	6.6	5.6	24.4
LiF	147.7	136.4	138.8
C_2H_2	422.8	414.2	405.3
C_2H_4	571.6	573.5	562.4
C_2H_6	711.7	718.1	710.7
CN	190.5	181.4	179.0
HCN	317.7	307.8	316.3
CO	283.2	270.5	259.2
HCO	307.4	301.4	278.3
H_2CO (formaldehyde)	394.4	391.2	373.4
H_3COH	521.9	527.0	511.6
N_2	215.0	203.6	228.5
N_2H_4	413.6	424.9	437.8
NO	163.2	155.3	152.9
O_2	157.5	154.0	120.4
H_2O_2	283.3	289.0	268.6
F_2	66.5	68.3	38.5
CO_2	456.5	437.9	388.9
$SiH_2(^1A_1)$	125.4	128.0	151.4
$SiH_2(^3B_1)$	121.2	122.6	130.7
SiH_3	194.7	196.3	226.7
SiH_4	281.7	284.1	321.4
PH_2	126.1	130.4	152.8
PH_3	203.4	210.0	242.0
H_2S	163.6	169.3	182.3
HCl	101.4	102.5	106.2
Na_2	4.9	5.2	16.8
Si_2	72.0	72.2	74.7
P_2	94.1	94.5	117.2
S_2	110.2	111.7	101.6
Cl_2	64.4	65.9	57.9
NaCl	88.8	90.6	97.8
SiO	199.0	198.5	191.2
CS	179.2	178.4	171.2
SO	142.6	147.8	125.1
ClO	79.5	85.6	64.3
ClF	80.5	85.7	61.4
Si_2H_6	475.4	480.3	529.5
CH_3Cl	400.5	404.8	393.6
H_3CSH	467.8	475.8	472.7
HOCl	174.5	182.1	164.3
SO_2	281.4	286.4	258.5

[a] The two basis sets are I: 6311G**/RIJ, II: DZVP2/A2. The last column contains the *exact* values.

Table 5. Mean and Mean Absolute Error (kcal/mol) in the Calculated *Atomization* Energy of 56 Molecules (Cf. Table 4) Relative to Their Experimental Values for Different Basis Sets[a]

OB	FB	SR (EA) mean	SR (EA) absolute	SR (HF) mean	SR (HF) absolute
6-311G**	RI-J	−1.9	17.3	−5.1	16.4
DGDZVP2	A2	−1.6	16.2	−4.8	16.1

[a] The SR (EA) are present calculations, while SR (HF) are SR calculation with α_{HF} values. The SR (HF) results are from ref 33. The mean absolute error for the standard Hartree−Fock theory and the local density approximation is 78 and 40 kcal/mol,[54,84] respectively. Abbreviations: OB, orbital basis; FB, fitting basis (pdfg-type).

uniformly overbinds, while the latter uniformly underbinds. A similar trend has been found when the HF values of α are used.[33] In general, the error in atomization energy is smaller when the molecules consist of atoms on the opposite sides of the periodic table, for example, HCl, H_2S, NaCl, HO, etc. The best agreement is observed for the NH_3 with the DGDZVP2 basis and OH with all basis. The mean and mean absolute errors in the atomization energies of these molecules calculated with respect to their experimental values is given in Table 5. These errors are somewhat larger than those obtained when the HF α values are used. We note here that the HF values of α, α_{HF}, are those that when used in the Xα model reproduce the HF exchange energies for atoms.

The α values obtained from the ϵ_{max} matching criteria can be used to obtain the first ionization energy of a molecule from its highest occupied eigenvalue. Here, we demonstrate their use for calculations of the first ionization potential of N_2 and CO molecules. For the N_2 molecule, using the α values from Table 2 for basis sets I−IV, the ionization potential is 15.38, 15.34, 15.40, and 15.20 eV, respectively. These values are in good agreement with the experimental value of 15.58 eV.[71] For the CO molecule, the ionization potential is 14.83, 14.83, 14.53, and 14.53 eV for the basis sets I−IV. These values are also in very good agreement with CO's experimental ionization energy (14.01 eV).[72] We expect that the use of these α values gives better eigenspectra, which could be used in the polarizability calculations by the sum over states method.[73] The atomic or Hartree−Fock α values could be used to optimize the geometry of a molecule and then to calculate the matrix elements required in the sum over states expression. The eigenvalues required in the sum over states method can be obtained by performing one more self-consistent calculation using the α values from Table 2. The use of the improved eigenspectrum in the sum over states method has been found to give good estimates of polarizability.[73] Calculations for the set of 54 molecules using molecular α values (cf. Table 3) are not performed because the SR method as implemented now[33] cannot handle a wide variation in the molecular α values. Moreover for the reasons mentioned earlier these α values (particularly for the molecules containing *F*) necessitate bigger orbital basis sets. The trend in the bond lengths in Table 3 suggests that for these dimers better values of α may be obtained by minimizing the deviation of both the dissociation energy and the bond length, from their experimental values. There are several other possibilities for determination of α which may be useful in improving the performance of the SR method. As mentioned earlier the differences in the Slater and the Gáspár−Kohn−Sham value of α are the result of a different averaging process employed in the derivation of exchange potential. This suggests the possibility that different α values for the description of the core (say α_c) and valence (α_v) electrons may provide a better description of the exchange correlations. The additional parameter in this case will provide greater flexibility in the fitting procedure. Alternatively, the Xα exchange functional can be augmented with suitable functional forms that allows the analytic solution of the problem. The local functional form by Liu and co-workers that consists of the sum of 0, 1, 1/3, and 2/3 powers of electron density and originates from the adiabatic connection formulation appears to be particularly suitable for this purpose.[74,75] We are currently exploring these possibilities.

To summarize three sets of α values are obtained on the different criteria for the use in the analytic SR method. The first set of α values is determined by equating the self-consistent total energy of atoms to the "exact" atomic energy. Two other sets are determined using different constraints such as equating the negative of the eigenvalue of the highest occupied orbital to the first ionization potential and reproduction of exact atomization energies for the diatomic molecules. The examination of the performance of the SR method for the atomization energies of 56 molecules with the first set of α values gives the mean absolute error to be about 17 kcal/mol. This value is far larger than the generally accepted chemical accuracy of 2 kcal/mol. The tabulation of the MAE for the G2 set and extended G2 set for more sophisticated functionals is given in ref 77. The MAE is within 4−9 eV for the parametrizations at the level of the generalized-gradient approximation (GGA).[78−81] The meta-GGA[82] or hybrid GGA functionals[54,83] perform even better with MAE of about 3 eV. The MAE is 37−40 kcal/mol for the local spin density approximation.[84] The SR method's performance is intermediate between the local spin density approximation and the GGA. We note that we have not optimized the method to give an accurate estimate of any particular property. The α parameters could be chosen to minimize the MAE in atomization energies such as the other density functional models, including the GGA, do. Such an optimization process would necessarily lower the error in Table 5. Our goal in this work is to examine simple schemes for extrapolating elemental properties to heterogeneous molecules, without any additional massaging of the results. For example, the set of α values from Table 1 can be used to extrapolate atomic energies to molecular energies by the SR method. Indeed, such an extrapolation gives remarkably accurate total energies (see Table 6), that are comparable to or better than those obtained by some popular, sophisticated pure and hybrid density functional models.[36] By construction, using the α values of this table, the MAE in total energies is the same as that in atomization energy. This is in contrast to many popular density functional models which seem to give better atomization energies due to cancellations of errors in total energies of atoms and molecules. Also, the SR method is unique in that molecules dissociate correctly in the separated atom limit. It is therefore not unreasonable to use α values adjusted for atoms. The present work shows

Table 6. Mean Absolute Error (MAE) in Calculated *Total* Energies of G2 Set of 56 Molecules for Different Models[a]

model	MAE
M1	17.3
M2	16.2
LDA	532
PBE-GGA	101
B3LYP-hybrid GGA	15

[a] The errors are in kcal/mol and are at optimized geometries in the respective model. M1: SR(EA)/6-311G**/RI-J, M2: SR(EA)/DGDZVP2/A2, LDA: local density approximation (see ref 36 for more details).

that fitting perhaps any molecular property can be done quantum mechanically through density functional theory. If the calculations are analytic and variational, then small basis sets can be used to generate unique, stable, and reliable energies using minimal computer time. Thus one can envision embedding quantum-mechanical calculations, with full geometry optimization where appropriate, to optimize very sophisticated quantum-mechanical calculations of molecular properties over the G2 and larger sets of molecules using PERL scripts to control the optimization on a single processor or to farm out independent suboptimizations on multiple processors. Our toolbox of functionals that can be treated analytically contains more[85] than the cube-root functional used in this work. Thus one can expect that as more experience is gained in analytic DFT we can better approximate the best exchange and correlation functionals that currently require limited-precision numerical integration. The process of driving MAEs lower and lower through better and better analytic functionals need never end, short of perfect agreement.

Acknowledgment. The Office of Naval Research, directly and through the Naval Research Laboratory, and the Department of Defense's High Performance Computing Modernization Program, through the Common High Performance Computing Software Support Initiative Project MBD-5, supported this work.

References

(1) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

(2) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

(3) Slater, J. C. *Phys. Rev.* **1951**, *81*, 385.

(4) Ratner, M. A.; Schatz, G. C. *Introduction to Quantum Mechanics in Chemistry*; Prentice-Hall: Upper Saddle River, New Jersey, 2001.

(5) Dirac, P. A. M. *Proc. Cambridge Philos. Soc.* **1930**, *26*, 376.

(6) Gáspár, R. *Acta Phys. Hung.* **1954**, *3*, 263.

(7) Slater, J. C.; Wilson, T. M.; Wood, J. H. *Phys. Rev.* **1969**, *179*, 28.

(8) Schwarz, K. *Phys. Rev. B* **1972**, *5*, 2466.

(9) McNauhton, D. J.; Smith, V. H. *Int. J. Quan. Chem.* **1970**, *3S*, 775.

(10) Smith, V. H.; Sabin, J. R. *J. Phys. B* **1978**, *11*, 385.

(11) Lindgren, I. *Phys. Lett.* **1965**, *19*, 382.

(12) Kmetko, E. A. *Phys. Rev. A* **1970**, *1*, 37.

(13) Wood, J. H. *Int. J. Quan. Chem.* **1970**, *3S*, 747. Wilson, T. M.; Wood, J. H.; Slater, J. C. *Phys. Rev. A* **1970**, *2*, 620.

(14) Berrondo, M.; Goscinski, O. *Phys. Rev.* **1969**, *184*, 10.

(15) Gopinathan, M. S.; Whitehead, M. A.; Bogdanovic, R. *Phys. Rev. A* **1976**, *14*, 1.

(16) Olevano, V.; Onida, G.; Del Sole, R. *Phys. Rev. B* **2000**, *61*, 1912.

(17) Slater, J. C.; Johnson, K. H. *Phys. Rev. B* **1972**, *5*, 844.

(18) Rösch, N.; Rhodin, T. N. *Phys. Rev. Lett.* **1974**, *32*, 1189.

(19) Connolly, J. W. D. In *Modern Theoretical Chemistry*; Segal, G. A., Ed.; Plenum: New York, 1977; Vol. 7, p 105. Danese, J. B.; Connolly, J. W. D. *J. Chem. Phys.* **1974**, *61*, 3063.

(20) Fischer, T. E.; Kelemen, S. R.; Wang, K. P.; Johnson, K. H. *Phys. Rev. B* **1979**, *20*, 3124.

(21) Messmer, R. P.; Lamson, S. H.; Salahub, D. R. *Phys. Rev. B* **1982**, *25*, 3576.

(22) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41.

(23) Fliszár, S.; Desmarais, N.; Comeau, M. *J. Mol. Struct. (THEOCHEM)* **1991**, *251*, 83.

(24) Vauthier, E.; Blain, M.; Odiot, S.; Barone, V.; Comeau, M.; Fliszár, S. *J. Mol. Struct. (THEOCHEM)* **1995**, *340*, 63.

(25) Vauthier, E.; Odiot, S.; Blain, M.; Fliszár, S. *J. Mol. Struct. (THEOCHEM)* **1998**, *423*, 195.

(26) Vauthier, E. C.; Cossé-Barbi, A.; Blain, M.; Fliszár, S. *J. Mol. Struct. (THEOCHEM)* **1999**, *492*, 113.

(27) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 3396; 4993.

(28) Werpetinski, K. S.; Cook, M. *Phys. Rev. A* **1995**, *52*, 3397.

(29) Werpetinski, K. S.; Cook, M. *J. Chem. Phys.* **1997**, *106*, 7124.

(30) Dunlap, B. I.; Cook, M. *Int. J. Quantum Chem.* **1986**, *29*, 767.

(31) Dunlap, B. I.; Rösch, N. *J. Chim. Phys.* **1989**, *86*, 671.

(32) Boys, S. F. *Proc. R. Soc. London, Ser. A* **1950**, *200*, 542.

(33) Dunlap, B. I. *J. Phys. Chem.* **2003**, *107*, 10082.

(34) Johnson, K. H. *Annu. Rev. Phys. Chem.* **1975**, *26*, 39.

(35) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622.

(36) Zope, R. R.; Dunlap, B. I. *Phys. Rev. B* **2005**, *71*, 193104.

(37) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L., Jr. *Phys. Rev. Lett.* **1982**, *49*, 1691. Levy, M.; Perdew, J. P.; Sahni, V. *Phys. Rev. A* **1984**, *30*, 2745. Almbladh, C. O.; von Barth, U. *Phys. Rev. B* **1985**, *31*, 3231.

(38) Kleinman, L. *Phys. Rev. B* **1997**, *56*, 16029.

(39) Perdew, J. P.; Levy, M. *Phys. Rev. B* **1997**, *56*, 16021.

(40) Kleinman, L. *Phys. Rev. B* **1997**, *56*, 12042.

(41) Harbola, M. K. *Phys. Rev. B* **1999**, *60*, 4545.

(42) Casida, M. E. *Phys. Rev. B* **1999**, *59*, 4694.

(43) Stewart, J. J. P. *J. Mol. Struct.* **2000**, *559*, 59.

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

(45) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907.

(46) Daw, M. S.; Baskes, M. I. *Phys. Rev. Lett.* **1983**, *50*, 1285.

(47) Brenner, D. W. *Phys. Rev. B* **1990**, *42*, 9458.

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

(49) Zope, R. R.; Mishin, Y. *Phys. Rev. B* **2003**, *68*, 024102.

(50) Cohen, R. E.; Mehl, M. J.; Papaconstantopoulos, D. *Phys. Rev. B* **1994**, *50*, 14694.

(51) Zope, R. R.; Dunlap, B. I. *Chem. Phys. Lett.* **2004**, *386*, 403.

(52) Zope, R. R.; Dunlap, B. I. *Phys. Rev. B* **2005**, *72*, 045439.

(53) Dunlap, B. I. *J. Phys. Chem.* **1986**, *90*, 5524.

(54) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(55) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(56) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.

(57) Andzelm, J.; Wimmer, E. *J. Phys. B* **1991**, *172*, 307; *J. Chem. Phys.* **1992**, *96*, 1280.

(58) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560.

(59) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119.

(60) Broyden, G. *J. Inst. Math. Appl.* **1970**, *6*, 222.

(61) Fletcher, R. *Comput. J.* **1970**, *13*, 317.

(62) Goldfarb, D. *Math. Comput.* **1970**, *24*, 23.

(63) Shanno, D. F. *Math. Comput.* **1970**, *24*, 647.

(64) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes The Art of Scientific Computing*; Cambridge University Press: Cambridge, England, 1986; p 309.

(65) Dunlap, B. I. *Phys. Rev. A* **2002**, *66*, 032502.

(66) Dunlap, B. I. *Comput. Phys. Comm.* **2005**, *165*, 18.

(67) Veillard, A.; Clementi, E. *J. Chem. Phys.* **1968**, *49*, 2415.

(68) Moore, C. E. *Ionization Potentials and Ionization Limits Derived from the Analysis of Optical Spectra;* Natl. Bur. Stand. Ref. Data Ser., Nat. Bur. Stand. (U.S.) U.S. GPO, Washington, DC, 1970; Vol. 34.

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

(70) Teller, E. *Rev. Mod. Phys.* **1962**, *34*, 627.

(71) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42.

(72) Erman, P.; Karawajczyk, A.; Rachlew-Kallne, E.; Stromholm, C.; Larsson, J.; Persson, A.; Zerne, R. *Chem. Phys. Lett.* **1993**, *215*, 173.

(73) Dunlap B. I.; Karna S. P. In *Theoretical and Computational Modeling of NLO and Electronic Materials*; Karna, S. P., Yeates, A. T., Eds.; ACS Symposium Series 628; American Chemical Society: Washington, DC, 1996; p 164.

(74) Liu, S.; Parr, R. G. *Phys. Rev. A* **1996**, *53*, 2211.

(75) Liu, S.; Süle, P.; Lopez-Boada, R.; Nagy, A. *Chem. Phys. Lett.* **1996**, *257*, 68.

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

(77) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029.

(78) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(79) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865; **1997**, *78*, 1396(E).

(80) 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; **1993**, *48*, 4978 (E).

(81) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.

(82) Adamo, C.; Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **2000**, *112*, 2643.

(83) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *298*, 113.

(84) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063.

(85) Dunlap, B. I. *J Mol. Struct. (THEOCHEM)* **2000**, *259*, 37.

CT050166W

JCTC Journal of Chemical Theory and Computation

Effects of Peripheral Substituents on the Electronic Structure and Properties of Unligated and Ligated Metal Phthalocyanines, Metal = Fe, Co, Zn

Meng-Sheng Liao,[†,§] John D. Watts,[†] Ming-Ju Huang,*[,†] Sergiu M. Gorun,[‡] Tapas Kar,[§] and Steve Scheiner*[,§]

Department of Chemistry, P.O. Box 17910, Jackson State University, Jackson, Mississippi 39217, Department of Chemistry and Environmental Science, New Jersey Institute of Technology, Newark, New Jersey 07102, and Department of Chemistry and Biochemistry, Utah State University, Logan, Utah 84322-0300

Received April 18, 2005

Abstract: The effects of peripheral, multiple −F as well as $-C_2F_5$ substituents, on the electronic structure and properties of unligated and ligated metal phthalocyanines, PcM, PcM(acetone)$_2$ (M = Fe, Co, Zn), PcZn(Cl), and PcZn(Cl^-), have been investigated using a DFT method. The calculations provide a clear explanation for the changes in the ground state, molecular orbital (MO) energy levels, ionization potentials (IP), electron affinities (EA), charge distribution on the metal (Q_M), axial binding energies, and in electronic spectra. While the strongly electron-withdrawing $-C_2F_5$ groups on the Pc ring change the ground state of PcFe, they do not influence the ground state of PcCo. The IP is increased by ∼1.3 eV from H_{16}PcM to F_{16}PcM and by another ∼1.1 eV from F_{16}PcM to F_{48}PcM. A similar increase in the EA is also found on going from H_{16}PcM to F_{48}PcM. Substitution by the $-C_2F_5$ groups also considerably increases the binding strength between PcM and the electron-donating axial ligand(s). Numerous changes in chemical and physical properties observed for the F_{64}PcM compounds can be accounted for by the calculated results.

1. Introduction

Metal phthalocyanines (PcMs), with their planar square structure, are interesting molecules that have been studied extensively in the literature. (Pc is used here to refer to any phthalocyanine, regardless of substituents.) They provide a versatile chemical system: almost every metal in the periodic table can combine with the Pc ring, and most of these compounds are very stable. The diversity of phthalocyanines, together with their high thermal and chemical stability, has made them suitable for many technological applications such as dyes, pigments, semiconductors,[1] energy conversion (photovoltaic and solar cells),[2] electrophotography,[3] photosensitizers,[4] gas sensors,[5] low-dimensional metals,[6] electrochromism,[7] Langmuir−Blodgett (LB) films,[8] liquid crystals,[9] and nonlinear optics.[10]

Consequently, a great deal of effort has been expended to develop new phthalocyanine materials that display certain characteristics useful for the particular application. It has been shown that the properties of phthalocyanines can be effectively modulated with different substituents on the periphery of the macrocycle. A useful methodology is the replacement of the hydrogens (Hs) of Pc with halogens (Xs), which can greatly increase the catalytic activity and stability of PcMs.[11,12] However, halogenated PcMs have the disadvantage of poor solubility, which impedes many new applications of the materials. Recently, several novel octakis-(perfluoro *i*-C_3F_7)(perfluoro)-PcM compounds (F_{64}PcM) (M = Zn, Co, Fe) have been synthesized and characterized by one of us,[13−16] where the aromatic X-atoms of halogenated

* Corresponding author e-mail: scheiner@cc.usu.edu.
† Jackson State University.
‡ New Jersey Institute of Technology.
§ Utah State University.

10.1021/ct050105y CCC: $30.25
Published on Web 08/27/2005

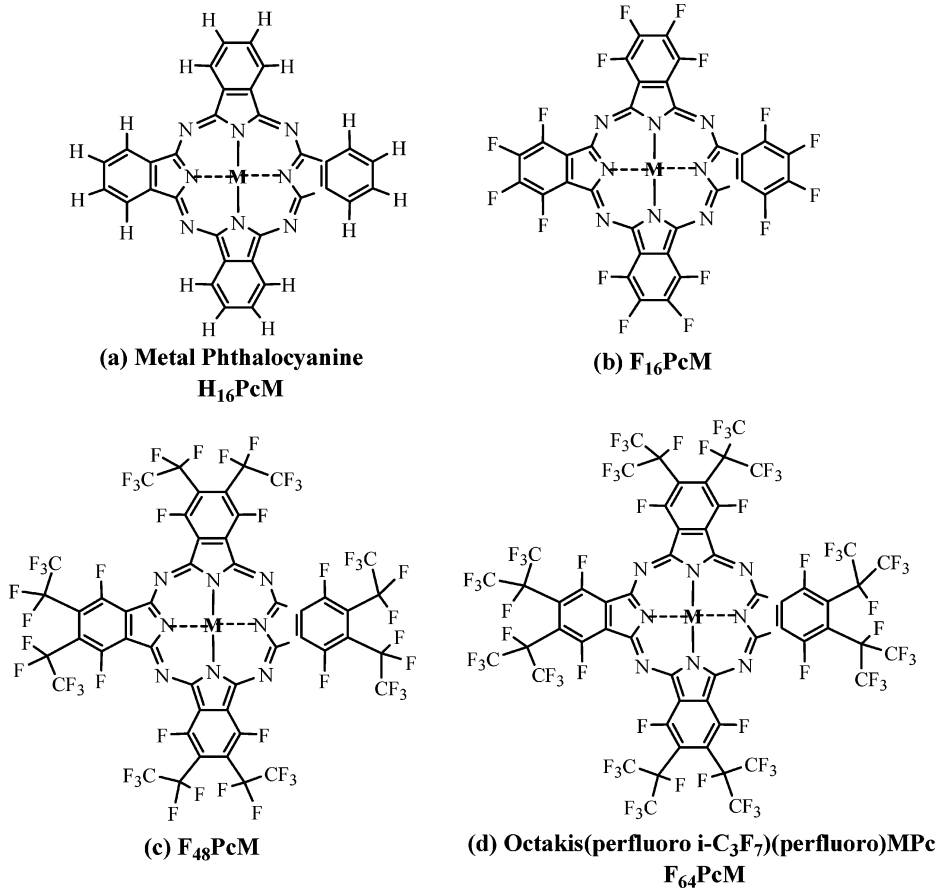


Figure 1. Molecular structures of metal phthalocyanine (H_{16}PcM) and its fluorosubstituted derivatives.

PcM (X_{16}PcM) are partially replaced by perfluoro isopropyl (*i*-C_3F_7) groups (R_f). These R_f groups are much more electronegative than the F atoms and were shown to increase the solubility of the compounds and also promote novel catalytic oxidations while resisting self-oxidation.[13]

In a recent paper,[17] density functional theory (DFT) calculations were carried out to investigate the electronic structure and properties of unligated and ligated F_{64}PcFe complexes. One important observation was the change of the ground state of Fe^{II} caused by the strongly electron-withdrawing peripheral substituents at the Pc ring. To shed more light on the effects of the R_f substituents, the previous theoretical study is extended in this report to include systems where M, Co, or Zn is four-, five-, or six-coordinated.

F_{64}PcZn has been studied in detail using single-crystal X-ray diffraction, optical and photoacoustic spectroscopy, and cyclic voltammetry, and it exhibits a number of interesting properties. (1) The ultraviolet−visible (UV−vis) spectra of F_{64}PcZn differ significantly from those of the parent, unsubstituted H_{16}PcZn species,[13] indicating that the electron-withdrawing property of the peripheral substituents affects the ring π molecular orbitals (MOs) to differing extents. (2) There are large shifts in the ring redox (oxidation and reduction) potentials on going from H_{16}PcZn to F_{64}PcZn, and the increase in oxidation potentials provides extra stability toward oxidative destruction for the R_f-substituted phthalocyanine.[13] (3) It is difficult to photochemically oxidize F_{64}PcZn, in stark contrast with the behavior of H_{16}PcZn, for which the photooxidation reaction results in the formation of a $[H_{16}PcZn]^+$ species.[13] (4) Axial ligation to F_{64}PcZn is favored relative to H_{16}PcZn; the X-ray structure of the former complex reveals the formation of $F_{64}PcZn(Ace)_2$ (Ace = acetone), but H_{16}PcZn as well as F_{16}PcZn do not retain solvent when crystallized.[14] Anionic ligands such as Cl^-, coordinate even more strongly. Thus, electrospray ionization mass spectrometry data for a DMF solution of F_{64}PcZn indicated the existence of F_{64}PcZn(Cl) and $F_{64}PcZn(Cl^-)$.[13]

The main goals of the present report are to examine the effects of F^- and R_f-substituents on various properties that include electronic structure, oxidation/reduction properties, ionization potentials, electron affinities, etc. This work extends previous semiempirical ZINDO and MOPAC calculations on F_{64}PcZn,[13] which were aimed mainly at interpretation of the absorption spectra of the metal complexes. The axial bonding properties of acetone, Cl, or Cl^- to the metal complexes are also examined. For comparison among the different metals, the results of iron phthalocyanines are also reported here.

2. Computational Details

The molecular structure of the parent metal phthalocyanine H_{16}PcM is illustrated in Figure 1a and that of (perfluoro *i*-C_3F_7)-(perfluoro)phthalocyanine, F_{64}PcM, is in Figure 1d. For computational convenience, the isopropyl $-C_3F_7$ substituent was modeled by ethyl $-C_2F_5$. The use of the smaller F_{48}PcM (Figure 1c) as an accurate electronic mimic of the larger F_{64}PcM was demonstrated in our previous calculations.[17] This simplification is also supported by earlier

semiempirical ZINDO calculations,[13] which show that the spectrum predicted for F_{48}PcZn is almost identical to that of F_{64}PcZn. For the sake of comparison and completion, the parent perfluorinated metal phthalocyanines, F_{16}PcM (Figure 1b), have also been considered here.

All calculations were carried out using the Amsterdam Density Functional (ADF) program package (version 2000.02) developed by Baerends and co-workers.[18−21] A triple-ζ STO basis was used for the metal 3s−4s shells plus one 4p polarization function, a triple-ζ basis for C/N/O 2s−2p and Cl 3s−3p shells plus one 3d polarization function, a double-ζ basis for F 2s−2p shells, and a double-ζ basis for the H 1s shell. It has been shown that high-quality basis sets (triple-ζ plus one polarization function) are required for the atoms within the macrocycle ring of the phthalocyanine in order to obtain the correct ground states of H_{16}PcFe and its derivatives.[22] The inner orbitals, i.e., 1s−2p for Fe/Cl and 1s for C/N/O/F, were considered as core and kept frozen according to the frozen-core approximation.[18] Among the various exchange-correlation potentials available, the density-parametrization form of Vosko, Wilk, and Nusair (VWN)[23] plus Becke's gradient correction for exchange (B)[24] and Perdew's gradient correction for correlation (P)[25] were employed. The combined VWN-B-P functional has been shown to provide accurate bonding energies for both main-group[26] and transition metal[27] systems. Relativistic corrections of the valence electrons were calculated by the quasi-relativistic (QR) method.[28] For the open-shell states, the unrestricted Kohn−Sham (UKS) spin-density functional approach was adopted.

Electron excitation energies related to the electronic absorption spectra were calculated using the time-dependent density functional response theory (TDDFT)[29] as implemented in the ADF program. TDDFT provides a first-principles method for the calculation of excitation energies and presents an excellent alternative to the conventional highly correlated configuration interactions (CI) method. Applications of TDDFT to excitation energy calculations can be found in recent work.[30,31]

3. Results and Discussion

Unsubstituted transition-metal H_{16}PcMs have been shown to have square planar D_{4h} symmetry.[30] X-ray crystal structure data[13−16] indicate that the Pc ring in substituted F_{64}PcM is quite planar, and this ring planarity is maintained in the solid state even in the presence of axial acetone ligands. The solution UV−vis spectra[32a−c,33] are also consistent with D_{4h} symmetry. The invariability of the MN_4 chromophore geometry upon substitution of the peripheral H-atoms by *i*-C_3F_7 groups strongly suggests that the geometry will not change for steric reasons when the *i*-C_3F_7 groups are replaced by less bulky groups in F_{16}PcM and F_{48}PcM. This argument is indeed supported by calculations.[13] Placing the molecule in the *xy* plane, the five metal 3d-orbitals transform as a_{1g} (d_{z^2}), b_{1g} ($d_{x^2-y^2}$), b_{2g} (d_{xy}), and e_g (d_π, i.e., d_{xz} and d_{yz}). For Fe^{II} and Co^{II}, different occupations of electrons in these d-orbitals may yield a number of possible low-lying states. To determine the ground state, relative energies of several selected configurations in H_{16}PcM, F_{16}PcM, and F_{48}PcM (M = Fe, Co) were calculated, wherein geometry optimization was performed separately for each state considered. These energies are reported in Table 1, along with the optimized M−N bond lengths.

Table 1. Calculated Relative Energies (*E*, eV) for Selected Configurations in H_{16}PcM, F_{16}PcM, and F_{48}PcM (M = Fe, Co)

configuration[a]				*E* (R)[b]		
b_{2g}/d_{xy}	a_{1g}/d_{z^2}	$1e_g$/d_π	state	H_{16}PcM	F_{16}PcM	F_{48}PcM
				M = Fe[c]		
2	2	2	$^3A_{2g}$	**0** (1.916)	**0** (1.918)	0 (1.919)
2	1	3	^3E_g (A)	0.05 (1.923)	0.01 (1.922)	−0.04 (1.921)
1	1	4	$^3B_{2g}$	0.08 (1.921)	0.05 (1.919)	**−0.06** (1.917)
1	2	3	^3E_g (B)	0.52 (1.910)	0.52 (1.911)	0.45 (1.910)
2	0	4	$^1A_{1g}$	1.43 (1.935)	1.37 (1.934)	1.28 (1.933)
				M = Co		
2	1	4	$^2A_{1g}$	**0** (1.922)	**0** (1.924)	**0** (1.931)
2	2	3	^2E_g	0.21 (1.909)	0.24 (1.911)	0.28 (1.916)
1	2	4	$^2B_{2g}$	0.95 (1.901)	0.94 (1.904)	0.88 (1.904)

[a] Orbital energy levels illustrated in Figure 2. [b] Values in parentheses refer to optimized M−N(Pc) bond length (in Å) for the pertinent state. [c] The calculated *E*'s and *R*'s with the present version ADF program (2000.02) may be slightly different from those obtained with the older version ADF (2.0.1), ref 17, but the trends of the results are not changed.

Table 2 displays the gross populations of M 3d, 4s, and 4p orbitals (in the ground state), along with the metal's Mulliken atomic charge. Table 3 lists the calculated M−Pc binding energies (E_{bind}), ionization potentials (IP) for several outer MOs, and the electron affinities (EA). E_{bind} is defined as the energy required to pull the metal apart from the Pc ring

$$-E_{bind} = E(\text{PcM}) - \{E(\text{M}) + E(\text{Pc})\}$$

where *E*(PcM), *E*(M), and *E*(Pc) represent the total energies of the indicated species. (The geometries of PcM and Pc are independently optimized.) The IPs and EAs were calculated by the so-called ΔSCF method which carries out separate SCF (self-consistent field) calculations for the neutral molecule and its ion, where EA = $E(X^-) - E(X)$. The computed relative energies of selected configurations of the ligated iron phthalocyanines are contained in Table 4, and their properties are presented in Table 5.

The electronic structure of H_{16}PcFe has been the subject of several experimental studies.[34−36] A $^3B_{2g}$ ground state was originally suggested for H_{16}PcFe on the basis of magnetic work,[35] but later magnetic circular dichroism spectra have shown that the ground state is in fact $^3A_{2g}$.[36] Our previous calculations[37] support the latter assignment. In a recent paper of ours,[17] we further calculated F_{16}PcFe, F_{48}PcFe, and their complexes with two axial ligands L (L = Ace, H_2O, pyridine). It is shown that the electronic configuration of each ligated PcFe is determined mainly by the axial ligand-field strength but can also be affected by peripheral substituents. Figure 2 illustrates the changes of the electronic structure from H_{16}PcFe to F_{16}PcFe to F_{48}PcFe. A detailed discussion of the results for the iron−phthalocyanine complexes is reported in ref 17.[38]

Table 2. Mulliken Orbital Populations and Atomic Charges on Metal (Q_M)

	H_{16}PcFe	F_{16}PcFe	F_{48}PcFe	H_{16}PcCo	F_{16}PcCo	F_{48}PcCo	H_{16}PcZn	F_{16}PcZn	F_{48}PcZn
3d	6.58	6.58	6.58	7.60	7.60	7.60	10.00	10.00	10.00
4s	0.42	0.38	0.27	0.32	0.28	0.27	0.55	0.53	0.53
4p	0.25	0.25	0.26	0.42	0.42	0.42	0.82	0.81	0.80
Q_M	0.75	0.80	0.88	0.65	0.70	0.70	0.63	0.67	0.67

Table 3. Calculated M−N Bond Lengths (R_{M-N}), M−Pc Binding Energies (E_{bind}), Ionization Potentials (IP), and Electron Affinities (EA) in the Ground State of the Systems

		H_{16}PcM	F_{16}PcM	F_{48}PcM[a]
		M = Fe		
R_{Fe-N}, Å		1.916	1.918	1.917
E_{bind}, eV		9.77	9.53	9.95
IP, eV	a_{1u}	6.39 (first)	7.67 (first)	8.76 (first)
	a_{1g}/d_{z^2}	6.42	7.70	10.63
	b_{2g}/d_{xy}	6.58	7.86	9.91
	$1e_g/d_\pi$	7.23	8.48	8.87
	b_{1u}	8.00	8.88	9.75
EA, eV	$1e_g$	−2.55	−3.90	−5.47 (b_{2g})
		M = Co		
R_{Co-N}, Å		1.922	1.924	1.931
E_{bind}, eV		11.51	11.29	11.71
IP, eV	a_{1u}	6.42 (first)	7.69 (first)	8.72 (first)
	$1e_g/d_\pi$	7.19	8.45	9.42
	b_{2g}/d_{xy}	7.51	8.73	9.53
EA, eV	a_{1g}/d_{z^2}	−2.94	−4.26	−5.35
	$2e_g$	−2.09	−3.44	−4.68
		M = Zn		
R_{Zn-N}, Å		2.000	2.001	2.006
E_{bind}, eV		5.64	5.50	5.89
IP, eV	a_{1u}	6.44 (first)	7.70 (first)	8.72 (first)
EA, eV	$2e_g$	−2.18	−3.52	−4.74

[a] Note that the ground state of F_{48}PcFe is different from those of H_{16}PcFe and F_{16}PcFe.

3.1. Cobalt Phthalocyanines. While the peripheral substituents have substantial influence on the electronic structure of the iron phthalocyanines, the Co analogues are much less sensitive. H_{16}PcCo, F_{16}PcCo, and F_{48}PcCo complexes all have a $^2A_{1g}$ ground state, with a fully occupied $1e_g/d_\pi$ level. The ^2E_g state, arising from the $(a_{1g})^2(1e_g)^3$ configuration, lies 0.2−0.3 eV higher in energy. The $^2B_{2g}$ state lies considerably higher, nearly a full eV above the ground state. The $^2E_g - {}^2A_{1g}$ energy gap increases gradually from H_{16}PcCo to F_{16}PcCo to F_{48}PcCo. This comparison between Fe and Co is illustrated more explicitly in Figure 3, where it may be noted first that the metal d-orbitals lie much lower in H_{16}PcCo than in H_{16}PcFe. Unlike the Fe case, the HOMO in PcCo (Pc a_{1u}) is no longer a metal 3d-orbital, and the unoccupied b_{1g} ($d_{x^2-y^2}$) lies below the b_{1u} orbital of the ring.

Similar to H_{16}PcFe, the first ionization of PcCo also occurs from the Pc a_{1u} orbital, and the first IP (6.42 eV) is very close to that obtained for H_{16}PcFe (6.39 eV). The calculated results are in agreement with experimental gas-phase photoelectron spectra (PES) of a series of metal phthalocyanines H_{16}PcM with M = Mg, Fe, Co, Ni, Cu, and Zn,[39] which find a sharp first IP at ∼6.4 eV for all the H_{16}PcMs and conclude that the orbitals of the first ionization are ringlike and not metal 3d-like in all cases. While for H_{16}PcFe the IPs from a_{1u} and a_{1g}/d_{z^2} are very close, for H_{16}PcCo the IP from a_{1u} is 0.77 eV smaller than that from a metal 3d-orbital (d_π), and so the one-electron oxidation of unligated H_{16}PcCo clearly occurs from the Pc ring. On the other hand, electrochemical studies of cobalt phthalocyanines in solution[40,41] find that $PcCo^{II}$ is oxidized to $[PcCo^{III}]^+$; i.e., one-electron oxidation of PcCo occurs from the metal. For example, Lever et al.[40] observed that oxidation of Co^{II} phthalocyanine by halogen (X) leads to the formation of a complex containing the trivalent metal $PcCo^{III}(X)_2$. The disagreement between the calculation and these sorts of experiments may be attributed to the effect of solvent. In solution, the cation species is ligated. The axial ligands raise the energy of the metal a_{1g}/d_{z^2} orbital (see Figure 3) so that the electron in this orbital may be ionized first. Our calculations on $H_{16}PcCo(Py)_2$ (Py = pyridine) show that the first ionization for such a complex indeed occurs from the metal a_{1g}/d_{z^2} orbital (see Table 5). Our further calculations on $H_{16}PcCo(Cl)_2$ show that the latter complex has a ground state of $^2A_{1u}$ $[(1e_g/d_\pi)^4(a_{1u})^1(a_{1g}/d_{z^2})^0]$, where the oxidation state of Co is III, in agreement with experiment;[40,41] a configuration of $(1e_g/d_\pi)^3(a_{1u})^2(a_{1g}/d_{z^2})^0$ is 0.57 eV higher in energy than the ground state.

The similarities between Fe and Co extend (trends are preserved upon fluorination) to the various degrees of fluorosubstitution reported in Table 3. On the other hand, the calculated EAs of PcCo are generally more negative than those of PcFe. For the reduction of PcCo, the added electron goes into the a_{1g}/d_{z^2} orbital, while it enters the $1e_g/d_\pi$ orbital for H_{16}PcFe and F_{16}PcFe. Since F_{48}PcFe has a different electronic structure (as compared to H_{16}PcFe/F_{16}PcFe), the calculated EA of this molecule is even slightly larger (0.12 eV) than that of F_{48}PcCo. The calculated Co−Pc binding energy is significantly larger (∼1.7 eV) than that of Fe−Pc, despite the fact that the Co−N bond lengths are somewhat (0.01 Å) longer than Fe−N. Note that for both Fe and Co, it is the F_{16}PcM species that has the lowest M−Pc binding energy.

The perturbations caused by the Ace ligands in the MO energy diagrams of PcM are illustrated in Figure 3 with $PcFe(Ace)_2$ on the left and $PcCo(Ace)_2$ on the right. The coordination of the two axial ligands lowers the symmetry of the system from D_{4h} to D_{2h} and splits the d_{xz}–d_{yz} degeneracy. Acetone is an electron-donating ligand, shifting the MOs upward. The a_{1g}/d_{z^2} orbital is particularly raised, owing to the strong repulsive interaction between the ligand HOMO and the metal d_{z^2}. As a result, the IPs of $PcM(Ace)_2$ are notably decreased as compared to those of PcM, suggesting that the former will be easier to oxidize than unligated PcM. Unlike $PcFe(Ace)_2$ where the first ionization

Table 4. Calculated Relative Energies (E, eV) for Selected Configurations in $H_{16}PcFe(Ace)_2$, $F_{16}PcFe(Ace)_2$, and $F_{48}PcFe(Ace)_2$

configuration					$E^{relative}$		
d_{xy}	d_{z^2}	d_{yz}	d_{xz}	state[a,b]	$H_{16}PcFe(Ace)_2$	$F_{16}PcFe(Ace)_2$	$F_{48}PcFe(Ace)_2$
2	1	2	1	$^3B_{2g}$ [^3E_g(A)]	**0** (1.933/2.373)[c]	**0** (1.939/2.300)	**0** (1.937/2.279)
2	1	1	2	$^3B_{3g}$ [^3E_g(A)]	0.05 (1.928/2.483)	0.07 (1.932/2.452)	0.09 (1.933/2.405)
1	1	2	2	$^3B_{1g}$ ($^3B_{2g}$)	0.15 (1.925/2.503)	0.18 (1.926/2.468)	0.17 (1.924/2.437)
2	0	2	2	$^1A_{1g}$ ($^1A_{1g}$)	0.16 (1.946/1.969)	0.12 (1.947/1.973)	0.06 (1.942/1.974)

[a] States in parentheses are the corresponding states in unligated species. [b] No minimum or very long Fe−O(Ace) distance was found for the $(d_{xy})^2(d_{z^2})^2(d_{yz})^1(d_{xz})^1$ − $^3B_{1g}$ ($^3A_{2g}$) and $(d_{xy})^1(d_{z^2})^2(d_{yz})^2(d_{xz})^1$ − $^3B_{3g}$ [^3E_g(B)] states. [c] The values in parentheses represent the optimized Fe−N(Pc) and Fe−O(Ace) bond lengths (in Å), respectively.

Table 5. Calculated Properties[a] of $H_{16}PcM$, $F_{16}PcM$, and $F_{48}PcM$ with Two Axial Ligands (L) at the Ground State (M = Fe, Co, Zn; L = Ace, Py)

		$H_{16}PcM(L)_2$ L = Ace	$F_{16}PcM(L)_2$ L = Ace	$F_{48}PcM(L)_2$ L = Ace	$H_{16}PcCo(L)_2$ L = Py
		M = Fe			
$R_{M-N(Pc)}$ (Å)		1.933	1.939	1.937	
$R_{M-O(Ace)}$ (Å)		2.373	2.300	2.279	
E_{bind}[PcM-$(L)_2$] (eV)		0.25	0.63	0.94	
Q_M (e)		0.79	0.81	0.80	
Q_{Ace} (e)		0.15	0.19	0.22	
IP (eV)	a_{1g}/d_{z^2}	6.97	8.01	8.86	
	$1b_{3g}/d_{yz}$	**5.74** (first)	**6.99** (first)	**7.94** (first)	
	b_{1g}/d_{xy}	6.01	7.26	8.14	
	a_{1u}	6.00	7.21	8.27	
EA (eV)	$1b_{2g}/d_{xz}$	−2.29	−3.55	−4.72	
	$2b_{3g}$	−1.66	−2.96	−4.22	
		M = Co			
$R_{M-N(Pc)}$ (Å)		1.926	1.929	1.932 (1.924)[c]	1.934
$R_{M-O(Ace)}$ (Å)		2.401	2.366	2.331 (2.314)	2.318
E_{bind}[PcM-$(L)_2$] (eV)		0.16	0.49	1.06	0.77
Q_M (e)		0.67	0.70	0.69	0.90
Q_{Ace} (e)		0.13	0.14	0.16	0.18
IP (eV)	a_{1g}/d_{z^2}	7.56	8.69	9.45	**5.85** (first)
	a_{1u}	**6.03** (first)	**7.25** (first)	**8.31** (first)	6.11
EA (eV)	$2b_{3g}$	−1.67	−2.98	−4.25	−1.82
	a_{1g}/d_{z^2}	−1.63	−2.82	−3.87	−0.95
E_{dis}[d](eV)		−1.11	−0.79	−0.33	−0.39
		M = Zn			
$R_{M-N(Pc)}$ (Å)		2.010	2.014	2.012 (2.006)[c]	
$R_{M-O(Ace)}$ (Å)		2.505	2.459	2.414 (2.411)	
E_{bind}[PcM-$(L)_2$] (eV)		0.12	0.44	0.73	
Q_M (e)		0.64	0.67	0.65	
Q_{Ace} (e)		0.10	0.12	0.13	
IP (eV)	a_{1u}	**6.05** (first)	**7.26** (first)	**8.30** (first)	
EA (eV)	$2b_{3g}$	−1.78	−3.08	−4.32	

[a] R: distance, E_{bind}[PcM-$(Ace)_2$]: binding energy between PcM and two Ace ligands, Q: charge distribution, IP: ionization potential, EA: electron affinity. [b] The first IP is indicated in bold. [c] The values in parentheses are experimental distances for $F_{64}PcM(Ace)_2$ (refs 14 and 15). [d] Dissociation energy for $[PcCo(L)_2]^- \rightarrow [PcCo]^- + 2L$.

occurs from the central metal ($1b_{3g}/d_{yz}$), Table 5 reveals that of $PcCo(Ace)_2$ takes place still from the Pc a_{1u} orbital, similar to unligated PcM. As mentioned above, in the presence of relatively strong field ligands L (e.g. L = Py), the first ionization of $PcCo(L)_2$ takes place from the central metal. For $PcCo(Ace)_2$, the IP from a_{1g}/d_{z^2} is more than 1 eV larger than that from a_{1u}. There is about a 0.4 eV decrease of IP from PcCo to $PcCo(Ace)_2$, in contrast to ∼0.8 eV decrease of IP from PcFe to $PcFe(Ace)_2$. Note also that for any degree of substitution, the IP of the $PcCo(Ace)_2$ is roughly 0.3 eV higher than the corresponding quantity for the Fe analogue.

The axial ligands reduce the electron affinity of PcM considerably (by 1.1−1.3 eV) when M is Co. In contrast to PcCo, the added electron in $PcCo(Ace)_2$ now occupies a high-lying antibonding Pc $2b_{3g}$ orbital. Since the added electron in PcCo goes into a low-lying metal orbital, $[PcCo]^-$ has a relatively low energy. Unless the ligand field of L is very strong, the $[PcM(L)_2]^-$ complex is expected to be unstable

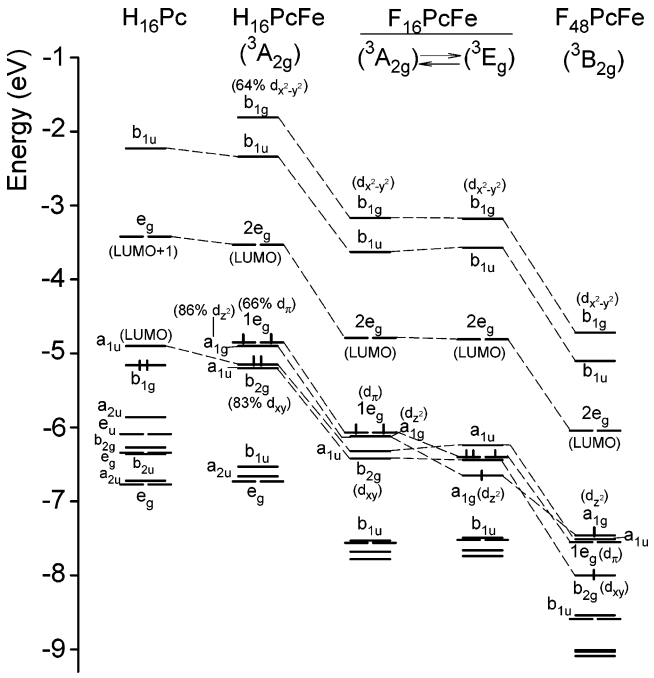


Figure 2. Orbital energy levels of $H_{16}Pc$ (on left, with no H atoms in the ring cage) and the various iron phthalocyanines.

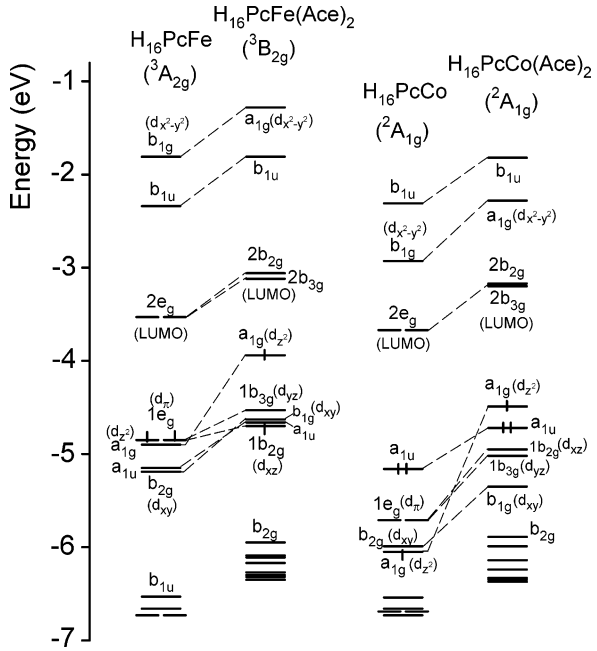


Figure 3. Orbital energy levels of $H_{16}PcM$ (M = Fe, Co) when complexed with two axial acetone (Ace) ligands.

and to dissociate into $[PcCo]^- + 2L$. Our calculations show that the dissociation energy E_{dis} for $[H_{16}PcCo(Ace)_2]^-$ is strongly exothermic (by more than 1 eV). This may account for the fact that previous experiments[41,42] have always detected Co^I and no Co^{II} when cobalt phthalocyanines are reduced. We note that the EA to the $2b_{3g}$ orbital is not significantly larger than that to a_{1g}/d_{z^2}, especially for $H_{16}PcCo(Ace)_2$. When a_{1g}/d_{z^2} is doubly occupied, the two electrons in this orbital would result in a very strong repulsion to axial ligands. This small difference in EA between $2b_{3g}$ and a_{1g}/d_{z^2} also leads us to expect that $[PcM(L)_2]^-$ is unstable

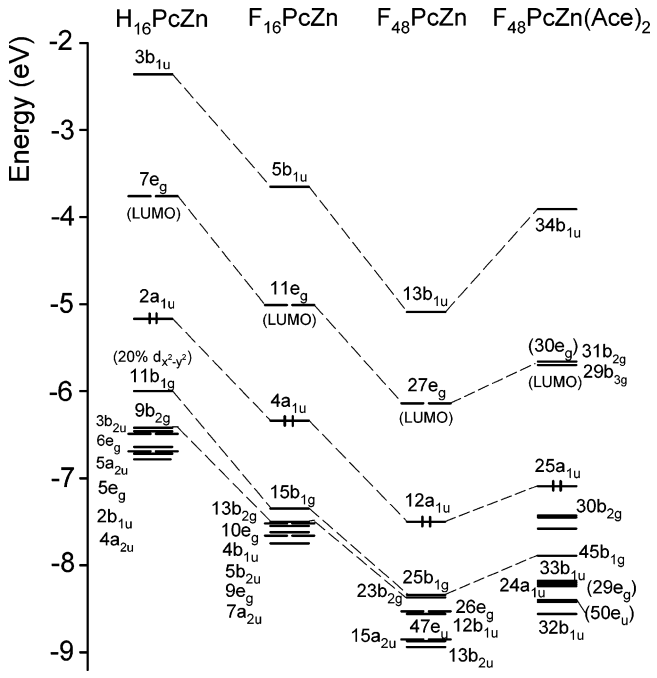


Figure 4. Orbital energy levels of the various zinc phthalocyanines.

and will dissociate. In fact, E_{dis} decreases to −0.33 eV when the $-C_2F_5$ substituents are introduced on the Pc ring. From the calculated E_{dis} value for $[H_{16}PcCo(Py)_2]^-$ (−0.39 eV) and the trend in E_{dis} from $[H_{16}PcCo(Ace)_2]^-$ to $[F_{48}PcCo(Ace)_2]^-$, the $[F_{48}PcCo(Py)_2]^-$ complex is expected to be stable against dissociation.

For both PcFe and $PcFe(Ace)_2$, however, the EAs all correspond to the addition of an electron to a low-lying metal d-orbital, and so there is a smaller decrease of EA from PcFe to $PcFe(Ace)_2$ as compared to the cobalt systems.

The binding energy (E_{bind}) between $H_{16}PcCo$ and a pair of Ace molecules is rather small, 0.16 eV. This quantity rises to more than 1 eV for the heavily substituted $F_{48}PcCo$. Corresponding to the increase of E_{bind}, the Co−O(Ace) distance is shortened as more F-atoms are added to the system. The Co−N(Pc) and Co−O(Ace) bond lengths (1.932 Å, 2.331 Å) calculated in $F_{48}PcCo(Ace)_2$ are in excellent agreement with the X-ray structural data on $F_{64}PcCo(Ace)_2$ (1.924 Å, 2.314 Å).[15] The axial ligation does not affect the equatorial Co−N distance very much (<0.005 Å).

3.2. Zinc Phthalocyanines. *3.2.1. Electronic Structure and Properties.* The perturbations in the orbital levels that accompany the various changes in the Zn systems are illustrated in Figure 4. The 3d-orbitals are particularly low in energy for Zn; b_{1g} has lost all but 20% of its metal contribution and is largely Pc σ in character. As was true for the Fe systems in Figure 2, fluorosubstitution has a strong lowering effect on the MOs. The magnitude of this shift is quite uniform for most MOs including the HOMO and LUMO. The main exceptions are some orbitals below the HOMO-1. For instance, the energy gap between b_{1g} (HOMO-1) and b_{2g} (HOMO-2) is large in $H_{16}PcZn$ but nearly disappears in $F_{48}PcZn$.

The first oxidation of PcZn clearly leads to a π-cation radical, where the electron is removed from the HOMO a_{1u},

yielding a $^2A_{1u}$ ground state. This IP is strikingly similar in magnitude to the Co cases. In the case of reduction, the electron is accommodated in the LUMO e_g (Pc π^*). Both IP and EA are increased considerably by the $-C_2F_5$ substituents, in agreement with the trend in the electrochemical oxidation and reduction potentials.[13] The EA values are slightly more negative for the Zn molecules than for Co.

With two electrons in the b_{1g} ($d_{x^2-y^2}$) orbital in PcZn, the repulsive interaction between these electrons and those on the pyrrole nitrogens lengthens the M−N bond length by ∼0.08 Å relative to Co and considerably decreases the M−Pc binding energy, as is evident by the E_{bind} entries in Table 3 for M = Zn.

On the other hand, the presence of two electrons in the a_{1g}/d_{z^2} orbital in PcZn(Ace)$_2$ gives rise to longer axial Zn−O(Ace) bond lengths as compared to Co−O(Ace). Therefore the binding energy of PcZn−(Ace)$_2$ is smaller than the PcCo−(Ace)$_2$ value. Also, there is a large increase of the PcZn−(Ace)$_2$ binding energy when more F atoms are added to the system. It was argued that the perfluoro peripheral substituents withdraw electron density from the inner ring and the central metal so that the axial ligation of F_{64}PcZn is enhanced relative to the parent H_{16}PcZn complex.[13] According to the calculation, the presence of the F-atoms in F_{16}PcZn makes the atomic charge on Zn more positive by 0.04 (see Table 2), but no change in Q_M is found from F_{16}PcZn to F_{48}PcZn. The same is true for the cobalt systems. There is a more pronounced shortening of the axial M−O(Ace) bond distance from H_{16}PcZn to F_{48}PcZn than from H_{16}PcCo to F_{48}PcCo. Again the calculated bond lengths $R_{Zn-N(Pc)}$ = 2.012 Å and $R_{Zn-O(Ace)}$ = 2.414 Å in F_{48}PcZn agree very well with the 2.006 Å and 2.411 Å measured in the F_{64}PcZn crystal.[14] An addition of (Ace)$_2$ to PcZn produces little geometric change, similar to the PcCo case.

The MO levels are shifted up from F_{48}PcZn to F_{48}PcZn(Ace)$_2$. The increase in the energy of a_{1u} leads to a relatively small IP from this orbital. Therefore, solvent coordination to the central metal is able to decrease the oxidation potential of PcM, as observed experimentally.[13] While attempts to oxidize F_{64}PcZn were unsuccessful, there is evidence that in the presence of strong electron-donating ligands such as imidazole, F_{64}PcZn can oxidized electrochemically.[13] The experimental observation can be accounted for by the calculated results. Here the IP and EA of PcZn(Ace)$_2$ are about 0.4 eV smaller (lower) than those of PcZn.

3.2.2. Electronic Spectra. Electronic spectra of both H_{16}PcZn and F_{48}PcZn have been calculated using the semiempirical ZINDO method,[13] with the aim of understanding the changes in energy of the spectral features when the electron-withdrawing peripheral substituents ($-C_2F_5$) are introduced. However, the semiempirical method yields some excitation energies which are considerably too large (more than 1 eV) as compared to the experimental spectra. In this study, we describe the results of TDDFT calculations for the series of PcZn complexes from H_{16}PcZn to F_{16}PcZn to F_{48}PcZn to F_{48}PcZn(Ace)$_2$. Our calculated excitation energies and oscillator strengths for several lowest, spin-allowed singlet ^1E_u excited states are displayed in Table 6, together with available experimental data.[13,33]

The experimental spectrum of PcZn is characterized by an intense absorption band in the visible (Q-band) and two other strong absorption bands in the near-UV (B_1 and B_2 bands).[33] The multiple bands in the B region is supported by the calculations. The Q-band is assigned to the 1^1E_u state, which is nearly pure (90%) HOMO → LUMO transition. For H_{16}PcZn, 1^1E_u is calculated to be 1.96 eV, in good agreement with the experimental value of 1.85 eV.[33] The close-lying 6^1E_u and 7^1E_u states are responsible for the B_2 and B_1 bands, respectively; their oscillator strengths are both quite large. Different from 1^1E_u, the latter two states involve significant mixture of several transitions. Again, the calculated E^{exc} values (3.40 and 3.74 eV) for the B bands agree very well with experiment (3.16 and 3.71 eV).[33] The same is true for F_{48}PcZn.

The 1^1E_u state (Q-band) in F_{16}PcZn is predicted at 1.84 eV, a red shift of 0.12 eV relative to that of H_{16}PcZn. For the B bands, the red shift is somewhat larger, about 0.2 eV. These results reveal that the introduction of the electron-withdrawing F-atoms into the Pc ring shifts the Q and B bands to the red.

Turning to F_{48}PcZn, its 1^1E_u has nearly the same E^{exc} as that of F_{16}PcZn. So the partial replacement of eight F-atoms by eight $-C_2F_5$ groups does not shift the Q-band. Here the main effect of the $-C_2F_5$ substituents is a notable red shift of the B_1 band.

From F_{48}PcZn to F_{48}PcZn(Ace)$_2$, the E^{exc} values for the various ^1E_u states remain nearly unchanged, indicating that the presence of axial acetone ligands does not change the positions of the absorption bands. These results are also in accord with experimental observations.[13] Note that in Figure 4, the LUMO−HOMO energy gap is not shown to be changed from one system to another, but the peripheral substituents may still change the calculated E^{exc} for the HOMO → LUMO transition. This is because the 1^1E_u state is not a pure HOMO → LUMO transition; it contains some contributions (∼10%) from other transitions owing to configuration interaction. As a matter of fact, the change in E^{exe} is insignificant from one system to another.

Similar to the calculations on H_{16}PcNi by Rosa et al.,[31] our TDDFT calculations on PcZn also predict several allowed $\pi \rightarrow \pi^*$ transitions located between the Q and B bands. These $\pi \rightarrow \pi^*$ transitions may be responsible for the intensity and broadness of the B bands. The ZINDO calculations predicted a HOMO → LUMO+1 transition (called second $\pi \rightarrow \pi^*$ transition) located to the red of the B band for both H_{16}PcZn and F_{48}PcZn. Our results show that the HOMO → LUMO+1 transition is located to the blue of the B band for F_{48}PcZn. No analogous transition is found for F_{16}PcZn.

3.2.3. PcZn(Cl) and PcZn(Cl$^-$) Complexes. Finally, calculations were also performed on the various PcZn(Cl) and PcZn(Cl$^-$) complexes. When only a single axial ligand is attached to the system, significant out-of-plane displacement of the metal is expected and in fact observed. For each molecule, the geometry was optimized assuming C_{4v} symmetry. The calculated PcZn-L binding energies (L = Cl, Cl$^-$) and three critical coordination parameters in PcZn(L) are displayed in Table 7. $R_{Ct\cdots N}$ (distance between the center of the ring and pyrrole nitrogen atom) is a measure of the ring

Table 6. Calculated Excitation Energies (E^{exc}, in eV) and Oscillator Strengths (f)

system	state	contribution[a]	E^{exc}, eV calcd	E^{exc}, eV exptl[c]	f	assignment
$H_{16}PcZn$	1^1E_u	91% ($2a_{1u} \rightarrow 7e_g$)	1.96	1.85	0.6482	Q
	2^1E_u	97% ($3b_{2u} \rightarrow 7e_g$)	2.80		0.0280	$\pi \rightarrow \pi^*$
	3^1E_u	39% ($2b_{1u} \rightarrow 7e_g$); 33% ($5a_{2u} \rightarrow 7e_g$); 25% ($5a_{2u} \rightarrow 7e_g$)	3.02		0.0014	$\pi \rightarrow \pi^*$
	4^1E_u	54% ($2b_{1u} \rightarrow 7e_g$); 23% ($4a_{2u} \rightarrow 7e_g$); 18% ($5a_{2u} \rightarrow 7e_g$)	3.07		0.3390	$\pi \rightarrow \pi^*$
	5^1E_u	76% ($2a_{1u} \rightarrow 8e_g$)	3.35		0.0242	$\pi \rightarrow \pi^*$
	6^1E_u	36% ($1a_{1u} \rightarrow 7e_g$); 23% ($4a_{2u} \rightarrow 7e_g$); 18% ($2a_{1u} \rightarrow 8e_g$); 17% ($5a_{2u} \rightarrow 7e_g$)	3.40	3.16	0.6840	B_2
	7^1E_u	53% ($1a_{1u} \rightarrow 7e_g$); 17% ($4a_{2u} \rightarrow 7e_g$); 16% ($5a_{2u} \rightarrow 7e_g$)	3.74	3.71	1.0402	B_1
	8^1E_u	95% ($6e_g \rightarrow 3b_{1u}$)	4.22		0.0332	
$F_{16}PcZn$	1^1E_u	91% ($4a_{1u} \rightarrow 11e_g$)	1.84		0.6332	Q
	2^1E_u	84% ($4b_{1u} \rightarrow 11e_g$); 11% ($5b_{2u} \rightarrow 11e_g$)	2.60		0.1442	$\pi \rightarrow \pi^*$
	3^1E_u	87% ($5b_{2u} \rightarrow 11e_g$); 10% ($4b_{1u} \rightarrow 11e_g$)	2.71		0.1360	$\pi \rightarrow \pi^*$
	4^1E_u	80% ($7a_{2u} \rightarrow 11e_g$); 15% ($6a_{2u} \rightarrow 11e_g$)	2.88		0.2928	$\pi \rightarrow \pi^*$
	5^1E_u	67% ($3a_{1u} \rightarrow 11e_g$); 29% ($6a_{2u} \rightarrow 11e_g$)	3.16		0.1038	B_2
	6^1E_u	43% ($6a_{2u} \rightarrow 11e_g$); 23% ($3a_{1u} \rightarrow 11e_g$)	3.58		1.4316	B_1
	7^1E_u	83% ($10e_g \rightarrow 5b_{1u}$)	4.00		0.0220	
	8^1E_u	89% ($9e_g \rightarrow 5b_{1u}$)	4.15		0.0786	
$F_{48}PcZn$	1^1E_u	91% ($12a_{1u} \rightarrow 27e_g$)	1.83	1.80	0.7984	Q
	2^1E_u	95% ($12b_{1u} \rightarrow 27e_g$)	2.51		0.3138	$\pi \rightarrow \pi^*$
	3^1E_u	75% ($13b_{2u} \rightarrow 27e_g$); 13% ($11a_{1u} \rightarrow 27e_g$)	2.89		0.1374	$\pi \rightarrow \pi^*$
	4^1E_u	59% ($11a_{1u} \rightarrow 27e_g$); 22% ($13b_{2u} \rightarrow 27e_g$); 17% ($15a_{2u} \rightarrow 27e_g$)	2.95		0.0086	$\pi \rightarrow \pi^*$
	5^1E_u	91% ($14a_{2u} \rightarrow 27e_g$)	3.13	2.89	0.0124	B_2
	6^1E_u	48% ($15a_{2u} \rightarrow 27e_g$); 20% ($12a_{1u} \rightarrow 28e_g$); 12% ($11a_{1u} \rightarrow 27e_g$)	3.33	3.10	1.7998	B_1
	7^1E_u	62% ($12a_{1u} \rightarrow 28e_g$); 27% ($26e_g \rightarrow 13b_{1u}$)	3.41		0.0356	$\pi \rightarrow \pi^*$
	8^1E_u	60% ($26e_g \rightarrow 13b_{1u}$); 12% ($12a_{1u} \rightarrow 28e_g$)	3.62		0.0058	
	9^1E_u	70% ($25e_g \rightarrow 13b_{1u}$); 22% ($12a_{1u} \rightarrow 29e_g$)	4.02		0.0700	
$F_{48}PcZn(Ace)_2$[b]	1^1E_u	90% ($25a_{1u} \rightarrow 30e_g$)	1.87		0.8075	Q
	2^1E_u	91% ($24a_{1u} \rightarrow 30e_g$)	2.65		0.2024	$\pi \rightarrow \pi^*$
	3^1E_u	62% ($33b_{1u} \rightarrow 30e_g$); 22% ($23a_{1u} \rightarrow 30e_g$)	2.86		0.5055	$\pi \rightarrow \pi^*$
	4^1E_u	89% ($32b_{1u} \rightarrow 30e_g$)	2.99		0.0021	$\pi \rightarrow \pi^*$
	5^1E_u	56% ($23a_{1u} \rightarrow 30e_g$); 27% ($31b_{1u} \rightarrow 30e_g$)	3.13		0.1025	B_2
	6^1E_u	51% ($25a_{1u} \rightarrow 31e_g$); 38% ($31b_{1u} \rightarrow 30e_g$)	3.30		0.7449	B_1
	7^1E_u	41% ($25a_{1u} \rightarrow 31e_g$); 28% ($31b_{1u} \rightarrow 30e_g$)	3.37		0.3465	$\pi \rightarrow \pi^*$
	8^1E_u	75% ($28e_g \rightarrow 26a_{1u}$)	3.59		0.1101	

[a] Contribution of less than 10% is not listed. [b] For convenience, we also use the label e_g (of D_{4h} symmetry) for the lower symmetry system. [c] References 13 and 33.

Table 7. Calculated Properties of $H_{16}PcZn$, $F_{16}PcZn$, and $F_{48}PcZn$ with Axial Ligands of Cl and of Cl^-

	$H_{16}PcZn(Cl)$	$F_{16}PcZn(Cl)$	$F_{48}PcZn(Cl)$	$H_{16}PcZn(Cl^-)$	$F_{16}PcZn(Cl^-)$	$F_{48}PcZn(Cl^-)$
$R_{Ct\cdots N(Pc)}$[a] (Å)	1.976	1.971	1.978	1.987	1.986	1.986
$R_{Ct\cdots Zn}$[b] (Å)	0.761	0.774	0.736	0.710	0.695	0.695
R_{Zn-Cl} (Å)	2.232	2.214	2.200	2.269	2.250	2.249
Q_{Zn} (e)	0.54	0.54	0.52	0.57	0.56	0.55
Q_{Cl} (e)	−0.45	−0.41	−0.44	−0.51	−0.45	−0.44
$E_{bind}(PcZn-Cl)$[c] (eV)	1.77	1.62	1.32	1.81	2.96	3.79

[a] Ct denotes the center of the Pc ring and $R_{Ct\cdots N(Pc)}$ denotes the distance between Ct and N(Pc). [b] $R_{Ct\cdots Zn}$ denotes displacement of the Zn atom out of the Pc plane. [c] Binding energy between PcZn and Cl or Cl^-.

core size, $R_{Ct\cdots Zn}$ represents the displacement of the Zn atom out of the N_4-plane toward the L ligand, and R_{Zn-Cl} refers to the Zn−Cl bond length.

The calculated $H_{16}PcZn-Cl$ binding energy E_{bind} is large, 1.77 eV, indicating a high affinity of $H_{16}ZnPc$ for Cl. As an electron-withdrawing ligand, the binding of neutral Cl to F_{48}-PcZn becomes significantly weaker. A large Zn out-of-plane displacement is found when Cl is coordinated to the system. In this case, the Ct···N distance becomes somewhat shorter than R_{Zn-N} of unligated PcZn.

For the chloride ion, there is a clear trend of increasing binding energy in the order $H_{16}PcZn-Cl^- < F_{16}PcZn-Cl^-$

< $F_{48}PcZn-Cl^-$, where the $F_{48}PcZn-Cl^-$ binding energy is much larger than that of $H_{16}PcZn-Cl^-$. The values of E_{bind} correlate well with the relative stabilities of these PcZn(Cl^-) complexes.[13]

4. Conclusions

The following main conclusions may be drawn from the calculated results.

(1) While the strongly electron-withdrawing $-C_2F_5$ groups on the Pc ring are able to change the ground state of PcFe, they do not influence the ground state of PcCo.

(2) Corresponding to the downshifts of the valence MOs caused by the peripheral substituents, the first ionization potential (IP) is increased by ~1.3 eV from $H_{16}PcM$ to $F_{16}PcM$ and by another ~1.1 eV from $F_{16}PcM$ to $F_{48}PcM$. A similar increase in the EA is also found on going from $H_{16}PcM$ to $F_{48}PcM$. These results account for the fact that the $F_{64}PcM$ compounds are difficult to oxidize but easy to reduce.[13] The change from $H_{16}PcZn$ to $F_{48}PcZn$ does not produce an obvious change in the LUMO−HOMO energy gap, but large changes in the relative MO energies are found between the b_{1g} (HOMO-1) and b_{2g} (HOMO-2) orbitals.

(3) The axial acetone (Ace) ligands in $PcM(Ace)_2$ shift the MO energy levels upward. There is a decrease of ~0.8 eV in IP from PcFe to $PcFe(Ace)_2$, and a decrease of ~0.4 eV when the metal M is Co and Zn. Therefore, axial coordination by (strong) σ-donors renders easier the oxidation of the compound, as observed experimentally.[13]

(4) The calculated $H_{16}PcM$-$(Ace)_2$ binding energy (E_{bind}) is small (<0.2 eV), particularly for M = Zn, indicating a low affinity of the parent metal phthalocyanine for acetone. The E_{bind} is increased to nearly 1 eV when the $-C_2F_5$ groups are introduced at the periphery of the Pc ring. This rise accounts for the experimental result[13] that there is formation of a $F_{64}PcZn(Ace)_2$ compound, but no $H_{16}PcZn(Ace)_2$ was detected. On the other hand, the calculated $PcZn-Cl^-$ binding energies can also account for the relative stabilities of those PcZn(Cl^-) compounds.

(5) The TDDFT calculated excitation energies for the PcZn complexes are in quantitative agreement with available experimental data. The substitution of 16 H-atoms in $H_{16}PcZn$ by F-atoms (i.e. from $H_{16}PcZn$ to $F_{16}PcZn$) shifts the Q and B bands to the red. A further red shift in the B_1 band occurs with partial replacement of eight F-atoms in $F_{16}PcZn$ by $-C_2F_5$ groups, but there is no shift in the Q-band from $F_{16}PcZn$ to $F_{48}PcZn$. On the other hand, the calculated electronic spectrum of $F_{48}PcZn$ is nearly the same as that of $F_{48}PcZn(Ace)_2$, supporting an assumption that the results (spectra) for the fluorinated phthalocyanines are virtually independent of the presence or absence of axial ligands.[13]

(6) The calculated redox properties of gas-phase cobalt phthalocyanines may be different from those measured in solution. These differences are likely due to association of specific solvent molecules with the complex.

Acknowledgment. This work was supported by grant DAAD19-99-1-0206 (to S.S.) from the Army Research Office and by grant S06 GM08047 from the National Institutes of Health (to M.J.H. and J.D.W.).

References

(1) Simon, J.; Andre, J.-J. *Molecular Semiconductors*; Springer: Berlin, 1985.

(2) Tang, C. W. *Appl. Phys. Lett.* **1986**, *48*, 183−185.

(3) Loutfy, R. O.; Hor, A. M.; Hsiao, C. K.; Baranyi, G.; Kazmaier, P. *Pure Appl. Chem.* **1988**, *60*, 1047−1054.

(4) Kato, M.; Nishioka, Y.; Kaifu, K.; Kawamura, K.; Ohno, S. *Appl. Phys. Lett.* **1985**, *46*, 196−197.

(5) Temofonte, T.; Schoch, K. F. *J. Appl. Phys.* **1989**, *65*, 1350−1355.

(6) Marks, T. J. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 857−879.

(7) Toshima, N.; Tominage, T. *Bull. Chem. Soc. Jpn.* **1996**, *69*, 2111−2122.

(8) Palacin, S.; Lesieur, P.; Stefanelli, I.; Barraud, A. *Thin Solid Films* **1988**, *159*, 83−90.

(9) Simon, J.; Sirlin, C. *Pure Appl. Chem.* **1989**, *61*, 1625−1629.

(10) Gasstevens, M. K.; Samoc, M.; Pfleger, J.; Prasad, P. N. *J. Chem. Phys.* **1990**, *92*, 2019−2024.

(11) Mckeown, N. B. *Phthalocyanine Materials*; Cambridge University Press: Cambridge, 1998.

(12) *Phthalocyanines: Properties and Applications*; Leznoff, C. C., Lever, A. B. P., Eds.; VCH Publishers: New York, 1996; Vol. 4.

(13) Keizer, S. P.; Mack, J.; Bench, B. A.; Gorun, S. M.; Stillman, M. J. *J. Am. Chem. Soc.* **2003**, *125*, 7067−7085.

(14) Bench, B. A.; Beveridge, A.; Sharman, W. M.; Diebold, G. J.; van Lier, J. E.; Gorun, S. M. *Angew. Chem., Int. Ed.* **2002**, *41*, 748−750.

(15) Bench, B. A.; Brennessel, W. W.; Lee, H.-J.; Gorun, S. M. *Angew. Chem., Int. Ed.* **2002**, *41*, 750−754.

(16) Lee, H.-J.; Brennessel, W. W.; Brucker, W. J.; Lessing, J. A.; Young, V. G., Jr.; Gorun, S. M. To be submitted.

(17) Liao, M.-S.; Kar, T.; Gorun, S. M.; Scheiner, S. *Inorg. Chem.* **2004**, *43*, 7151−7161.

(18) Baerends, E. J.; Ellis, D. E.; Roos, P. *Chem. Phys.* **1973**, *2*, 41−51.

(19) Versluis, L.; Ziegler, T. *J. Chem. Phys.* **1988**, *88*, 322−328.

(20) te Velde, G.; Baerends, E. J. *J. Comput. Phys.* **1992**, *99*, 84−98.

(21) Fonseca-Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391−403.

(22) Liao, M.-S.; Scheiner, S. Unpublished results.

(23) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.

(24) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(25) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(26) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612−5626.

(27) Li, J.; Schreckenbach, G.; Ziegler, T. *J. Am. Chem. Soc.* **1995**, *117*, 486−494.

(28) Ziegler, T.; Tschinke, V.; Baerends, E. J.; Snijders, J. G.; Ravenek, W. *J. Phys. Chem.* **1989**, *93*, 3050−3056.

(29) van Gisbergen, S. J. A.; Snijders, J. G.; Baerends, E. J. *Comput. Phys. Commun.* **1999**, *118*, 119−138.

(30) Nguyen, K. A.; Pachter, R. *J. Chem. Phys.* **2001**, *114*, 10757−10767.

(31) Rosa, A.; Ricciardi, G.; Baerends, E. J.; van Gisbergen, S. J. A. *J. Phys. Chem. A* **2001**, *105*, 3311−3327.

(32) (a) Kahl, J. L.; Faulkner, L. R.; Dwarakanath, K.; Tachikawa, H. *J. Am. Chem. Soc.* **1986**, *108*, 5434−5440. (b) Williamson, B. E.; VanCott, T. C.; Boyle, M. E.; Misener, G. C.; Stillman, M. J.; Schatz, P. N. *J. Am. Chem. Soc.* **1992**, *114*, 2412−2419. (c) Janczak, J.; Kubiak, R. *Inorg. Chim. Acta* **2003**, *342*, 64−76.

(33) Stillman, M. J.; Nyokong, T. In *Phthalocyanine: Principles and Properties*; Leznoff, C. C., Lever, A. B. P., Eds.; VCH Publishers: New York, 1993; Vol. 1, Chapter 3, pp 133−290.

(34) Dale, B. W.; Williams, R. J. P. *J. Chem. Phys.* **1968**, *49*, 3445−3449.

(35) Barraclough, C. G.; Martin, R. L.; Mitra, S. *J. Chem. Phys.* **1970**, *53*, 1643−1648.

(36) Stillman, M. J.; Thomson, A. J. *J. Chem. Soc., Faraday Trans. 2* **1974**, *70*, 790−804.

(37) Liao, M.-S.; Scheiner, S. *J. Chem. Phys.* **2001**, *114*, 9780−9791.

(38) The calculated results of the iron phthalocyanines with the present version of the ADF program may be slightly different from those obtained with an older version,[17] but the trends remain unchanged.

(39) Berkowitz, J. *J. Chem. Phys.* **1979**, *70*, 2819−2828.

(40) Myers, J. F.; Canham, G. W. R.; Lever, A. B. P. *Inorg. Chem.* **1975**, *14*, 461−468.

(41) Minor, P. C.; Gouterman, M.; Lever, A. B. P. *Inorg. Chem.* **1985**, *24*, 1894−1900.

(42) Clack, D. W.; Yandle, J. R. *Inorg. Chem.* **1972**, *11*, 1738−1742.

CT050105Y

JCTC Journal of Chemical Theory and Computation

Stability of K-Montmorillonite Hydrates: Hybrid MC Simulations

G. Odriozola* and J. F. Aguilar

Programa de Ingeniería Molecular, Instituto Mexicano del Petróleo, Lázaro Cárdenas 152, 07730 México, D. F., México

Received March 9, 2005

Abstract: $NP_{zz}T$ and $\mu P_{zz}T$ simulations of K-montmorillonite hydrates were performed employing hybrid Monte Carlo simulations. Two condition sets were studied: $P = 1$ atm and $T = 300$ K (ground level conditions) and $P = 600$ atm and $T = 394$ K; this last condition mimics a burial depth close to 4 km. For these conditions, swelling curves as a function of the reservoir water vapor pressure were built. We found the single layer K-montmorillonite hydrate stable for high vapor pressures for both burial and ground level conditions. A simple explanation for this high stability is given.

I. Introduction

Clays are layer type aluminosilicate minerals, existent everywhere in nature and industry, hence the importance of a detailed understanding of their physics and chemistry. They are used as building materials, ceramics, and catalysts; they are employed in cosmetics, as rheological modifiers for paints, and in technological processes such as oil well drilling. In this last application, the control of their stability is a key for drilling success. During it, the use of water based muds induce destabilization of shale and clay formations that would disintegrate, or heave, upon contact with water.

One way to maintain stability of shales during the drilling process is by addition of potassium salts to drilling muds. This helps to avoid fluid loss and water infiltration. These kind of muds, that contain potassium ions dissolved in the water phase, are widely used for drilling water-sensitive shales, specially hard, brittle shales. Potassium cations in these systems replace ions such as sodium found in most shales to produce less hydrated clays with significantly reduced swelling potential. These ions also help to hold the cuttings together, minimizing their dispersion into finer particles. From all of the previous facts, a good understanding of the role of potassium in the swelling clays such as montmorillonite is mandatory, in particular at basin conditions of hard experimental implementation.

Besides the experimental studies, computer simulations are essential components of research on clay−water−cation systems.[1−13] These simulations give microscopic insights that are difficult to access experimentally. There are, however, not many computational studies on the swelling of montmorillonite hydrates for potassium interlayer cations.[14−20] In addition, excluding Hensen et al.[7,17] and Tambach et al.[20] works, these papers report solely $NP_{zz}T$ and NVT simulations, where the number of water molecules per interlaminar space is someway arbitrarily fixed, and hence, they do not show the whole picture of swelling. A good example of this is that hysteresis is naturally predicted by sampling in an open ensemble,[4,7,21] without the need of measuring properties such as water chemical potential,[9] immersion energies,[13] or swelling free energies.[12,22] Finally, this paper focuses on the stability of the different hydrates in contact with several reservoirs, which differ in temperature, pressure, and water activity.

The importance of potassium as swelling inhibitor of clays and the above-mentioned reasons motivated us to study the microscopic mechanisms underlying the behavior of K-montmorillonite hydrates at equilibrium with different reservoirs. We performed simulations of these systems in the $NP_{zz}T$ and $\mu P_{zz}T$ ensembles, considering explicitly two clay layers in the simulation box to avoid finite size effects. The simulations were carried out for two condition sets. One at ground level, with $P = 1$ atm and $T = 298$ K, and the other one with $P = 600$ atm and $T = 394$ K, which corresponds to an average burial depth close to 4 km.

* Corresponding author phone: +5255 91758176; e-mail: godriozo@imp.mx.

10.1021/ct050062l CCC: $30.25
Published on Web 09/20/2005

The paper is organized as follows. In section II, we briefly describe the models and the methodology employed for performing the simulations. The results are shown in section III. Finally, section IV discusses the main results and extracts some conclusions.

II. Methodology

A. The Model. A 4 × 2 layer of Wyoming type montmorillonite clay was built up by replication of the unit cell given by Skipper et al.[23] This layer has $L_x = 21.12$ Å, $L_y = 18.28$ Å, and $L_z = 6.56$ Å dimensions. The Wyoming type montmorillonite was obtained by isomorphous substitutions of trivalent Al atoms of the octahedral sites by divalent Mg atoms and tetravalent Si by trivalent Al atoms. The unit cell formula of this clay is given by $K_{0.75}nH_2O(Si_{7.75}Al_{0.25})(Al_{3.5}Mg_{0.5})O_{20}(OH)_4$. Size effects were avoided by considering two layers in the simulation box.[4] Periodic boundary conditions were imposed on the three space directions. The initial configuration consists of water molecules randomly placed in the interlaminar spaces and six potassium ions distributed in the interlayer midplanes. These counterions balance the negative charge of the clay framework keeping the system electroneutral.

The rigid TIP4P model was used for water molecules,[20,24] and the water clay interactions were taken from Boek et al.[1] Here, site-to-site intermolecular interactions are given by electrostatic and Lennard-Jones contributions

$$U_{ij} = \sum_{a,b}\left[\frac{q_a q_b}{r_{ab}} + 4\epsilon_{ab}\left[\left(\frac{\sigma_{ab}}{r_{ab}}\right)^{12} - \left(\frac{\sigma_{ab}}{r_{ab}}\right)^{6}\right]\right] \quad (1)$$

where subindexes i and j are for molecules, and a and b run over all sites of each molecule. q_a and q_b are the corresponding site charges, ϵ_{ab} and σ_{ab} are site-to-site specific Lennard-Jones parameters, and r_{ab} is the intersite distance. The Lennard-Jones parameters for single sites are shown in Table 1. Here, those parameters for Si were taken from Marry et al.,[25] and parameters for Al and Mg were assumed to be equal to those of Si. The site-to-site Lennard-Jones parameters are given by the Lorentz−Berthelot rules

$$\sigma_{ab} = \frac{\sigma_a + \sigma_b}{2} \quad (2)$$

$$\epsilon_{ab} = \sqrt{\epsilon_a \epsilon_b} \quad (3)$$

On the other hand, the K−H_2O interactions and those between the oxygens of the clay and potassium ions are based on the ones proposed by Bounds.[26] Bounds potential is chosen since, while simple, it produces K-TIP4P radial distribution functions in agreement with available experimental data and close to those obtained by hybrid quantum mechanics/molecular mechanics (QM/MM) simulations, which naturally account for the many body contributions to the potential.[27] That is, the K−O radial distribution function peaks at 2.86 Å leading to a first shell oxygen coordination number of 7.6, while high accuracy QM/MM simulations performed at density functional theory level (LANL2DZ basis set) give 2.81 Å of K−O distance and 8.3 of coordination number. Experimental results give K−O distances between 2.7 and 3.1 Å and coordination numbers in the wide range of 4−8.[28,29] He fitted the following pair potential for the K−H_2O dispersion−repulsion contribution obtained from ab initio calculations

Table 1: Lennard-Jones Parameters for H_2O−Clay−K Interactions[a]

sites	ϵ (kcal/mol)	σ (Å)
O	0.155	3.1536
H	0.000	0.0000
K	3.630	2.4500
Si	3.150	1.8400
Al	3.150	1.8400
Mg	3.150	1.8400

[a] Exceptions are the K−O and K−H interactions.

$$U_{K-H_2O} = A_{KO}\exp(-b_{KO}r_{KO}) - C_{KO}/r_{KO}^{4} - D_{KO}/r_{KO}^{6} + A_{KH}\exp(-b_{KH}r_{KH_1}) + A_{KH}\exp(-b_{KH}r_{KH_2}) \quad (4)$$

yielding $A_{KO} = 53884.0$ kcal/mol, $b_{KO} = 3.3390$ Å^{-1}, $C_{KO} = 438.0$ kcal Å^4/mol, $D_{KO} = -638.0$ kcal Å^6/mol, $A_{KH} = 5747.0$ kcal/mol, and $b_{KH} = 3.4128$ Å^{-1}. This intersite potential, although somewhat more complicated than the Lennard-Jones type, produces a much better match to the ab initio data.[26] This is clearly seen if one tries to fit eq 4 with a Lennard-Jones type potential. In fact, parameters shown in Table 1 for potassium were obtained by this way, yielding a poor match although reproducing the depth and position of the pair-potential minimum. We observed that the discrepancies are very pronounced at short distances, where the Lennard-Jones potential shows a much harder behavior. This explains why Boek et al. found a very large dehydrated interlaminar space when they employed a Lennard-Jones type pair potential for K−O.[14] Naturally, they overcome this difficulty by employing the pair potential proposed by Bounds.[26]

Nevertheless, since it is crucial for the hybrid Monte Carlo simulations to keep the energy fluctuations as low as possible in order to enlarge the acceptation rate,[30] it is convenient to avoid employing relatively long-range pair potential contributions such as $\sim r^{-4}$, if no Ewald treatment is applied on them. Hence, we refitted to eq 4 the following expression

$$U_{K-H_2O} = A_{KO}\exp(-b_{KO}r_{KO}) - C_{KO}/r_{KO}^{6} + A_{KH}\exp(-b_{KH}r_{KH_1}) + A_{KH}\exp(-b_{KH}r_{KH_2}) \quad (5)$$

by employing a Levenberg−Marquardt algorithm and considering several K−H_2O configurations. The procedure yields $A_{KO} = 120750.2$ kcal/mol, $b_{KO} = 3.4110$ Å^{-1}, $C_{KO} = 5153.8$ kcal Å^6/mol, $A_{KH} = 2109.4$ kcal/mol, and $b_{KH} = 2.8515$ Å^{-1}. We observed that avoiding the $\sim r^{-4}$ term the acceptance rate enlarges more than three times for a small (inner) time step of 0.8 fs. In general, both functions yield similar values of the K−H_2O potential energy. Minima are located practically at the same distance although the depth of the fitted function is 5% larger. For larger distances this difference decreases. To check the obtained interaction potential, a *NPT* simulation containing 216 water molecules, a potassium cation, and a chloride anion was performed at $P = 1$ atm and $T = 293$ K.

The corresponding K−O and K−H radial distribution functions, $g(r)$, and coordination numbers, $n(r)$, were studied. These functions were observed to be very similar to those reported by Bounds.[26] That is, the K−O and K−H main peaks are located at 2.83 and 3.29 Å, respectively, which compare well with their corresponding values of 2.86 and 3.32 Å.[26] The coordination number for the $g(r)$ minimum was found at 7.7 Å, in agreement with his value of 7.6 Å.[26] Moreover, all these values are even closer to the results obtained by QM/MM simulations.[27] Hence, expression 5 seems to be suitable for our purposes.

Finally, we should mention that electrostatic contributions, $\sim r^{-1}$, were treated through the implementation of the Ewald summation formalism. Here the convergence factor was fixed to $5.6/L_{min}$, where L_{min} is the minimum simulation box side. There were set five reciprocal lattice vectors for the directions along the shortest sides and six vectors for the direction along the largest side.[31] The dispersion−repulsion contributions were corrected using the standard methods for homogeneous fluids,[32] and a spherical cutoff of $L_{min}/2$ was imposed.

B. Simulations. Simulations were performed employing the hybrid Monte Carlo (HMC) method.[21,30] This technique allows making global moves while keeping a high average acceptance probability. Global moves are done as follows from molecular dynamics (MD), i.e., by assigning velocities and by using a particular scheme for integrating the Newton's equations of motion. Velocities are assigned randomly from a Gaussian distribution in correspondence with the imposed temperature and in such a way that total momentum equals zero for both interlaminar spaces. To fulfill detail balance condition, the discretization scheme must be time reversible and area preserving.[30] In particular, we employed the multiple time scale algorithm given by Tuckerman and Berne.[33] This algorithm has the property of splitting the forces into short and long range. The Lennard-Jones contribution plus the real part of the electrostatic forces are set as short range, and the reciprocal space contribution of the electrostatic forces is set as long range. To decrease time correlations a new configuration is generated each 10 integration steps. The probability to accept this new configuration is given by

$$P = \min\{1, \exp(-\beta\Delta H)\} \quad (6)$$

where ΔH is the difference between the new and previous configuration Hamiltonians, and β is the inverse of the thermal energy. The long time step is set to 8 times the short time step, and the short time step is chosen to obtain an average acceptance probability of 0.7.[30] This way, we obtained short time steps close to 1.0 and 0.5 fs for systems containing 10 and 100 water molecules per interlaminar space, respectively. As can be seen, the time step shortens with increasing the system size, since energy fluctuations enlarge. This is why HMC is not very efficient for systems counting on a large number of movable sites. This is not our case, since few ions and water molecules are the only contributors to energy fluctuations. This makes HMC a reasonable choice. In fact, the time steps we are obtaining are similar to those usually employed for typical MD calculations.[5,11,20,25]

For sampling in the $NP_{zz}T$ ensemble, after a trial change of particles' positions, a box change is attempted in such a way that the stress normal to the surface of the clays, P_{zz}, is kept constant. For this purpose, box fluctuations are allowed only in the z-direction, and the probability for accepting the new box configuration is given by

$$P = \min\{1, \exp[-\beta(\Delta\mathcal{U} + P_{zz}\Delta V - N\beta^{-1}\ln(V_n/V_o))]\} \quad (7)$$

Here, $\Delta\mathcal{U}$ is the change in the potential energy, ΔV is the volume change, N is the total number of molecules, and V_n and V_o are the new and old box volumes, respectively.[4]

For sampling in an open ensemble, the possibility of insertions and deletions of water molecules has to be considered. Water insertions and deletions were performed by Rosenbluth sampling.[17,21] The $\mu P_{zz}T$ ensemble[21] was used to obtain the equilibrium states when the system is in contact with a reservoir at certain temperature, pressure, and water chemical potential. For this purpose, the algorithm must sample the probability density of finding the system in a particular configuration, i.e.,

$$\mathcal{N}_{\mu PzzT} \propto \frac{V^N \exp\{-\beta[\mathcal{U} - \mu N + P_{zz}V]\}}{\Lambda^{3N}N!} \quad (8)$$

Hence, particle movements, insertions, deletions, and box changes must be done as in a typical NVT, μVT, and $NP_{zz}T$ sampling.[34] In particular, after trying a change of particles' positions, we performed tries of inserting−deleting water molecules. This is done by randomly calling both possible trials in such a way that calls are equally probable. Since accepting insertions or deletions are rare, we repeat this step 10 times or until any insertion or deletion is accepted. In case of refusing the 10 insertion−deletion trials, we performed a box trial move.[35] In this way, the system rapidly evolves to an equilibrium state. For some conditions, however, two free energy local minima appear, which are accessed by handling initial conditions.

III. Results

Results are presented in two subsections. These are *Sampling in the $NP_{zz}T$ Ensemble* and *Sampling in the $\mu P_{zz}T$ Ensemble*. Each part presents the results for ground level conditions, i.e, $T = 298$ K and $P = 1$ atm, and for 4 km of burial depth, i.e, $T = 394$ K and $P = 600$ atm, assuming average gradients of 30 K/km and 150 atm/km.

A. Sampling in the $NP_{zz}T$ Ensemble. Let us first focus on the swelling behavior of the K-montmorillonite hydrates under ground level conditions. This is shown as a plot of the interlaminar distance as a function of the number of water molecules obtained by $NP_{zz}T$ simulations. This curve starts from a dehydrated state having an interlaminar distance of 10.4 Å, which is somewhat larger than the experimental value of 10.1 Å.[36] For a small number of water molecules, the system reaches interlaminar distances above 11 Å, which slightly increases with increasing the number of water molecules producing a plateau. This plateau shows interlaminar distances in the range 11.3−12.3 Å, which are clearly lower than those produced by Na-montmorillonite.[4,9]

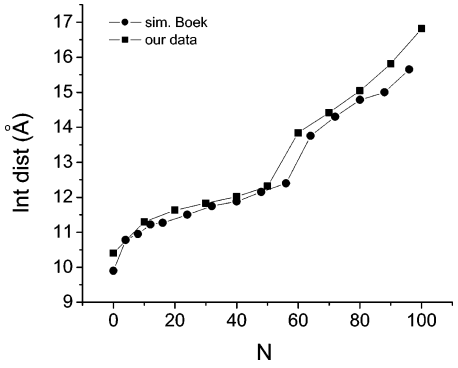


Figure 1. Interlaminar distance as a function of the number of water molecules per interlaminar space.

This might be seen as something unexpected, since potassium ions are much larger than sodium ions. As we comment ahead, this is due to the different interlayer structures that sodium and potassium ions generate. For 60 water molecules per interlaminar space, there is a jump from a single water layer to a double water layer, which yields an interlaminar distance of 13.9 Å. For larger amounts of water the system increases almost linearly, producing a very small step when jumping from a double to a triple water layer structure.

Our findings are similar to those reported by Boek et al.[14] We included them in Figure 1 to make the comparison easy. It is seen that trends are practically equal. This indicates that differences in models and methods are not very important. Nevertheless, they always obtain slightly smaller interlaminar distances for a given number of water molecules. This difference is always lower than 2%, except for the dehydrated state and the points close to 100 water molecules, where differences close to 5% are observed. Since in our case the jump from a single to a double layer is obtained for a lower amount of interlaminar water, the points that correspond to 55−60 water molecules also show a larger difference. Anyway, differences seem reasonable taking into account the differences in pair potential definitions, box sizes, and methodologies.

The structure of the interlaminar space for the system containing 40 water molecules per interlaminar space is shown in Figure 2. A high and narrow oxygen peak at the interlaminar midplane, three hydrogen peaks, one coinciding with the oxygen peak and the other two symmetrically situated at both sides, and a single potassium ion peak, also at the interlaminar midplane, are observed. Most of these results agree with others previously reported.[14,16,17] It should be noted that this structure contrasts with that one of the Na-montmorillonite single layer hydrate. In this case, sodium ions are distributed at the sides of the oxygen peaks, closer to the clay sheets, some of them being strongly attached to the clay surface and so forming inner-sphere surface complexes.[15] This makes water molecules to cluster in two layers joined at the interlaminar midplane, producing an oxygen double peak.[21,37] Naturally, this effect widens the interlaminar space, despite the smaller size of sodium ion.

The radial distribution functions and coordination numbers for K−O sites were also studied. Main $g(r)$ peaks are situated at 2.77, 2.90, and 2.83 Å for water, clay, and total oxygen

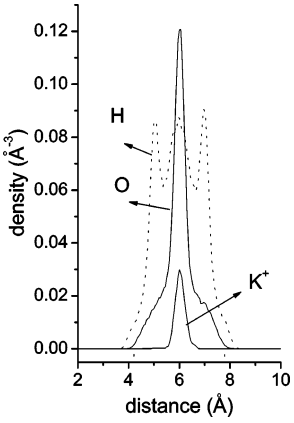


Figure 2. Oxygen, hydrogen, and potassium density profiles of the interlaminar space. The water amount was fixed to 40 molecules per interlaminar space.

sites, respectively. Although they are close to the one observed for bulk potassium water solution, a contraction is seen for the K−O water distance as a consequence of confinement. On the other hand, the relatively large K−O clay distance is due to the distribution of the potassium ions around the midplane of the interlayer space. The coordination numbers for the first shell of oxygen atoms are 5.1, 5.0, and 10.1 for water, clay, and total oxygen sites, respectively. The clay and total coordination numbers are somewhat inflated due to the smaller separation of the O−O sites of the clay. This explains why the total coordination number is larger than the one found for bulk. It should be noted the large contribution of the clay to the total coordination number. This suggests that potassium ions are interacting with both clay sheets at the same time. In addition, the large K−O clay distance may be indicating that potassium ions are directly contributing to attract both layers toward the midplane and hence holding the sheets close to one another.

All the already pointed features are illustrated in Figure 3. For instance, it is observed that potassium ions are practically centered on the interlayer midplane with an approximate average of 5 water molecules surrounding each ion. It is also possible to see how ions tend to separate from each other, and that one of them is always close to a tetrahedral aluminum (these sites are not highlighted in the figure). Moreover, it is shown that there are no water molecules interposed between the potassium ions and the clay sheets. Hence, potassium ions are behaving as inner-sphere complexes but simultaneously with both clay layers. To see this even clearer, Figure 4 was built by rotating and zooming in Figure 3. Here, a potassium ion and its inner water shell are shown. As can be seen, only 5 water molecules surround the ion that coordinates with two oxygen atoms from each clay. For this particular case, are also seen average K−O distances of 2.81 and 3.04 Å for water and clay, respectively.

As mentioned before, for 60 water molecules a double layer hydrate is formed. This double layer hydrate becomes fully developed for 80 water molecules, where an interlayer distance close to 15.0 Å is observed. For this system, the oxygen, hydrogen, and potassium profiles are shown in Figure 5. Here, the two oxygen peaks are a signature of the

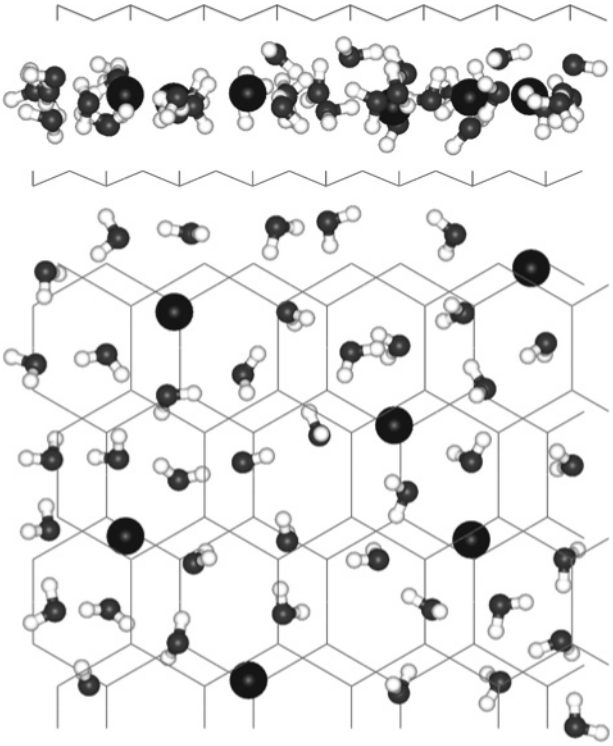


Figure 3. Snapshot of an equilibrated system having 40 water molecules. H sites are white, O are gray, and K are black. The wireframe represents the clay structure. The topmost image is a side view, and the lower image is the corresponding top view.

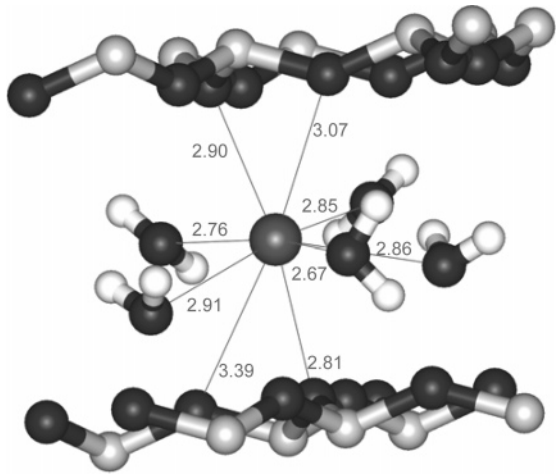


Figure 4. Zoom in of a potassium ion and its coordination shell taken from Figure 3. Only water molecules having K−O distances smaller than 3.7 Å are shown. Distances in the figure are given in Å.

double water layer structure. For each oxygen peak a potassium peak and two hydrogen peaks are found. The potassium peak is found almost coinciding with the oxygen peak but slightly displaced at the side closer to the clay sheet. This suggests the formation of inner-sphere complexes. On the other hand, a small hydrogen peak is found at the side closer to the clay sheet and a larger one closer to the interlaminar midplane. Hence, the third hydrogen peak found for the one layer case is missing here.

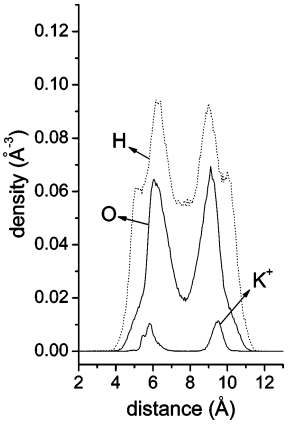


Figure 5. Oxygen, hydrogen, and potassium density profiles of the interlaminar space. The water amount was fixed to 80 molecules per interlaminar space.

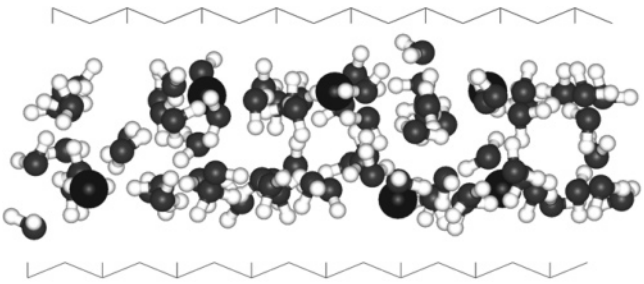


Figure 6. Snapshot of an equilibrated system having 80 water molecules per interlaminar space. H sites are white, O are gray, and K are black. The wireframe represents the clay structure.

For this configuration, main $g(r)$ peaks are located at 2.83, 2.75, and 2.79 Å for water, clay, and total oxygen sites, respectively. Hence, the K−O distance for water is not contracted anymore but equal to the bulk water K−O distance. On the contrary, the K−O distance for the clay decreased 0.17 Å. Moreover, this distance is found to be almost constant for systems having more than 60 water molecules, strongly suggesting that this is the natural average K−O distance for a potassium ion attached to the siloxane surface. Hence, the distance found for the single layer hydrate would be elongated as a result of the pressure the water molecules produce on the clay surfaces. The corresponding coordination numbers are 6.8, 2.8, and 9.6 for water, clay, and total oxygen sites, respectively. As can be seen, the coordination number for the clay is much lower than the value found for the single layer hydrate. This means that potassium ions are coordinated to only one clay layer for the double layer hydrate. Finally, and as expected, K−O water coordination numbers increase with the number of water molecules, whereas K−O clay and K−O total coordination numbers decrease. A snapshot of this two layer hydrate is shown in Figure 6. We should mention that the structure of the inner-sphere complexes we found in this case are very similar to those already reported by Sposito et al.[19] (not shown here).

For burial conditions the interlaminar distance increases for a given number of water molecules per interlaminar space. This is shown in Figure 7, where it shows the

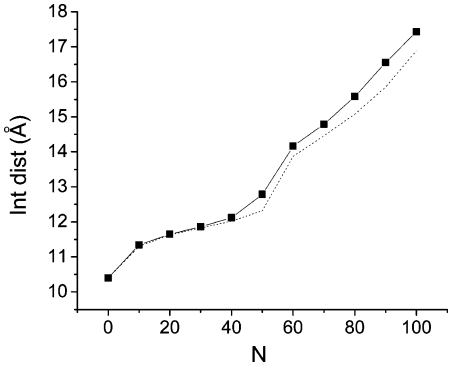


Figure 7. Interlaminar distance as a function of the number of water molecules per interlaminar space for burial conditions. For comparison, the dotted line corresponds to ground conditions (from Figure 1).

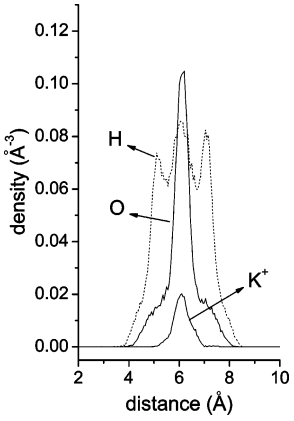


Figure 8. Oxygen, hydrogen, and potassium density profiles of the interlaminar space. The water amount was fixed to 40 molecules per interlaminar space for burial conditions.

interlaminar distance as a function of the number of water molecules for burial and for ground level conditions. This increment is more pronounced for large amounts of water. This is an expected behavior since water molecules occupy larger effective volumes at burial conditions.[9] Consequently, profiles turn less sharp, i.e, peaks broad and shorten, as it is shown in Figure 8. This was seen experimentally by Skipper et al., although for a Na-montmorillonite system.[38] On the other hand, trends for both the ground and burial data are very similar, i.e, jumps are found at equal numbers of water molecules. This last finding differs from the one found for Na-montmorillonite, where the single layer to double layer jump occurs for 60 or 50 water molecules, depending on the burial depth.[9]

B. Sampling in the $\mu P_{zz}T$ Ensemble. Sampling in this ensemble allows the system to reach equilibrium with a reservoir whose temperature, pressure, and, in our case, water chemical potential are fixed. We employed the expression $\beta\mu = \beta\mu_0 + \ln(p/p_0)$, where p_0 is the vapor pressure at equilibrium with liquid water whose chemical potential is μ_0, and p is the vapor pressure. For the TIP4P water model with $T = 298$ K and $P = 1$ atm, we employed $\beta\mu_0 = -17.4$. This value was obtained by *NPT* simulations of bulk TIP4P water[21] and using the Rosenbluth sampling method explained elsewhere.[17,34] It is in good agreement with the value reported by Chávez-Páez et al.,[4] and it is 2.5% larger than the one

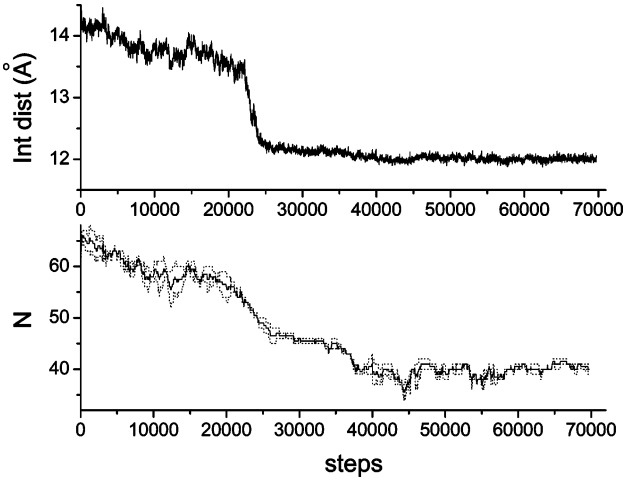


Figure 9. Step evolution of the interlaminar distance and number of water molecules obtained by $\mu P_{zz}T$ sampling. In the lower plot, dashed lines correspond to each interlaminar space, whereas the solid line is the average. Initial conditions were 16 Å and 60 water molecules per interlaminar space. Established conditions were $T = 298$ K, $P = 1$ atm, and $p/p_0 = 0.4$.

reported by Tambach et al.[20] For $T = 394$ K and $P = 600$ atm, $\beta\mu_0 = -13.4$ was obtained.[21]

The evolution of the interlaminar distance and number of water molecules for ground level conditions and for $p/p_0 = 0.4$ is shown in Figure 9. Here, the initial conditions were 16 Å for the interlaminar distance and 60 water molecules per interlaminar space. It is observed that the system reaches approximately 65 water molecules and 14.3 Å of interlaminar distance at the initial simulation steps. This happens so fast since the system is initially very far from equilibrium. Immediately after, the system starts slowly losing water molecules and decreasing its interlaminar distance on its way toward equilibrium. It is observed that, once the system reaches 13.5 Å of interlaminar space, it quickly falls down to 12.2 Å, losing many water molecules during the process. This is a signature of the transition from a double layer hydrate to a single layer hydrate. The amount of water that signals the transition goes from 55 to 45 molecules. This agrees with the $NP_{zz}T$ results, where a single to a double layer transition is observed in the range of 40−60 water molecules and 12.2−14.3 Å. After the transition, it is observed that the system takes several steps to finally reach equilibrium at approximately 40 000 steps. In this case, sampling was performed in the step range of 40 000−70 000. For this particular run 12.0 Å of interlaminar space and 39.3 water molecules were obtained.

Many runs were performed in order to build up Figure 10. For ground level conditions, this figure shows the interlaminar distance and number of water molecules at equilibrium with reservoirs having different water vapor pressures. For zero vapor pressure, no matter what the established initial conditions, the system is forced to eliminate all of its water content, and so, the dehydrated state is yielded. This has 10.40 Å of interlaminar space, as was found in the preceding section. For increasing the vapor pressure, the system starts to uptake water molecules from the reservoir

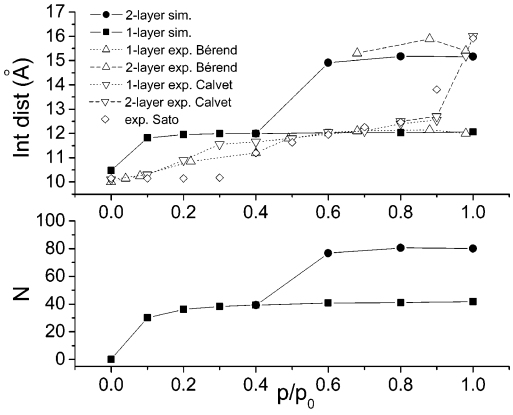


Figure 10. Interlaminar distance and number of water molecules per interlaminar space as a function of the vapor pressure for ground level conditions. Symbols □ and ○ correspond to initial conditions of 10 water molecules and 11.5 Å of interlaminar distance and 60 water molecules and 16.0 Å of interlaminar distance, respectively.

producing larger interlaminar distances. Nevertheless, the first water layer saturates for vapor pressures over $0.2p_0$, and so, a plateau is generated. Again, this single layer hydrate is produced for $p/p_0 \leq 0.4$ for all initial conditions.

For higher vapor pressures, however, at least two equilibrium states are yielded. That is, depending on the established initial conditions, the system may produce a double or a single layer hydrate. Hence, the topmost lines of both plots of Figure 10 correspond to an initial configuration of 60 water molecules and the others to an initial configuration of 10 water molecules. These two equilibrium states are stable up to water vapor saturation, producing an open hysteresis cycle. The single layer state yields interlaminar distances ranging in 11.82−12.06 Å and amounts of water in the range of 30.1−41.8 molecules. The double layer state produces interlaminar distances in the range 14.91−15.16 Å and numbers of water molecules ranging in 76.7−80.6. These data contrast with those obtained for Na-montmorillonite.[21] In this case, both initial conditions predict only a single layer hydrate for $p/p_0 \leq 0.2$. In addition, this single layer hydrate yields values of the interlaminar distance in the range of 12.34−12.73 Å, having 34.6−47.5 water molecules.[21] Therefore, it becomes evident the K-montmorillonite tendency to produce single layer hydrates even for relatively high vapor pressures and the small interlaminar distance and amounts of water it yields. On the other hand, the double layer plateaus for potassium and sodium do not behave very differently for high vapor pressures.

We should also compare our results with the experimental data obtained by Bérend et al.,[39] Calvet,[36] and Sato et al.[40] For this purpose, these data are also included in Figure 10. It should be pointed out that these data are direct measurements of interlaminar distances against the relative vapor pressure. Here, Bérend et al.'s data were obtained as the evolution of the interlaminar distance for a dehydration process, by starting from a double or a single layer hydrate. Similarly, Calvet presents his data by starting from a double or a single layer hydrate, but for a hydration process. Sato et al.'s data are also obtained by hydration.

The best agreement between our simulations and experiments is produced when comparing with the data of Bérend et al.[39] We highlight the double layer hydrate formation for the range of relative vapor pressures of 0.6−1.0; the single layer hydrate which keeps stable even for saturated vapor pressures; the very good agreement for the basal space for the single layer hydrate; and the relative good agreement for the double layer hydrate interlaminar distance, close to 15.5 Å. Calvet's data also are in general agreement with our simulations. He observed a very stable single layer hydrate, with interlaminar distances ranging in 11.8−12.5 Å. Nevertheless, he obtained a very unstable double layer hydrate. Only for relative vapor pressures of 0.9−1.0 this hydrate was observed. Its interlaminar distance is in the range of 15.2−16.0 Å. Sato et al.'s data also lead to similar interlaminar distances for the single and double layer hydrate. Finally, the single layer hydrate interlaminar distances also well agree with data reported by Brindley and Brown.[41] For 0.32, 0.52, and $0.79p_0$ of vapor pressure they reported 11.9, 11.9, and 12.1 Å, respectively.

On the other hand, the most important discrepancy with experiments is that they obtained interlaminar distances close to 10.0 Å for the dehydrated state, which seems to be stable up to a relative vapor pressure ranging in 0.1−0.3. We think that if we were capable of reproducing the correct dehydrated distance, i.e., 10.0 instead of 10.4 Å, this state would probably become stable for small vapor pressures, as they found.

Experimentalists agree that for relative vapor pressures ranging from 0.0 to 0.4, a mixture of the dehydrated state and the single layer hydrate coexists, i.e., there is interstratification. Similarly, they observed the coexistence of double and single layer hydrates, for high relative vapor pressures. In fact, Calvet refers to his own data as "apparent distances"; Bérend et al. conclude that all montmorillonites (they studied Li, Na, K, Rb, and Cs-montmorillonites) form interstratified hydrates; and Sato et al. observed several "nonintegral basal reflections", which are interpreted as "random or segregated-type interstratification of collapsed and expanded layers". An example is the 13.81 Å of interlaminar distance he obtained for the relative vapor pressure of 0.9. This point does not match either a single or a double hydrate interlaminar distance, as it is clearly seen in Figure 10. It is important to mention that the general belief is that interstratification occurs due to chemical heterogeneities of the clay layers. It was already proven by simulations that changes on the positions of the clay substitutions produce different interlaminar distances (although differences are not very pronounced). Hence, these heterogeneities surely lead to quasihomogeneous states and, probably, to interstratified ones. On the other hand, a perfect system like ours shows the double and the single layer hydrates to be stable for identical conditions. Hence, we do not see any reason for this not to occur in a real system. In other words, we think that this is another source of interstratification, which arises just as a consequence of the inherent thermodynamics of the perfect system. In fact, similar conclusions are deduced by Tambach et al.[22]

Table 2: $g(r)$ Main K−O Peak Positions, p, and First Shell Coordination Numbers, n, for Systems under Different Water Vapor Pressures, p/p_0[a]

	single water layer				double water layer			
p/p_0	p_w	p_c	n_w	n_c	p_w	p_c	n_w	n_c
0.1	2.77	2.83	4.83	5.32				
0.2	2.77	2.85	4.96	5.11				
0.3	2.77	2.88	5.19	4.94				
0.4	2.77	2.90	5.17	4.96				
0.6	2.76	2.89	5.15	4.93	2.80	2.75	6.62	3.16
0.8	2.76	2.88	5.24	4.85	2.82	2.77	6.77	2.82
1.0	2.76	2.90	5.19	4.89	2.81	2.75	6.71	2.95

[a] Subindexes w and c refer to water and clay, respectively. Peak positions are given in Å.

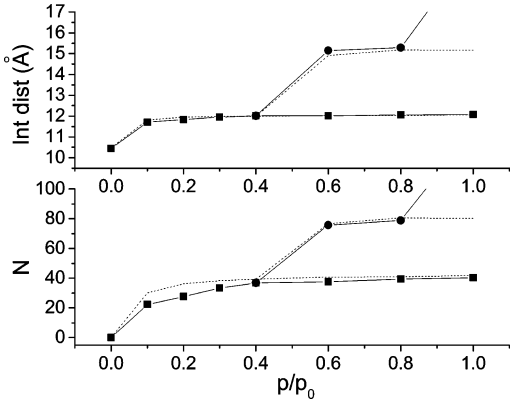


Figure 11. Interlaminar distance and number of water molecules per interlaminar space as a function of the vapor pressure for burial conditions. For comparison, the dotted line corresponds to ground conditions (from Figure 10).

Table 2 shows the main peak positions and the first shell coordination numbers for the systems equilibrated at different relative vapor pressures and for a single and a double water layer configurations. This was done to prove that the shifts of the water and clay peaks mentioned in the preceding section are indeed significant over a wide range of vapor pressures. As can be seen for the single layer hydrate, the K−O water clay position is practically constant and equal to the one found from the $NP_{zz}T$ sampling. Also the double layer p_w is close to the value found in the previous section, i.e, similar to the K−O bulk water position. On the other hand, p_c increases for increasing the relative water pressure, reaching a plateau at $p/p_0 \geq 0.3$. The plateau value is close to 2.89 Å, which is also similar to the value found in the preceding section. The decrease of this distance with p/p_0 is a consequence of the decrease of the interlaminar distance with it. For the double layer configuration it also confirmed a value close to 2.76 Å. Hence, the shifts of these peaks are relevant and not just casual values obtained for some particular conditions. From Table 2 it is seen that the values for the coordination numbers also agree with those presented in the previous section.

To evaluate the effect of the burial depth on the K-montmorillonite swelling curves, Figure 11 was built for $T = 394$ K and $P = 600$ atm. Additionally, results for ground level conditions were included to make the comparison easy. As can be seen, for initial conditions of 10 water molecules and a small interlaminar space, the interlaminar distance is slightly smaller than those obtained for ground level conditions. This combines with the fact that water molecules occupy a larger effective volume for burial conditions, and so, the number of water molecules decreases. In other words, the single layer hydrate of K-montmorillonite dehydrates under burial conditions. Furthermore, this single layer hydrate is stable even for a saturated water vapor pressure. Again, this contrasts with the results we obtained for Na-montmorillonite,[21] where the single layer hydrate was found to be unstable for large water activities. Moreover, for $p/p_0 = 1.0$, the K-montmorillonite single layer hydrate yields 12.07 Å of interlaminar space and 40.4 water molecules, which seems to be far from the transition range of 45−55 water molecules previously found.

The double layer hydrate behaves differently. That is, it keeps constant its water content, and it slightly increases its interlaminar distance. As found for ground level conditions, it collapses forming the single layer hydrate for $p/p_0 < 0.6$. On the other hand, for a saturated vapor pressure the system monotonically increases its amount of water as the simulation evolves. In this last case we stopped the simulations when reaching 180 water molecules, and thus, we assumed that the system entered the osmotic regime.

In summary, these results indicate that once formed the K-montmorillonite single water hydrate will not be destabilized neither at ground level conditions nor at burial depths. In addition, if the reservoir vapor pressure fells down $0.4p_0$ the single water hydrate will form.

IV. Discussion and Conclusions

Single and double layer hydrates of K-montmorillonite were studied by means of $NP_{zz}T$ and $\mu P_{zz}T$ simulations. For that purpose a hybrid Monte Carlo scheme was employed. Most results from the $NP_{zz}T$ sampling just confirm those previously reported elsewhere,[14,15,17−19] suggesting that differences in models and methods are not very important.

We found different clay−water−ion complexes for the single water hydrate. They consist of an average of 5 water molecules surrounding a potassium ion that additionally coordinates with two oxygen atoms of each closest clay layer. These potassium ions are placed at the interlayer midplane, and so, distances between them and the oxygens of the clay are slightly larger than those observed for the double layer hydrate. It should be mentioned that these midplane ions were reported by previous works as forming outer-sphere complexes.[14,16] In fact, Chang et al.[16] found that these midplane ions have a relatively large mobility, since potassium ions neither strongly coordinate to water molecules nor strongly interact with the clay surfaces. This is, indeed, characteristic of an outer-sphere complex. This explains why they call them in this way. Nonetheless, the inner-sphere definition given elsewhere[15] says literally "The surface complex is inner-sphere if the cation is bound directly to a cluster of surface oxygen ions, with no water molecules interposed". Accordingly, our midplane potassium ions form inner-sphere complexes but simultaneously with both clay layers. Nevertheless, we expect them to have a diffusivity similar to that reported by Chang et al.[15] This is a clear difference

between this kind of inner-sphere complex and the typical inner-sphere complex, where ions are strongly coordinated to one of the surfaces.[15] This simultaneous coordination of potassium ions with oxygen atoms of the two adjacent clay layers plus their relatively large coordination distances suggests to us that these complexes are attracting the clay layers toward the interlayer midplane, aiding to keep them together.

Although the previous finding seems to explain the stability of the single layer hydrate found experimentally, we should also reproduce computationally this behavior. Hence, $\mu P_{zz}T$ simulations were carried out. This ensemble lets interchange water with a given reservoir and volume fluctuations. So, for any given water vapor pressure of the reservoir, an average amount of water and an average interlaminar distance are found. However, two different equilibrium states may be produced, pointing to the formation of free energy local minima,[20] which are accessed by handling initial conditions. This produces hysteresis loops.[20] We found that for initial conditions close to the dehydrated state, a single water layer is always obtained for any nonzero vapor pressure, signaturing its high stability. The amount of water of this single layer slightly increases as the vapor pressure increases. This was obtained for ground level and burial conditions. The interlaminar distance for the ground level state is always close to 12.0 Å. The number of water molecules was found to be close to 38 for ground level conditions and about 36 for burial conditions. This is due to the larger effective volume the water molecules occupy at higher temperatures.

On the other hand, the stability of the double layer hydrate differs from that one of the single layer. For both conditions studied, it was seen that the double layer collapses to form the single layer for vapor pressures under $0.4p_0$. In addition, this double layer was found to be unstable for burial depth and for saturated vapor pressures, producing a hydrated state in the osmotic regime.

We should point out that these results agree with those of Boek et al.[14] In their Figure 3 is shown the potential energy of the interlayer water as a function of the number of water molecules. It can be seen that the potassium curve is very different than the sodium and lithium ones only for small amounts of water (it shows much higher energy values, even well above the water bulk reference). This supports our finding of an extremely stable single water hydrate. Nevertheless, they claim that potassium is a good swelling inhibitor due to its ability to migrate and bind to the clay surfaces. Hence, the negatively charged surface becomes screened, making its inherent repulsion less effective. We agree that this mechanism explains the relatively high stability of double layer hydrates and even explains the stability of K-montmorillonite hydrates in the osmotic regime. However, it cannot explain the remarkable stability of the single layer hydrate, since at least a double layer is needed to obtain the binding between ions and surfaces. This fact makes us think that midplane potassium ions are playing an important role in the stability of single layer hydrates.

Despite the thinking that the potassium simultaneously binding to adjacent layers is the key to the single layer hydrate stability, we do not think it is the only factor. Lowest interaction energy for the pair K-TIP4P water is close to −20 kcal, which is higher than that for Na-TIP4P of approximately −25 kcal. This means that potassium ion does not strongly interact with water, and, therefore, it easily loses water from its coordination shell. In other words, the water chemical potential of the interlayer is not very negative. That is, water may be more comfortable in bulk than in the interlayer, surrounding the ions. As mentioned by Sposito et al.,[19] potassium ions show a kind of "hydrophobic character", in the sense that they tend to interact with water molecules not only through their positive charge but also through solvent cage formation. Consequently, some of the water oxygen atoms may be easily exchanged by oxygen atoms from the clay surface.

Finally, for drilling purposes and based on our results, it seems to not be enough to add potassium to the mud in order to guarantee the single layer water formation and, in this way, avoid swelling. One should simultaneously decrease the mud water activity to safely produce the single layer state. Once obtained, the clay will not swell at all.

Acknowledgment. This research was supported by Instituto Mexicano del Petróleo Grant D.00072.

References

(1) Boek, E. S.; Conveney, P. V.; Skipper, N. T. *Langmuir* **1995**, *11*, 4629−4631.

(2) Boek, E. S.; Sprik, M. *J. Phys. Chem. B* **2003**, *107*, 3251−3256.

(3) Chávez-Páez, M.; de Pablo, L.; de Pablo, J. J. *J. Chem. Phys.* **2001**, *114*, 10948−10953.

(4) Chávez-Páez, M.; Van Workum, K.; de Pablo, L.; de Pablo, J. J. *J. Chem. Phys.* **2001**, *114*, 1405−1413.

(5) Chou Chang, F.; Skipper, N. T.; Sposito, G. *Langmuir* **1995**, *11*, 2734−2741.

(6) Chou Chang, F.; Skipper, N. T.; Sposito, G. *Langmuir* **1997**, *13*, 2074−2082.

(7) Hensen, E. J. M.; Smit, B. *J. Phys. Chem. B* **2002**, *106*, 12664−12667.

(8) Marry, V.; Turq, P. *J. Phys. Chem. B* **2002**, *107*, 1832−1839.

(9) Odriozola, G.; Guevara, F. *Langmuir* **2004**, *20*, 2010−2016.

(10) Smith, D. E.; Wang, Y.; Whitley, H. D. *Fluid Phase Equilib.* **2004**, *222−223*, 189−194.

(11) Sutton, R.; Sposito, G. *J. Colloid Interface Sci.* **2001**, *237*, 174−184.

(12) Whitley, H. D.; Smith, D. E. *J. Chem. Phys.* **2004**, *120*, 5387−5395.

(13) Young, D. A.; Smith, D. E. *J. Phys. Chem. B* **2000**, *104*, 9163−9170.

(14) Boek, E. S.; Coveney, P. V.; Skipper, N. T. *J. Am. Chem. Soc*. **1995**, *117*, 12608−12617.

(15) Chang, F.; Skipper, N. T.; Refson, K.; Greathouse, J.; Sposito, G. In *ACS Symposium Series No. 715*; Schultz, L., van Olphen, H., Mumpton, F., Eds.; American Chemical Society: Washington, DC, 1999.

(16) Chou Chang, F.; Skipper, N. T.; Sposito, G. *Langmuir* **1998**, *14*, 1201−1207.

(17) Hensen, E. J. M.; Tambach, T. J.; Bliek, A.; Smit, B. *J. Chem. Phys*. **2001**, *115*, 3322−3329.

(18) Park, S.; Sposito, G. *J. Phys. Chem. B* **2000**, *104*, 4642−4648.

(19) Sposito, G.; Skipper, N. T.; Sutton, R.; Park, S. H.; Soper, A. K.; Greathouse, J. *Proc. Natl. Acad. Sci.* **1999**, *96*, 3358−3364.

(20) Tambach, T. J.; Hensen, E. J. M.; Smit, B. *J. Phys. Chem. B* **2004**, *108*, 7586−7596.

(21) Odriozola, G.; Aguilar, F.; López-Lemus, J. *J. Chem. Phys.* **2004**, *121*, 4266−4275.

(22) Tambach, T. J.; Bolhuis, P. G.; Smit, B. *Angew. Chem., Int. Ed.* **2004**, *43*, 2650−2652.

(23) Skipper, N. T.; Chou Chang, F.; Sposito, G. *Clays Clay Miner.* **1995**, *43*, 285−293.

(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys*. **1983**, *79*, 926−935.

(25) Marry, V.; Turq, P.; Cartailler, T.; Levesque, D. *J. Chem. Phys*. **2002**, *117*, 3454−3463.

(26) Bounds, D. G. *Mol. Phys*. **1985**, *54*, 1335−1355.

(27) Tongraar, A.; Liedl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1998**, *102*, 10340−10347.

(28) Neilson, G. W.; Skipper, N. T. *Chem. Phys. Lett*. **1985**, *114*, 35−38.

(29) Ohtaki, H.; Radnai, T. *Chem. Rev.* **1993**, *93*, 1157−1204.

(30) Mehlig, B.; Heermann, D. W.; Forrest, B. M. *Phys. Rev. B* **1992**, *45*, 679−685.

(31) Alejandre, J.; Tildesley, D. J.; Chapela, G. A. *J. Chem. Phys.* **1994**, *102*, 4574−4583.

(32) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids;* Clarendon: Oxford, 1986.

(33) Tuckerman, M.; Berne, B. J. *J. Chem. Phys.* **1992**, *97*, 1990−2001.

(34) Frenkel, D.; Smit, B. *Understanding molecular simulation;* Academic: New York, 1996.

(35) Different types of movements should be called randomly, otherwise detail balance is violated.[34] The data published here and in refs 9 and 21 were obtained by following this sequential algorithm. However, some data were checked by means of the corresponding random algorithm, and the same results were obtained.

(36) Calvet, R. *Ann. Agron*. **1973**, *24*, 77−133.

(37) de Pablo, L.; Chávez, M. L.; Sum, A. K.; de Pablo, J. J. *J. Chem. Phys*. **2004**, *120*, 939−946.

(38) Skipper, N. T.; Williams, G. D.; de Siqueira, A. V. C.; Lobban, C.; Soper, A. K. *Clay Miner*. **2000**, *35*, 283−290.

(39) Bérend, I.; Cases, J. M.; François, M.; Uriot, J. P.; Michot, L.; Masion, A.; Thomas, F. *Clays Clay Miner*. **1995**, *43*, 324−336.

(40) Sato, T.; Watanabe, T.; Otsuka, R. *Ann. Agron*. **1992**, *40*, 103−113.

(41) Brindley, G. W.; Brown, G. *Crystal Structures of Clay Minerals and their X-ray Identification*, 5th ed.; Mineralogical Society: London, 1984.

CT050062L

JCTC Journal of Chemical Theory and Computation

Hydrogen Bond Properties and Dynamics of Liquid−Vapor Interfaces of Aqueous Methanol Solutions

Sandip Paul and Amalendu Chandra*

Department of Chemistry, Indian Institute of Technology, Kanpur, India 208016

Received April 13, 2005

Abstract: The hydrogen bonded structure and dynamics of liquid−vapor interfaces of aqueous methanol solutions of varying compositions are investigated by means of molecular dynamics simulations. The dynamical aspects of the interfaces are investigated in terms of the single-particle dynamical properties such as the relaxation of velocity autocorrelation and the translational diffusion coefficients along the perpendicular and parallel directions and the dipole orientational relaxation of the interfacial water and methanol molecules and also in terms of the relaxation of water−water, water−methanol, and methanol−methanol hydrogen bonds at interfaces at 298 K. The results of the interfacial dynamics are compared with those of the corresponding bulk phases. The inhomogeneous density, anisotropic orientational profiles, surface tension, and the pattern of hydrogen bonding are calculated in order to characterize the location, width, microscopic structure, and the thermodynamic aspects of the interfaces and to explore their effects on the interfacial dynamical properties of water and methanol molecules.

1. Introduction

A detailed knowledge of the molecular properties of liquid−vapor interfaces of molecular liquids is important in the understanding of equilibrium and dynamical aspects of various chemical processes that occur at such interfaces. The present paper deals with the equilibrium and dynamical behavior of liquid−vapor interfaces of aqueous methanol solutions of varying composition. Monohydroxyl alcohol−water mixtures are interesting not only because of their ubiquitous nature but also due to their importance as model systems. The amphiphilic nature of alcohol molecules makes them excellent probes for studying the structure of aqueous solutions since they strongly interact with water through hydrogen bonding. Also, depending on the size of their alkyl groups, the alcohol molecules perturb the water structure through hydrophobic hydration as well. The molecular structure of liquid water consists of a tetrahedral network of hydrogen bonds. X-ray and neutron diffraction studies of liquid methanol[1,2] have shown that each methanol molecule engages in approximately two hydrogen bonds giving rise to a chainlike structure similar to that found in the solid state. The chainlike structure has also been found in recent first principles simulations of liquid methanol.[3] When mixed with water, it has been found that the water−methanol mixtures show many interesting nonideal behavior in terms of both their equilibrium and dynamical properties.[4−12] Mixing of methanol with water causes structural changes in the liquid due to rather strong interactions of water with the alcohol molecules. Several simulations have been carried out on dilute solutions of methanol in water.[13−16] In particular, we note the Monte Carlo study of Jorgensen and Madura[13] who used a methanol potential[16] with explicit methyl hydrogens to study the solvation and conformation of methanol in water. The preferred conformation of a methanol molecule in water was found to be staggered, and the water molecules were found to form a cage around the methyl group. Subsequently, experimental studies based on neutron diffraction and isotopic substitution provided a detailed description of the structure of a hydration shell around a methanol molecule.[17] The water molecules were found to form a loose hydrogen bonded shell around the methanol at a carbon-to-water distance of ≈3.7 Å supporting the picture of cagelike hydration of methanol in aqueous solutions.

There have also been a number of simulation studies on water−methanol mixtures of higher methanol concentrations in bulk liquid phases.[18−25] However, studies of liquid−vapor interfaces of water−methanol mixtures have received relatively less attention. Matsumoto et al.[26] studied the structural

10.1021/ct050098d CCC: $30.25
Published on Web 10/12/2005

and thermodynamical aspects of liquid−vapor interfaces of water−methanol mixtures for a wide range of concentrations by using molecular dynamics simulations. They reported that the outermost layer of the interface is saturated with methanol even at low bulk concentrations of methanol. They also found that the density of water is enhanced as compared to the bulk density just inside the adsorbed methyl layer. The work of Matsumoto et al.[26] involved nonpolarizable models for both water and methanol molecules. Very recently, polarizable models have also been employed to study the liquid−vapor interfaces of pure methanol[27] and also of water−methanol mixtures.[28] These studies provided information on the variation of dipole moment across the interfaces which otherwise could not be obtained by using nonpolarizable models.

The experimental studies of the molecular orientational behavior of liquid−vapor interfaces primarily involve the surface specific methods such as second harmonic generation (SHG) and sum frequency generation (SFG). These methods have been used to investigate molecular orientation at a wide variety of interfaces involving pure water or aqueous solutions[29−45] including water−methanol mixtures.[29−31] It was found that the orientational order of methanol molecules at the surface of water−methanol increases with a decrease of methanol concentration. Very recently, Rao and co-workers[46] investigated the surface composition of water−alcohol mixtures by a combination of molecular beam and mass spectroscopic measurements. In this study, which was the first direct experimental verification of surface enrichment, alcohols of varying chain length were considered including methanol. The surfaces were found to be enriched by alcohol as compared to the bulk liquid composition, and the extent of enrichment was found to depend on the chain length of the alcohol molecules. Even for the mixture of water with the shortest alcohol of methanol, the surfaces were found to be enriched by the alcohol molecules.

We note that all the existing theoretical and experimental studies on liquid−vapor interfaces of water−methanol mixtures deal with their structural and thermodynamical properties. The dynamical properties, both of single molecules and hydrogen bonded pairs, at the liquid−vapor interfaces of water−methanol mixtures have not yet been investigated. We note that, for water−methanol mixtures, the dynamics of the interfaces is expected to be intimately related to the structure and energetics of hydrogen bonds that are present at such interfaces. Thus, it would be worthwhile to make a detailed molecular-level investigation of the hydrogen bonding and dynamical properties of liquid−vapor interfaces of water−methanol mixtures. The present work makes a contribution toward this end.

In this work, we have carried out molecular dynamics simulations of liquid−vapor interfaces of water−methanol solutions at varying composition. Our main focus has been to calculate the hydrogen bond and dynamical properties such as the average number of hydrogen bonds per molecule in different regions, the lifetimes of hydrogen bonds, the relaxation of the perpendicular and parallel components of velocity autocorrelation function and diffusion, and also the relaxation of single-particle dipole orientational correlation function of the interfacial water and methanol molecules. The dynamical properties of the interfaces are compared with those of the bulk phases. In addition, we have also calculated the inhomogeneous density and orientational profiles, surface tension, and also the distribution of hydrogen bonds as the dynamical properties of the interfaces are intimately related to these equilibrium quantities. Also, the density profiles help us to characterize the location and thickness of the interfaces.

Table 1. Values of the Lennard-Jones and Electrostatic Interaction Potential Parameters of Water and Methanol

	atom/ion	σ (Å)	ϵ (kcal/mol)	charge (e)[a]
water (SPC/E model)	O	3.169	0.1554	−0.8476
	H			+0.4238
methanol (H1 model)	CH_3	3.861	0.1823	+0.297
	O	3.083	0.1758	−0.728
	H			+0.431

[a] e represents the magnitude of electronic charge.

The outline of the present paper is as follows. The details of the simulations including the construction of the interfaces and their characterization in terms of density profiles are presented in section 2. Calculations of interfacial structures and surface tensions are described in section 3. In section 4, we have discussed the hydrogen bond patterns in the bulk liquid and at the interfaces, and section 5 deals with the simulation results of the single-particle dynamics at the interfaces. The results of hydrogen bond dynamics are presented in section 6, and our conclusions are briefly summarized in section 7.

2. Details of Models, Simulation Method, and Construction of the Interfaces

We have carried out molecular dynamics simulations of the liquid−vapor interface of water−methanol mixtures at 298 K. We have used the H1 model for methanol molecules.[47] In this model, the methanol molecules are modeled as rigid objects with three interaction sites for short-range Lennard-Jones and long-range Coulomb interactions. The methyl group is considered as a united atom with a single interaction site. The water molecules are characterized by the SPC/E potential.[48] In these models, the interaction between atomic sites of two different molecules is expressed as

$$u_{\alpha\beta}(r_\alpha, r_\beta) = 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}}\right)^{6}\right] + \frac{q_\alpha q_\beta}{r_{\alpha\beta}} \qquad (1)$$

where $r_{\alpha\beta}$ is the distance between the atomic sites α and β, and q_α is the charge of the αth atom. The Lennard-Jones parameters $\sigma_{\alpha\beta}$ and $\epsilon_{\alpha\beta}$ are obtained by using the combination rules $\sigma_{\alpha\beta} = (\sigma_\alpha + \sigma_\beta)/2$ and $\epsilon_{\alpha\beta} = \sqrt{\epsilon_\alpha \epsilon_\beta}$. The values of the potential parameters q_α, σ_α, and ϵ_α for methanol and water are summarized in Table 1.

We first carried out a bulk simulation in a cubic box of 864 molecules, periodically replicated in all three dimensions. The box length L was adjusted in such a way that the pressure would be close to the atmospheric pressure at 298 K. After this bulk solution was properly equilibrated, two empty boxes of equal size were added on either side of the original simulation box along the z-direction, and this larger rectan-

gular box (of dimension $L \times L \times 3L$) was taken as the simulation box in the next phase of the simulation run, and the system was reequilibrated by imposing periodic boundary conditions in all three directions. This resulted in a liquid slab of approximate width L separated by vacuum layers of approximate width $2L$. Some of the molecules were found to vaporize to the empty space to form a liquid−vapor interface on both sides of the liquid slab. In the simulations, the long-range electrostatic interactions were treated by using the three-dimensional Ewald method.[49] The real space part of the Ewald summation was calculated by using the minimum image convention, and the short-range Lennard-Jones interactions were calculated by using a spherical cutoff at distance $L/2$. We employed the quaternion formulation of the equations of rotational motion, and, for the integration over time, we adapted the leapfrog algorithm with a time step of 10^{-15} s (1 fs). MD runs of 500 ps were used to equilibrate each system in the bulk phase, and then the liquid−vapor interfacial systems in rectangular boxes were equilibrated for 500 ps. During the equilibration, the temperature of the simulation system was kept at 298 K through rescaling of the velocities. The simulations of the interfacial systems were then continued in a microcanonical ensemble for another 1 ns. The interfaces were found to be stable over the simulation time. The density, orientational profiles, surface tensions, hydrogen bonds, and the dynamical properties of the interfaces were calculated during the last production phase of the simulations.

3. Interfacial Structure and Surface Tension

We calculated the number density profile of water−methanol mixtures for various concentrations of methanol as a function of z by computing the average number of molecules in slabs of thickness $\Delta z = 0.05$ Å lying on either side of the central plane at $z = 0$, and the results are shown in Figure 1. It is seen that the methanol molecules have a higher propensity of staying in the interfacial region than water molecules which is in agreement with the experimental results.[46] The preferential presence of methanol molecules at interfaces is further discussed in the later part of this section.

The thickness of a liquid−vapor interface has been defined in the literature in different ways: By fitting the inhomogeneous density distribution to a hyperbolic tangent function[50,51] or to an error function[52,53] or by finding the distance over which the number density decreases from 90% to 10% of the bulk liquid density.[54−56] The last definition is used in the present work to calculate the width of the liquid−vapor interfaces, and the results are shown in Figure 2 for varying mole fractions of methanol. The thickness of the interfaces is found to increase with increasing methanol mole fractions which can be attributed to weaker interactions and higher volatility of methanol molecules. For pure methanol, the present result of the interfacial thickness is very close to the result of ref 27 where a polarizable model of methanol was used. For water−methanol mixtures also, the present results of interfacial thickness agree rather well with the corresponding results of polarizable model calculations of ref 28. Note that the interfacial thicknesses of ref 28, which were obtained by fitting the density data to hyperbolic tangent

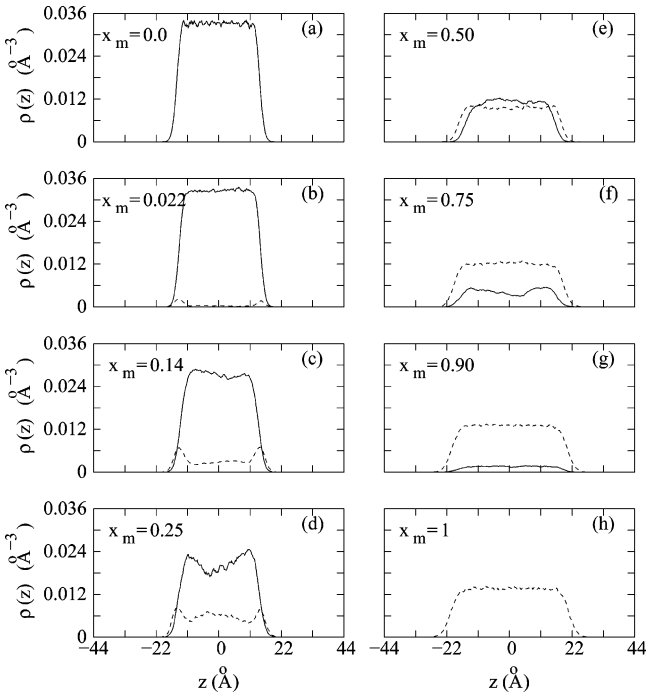


Figure 1. Number density profiles of liquid−vapor interfaces of water−methanol mixtures of varying composition. Solid curves representing water and dashed curves are for methanol.

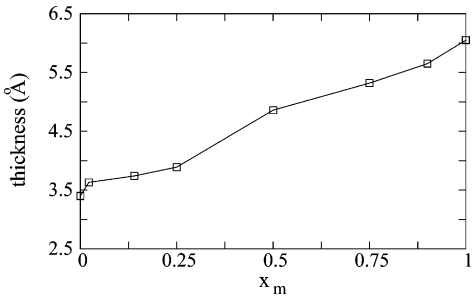


Figure 2. Interfacial thickness as a function of the mole fraction of methanol.

function, are to be multiplied by a factor of 2.1972[27,50,51] to get the corresponding results of the 90−10 interface thickness which are then readily comparable to the results of the present study. We also note that the observed changes of the number densities at interfaces have contributions both from intrinsic changes of density profiles and also from broadening due to capillary wave fluctuations as allowed by the length and time scales of the present simulations.[52,53,57,58] Thus, both intrinsic density changes and surface waves contribute to the lower number density in the interfacial regions. We can now make a more precise analysis of the enrichment of the interfaces by methanol or water molecules by plotting the mole fractions of water and methanol molecules at interfaces, defined as the 90−10 regions as described above, against that of the bulk liquid phases. Such plots are shown in Figure 3. It is clearly seen that that the interfacial regions are enriched by methanol molecules. We note that the results of Figure 3 can be readily compared with the experimental results of Figure 6 of ref 46 where a similar surface enrichment by methanol molecules was found.

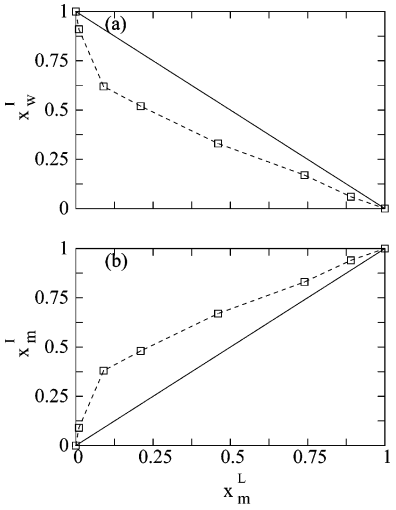


Figure 3. The variation of the mole fraction of (a) water and (b) methanol at interfaces with the mole fraction of methanol in the bulk liquid phases. The dashed lines are drawn only to connect the simulation results (squares). The solid lines correspond to the equal mole fractions of the interfacial and bulk liquid phases.

We note that, in Figure 3, x_m^L denotes the mole fraction of methanol in the bulk liquid region as obtained from the simulated densities. For low methanol mole fraction, the value of x_m^L is found to be slightly smaller than x_m which is the methanol mole fraction for the entire simulation system. This difference arises due to surface activity of methanol molecules and finite size of the simulation systems. For macroscopic systems, of course, this difference would be negligibly small.

We describe the orientation of a water molecule in terms of the angle θ that the molecular dipole vector makes with the surface normal along z-axis. We have calculated the orientational probability by dividing the angular region of $\cos\theta = -1$ to $+1$ into 60 bins of equal width and computing the average number of molecules found in a given bin. In Figure 4, we have shown the results of the normalized probability function $P(\cos\theta)$ as a function of $\cos\theta$ for both interfacial and bulk water molecules at various methanol concentrations. In the bulk phase, the probability function, as expected, is found to be uniform. Clearly, there is no preferred orientation of the water molecules in the bulk phase of the liquid slabs as one would expect. In the interfacial region, however, the probability function is nonuniform which shows an orientational structure of the interfacial molecules. For pure water, $P(\cos\theta)$ is maximum at around $\cos\theta = 0$ which means that water molecules at the interface prefer to orient with their dipoles parallel to the surface. One hydrogen atom projects into the liquid sides of the interface and other projects toward the vapor side of the interface. With an increasing methanol mole fraction, the angle θ increases which indicates that the dipole vector is tilted toward the liquid phase.

In Figure 5, we have shown the probability function of the orientation of the vector that bisects the C−O−H angle of a methanol molecule in the interfacial and bulk regions. The orientation of a methanol molecule is described by the

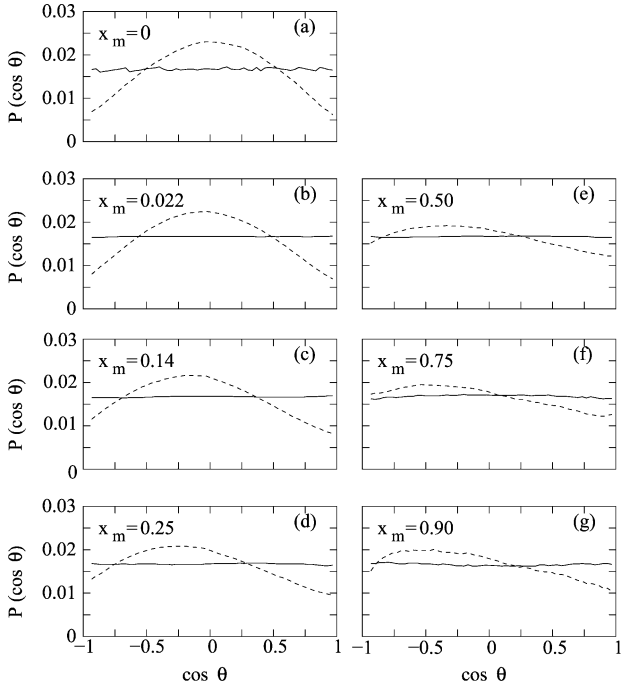


Figure 4. The probability function of the orientation of water dipole vectors in the interfacial (dashed) and bulk (solid) regions for different mole fractions of methanol. θ is the angle between the water dipole vector and the surface normal.

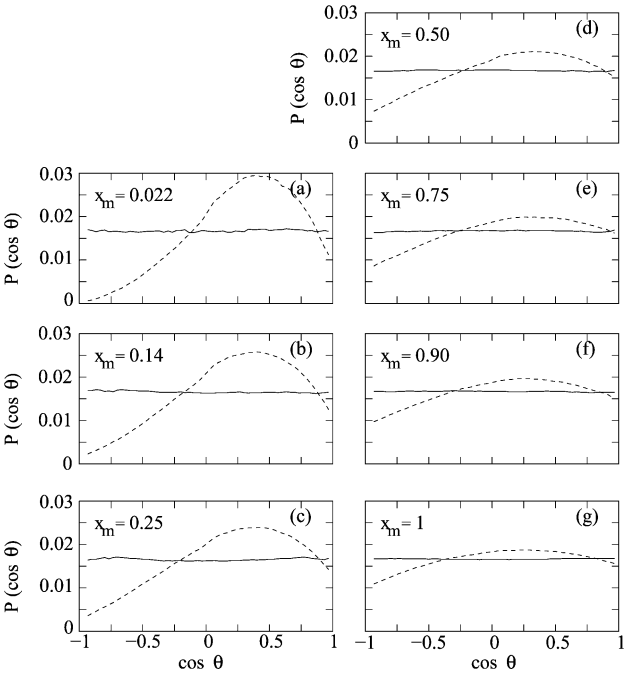


Figure 5. The probability function of the orientation of the bisector of the angle formed by the O−C and O−H bonds of a methanol molecule in the interfacial (dashed) and bulk (solid) regions for various methanol mole fraction. θ is the angle between the bisector and the surface normal.

angle θ that the above bisector vector makes with the surface normal. The orientational distribution of pure methanol is broad, and the maximum of $P(\cos\theta)$ appears at around $\cos\theta = 0.25$ which indicates that the angle between O−C vector and surface normal (θ') is about 21°. This means that the methyl group is projecting toward the vapor side and the

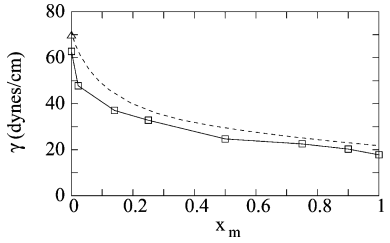


Figure 6. Surface tension vs mole fraction of methanol. Simulation data (open squares) are compared with experimental data (dashed curve) or refs 60 and 61. Diamond represents the simulated surface tension value of pure water with tail correction (ref 55).

hydroxyl hydrogen is pointed toward bulk liquid phases. This molecular orientation is also confirmed by a calculation of the orientation of the O−H vector of interfacial methanol molecules with respect to the surface normal. For example, for $x_m = 0.022$, the methanol O−H vector makes an angle of about 115° with the surface normal which clearly shows that the hydrogen is pointed inward. The ordering of surface methanol molecules increases with a decreasing methanol concentration. At very low concentration of methanol ($x_m = 0.022$) the value of the angle θ' is 5.7° which is consistent with the findings of experiments performed by Wolfrum, Graener, and Laubereau using the infrared-visible sum-frequency generation technique.[29] They reported that the value of θ' is less than 40° for pure methanol, and the polar order of methanol molecules at the surface of CH_3OH-H_2O increases with a decreasing methanol concentration. At low methanol concentration ($x_m < 0.1$) the value of θ' was found to be approximately zero.[29]

We calculated the surface tension by using the following virial expression which is obtained from the well-known Kirkwood−Buff theory[59]

$$\gamma = \frac{1}{2A}\left\langle \sum_{i<j}\sum_{\alpha,\beta} \frac{\partial u_{\alpha\beta}}{\partial r_{\alpha\beta}} \frac{1}{r_{\alpha\beta}} (r_{ij} \cdot r_{\alpha\beta} - 3 z_{ij} z_{\alpha\beta}) \right\rangle \qquad (2)$$

where $u_{\alpha\beta}$ is the interaction energy between sites α and β of molecules i and j, r_{ij} and z_{ij} are the centers of mass distance and the distance along z direction between molecules i and j, and $r_{\alpha\beta}$ and $z_{\alpha\beta}$ are the corresponding distances between sites α and β. A is the total surface area which is equal to $2L^2$. We calculated the summations in the above expression at each MD step, and finally the averaging was done over the total number of MD steps that were run during the production phase of the simulations. The standard deviations of the surface tension data, which were calculated by using block averages over 100 ps, are about 4.5% of the average values reported. It is found that the surface tension decreases with an increase of methanol concentration. At all the methanol concentrations, our calculated values of the surface tension are somewhat lower than the experimental values. In Figure 6, we compare the calculated surface tension values, γ, for various mixtures with the available experimental results[60,61] at 300 K. Although the calculated surface tension values are somewhat smaller than the experimental ones, the dependence on the composition of the mixtures is rather similar. Since the methanol molecules are strongly adsorbed at the surface, a small amount of methanol has a rather large effect in reducing the surface tension. We also note that in the present study, the Lennard-Jones interactions are truncated at $L/2$ where L is the box length. Previous studies[55,56,62] have shown that this potential truncation can underestimate the calculated surface tension to some extent. For example, the calculated surface tension of pure water at 298 K for SPC/E model was found to increase from about 62 to 70 upon inclusion of the tail corrections.[55] For the present systems also, we expect that similar tail corrections would improve the agreement between calculated and experimental results of the surface tension.

4. Hydrogen Bond Distribution at Interfaces

Following earlier studies on hydrogen bonds in water and aqueous solutions,[63−67] we have used a set of geometric criteria to define the presence of a hydrogen bond between two molecules. Two water molecules are taken to be hydrogen bonded if their interoxygen distance is less than 3.5 Å and simultaneously the hydrogen−oxygen distance is less than 2.45 Å and the oxygen−oxygen−hydrogen angle is less than a cutoff value of 45°. A hydrogen bond between a water molecule and a methanol molecule exists if their interoxygen distance is less than 3.3 Å and simultaneously the hydrogen−oxygen distance is less than 2.4 Å. The cutoff value of the oxygen−oxygen−hydrogen angle is again 45°. A hydrogen bond between two methanol molecules occurs if their interoxygen distance is less than 3.5 Å and simultaneously the hydrogen−oxygen distance is less than 2.6 Å. Again, the cutoff value of the oxygen−oxygen−hydrogen angle is the same as for the water−water and water−methanol hydrogen bonds. The oxygen−oxygen and hydrogen−oxygen distances are determined from the positions of the first minimum of the corresponding radial distribution functions of the liquid mixtures. Following previous nomenclature in the context of water−water hydrogen bonds,[64] we have used a "less strict" definition of the hydrogen bonds in the present study. The quantities of interest are the percentages f_n of water or methanol molecules that engage in n hydrogen bonds and the average number of hydrogen bonds per molecule n_{HB}. Here we have calculated the above quantities for three types of hydrogen bonds: water−water (ww), water−methanol (wm), and methanol−methanol (mm).

In Figure 7, we have shown the variation of the number of hydrogen bonds per molecule with a change of methanol concentration in bulk liquid phases and also at interfaces. For pure water, in the bulk liquid phase, the majority of water molecules participate in four hydrogen bonds, whereas, in the liquid−vapor interfacial region, most of the molecules are found to have either three or two hydrogen bonds. The average number of hydrogen bonds per water molecule is also significantly smaller than that in the bulk liquid phase. This smaller number of hydrogen bonds at the vapor−water interfaces is likely due to the lower density and the presence of vapor (essentially vacuum) on one side of the liquid. Now with the increase of methanol concentration, the numbers of ww (water−water) and wm/m (water−methanol per methanol) hydrogen bonds decrease and that of wm/w (water−methanol per water) and mm (methanol−methanol) increases. We note that the average number of mm hydrogen

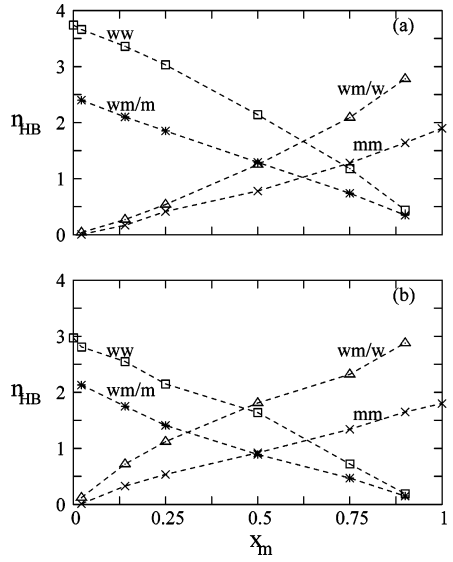


Figure 7. Variation of the number of hydrogen bonds per molecule at various methanol concentrations: (a) is for the bulk liquid region and (b) is for the interfacial region.

bonds in the interfacial region is higher than that in the bulk liquid up to $x_m = 0.90$ which clearly indicates that the methanol molecules are strongly adsorbed at the surface. Above $x_m = 0.90$ the average number of mm hydrogen bonds in the bulk liquid region is higher than that in the interfacial region. Figure 8 shows the fraction of molecules with n number of hydrogen bonds in bulk liquid and at the interfacial region for a 50:50 mixture of water and methanol. It is seen from Figure 8(b) that the number of wm hydrogen bonds per water molecule is somewhat higher in the interfacial region than that in the bulk liquid which can again be attributed to the higher population of methanol molecules in the surface region.

5. Diffusion and Orientational Relaxation of Interfacial Molecules

In this section, we report the single-particle dynamical properties of water and methanol molecules at the liquid−vapor interfaces that we have calculated in the present work. An important issue concerning the dynamics of an interface is as follows: How different the dynamics of the interface is compared to that of the corresponding bulk phases. The dynamical behavior of interfacial molecules is expected to be different from bulk molecules both translationally and rotationally. Also, the translational motion of molecules at the interface can be highly anisotropic in contrast to the bulk molecules which move in an isotropic environment. In view of this anisotropic aspect, we have separately calculated the perpendicular and parallel components of the translational diffusion. We have also calculated the dipole orientational relaxation of both the interfacial and bulk molecules.

We denote the αth component of velocity of a molecule by $v_\alpha(t)$ ($\alpha = x, y, z$), and its normalized autocorrelation function $C_{v;\alpha}(t)$ is defined by

$$C_{v;\alpha}(t) = \frac{\langle v_\alpha(t) v_\alpha(0) \rangle}{\langle v_\alpha^2 \rangle} \tag{3}$$

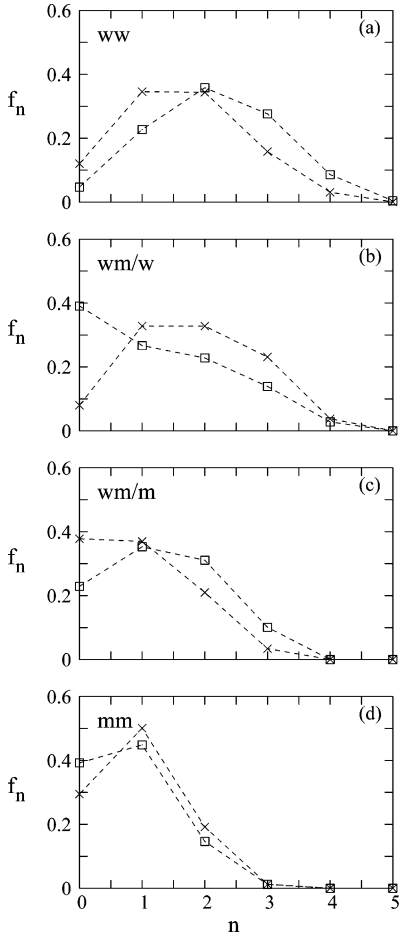


Figure 8. The fraction of molecules with n the number of hydrogen bonds in bulk liquid (open squares) and at the interfacial region (crosses) for a 50:50 mixture of water−methanol.

where $\langle \cdots \rangle$ denotes an equilibrium ensemble average. In these calculations, the average of eq 3 is carried out over those molecules which are found in the interfacial region at time 0 and also at time t.

We have also calculated the diffusion coefficient D_α ($\alpha = x, y, z$) from the velocity−velocity autocorrelation function by using the following relation

$$D_\alpha = \frac{k_B T}{m} \int_0^\infty C_{v;\alpha}(t)\, dt \tag{4}$$

where m is the mass of a molecule and k_B is the Boltzmann constant. The anisotropic nature of the translational motion is clearly illustrated in Figure 9 where we have shown the diffusion coefficient values in the parallel (x) and perpendicular (z) direction of the interface, and the values are compared with their corresponding bulk liquid values at different mole fractions of methanol. We have also shown the experimental bulk diffusion coefficient values for both water and methanol molecules as reported in refs 8−11. At low concentration of methanol (upto $x_m = 0.25$) the diffusion coefficients of both water and methanol in the bulk liquid region decreases which indicates that there is significant interaction between the water and methanol molecules. Above $x_m = 0.25$ both $D^{bulk}(H_2O)$ and $D^{bulk}(CH_3OH)$ begin to increase. At low methanol concentrations, the methanol

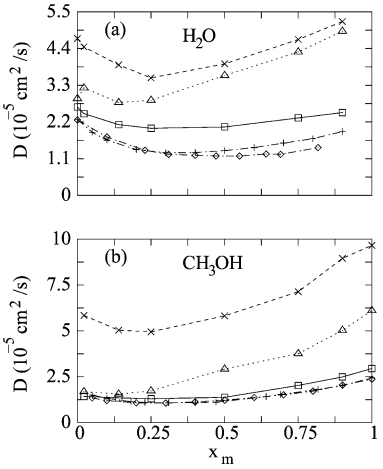


Figure 9. Diffusion coefficients as functions of methanol mole fraction: Open squares are for bulk liquids, open triangles are for diffusion coefficients at the interfacial *z*-direction, and crosses are for the interfacial *x*-direction, and plus signs and diamonds are experimental data from refs 8 and 10.

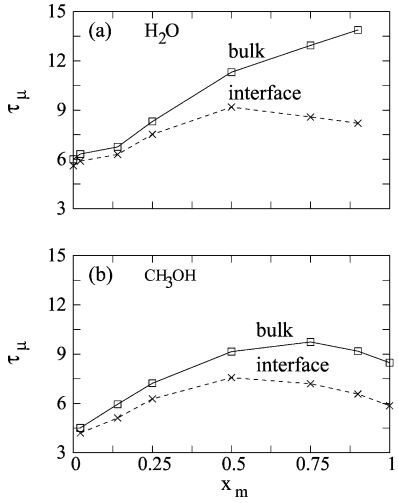


Figure 10. The dipole orientational relaxation times of (a) water and (b) methanol molecules in the interfacial (dashed line) and bulk regions (solid line) at various methanol mole fractions.

is rather strongly hydrated by a cage of water molecules[13,17] leading to a retardation of their diffusion. At a high concentration of methanol, there are an insufficient number of water molecules present to form the hydration shells of all the methanol molecules. Thus, the relative importance of the strength of water−methanol interactions decreases, and, as a result, we observe an increase of the diffusion coefficients. The diffusion coefficients of both water and methanol molecules along the *x*-direction at the interface (parallel diffusion) follow the same trend as that of the corresponding bulk liquid values with the minimum of the diffusion coefficient values appearing at $x_m = 0.25$. For the perpendicular diffusion at interfaces, however, the minimum occurs at $x_m = 0.14$ for both water and methanol molecules. We note that, for pure water, the relative changes of the diffusion coefficients between the bulk and interfacial regions are qualitatively similar to the results obtained by Liu et al.[68] who used the method of survival probabilities to calculate the diffusion coefficients.

The rotational motion of a molecule at liquid−vapor interfaces is investigated by calculating the time dependence of the self-dipole correlation function

$$C_\mu(t) = \frac{\langle \mu(t)\cdot\mu(0)\rangle}{\langle \mu(0)^2\rangle} \tag{5}$$

where $\mu(t)$ is the dipole vector of a molecule at time t. The results of $C_\mu(t)$ are shown in Figure 10 for both interfacial and bulk water and methanol molecules at varying methanol concentrations. It is seen that the orientational relaxation at the interface occurs at a faster rate than that in the bulk. We define the orientational relaxation time τ_μ as the time integral of the orientational correlation function

$$\tau_\mu = \int_0^\infty C_\mu(t)\, dt \tag{6}$$

where we have calculated the integral explicitly up to 4 ps by using the simulation data of $C_\mu(t)$, and the contribution of the tail part is obtained by using the fitted exponential functions. The orientational relaxation time of interfacial molecules is found to be shorter than that of bulk molecules for all methanol mole fractions considered here. We note that the density of molecules is low in the interfacial region, and also the interfacial molecules have less number of hydrogen bonds and that they essentially do not have any solvation shell on the vapor side of the interface. Because of these reduced density, less number of hydrogen bonds and incomplete solvation effects, a molecule in the interfacial region experiences less rotational friction than that in the bulk phase leading to a faster rotational relaxation of interfacial molecules. We note that similar faster relaxation of the interfacial molecules than the bulk ones were also observed for liquid−vapor interfaces of pure water and aqueous ionic solutions.[55] Figure 10 shows that the rotational correlation times of methanol molecules has a maximum at $x_m = 0.50$ for the bulk liquid region and at $x_m = 0.25$ at the interface. For water molecules, the maximum appears at $x_m = 0.50$ for the interfacial region, but, for the bulk liquid region, the orientational correlation time of water increases monotonically with an increase of methanol concentration.

6. Dynamics of Hydrogen Bonds at Interfaces

The dynamics of hydrogen bonds in the interfacial and bulk regions are investigated by calculating the so-called continuous hydrogen bond time correlation functions[64−67,69] and the average lifetimes of hydrogen bonds. We have calculated these dynamical properties for water−water, water−methanol, and methanol−methanol hydrogen bonds where both molecules are in the same region, interface or bulk, and also for inter-region hydrogen bonds where one molecule of the bonded pair is in the interface and the other is in the adjacent layer on the bulk side. For convenience, we denote the interfacial region as region I and the bulk region as region II.

To calculate the hydrogen bond dynamics, we first define two hydrogen bond population variables $h(t)$ and $H(t)$: $h(t)$

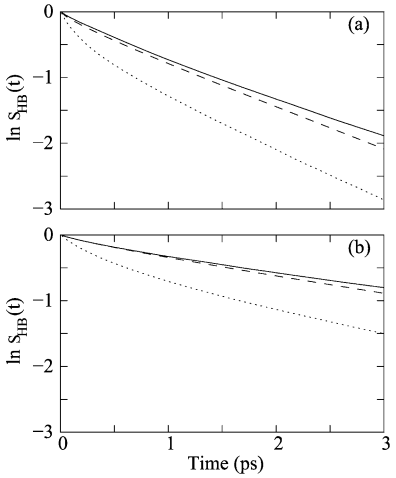


Figure 11. Time dependence of the correlation function $S_{HB}(t)$ for (a) pure water and (b) pure methanol. The solid, dashed, and dotted curves correspond to I−I, II−II, and I−II hydrogen bonds.

is unity when a particular tagged pair of water molecules is hydrogen bonded at time t, according to the adopted definition as described earlier and zero otherwise. The function $H(t)$ is unity if the tagged pair of water molecules remains continuously hydrogen bonded from $t = 0$ to time t, and it is zero otherwise. We define the continuous hydrogen bond time correlation function $S_{HB}(t)$ as[61−64,66]

$$S_{HB}(t) = \langle h(0)H(t)\rangle/\langle h\rangle \tag{7}$$

where $\langle\cdots\rangle$ denotes an average over all hydrogen bonds that are present at $t = 0$. Clearly, $S_{HB}(t)$ describes the probability that a water pair, which was hydrogen bonded at $t = 0$, remains continuously bonded up to time t. The time integral of this function describes the average time that a hydrogen bond survives after it is chosen at time $t = 0$. We denote the integral by τ_{HB} and call it the average hydrogen bond lifetime. We note, however, that since the hydrogen bonds are chosen randomly without keeping any condition on when they were created, the integral of $S_{HB}(t)$ should more appropriately be called the average persistence time of a randomly chosen hydrogen bond.[64]

In Figure 11, we have shown the decay of $S_{HB}(t)$ for different regions of the liquid−vapor systems for pure water and pure methanol. The corresponding results for water−water, water−methanol, and methanol−methanol hydrogen bonds in water−methanol mixture with $x_m = 0.25$ are shown in Figure 12. The results of τ_{HB} for all the systems studied here are included in Tables 2−4. The relaxation of $S_{HB}(t)$ is found to be somewhat slower at liquid−vapor interfaces compared to the corresponding relaxation in the bulk phases. An insight into this different relaxation behavior of interfacial and bulk hydrogen bonds can be obtained from the energetics of these hydrogen bonds which are also included in Tables 2−4. The hydrogen bond energies are found to be most negative when both molecules are at the interface, whereas they are least negative when one molecule is at the interface and the other one is in the bulk liquid. Thus, although the number of hydrogen bonds in the liquid−vapor interfacial region is less as reported in Figures 7 and 8, the hydrogen

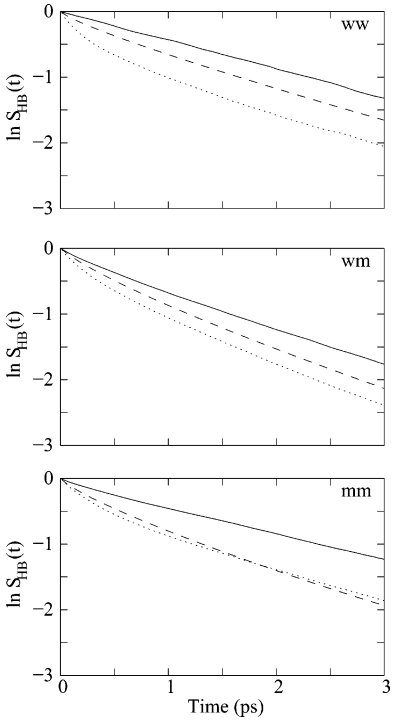


Figure 12. The time dependence of the correlation function $S_{HB}(t)$ for ww, wm, and mm hydrogen bonds for $x_m = 0.25$. Different curves are as in Figure 11.

Table 2. Lifetimes and the Energies of Water−Water (ww) Hydrogen Bonds for Different Mole Fractions of Methanol[a]

x_m	region	τ_{HB}	E_{HB}
0.0	I−I	1.50	−19.74
	I−II	0.80	−18.65
	II−II	1.35	−18.72
0.25	I−I	2.28	−20.20
	I−II	1.15	−19.20
	II−II	1.68	−19.28
0.50	I−I	2.75	−20.46
	I−II	1.35	−19.63
	II−II	2.10	−19.83
0.75	I−I	3.13	−20.60
	I−II	1.43	−20.45
	II−II	2.70	−20.60

[a] The relaxation times are expressed in units of ps, and the hydrogen bond energies are expressed in units of kJ/mol.

bonds in this region are found to be relatively stronger and hence live longer. Besides, as noted earlier,[70] the rate of relaxation of hydrogen bonds also depends on the number of adjacent but non-hydrogen-bonded water (or methanol) molecules. The higher the number of such non-hydrogen-bonded adjacent molecules, the faster would be the relaxation because these molecules can take a new hydrogen bond and thus help in breaking the original hydrogen bond. Since the number of such non-hydrogen-bonded adjacent molecules in the interfacial regions is smaller than that in the bulk liquids, the hydrogen bonds at the interfaces relax at a slower rate than those in the corresponding bulk liquid phases. An interesting dynamical behavior is found for the dynamics of those hydrogen bonds where one molecule of the bonded pair belongs to the interface (region I) and the second one

Table 3. Lifetimes and the Energies of Water−Methanol (wm) Hydrogen Bonds for Different Mole Fractions of Methanol[a]

x_m	region	τ_{HB}	E_{HB}
0.25	I−I	1.60	−20.20
	I−II	1.05	−19.05
	II−II	1.25	−19.13
0.50	I−I	1.92	−20.80
	I−II	1.10	−19.54
	II−II	1.50	−19.64
0.75	I−I	2.22	−21.60
	I−II	1.28	−20.07
	II−II	1.80	−20.12

[a] The relaxation times are expressed in units of ps, and the hydrogen bond energies are expressed in units of kJ/mol.

Table 4. Lifetimes and the Energies of Methanol−Methanol (mm) Hydrogen Bonds for Different Mole Fractions of Methanol[a]

x_m	region	τ_{HB}	E_{HB}
0.25	I−I	2.36	−20.02
	I−II	1.30	−18.97
	II−II	1.40	−19.05
0.50	I−I	2.98	−20.24
	I−II	1.40	−19.40
	II−II	1.95	−19.55
0.75	I−I	4.13	−20.32
	I−II	1.52	−19.72
	II−II	2.50	−19.98
1.0	I−I	3.57	−20.27
	I−II	1.75	−19.92
	II−II	3.25	−20.10

[a] The relaxation times are expressed in units of ps, and the hydrogen bond energies are expressed in units of kJ/mol.

to its adjacent region, i.e., region II. The dynamics of these inter-region hydrogen bonds are found to be faster than even those in the bulk phase. It is seen from Figures 7 and 8 that the hydrogen bond environment of different regions are different to some extent. The energetic data of Table 2 show that the energies of these inter-region or interenvironment hydrogen bonds are higher (least negative) than the corresponding intra-region hydrogen bonds for the bulk or interfacial zones. Thus, the hydrogen bonds that connect water molecules of two different regions or environments are found to be energetically weaker and hence relax at a faster rate than those belonging to a single region, either interface or the bulk. With an increase of methanol concentration, the average hydrogen bond lifetime (τ_{HB}) of ww and wm bonds increases in all three regions (I−I, I−II, and II−II), and this behavior is also supported by hydrogen bond energies. For mm hydrogen bonds, with increasing methanol mole fractions τ_{HB} increases in the II−II and I−II regions, but the τ_{HB} values at $x_m = 1.0$ are slightly lower than the corresponding values for $x_m = 0.75$ for the I−I region.

7. Conclusion

We have performed molecular dynamics simulations to investigate the various equilibrium and dynamical properties of liquid−vapor interfaces of water−methanol mixtures of varying composition. Simulations are carried out at room temperature, and various interfacial properties that are calculated include density and orientational profiles, intrinsic width of the interfaces, surface tension, structure and dynamics of hydrogen bonds, molecular diffusion, and orientational relaxation of both water and methanol molecules.

It is found that the orientational distribution of the dipole vector of interfacial water molecules changes with a change of methanol concentration. The orientational ordering of methanol molecules at the surface of H_2O-CH_3OH is found to increase with a decrease of methanol concentration. At the interface, methanol molecules orient in such a way that the hydrophobic group seems to be oriented out of the surface plane which is in good agreement with the experimental results. The width of the interfaces, which is calculated by using the 90%−10% rule, is found to decrease, and the surface tension is found to increase with a decrease of the methanol mole fraction. Although the surface tension values calculated in the present study are somewhat lower compared to the experimental values, the trend of its variation with composition is very similar to the experimental results. The distribution of different types of hydrogen bonds (ww, wm, and mm) and the total number of hydrogen bonds per molecule are also calculated in both bulk liquid and interfacial regions.

The translational motion of water and methanol molecules is found to be highly anisotropic in the interfacial regions. The nonideal diffusive behavior with a variation of the methanol concentration is also found for interfacial diffusion where the minimum appears at a lower bulk methanol concentration because of the enrichment of the interfaces by methanol molecules. The rotational correlation times of methanol has a maximum at $x_m = 0.75$ at the bulk regions and $x_m = 0.50$ at the interface. For water, the orientational relaxation time increases monotonically with an increase of methanol concentration in the bulk region with no maximum. However, a maximum of the orientational relaxation time of interfacial water molecules is found at $x_m = 0.50$. We have also investigated the dynamics of water−water, water−methanol, and methanol−methanol hydrogen bonds at liquid−vapor interfaces and also in the bulk liquid phases. For all the systems, the relaxation of hydrogen bonds in the interfacial region is found to occur at a somewhat slower rate than that in bulk liquid. However, the dynamics of those hydrogen bonds which connect interfacial molecules to its adjacent layer on the bulk side is found to be faster than even the bulk hydrogen bonds. Correlations of these dynamical results with the energies of hydrogen bonds in different regions of the liquid−vapor systems are also explored. In addition, variation of the average lifetimes of different hydrogen bonds with composition of the water−methanol mixtures is also investigated.

In the present work, both water and methanol molecules are modeled as nonpolarizable molecules with fixed partial charges on various atomic sites. The solvent and solute molecules at liquid−vapor interfaces, however, encounter varying environments, and hence the degree of polarization of individual molecules can vary significantly across such

interfaces. Thus, the polarization effects can play an important role in determining the equilibrium and dynamical properties of liquid−vapor interfaces as has been shown in some of the recent theoretical studies on aqueous ionic solutions with large halide ions.[71,72] In the context of the liquid−vapor interfaces of water−methanol systems, we note two recent studies[27,28] where structural and thermodynamics aspects of such interfaces were examined by using polarizable models of water and methanol molecules. The inhomogeneous density and orientational profiles of the interfaces, including their dependence on the composition of the mixtures, for the polarizable models are found to be similar to the results of nonpolarizable models presented here. The thickness of the interfaces as found in these studies[27,28] is also rather close to the results of the present study as discussed in section 3. The dynamical aspects of the liquid−vapor interfaces of water−methanol mixtures have not yet been investigated by using polarizable models. For the hydrogen bond dynamics at pure water surfaces, a recent study[70] has shown that many-body polarization effects can alter the relative time scales of relaxation of interfacial and bulk hydrogen bonds. Thus, it would be a worthwhile exercise to employ polarizable models to investigate the single-particle and hydrogen bond dynamical properties of liquid−vapor interfaces of water−methanol mixtures. There are several methods to account for the polarization effects in molecular dynamics simulations such as the fluctuating charge model,[73,74] the model of dipoles on atomic sites,[75] the self-consistent reaction field method,[76] and the charge-on-spring model.[77,78] We are currently developing polarizable models for different polyatomic molecules using the charge-on-spring model,[77,78] and we hope to address the issue of many-body polarization effects on interfacial dynamics of water−methanol mixtures in a future publication. Also, in the present study, we have looked at the self-diffusion coefficients of bulk and interfacial molecules. It would also be worthwhile to study the mutual and distinct diffusion coefficients[79−83] of water and methanol molecules in the bulk and interfacial regions. Such studies would reveal valuable information on the pair dynamics of these molecules which, in turn, would help us to investigate the tendency of these molecules to locally associate or demix, i.e., to create local heterogeneity, as one moves from the bulk to the surface regions.

Acknowledgment. One of the authors (A.C.) would like to thank Professor C. N. R. Rao, F.R.S., for helpful communications and enlightening discussions on surface enrichment in water−alcohol mixtures. The financial support from the Council of Scientific and Industrial Research (CSIR) and Department of Science and Technology (DST), Government of India, is gratefully acknowledged.

References

(1) Narten, A. H.; Habenschuss, A. *J. Chem. Phys.* **1984**, *80*, 3387.

(2) Yamaguchi, T.; Hidaka, K.; Soper, A. K. *Mol. Phys.* **1999**, *96*, 1159. Yamaguchi, T.; Hidaka, K.; Soper, A. K. *Mol. Phys.* **1999**, *97*, 603.

(3) Morrone, J. A.; Tuckerman, M. E. *J. Chem. Phys.* **2002**, *117*, 4403.

(4) Easteal, A. J.; Woolf, L. A. *J. Chem. Thermodyn.* **1985**, *17*, 69.

(5) Easteal, A. J.; Woolf, L. A. *J. Chem. Thermodyn.* **1985**, *17*, 49.

(6) Staveley, L. A. K.; Hart, K. R.; Tupman, W. I. *Discuss. Faraday Soc.* **1953**, *156*, 130.

(7) Onori, G. *J. Chem. Phys.* **1987**, *87*, 1251.

(8) Woolf, L. A. *Pure Appl. Chem.* **1985**, *57*, 1083.

(9) Roux, A. H.; Desnoyers, J. E. *Proc. Indian Acad. Sci. (Chem. Sci.)* **1987**, *98*, 435.

(10) Hawlicka, E. *Ber. Bunsen-Ges., Phys. Chem.* **1983**, *87*, 425.

(11) Versmold, H. *Ber. Bunsen-Ges., Phys. Chem.* **1980**, *84*, 168.

(12) Easteal, A. J.; Woolf, L. A. *J. Phys. Chem.* **1985**, *89*, 1066.

(13) Jorgensen, W. L.; Madura, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 1407.

(14) Okazaki, S.; Nakanishi, K.; Touhara, H. *J. Chem. Phys.* **1983**, *78*, 454. Okazaki, S.; Touhara, H.; Nakanishi, K. *J. Chem. Phys.* **1984**, *81*, 890.

(15) Bolis, G.; Corongiu, G.; Clementi, E. *Chem. Phys. Lett.* **1982**, *86*, 299.

(16) Jorgensen, W. L. *J. Am. Chem. Soc.* **1980**, *102*, 543. Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276. Jorgensen, W. L.; Ibrahim, M. *J. Am. Chem. Soc.* **1982**, *104*, 373.

(17) Soper, A. K.; Finney, J. L. *Phys. Rev. Lett.* **1993**, *71*, 4346.

(18) Derlacki, Z. J.; Easteal, A. J.; Edge, A. V. J.; Woolf, L. A.; Roksandic, Z. *J. Phys. Chem.* **1985**, *89*, 5318.

(19) Tanaka, H.; Gubbins, K. E. *J. Chem. Phys.* **1992**, *97*, 2626.

(20) Skaf, M. S.; Ladanyi, B. M. *J. Chem. Phys.* **1995**, *102*, 6542.

(21) Ferrario, M.; Haughney, M.; Mcdonald, I. R.; Klein, M. L. *J. Chem. Phys.* **1990**, *93*, 5156.

(22) Day, T. J. F.; Patey, G. N. *J. Chem. Phys.* **1999**, *110*, 10937.

(23) Wensink, E. J. W.; Hoffmann, A. C.; van Maaren, P. J.; van der Spoel, D. *J. Chem. Phys.* **2003**, *119*, 7308.

(24) Morita, A. *Chem. Phys. Lett.* **2003**, *375*, 1.

(25) Hawlicka, E.; Swiatla-Wojcik, D. *Phys. Chem. Chem. Phys.* **2000**, *2*, 3175.

(26) Matsumoto, M.; Takaoka, Y.; Kataoka, Y. *J. Chem. Phys.* **1993**, *98*, 1464.

(27) Patel, S.; Brooks, C. L., III *J. Chem. Phys.* **2005**, *122*, 24508.

(28) Chang, T.-M.; Dang, L. X. *J. Phys. Chem. B* **2005**, *109*, 5759.

(29) Wolfrum, K.; Graener, H.; Laubereau, A. *Chem. Phys. Lett.* **1993**, *213*, 41.

(30) Huang, J. Y.; Wu, M. H. *Phys. Rev. E* **1994**, *50*, 3737.

(31) Stanners C. D.; Du, Q.; Chin, R.; Cremer, P.; Somorjai, G. A.; Shen, Y. R. *Chem. Phys. Lett.* **1995**, *232*, 407.

(32) Eisenthal, K. B. *Chem. Rev.* **1996**, *96*, 1343 and references therein.

(33) Du, Q.; Freysz, E.; Shen, Y. R. *Science* **1994**, *264*, 826. Du, Q.; Superfine, R.; Freysz, E.; Shen, Y. R. *Phys. Rev. Lett.* **1993**, *70*, 2313.

(34) Goh, M. C.; Hicks, J. M.; Kemnitz, K.; Pinto, G. R.; Bhattacharyya, K.; Eisenthal, K. B.; Heinz, T. F *J. Phys. Chem.* **1988**, *92*, 5074.

(35) Sitzmann, E. V.; Eisenthal, K. B. *J. Phys. Chem.* **1988**, *92*, 4579.

(36) Kemnitz, K.; Bhattacharyya, K.; Hicks, J. M.; Pinto, G. R.; Eisenthal, K. B.; Heinz, T. F. *Chem. Phys. Lett.* **1986**, *131*, 285.

(37) Bhattacharyya, K.; Sitzmann, E. V.; Eisenthal, K. B. *J. Chem. Phys.* **1987**, *87*, 1442.

(38) Petralli-Mallow, T.; Wong, T. M.; Byers, J. D.; Yee, H. I.; Hicks, J. M. *J. Phys. Chem.* **1993**, *97*, 1383.

(39) Superfine, R.; Huang, J. Y.; Shen, Y. R. *Opt. Lett.* **1990**, *15*, 1276.

(40) Conboy, J. C.; Daschbach, J. L.; Richmond, G. L. *J. Phys. Chem.* **1994**, *98*, 9688.

(41) Raising, T.; Stehlin, T.; Shen, Y. R.; Kim, M. W.; Valint, P., Jr. *J. Chem. Phys.* **1988**, *89*, 3386.

(42) Shi, X.; Borguet, E.; Tarnovski, A. N.; Eisenthal, K. B. *Chem. Phys.* **1996**, *205*, 167.

(43) Wolfrum, K.; Lobau, J.; Laubereau, A. *Appl. Phys. A* **1994**, *59*, 605.

(44) Zimdars, D.; Eisenthal, K. B. *J. Phys. Chem. B* **2001**, *105*, 3993.

(45) Zhang, D.; Gutow, J. H.; Eisenthal, K. B.; Heinz, T. F. *J. Chem. Phys.* **1993**, *98*, 5099.

(46) Raina, G.; Kulkarni, G. U.; Rao, C. N. R. *J. Phys. Chem. A* **2001**, *105*, 10204.

(47) Haughney, M.; Ferrario, M.; McDonald, I. R. *J. Phys. Chem.* **1987**, *91*, 4934. Haughney, M.; Ferrario, M.; McDonald, I. R. *Mol. Phys.* **1986**, *58*, 849.

(48) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(49) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford, 1987.

(50) Chapela, G. A.; Saville, G.; Thompson, S. M.; Rowlinson, J. S. *J. Chem. Soc., Faraday Trans. 2* **1977**, *8*, 133.

(51) Rowlinson, J. S.; Widom, B. *Molecular theory of capillarity*; Clarendon: Oxford, 1982.

(52) Sides, S. W.; Grest, G. S.; Lacasse, M.-D. *Phys. Rev. E* **1999**, *60*, 6708.

(53) Senapati, S.; Berkowitz, M. L. *Phys. Rev. Lett.* **2001**, *87*, 176101.

(54) Taylor, R. S.; Dang, L. X.; Garrett, B. C. *J. Phys. Chem.* **1996**, *100*, 11720.

(55) Paul, S.; Chandra, A. *Chem. Phys. Lett.* **2003**, *373*, 87.

(56) Alejandre, J.; Tildesley, D. J.; Chapela, G. A. *J. Chem. Phys.* **1995**, *102*, 4574.

(57) Lacasse, M.-D.; Grest, G. S.; Levine, A. J. *Phys. Rev. Lett.* **1998**, *80*, 309.

(58) Werner, A.; Schmid, F.; Müller, M.; Binder, K. *Phys. Rev. E* **1999**, *59*, 728.

(59) Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1949**, *17*, 338.

(60) *CRC Handbook of Chemistry and Physics*, 72nd ed.; Lide, D. R., et al., Eds.; CRC: Boston, 1991.

(61) *International Critical Tables of Numerical Data, Physics, Chemistry and Technology*; Washburn, E. W., Ed.; McGraw-Hill Book Company, Inc.: New York and London, 1928; Vol. 4.

(62) Mecke, M.; Winkelmann, J.; Fischer, J. *J. Chem. Phys.* **1999**, *110*, 1188.

(63) Luzar, A.; Chandler, D. *Nature (London)* **1996**, *379*, 53. Luzar, A.; Chandler, D. *Phys. Rev. Lett.* **1996**, *76*, 928.

(64) Luzar, A. *J. Chem. Phys.* **2000**, *113*, 10663.

(65) Chandra, A. *Phys. Rev. Lett.* **2000** *85*, 768. Chandra, A. *J. Phys. Chem. B* **2003**, *107*, 3899.

(66) Balasubraminian, S.; Pal, S.; Bagchi, B. *Phys. Rev. Lett.* **2002**, *89*, 115505.

(67) Paul, S.; Chandra, A. *Chem. Phys. Lett.* **2004**, *386*, 218.

(68) Liu, P.; Harder E.; Berne, B. J. *J. Phys. Chem. B* **2004**, *108*, 6595.

(69) Rapaport, D. *Mol. Phys.* **1983**, *50*, 1151.

(70) Liu, P.; Harder E.; Berne, B. J. *J. Phys. Chem. B* **2005**, *109*, 2949.

(71) Jungwirth, P.; Tobias, D. J. *J. Phys. Chem. B* **2001**, *105*, 10468.

(72) Jungwirth, P.; Tobias, D. J. *J. Phys. Chem. B* **2002**, *106*, 6361.

(73) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141.

(74) Rick, S. W. *J. Chem. Phys.* **2001**, *114*, 2276.

(75) Koneshan, S.; Rasaiah J. C.; Dang, L. X. *J. Chem. Phys.* **2001**, *114*, 7544.

(76) Luque, F. J.; Bofill, J. M.; Orozco, M. *J. Chem. Phys.* **1995**, *103*, 10183.

(77) Straatsma, T. P.; McCammon, J. A. *Mol. Simul.* **1990**, *5*, 181.

(78) Yu, H.; Hansson, T.; van Gunsteren, W. F. *J. Chem. Phys.* **2003**, *118*, 221.

(79) Jacucci, G.; McDonald, I. R. *Physica A (Amsterdam)* **1975**, *80*, 607.

(80) Schoen, M.; Hoheisel, C. *Mol. Phys.* **1984**, *52*, 33.

(81) Ali, S. M.; Samanta, A.; Ghosh, S. K. *J. Chem. Phys.* **2001**, *114*, 10419.

(82) Chowdhuri, S.; Chandra, A. *Phys. Rev. E* **2002**, *66*, 41203.

(83) Kamala, C. R.; Ayappa, K. G.; Yashonath, S. *J. Phys. Chem. B* **2004**, *108*, 4411.

CT050098D

JCTC Journal of Chemical Theory and Computation

Modeling Proton Transfer in Zeolites: Convergence Behavior of Embedded and Constrained Cluster Calculations

Justin T. Fermann,[†] Teresa Moniz,[†] Oliver Kiowski,[†] Timothy J. McIntire,[†] Scott M. Auerbach,*[,†,‡] Thom Vreven,[§] and Michael J. Frisch[§]

Departments of Chemistry and Chemical Engineering, University of Massachusetts, Amherst, Massachusetts 01003, and Gaussian, Inc., 340 Quinnipiac Street, Building 40, Wallingford, Connecticut 06492

Received May 2, 2005

Abstract: We have studied the convergence properties of embedded and constrained cluster models of proton transfer in zeolites. We applied density functional theory to describe clusters and ONIOM to perform the embedding. We focused on converging the reaction energy and barrier of the O(1) to O(4) jump in H−Y zeolite as well as vibrational and structural aspects of this jump. We found that using successively larger clusters in vacuo gives convergence of this reaction energy to 14 ± 2 kJ mol^{-1} and the barrier to 135 ± 5 kJ mol^{-1} at a cluster size of 5 Å, which contains 11 tetrahedral (Si or Al) atoms. We embedded quantum clusters of various sizes in larger clusters with total radii in the range 7−20 Å, using the universal force field as the lower level of theory in ONIOM. We found convergence to the same values as the constrained clusters, without the use of reactive force fields or periodic boundary conditions in the embedding procedure. For the reaction energy, embedded cluster calculations required smaller clusters than in vacuo calculations, reaching converged reaction energies for quantum systems containing at least 8 tetrahedral atoms. In addition, optimizations on embedded clusters required many fewer cycles, and hence much less CPU time, than did optimizations on comparable constrained clusters.

I. Introduction

Zeolites offer a versatile class of shape-selective catalysts for important chemical processes such as petroleum cracking and reforming.[1,2] Many catalytic processes in zeolites are activated by proton transfer from Brønsted acid sites: Si−OH−Al. Progress in steering such reactions would be fueled by enhanced understanding of the mechanisms that control proton transfer in zeolites, which can be provided by molecular modeling.[3,4] Studying proton transfer in bare zeolites is important for several reasons: (i) trends in catalytic activity have been correlated with proton-transfer rates,[5,6] (ii) several intriguing discrepancies remain among experimental probes of proton transfer in bare zeolites,[7−10] and (iii) computational methods needed to model complex reactions in zeolites can be validated on this relatively simpler process.[11−14] An important issue in modeling zeolite electronic structures is how to represent the extended nature of zeolites with tractable calculations. In this article, we benchmark two approaches for computing the reaction energy and barrier for the O(1) to O(4) proton transfer in H−Y zeolite.

Quantum calculations on small clusters have long been the staple for modeling zeolite electronic structures.[3,4] However, even when clusters are constrained to mimic the target zeolite, qualitative errors can arise.[15−23] We found this in our study of the O(1) to O(4) jump in H−Y modeled with

* Corresponding author e-mail: auerbach@chem.umass.edu.
[†] Department of Chemistry, University of Massachusetts.
[‡] Department of Chemical Engineering, University of Massachusetts.
[§] Gaussian, Inc.

10.1021/ct0501203 CCC: $30.25
Published on Web 08/20/2005

a "3T" cluster [≡Si−O(1)−Al(OH)$_2$−O(4)−Si≡].[24−26] Although protons are observed by neutron diffraction at O(1) but not at O(4),[27] our O(1) to O(4) reaction energies were found to be negligible, signaling a qualitative failing of the small cluster model. In addition to ignoring long-range forces and some hydrogen bonding, small clusters lack the steric constraints that characterize adsorption in zeolites, because the cluster may not represent enough of the zeolite cavity.

Several remedies can be considered for solving the problems of small cluster models. The simplest is to use progressively larger clusters in vacuo until desired properties converge.[21,28] It is clear that with increasing cluster size, more of the important local electronic and steric effects are included, while the effect of the terminating region diminishes when it is further separated from the proton transfer site. In that approach, one has the choice of fully relaxing the clusters or imposing some geometrical constraints based on structural data.

Two promising avenues for including long-range forces are periodic quantum calculations[29−31] and embedded clusters.[32−34] The periodic approach, which is typically based on density functional theory (DFT), is relatively straightforward and provides the only reliable method for testing approximations other than those in DFT itself.[35−37] However, in the present context periodic DFT calculations suffer from two main drawbacks: they are expensive, treating all atoms equally, even distant spectator atoms; and they can be applied only to zeolites with sufficiently small unit cells, wherein too many acid sites may be formed. These limitations steer us to the embedded cluster approach.

The embedded cluster idea generally involves breaking a total system (S) into an inner region (I) of chemical interest and an outer environment (O). For zeolites and other covalent-network solids, this partitioning inherently leaves dangling bonds, requiring special methods to saturate the bonding in I. In many implementations of embedding,[38] link atoms are added to the inner region, yielding a cluster (C). One then endeavors to simulate the cluster C with chemical accuracy (hi), while modeling region O with much cheaper methods (lo). The total potential energy is then approximated by

$$E_{\text{embed}} = E_{\text{lo}}(\text{S}) + [E_{\text{hi}}(\text{C}) - E_{\text{lo}}(\text{C})] \quad (1)$$

$$= E_{\text{hi}}(\text{C}) + [E_{\text{lo}}(\text{S}) - E_{\text{lo}}(\text{C})] \quad (2)$$

Although eqs 1 and 2 are identical, they offer different viewpoints on embedding. Equation 1 suggests a low-level treatment of the total system corrected by a high-level cluster calculation, while eq 2 implies a high-level cluster calculation corrected by a low-level treatment of the environment. Although the embedding approach is appealing, two main issues about embedding in zeolite science remain generally unknown: optimal low- and high-levels of theory and optimal sizes of cluster C and total system S. Below we take a systematic approach at addressing this latter issue; in a forthcoming publication, we will report on optimal low- and high-levels of theory for modeling proton transfer in zeolites.

Sauer and co-workers have published many seminal calculations of proton-transfer energies in bare zeolites.[39] Their recent work on embedding clusters with their QM-Pot code is based on the following approach:[12,13,39,40] (i) use periodic boundary conditions for the total system; (ii) use a well-tuned molecular mechanics potential for the low level of theory, preferably one that has been fitted to electronic structure data obtained at the same level used for quantum cluster calculations; (iii) use a shell-model potential for the low level, to account for some electronic polarization;[41,42] (iv) use a reactive potential function for the low level, to mimic energies associated with making and breaking bonds;[43,44] and (v) do not electronically polarize the quantum cluster by its environment, to avoid double counting such polarization which is already treated approximately at the low level. Using this approach, Sierka and Sauer calculated reaction energies and barriers for specific proton jumps in acidic chabazite, H−Y and H-ZSM-5.[13] For the O(1) to O(4) jump in H−Y zeolite, they obtained reaction energies in the range 10−22 kJ mol^{-1} and a bare barrier of 95−100 kJ mol^{-1}, and found that proton transfer barriers are generally increased by long-range forces.[12,13]

We have initiated a research program modeling reactions catalyzed by zeolites of various chemical compositions.[35] Our present study of proton transfer in bare zeolites represents a base case for calibrating methods for future study. To promote the widest applicability by zeolite modelers, we seek to benchmark embedding methods suitable for use with all elements without further parametrization. This represents a departure from the philosophy behind methods such as QM-Pot, which employ reactive force fields finely tuned for specific systems. This paper is our first in a series using the ONIOM embedding procedure[38] in the Gaussian quantum chemistry code.[45,46] We believe this to be a versatile combination for modeling zeolites[47−49] because the implementation of ONIOM in Gaussian allows partitioning into more than two layers and provides molecular mechanics, semiempirical and ab initio methods for each level of theory. In the present article, we show that reliable energies of proton transfer in zeolites can be obtained using smaller clusters embedded in larger clusters and using the universal force field (UFF),[50] a generic nonreactive force field, as the low level of theory. This represents progress toward a simpler prescription for using embedding methods to model zeolite-guest systems of arbitrary composition.

We compare results from constrained and embedded quantum clusters of the same size. In constrained clusters, terminal atoms are fixed at positions suggested by diffraction experiments, while embedded clusters are connected to larger networks. We refer to these as 'hard' and 'soft' restraints, respectively. By comparing clusters of the same size with hard and soft restraints, we determine how various methods of terminating a quantum cluster influence proton-transfer properties. One might imagine that imposing hard restraints would perturb structural and vibrational properties of proton transfer. To address this issue, we compute proton-transfer attempt frequencies and O−Al−O angles in most systems studied.

In the present study we do not include electrostatic interactions between the quantum region and its environment. In a forthcoming publication we will include these electro-

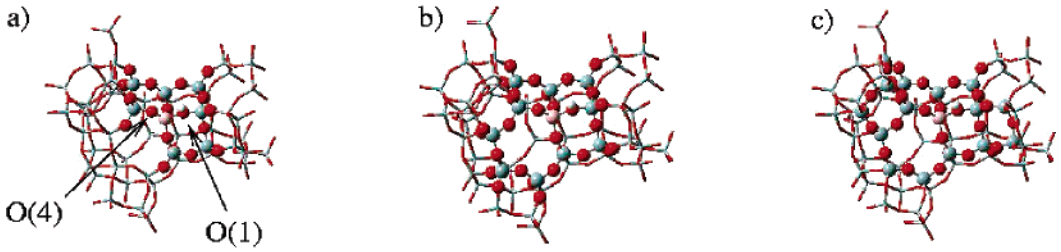


Figure 1. (a) 8T cluster embedded in a 53T system, also denoted 8T-16.95 Å, where 16.95 Å is the radius of the smallest sphere containing the entire system, with oxygens O(1) and O(4) labeled. (b) 11T cluster embedded in a 53T system (11T-16.95 Å). (c) 16T cluster embedded in a 53T system (16T-16.95 Å).

statics, incorporating charges from region O in the Hamiltonian of the quantum cluster.[51] Studying embedded clusters in this stepwise fashion allows us to disentangle mechanical and electronic embedding effects, thereby shedding further light on the energetics of reactions in zeolites.

The remainder of this article is organized as follows: in section II, we describe the zeolite clusters, electronic structure methods, and types of calculations performed. Section III details the results and discusses their implications, and in section IV we give concluding remarks.

II. Computational Methods

In this section, we describe the methods used to study the O(1) to O(4) proton jump in H−Y zeolite. First, we detail the zeolite clusters investigated and the methods used to incorporate properties of an extended system. We then review the electronic structure methods and geometry optimization techniques employed to compute energies, structures, and vibrational frequencies.

A. Zeolite Cluster Models. The chemical system under investigation is H−Y zeolite in the low-Al limit, where a single Si atom in the entire framework is replaced with an Al and charge balanced with a proton. The reaction in question is an internal proton transfer between crystallographically distinct oxygen atoms, O(1) and O(4). All of our models therefore feature a single Al tetrahedral site surrounded by a siliceous framework of varying size.

We study this internal proton transfer in quantum mechanically modeled clusters ranging in size from 3 tetrahedral centers (3T = Al + 2Si) to 22 tetrahedral centers (22T = Al + 21Si). We construct each cluster by clipping pieces from a crystal structure[27] of H−Y zeolite and terminating with hydrogen atoms at appropriate distances from atoms with dangling bonds. In constrained clusters, terminal hydrogens are placed along the vector toward the next atom in the zeolite framework at distances of 0.9 and 1.4 Å for Si−H and SiO−H termination, respectively. If a cluster is terminated primarily with Si−H bonds, the cluster is denoted as a "T-H" cluster; the O−H cluster termination is marked with "T-OH." Atoms were included in a cluster based on their distance from the Al. When necessary, atoms were added to complete a zeolite ring, which avoids placing terminal atoms in too close proximity when terminating dangling bonds. Figure 1 shows examples of various cluster sizes.

Terminating a molecular cluster fails to include the effects of excluded atoms; we attempt to include the most important of these effects through the use of geometric constraints placed on the cluster exterior. As outlined in the Introduction, we compare the effect of hard and soft restraints to gauge their influence on reaction energies. (We do not include fully relaxed clusters in our study, because they relax to rather nonzeolitic structures, especially for small clusters.) In cluster calculations with hard restraints (hereafter denoted "constrained"), the terminal hydrogen atoms are frozen in place, while all other atoms are allowed to relax. As in our previous work,[24] this serves to mimic the covalent "footprint" of the extended zeolite from which the cluster was clipped. Soft restraints are enforced by embedding a smaller cluster in a larger one (hereafter denoted "embedded"), using the Own N-Layer Integrated Molecular Orbital − Molecular Mechanics (ONIOM) method in GAUSSIAN quantum chemistry codes.[45,46] The embedded clusters differ from those in the constrained cluster calculations in the placement of the terminal hydrogens. In the case of constrained clusters, the terminal hydrogens are fixed in place through the entire optimization procedure, at positions determined from the crystal structure. In the case of embedded clusters, terminal hydrogens, also referred to as link atoms, are placed along the bond vectors pointing from the last shell of atoms in the cluster to the first shell of atoms in the outer ONIOM layer. Link atom positions vary during optimization because the outer ONIOM layer is itself flexible, except for terminal atoms on the boundary of the total system, whose positions are fixed at crystallographic locations.

B. Electronic Structure Methods. In an earlier study, we reported that constrained cluster models treated at the B3LYP/6-311G(d,p) level of theory yield accurate results for geometrical parameters and vibrational frequencies.[24] We also found that convergence of electronic energies is best obtained through the focal point method by correcting MP2/6-311G(d,p) energies with the difference {E[MP4/6-31G(d)] − E[MP2/6-31G(d)]}. In that same publication, we found that B3LYP/6-311G(d,p) underestimates reaction barriers by about 10 kJ mol^{-1} for this zeolite system, which is about 10% of the classical barrier height. This accuracy is deemed acceptable for the purposes of this study because such calculations provide a very good compromise between computational efficiency and accuracy. We thus employed the B3LYP/6-311G(d,p) level of theory as our high level throughout this work. Our ONIOM calculations use this basis set and level of theory to compute E_{hi}(C), while the Universal Force Field[50] (UFF) level of theory is used to compute E_{lo}(S) and E_{lo}(C) in eqs 1 and 2.

The choice of UFF for the low level of theory in our ONIOM calculations is justified in several ways. First, UFF contains parameters for all elements in the periodic table and as such provides a broadly applicable low level of theory. Though we do not exploit this generality in the present study, it may prove useful in future work. Second, using UFF allows us to test whether a nonreactive force field can produce accurate proton-transfer energies within ONIOM. The form of the ONIOM energy in eq 2 suggests this is possible, because unphysical energy terms that are introduced by bond breaking events identically cancel in the correction term $[E_{lo}(S) - E_{lo}(C)]$. Third, UFF as parametrized for silicates lacks partial charges. As such, this low level of theory is essentially a ball-and-spring model with minimal electronic effects, making it ideal for determining the influence of only network constraints, without additional complexities associated with long-range forces. This is consistent with our overall stepwise investigation into the various effects of the environment on proton transfer.

C. Calculation Details. We investigated convergence of four reaction parameters with respect to cluster sizes. These parameters are the reaction energy, activation energy, optimized O(1)−Al−O(4) angles for protons at O(1) and at O(4), and the vibrational frequency most closely associated with the O(1) to O(4) proton-transfer reaction coordinate. To compute these parameters, geometry optimizations of each cluster model were performed, followed by a vibrational frequency analysis.

The reaction energy and activation energy were computed as $\Delta E_{rxn} = E_{HO(4)} - E_{HO(1)}$ and $\Delta E_{act} = E_{TS} - E_{HO(1)}$, respectively. The reported energies do not include zero-point or thermal corrections. From these calculations, the O(1)−Al−O(4) angles at the reactant and product geometries were also recorded. In each individual geometry optimization, the cluster energy was converged to 10^{-6} Hartrees and the RMS force to 10^{-4} Hartrees a_o^{-1}. All transition states were confirmed to be first-order saddle points by computing and diagonalizing the mass-weighted Hessian matrix. We calculated vibrational frequencies with the masses of terminal atoms set to 10^6 au to mimic the macroscopic mass of the zeolite framework. For a subset of optimized constrained and embedded clusters, we report vibrational frequencies associated with the early stages of proton transfer, i.e., proton transfer "attempt frequencies".

Most embedding or QM/MM packages use a two-step or double-iteration method for geometry optimization.[34,52] In the first step, a geometry displacement is made only in the QM region (region I), while the MM region (region O) is kept frozen. For this step all the usual "small molecule" geometry optimization techniques are employed, such as a second-order approximation of the potential surface, Hessian update mechanisms, and the use of redundant internal coordinates. We note that QM/MM forces are used to determine the QM step, not just the QM forces. In the second step, the MM region is fully optimized, while the QM region is kept frozen. For this step we use a conjugate gradient optimizer in Cartesian coordinates, which is suitable for large systems. The two steps are repeated until convergence is reached.

The main reason for using this double-iteration method is that it remains intractable to use a second-order optimizer in redundant internal coordinates for very large systems. The required inversion or diagonalization of the Hessian is a computational bottleneck for larger systems, and the coordinate transformations become prohibitively expensive. With the double-iteration method, however, we still use a very accurate optimization method for the QM region, keeping the number of (expensive) QM energy and gradient evaluations to a minimum. The number of MM energy and gradient evaluations will increase significantly, but since these are several orders of magnitude cheaper than QM calculations, they have little effect on the total computational time.

The method outlined above, which uses relatively few "macroiterations" for the QM region and several more "microiterations" for the MM region, has been implemented in several QM/MM packages and has been used in numerous QM/MM studies. There are, however, several serious drawbacks of this scheme, related to the fact that the Hessian is only updated for the QM region. First, large displacements in the MM region, especially when a different local minimum is found, can lead to numerical instability of the Hessian update. Second, in the QM step, there is no direct coupling with the MM region, which deteriorates the quality of the QM step and increases the number of steps needed to reach convergence. Recently, we have addressed these issues by developing techniques that explicitly include coupling between the two regions in the QM step, while still using the double-iteration scheme.[53] This has significantly improved convergence behavior; indeed, we find no significant difference in performance between regular QM geometry optimization and QM/MM geometry optimizations. All the calculations were performed with GAUSSIAN03[45] and Gaussian Development Version[46] on Intel Linux workstations.

III. Results and Discussion

In this section, we describe the optimized geometries, reaction and activation energies, and vibrational frequencies obtained by the computational methods described above for modeling proton transfer from O(1) to O(4) in H−Y zeolite. We note that both experimental data and previous calculations agree that the reaction energy for this transfer should be endothermic, by 10−22 kJ mol^{-1} according to Sauer et al.,[12] which provides a broad target for our convergence studies.

A. Geometric Parameters. The O(1)−Al−O(4) angle affects proton-transfer kinetics by determining the jump distance and hence the activation energy of the reaction. Tables 1 and 2 show the values of O(1)−Al−O(4) angles from the constrained and embedded clusters for a variety of sizes. In Table 1, the O(1) results show reasonable convergence, and only the smallest (5T-H and 8T-H) constrained clusters are slightly off. In contrast, the angles of the embedded clusters (Table 2) are constant throughout the series, including the smallest clusters.

B. Reaction Energies. Table 3 summarizes our results for the proton-transfer reaction energy for both the con-

Table 1. OAlO Angle at Optimized Geometry from Constrained Cluster

	H on O(1)	H on O(4)
5T-H	93	93
8T-H	85	86
8T-OH	105	98
11T-OH	98	96
16T-OH	99	100
22T-OH	102	103

Table 2. OAlO Angle at Optimized Geometry from Embedded Cluster

	H on O(1)	H on O(4)
5T(H)-53T	99	99
8T(OH)-53T	101	105
16T(OH)-53T	100	98
16T(OH)-259T	100	97
22T(OH)-259T	101	100

strained and embedded clusters. Reaction energies of constrained zeolite clusters show stable convergence, reaching a value of 14 ± 2 kJ mol^{-1} at the 11T cluster size. The anomalously large reaction energy found for both 8T-H and 8T-OH clusters is due to a significant distortion of the reactant optimized structure allowed by floppiness of dihedral angles in terminal OSiOH groups. This effect is lessened both in larger clusters and in the ONIOM calculations. All constrained calculations correctly show that the transfer is endothermic. Extending the constrained cluster to include 22T atoms does not significantly affect ΔE_{rxn}, suggesting that convergence with respect to cluster size is reached. This suggestion is further corroborated by the ONIOM results below.

Using the ONIOM method, we seek convergence with respect to both the quantum cluster size and the total system size, which are represented as the rows and columns of Table 3, respectively. By scanning across any row, it is clear that changing the total system size has little effect on the reaction energy. At first this may be surprising, because the larger calculations include an order of magnitude more atoms. However, because the lower level of theory employed is UFF and therefore lacks long range forces in our calculations, there is only a local mechanical coupling between the inner and outer layers of the ONIOM model.

The size of the inner layer does, however, strongly affect the calculated reaction energy in approximately the same fashion as seen in the constrained cluster calculations. Here, the values for each total system size approach convergence at the 8T-OH quantum cluster size. Again, we observe that describing larger sections of the total system with a quantum mechanical method does not markedly improve the final result.

In our work all the larger clusters are terminated with hydroxyl groups; only for 8T can we compare the effect of hydrogen termination vs hydroxyl termination. We see indeed a large difference in the embedded 8T data, which was also reported by Brand et al. for proton affinities of acid sites in ZSM-5.[54]

In general, we find that a finite cluster, whether constrained or embedded, appears to be sufficient for predicting proton transfer reaction energies in zeolites. The use of methods based on periodic boundary conditions actually masks the length scales beyond which interactions no longer influence proton transfer. Determining such length scales is of fundamental importance for understanding how much of the zeolite actually controls reaction dynamics. The concept of a finite interaction length may come as a surprise considering that many researchers employ Ewald sums when simulating Coulombic interactions in partially ionic media such as zeolites.[55] Such long-range summations are crucial for calculating *absolute* adsorption energies of molecules in zeolites relative to vacuum. However, for calculating energy differences between nearby configurations of an adsorbed species, for example a reaction energy or barrier, the present results suggest that long-range interactions essentially cancel, leaving an energy difference controlled by local electronic interactions.

Our data suggest that this proton-transfer reaction is sensitive to atoms within ca. 5 Å of the reaction center. The presence of guest molecules will likely change this cutoff distance, but even in this more complex scenario we still expect a finite cluster to capture the reaction energetics. It is reasonable to surmise that using a low level of theory that includes electronic interactions would enable the use of even smaller inner regions in embedded calculations. We explore this possibility in future work.

C. Activation Energies. We calculated activation energies for constrained clusters and a subset of embedded clusters (Table 4). As discussed above, all transition states were confirmed to be first-order saddle points by the usual Hessian analysis. For the embedded clusters we considered total sizes of 53T and 166T. Although the number of data points is

Table 3. ΔE_{rxn} for O(1) to O(4) Proton Transfer[a]

		embedded (with size of full system in parentheses)					
cluster size	constrained	23 T (7 Å)	53 T (10 Å)	98 T (12.44 Å)	166T (14.8 Å)	259T (16.95 Å)	439T (19.91 Å)
3T−H	2.5	1.3	1.3	2.0	1.8	1.8	1.6
5T−H	4.3	5.4	5.9	6.9	6.5	6.4	6.3
8T−H	22.6	8.4	9.9	10.7	10.3	10.7	10.6
8T-OH	26.7		14.7	15.4	15.4	15.2	
11T-OH	13.6		14.4	14.9	15.0	14.8	
16T-OH	14.6					18.3	
22T-OH	13.8					15.4	

[a] In kJ mol^{-1}.

Table 4. ΔE_{act} for O(1) to O(4) Proton Transfer[a]

		embedded (with total size)	
cluster size	constrained	53T (10 Å)	166T (14.8 Å)
3T−H	90.2		118.9
5T−H	93.6	128.3	126.8
8T−H	57.3		134.7
8T-OH	138.8	137.8	139.2
11T-OH	133.4	133.7	136.5
16T-OH	131.6		132.7
22T-OH	138.8		

[a] In kJ mol^{-1}.

smaller than for the reaction energy, this still allows us to investigate convergence for both the total and quantum cluster size. From Table 4 we see that for embedded clusters, independent of total cluster size, the barrier converges at quantum cluster size 8T-OH. This was also the case for the reaction energy treated by embedded clusters (Table 3). The constrained barrier calculations also converge at quantum cluster size 8T-OH. These results further suggest that finite clusters, either constrained or embedded, can produce stable energetics of proton transfer in zeolites.

Our converged barrier is about 135 kJ mol^{-1} for both embedded and constrained clusters. For comparison, Sierka and Sauer obtained a barrier of about 100 kJ mol^{-1} for the same proton jump.[13] About 10 kJ mol^{-1} of this difference is a basis set effect. This was determined by repeating our computation of the barrier height in the 8T(OH)-53T embedded cluster using the Alrichs basis set used by Sierka and Sauer. This gave a barrier of 127.0 kJ mol^{-1}, as compared to 137.8 kJ mol^{-1} obtained using the 6-311G(d,p) basis set (see Table 4).

The remaining difference in barriers is harder to pin down. Our calculations and those of Sierka and Sauer differ mainly in the way volume is constrained and in the treatment of outer layer atoms. Regarding the former, we have built embedded and constrained zeolite clusters from diffraction data, fixing terminal atoms at experimentally determined locations. In contrast, Sierka and Sauer apply periodic boundary conditions at constant volume, with the lattice parameter determined from an initial force field optimization of the H−Y system at constant pressure. It is possible that our method of fixing terminal atoms can impose strain on reactant and transition state configurations. Such strain is expected to diminish as the total system size increases, thus placing terminal atom constraints farther away from the reaction center. However, the fact that barriers from our 8T-OH and 11T-OH quantum clusters (constrained and embedded) remain stable with respect to total system size suggests that our method of constraining volume does *not* impose unphysical strain on this proton-transfer process.

Our calculations and those of Sauer and Sierka also differ in the treatment of outer layer atoms. Our calculations represent electronic effects only in the quantum cluster, while those of Sierka and Sauer include classical electrostatics in the outer layer as well. It is possible that our local treatment of electronic effects can introduce errors into reactant and transition state energies. Such errors are expected to diminish as the quantum cluster size increases. The fact that barriers from quantum clusters 8T-OH and larger (constrained and embedded in Table 4) remain stable with respect to quantum cluster size suggests that our calculations include the electronic effects relevant for this proton transfer process. To pursue this point further, we will report in a forthcoming publication the results of fully periodic quantum calculations on this system.[15]

Table 5. Vibrational Frequency in Wavenumbers (cm^{-1}) for Constrained Cluster

	H on O(1)
5T−H	1054
8T-OH	1039
11T-OH	1011

Table 6. Vibrational Frequency in Wavenumbers (cm^{-1}) for Embedded Cluster

	H on O(1)	H on O(4)
5T(H)-23T	1064	1041
8T(H)-53T	1037	1030

Table 7. Calculation Times for Optimizing the System with the Proton at O(1)

system	av time per optimization cycle (h)	no. of optimization cycles	total CPU time (h)
22T-OH *constrained*	10.5	60	630
22T(OH)-259T *embedded*	10.7	20	213
11T-OH *constrained*	4.5	35	155
11T(OH)-259T *embedded*	4.9	20	98.8
8T-OH *constrained*	2.5	60	153
8T(OH)-53T *embedded*	2.6	20	51.8

D. Vibrational Frequencies. Vibrational frequencies are important dynamical parameters for quantifying activation entropies and attempt frequencies. Vibrational frequency analyses were performed on selected constrained and embedded clusters to test the convergence behavior of frequencies with respect to cluster size. Tables 5 and 6 show the vibrational frequency of the normal mode with the largest component of Al−O−H wag for the proton situated at O(1), which corresponds closely with the proton-transfer reaction coordinate.

Both sets of calculations show very similar vibrational frequencies for the H wag. The variation from one cluster size to another is smaller than the expected uncertainty of the electronic structure method. This is a bit surprising, considering that the hard termination inherent in constrained clusters might be expected to shift these to higher frequencies. However, our results suggest that either method of applying geometric constraints to the clusters is sufficient to achieve convergence of this particular dynamical parameter.

E. Computational Time Comparisons. In Table 7, we show CPU time comparisons for pairs of constrained and embedded calculations that contain identical quantum clusters. The times presented are for geometry optimizations starting from the crystal structure, with the acidic hydrogen added in a reasonable position relative to O(1). All timing

calculations were performed on dual processor 2.8 GHz PIV computers with 1 GB of core memory running Linux RedHat 9. Embedded cluster timings were obtained with the quadratically coupled QM/MM geometry optimization method.[45,46,53] For purposes of comparing CPU times, the optimizations were considered converged when the energy is stable to 0.00005 H (0.13 kJ mol^{-1}), which is the degree of precision we seek in our reaction energies. We allowed the optimizations to run significantly longer to test the robustness of this criterion and found only negligible changes in structure or energy. In Table 7, we also show the number of macroiterations required for geometry optimizations and the average CPU time required for each macroiteration.

In principle, we expect very similar times per optimization cycle for identical quantum clusters regardless of the method of constraint, because in the embedded calculations, the time used to compute energies and forces using UFF is negligible compared to the time spent in the quantum part of the calculation. This is precisely what we observe in Table 7. These timing data suggest that the principal difference between a constrained calculation and an embedded calculation on identical quantum clusters is simply the number of optimization cycles required to achieve convergence. In this respect, the constrained calculations appear to be *much slower* since they require many more optimization cycles. This poor convergence behavior may arise from a mismatch between floppy OSiOH dihedral angles and fixed terminal hydrogen atoms. More work is required to better understand this phenomenon.

Perhaps the most straightforward and important comparison in Table 7 is between the constrained and embedded cluster calculations that give converged reaction energies in the least time. These are the constrained 11T-OH and embedded 8T(OH)-53T systems, which converge in 155 and 51.8 CPU h, respectively. Thus, embedding speeds up these geometry optimizations by a factor of 3, despite the use of a generic low-level of theory (UFF) in ONIOM calculations.

IV. Concluding Remarks

We have studied the convergence properties of embedded and constrained cluster models of proton transfer in zeolites. We applied density functional theory to describe clusters and ONIOM to perform the embedding. We focused on converging the reaction energy and barrier of the O(1) to O(4) jump in H−Y zeolite as well as vibrational and structural aspects of this jump. We found that using successively larger clusters in vacuo gives convergence of the reaction energy to 14 ± 2 kJ mol^{-1}, and the barrier to 135 ± 5 kJ mol^{-1}, as long as the clusters are constrained to mimic zeolitic structures. These calculations converged for clusters with radii larger than 5 Å, containing at least 11 tetrahedral (Si or Al) atoms. We embedded quantum clusters of various sizes in larger clusters with total radii in the range 7−20 Å, using the universal force field (UFF) as the lower level of theory in ONIOM. We found convergence of the proton-transfer energy without the use of reactive force fields or periodic boundary conditions in the embedding procedure. Embedded cluster calculations gave converged reaction energies for quantum clusters containing at least 8 tetrahedral atoms. Optimizations on embedded clusters required many fewer cycles, and hence much less CPU time, than did optimizations on comparable constrained clusters.

Our present results suggest a reaction energy of 14 ± 2 kJ mol^{-1} and a barrier of 135 ± 5 kJ mol^{-1} for the O(1) to O(4) proton jump in H−Y zeolite. The smallest systems that yield reaction energies converged to within 1 kJ mol^{-1} of this value are the 11T-OH constrained cluster and the 8T(OH)-53T embedded cluster. These quantum clusters include atoms within ca. 5 Å of the reactive center. In future work, we will explore whether greater computational efficiency can be obtained by splitting the reactive region into two layers described by an accurate ab initio theory and a cheaper electronic theory, thereby giving a three-layer ONIOM calculation.[56] In the end, we aim for a simple and user-friendly method for modeling a wide variety of reactions in zeolites.

Acknowledgment. S.M.A. thanks the UMass Amherst Department of Chemistry for generous funding of our computer laboratory.

References

(1) *Handbook of Zeolite Science and Technology*; Auerbach, S. M., Carrado, K. A., Dutta, P. K., Eds.; Marcel Dekker: New York, 2003.

(2) Corma, A. *Chem. Rev.* **1995**, *95*, 559.

(3) Van Santen, R. A.; Kramer, G. J. *Chem. Rev.* **1995**, *95*, 637.

(4) Sauer, J. *Chem. Rev.* **1989**, *89*, 199.

(5) Farneth, W. E.; Gorte, R. J. *Chem. Rev.* **1995**, *95*, 615.

(6) Baba, T.; Ono, Y. *Appl. Catal. A* **1999**, *181*, 227.

(7) Baba, T.; Komastsu, N.; Ono, Y.; Sugaisawa, H. *J. Phys. Chem. B* **1998**, *120*, 804.

(8) Baba, T.; Komatsu, N.; Ono, Y.; Sugisawa, H.; Takahashi, T. *Microporous Mesoporous Mater.* **1998**, *22*, 203.

(9) Ernst, H.; Freude, D.; Mildner, T.; Pfeifer, H. High temperature ^{1}H MAS NMR studies of proton mobility in zeolites. In *Proceedings of the 12th International Zeolites Conference*; Treacy, M. M. J., Marcus, B. K., Bisher, M. E., Higghins, J. B., Eds.; MRS: Warrendale, PA, 1999; Vol. IV, p 2955.

(10) Sarv, P.; Tuherm, T.; Lippmaa, E.; Keskinen, K.; Root, A. *J. Phys. Chem.* **1995**, *99*, 13763.

(11) Brandle, M.; Sauer, J.; Dovesi, R.; Harrison, N. M. *J. Chem. Phys.* **1998**, *109*, 10379.

(12) Sauer, J.; Sierka, M.; Haase, F. Acid Catalysis by Zeolites: Ab Initio Modeling of Transition Structures. In *Transition State Modeling for Catalysis, Chapter 28*; Truhlar, D. G., Morokuma, K., Eds.; American Chemical Society: Washington, 1999; p 358.

(13) Sierka, M.; Sauer, J. *J. Phys. Chem. B* **2001**, *105*, 1603.

(14) Ryder, J. A.; Chakraborty, A. K.; Bell, A. T. *J. Phys. Chem. B* **2000**, *104*, 6998.

(15) Hill, J. R.; Freeman, C. R.; Delley, B. *J. Phys. Chem. A* **1999**, *103*, 3772.

(16) Limtrakul, J.; Khongpracha, P.; Jungsuttiwong, S.; Truong, T. N. *J. Mol. Catal. A* **2000**, *153*, 155.

(17) Limtrakul, J.; Nanok, T.; Jungsuttiwong, S.; Khongpracha, P.; Truong, T. N. *Chem. Phys. Lett.* **2001**, *349*, 161.

(18) Boronat, M.; Zicovich-Wilson, C. M.; Viruela, P.; Corma, A. *J. Phys. Chem. B* **2001**, *105*, 11169.

(19) Sinclair, P. E.; Vries, A. d.; Sherwood, P.; Catlow, C. R. A.; Van Santen, R. A. *J. Chem. Soc., Faraday Trans* **1998**, *94*, 3401.

(20) Rozanska, X.; Van Santen, R. A.; Hutschka, F.; Hafner, J. *J. Am. Chem. Soc.* **2001**, *123*, 7655.

(21) Zygmunt, S. A.; Curtiss, L. A.; Zapol, P.; Iton, L. E. *J. Phys. Chem. B* **2000**, *104*, 1944.

(22) Khaliullin, R. Z.; Bell, A. T.; Kazansky, V. B. *J. Phys. Chem. A* **2001**, *105*, 10454.

(23) Martinez-Magadan, J. M.; Cuan, A.; Castro, M. *Int. J. Quantum Chem.* **2002**, *88*, 750.

(24) Fermann, J. T.; Blanco, C.; Auerbach, S. M. *J. Chem. Phys.* **2000**, *112*, 6779.

(25) Fermann, J. T.; Auerbach, S. M. *J. Chem. Phys.* **2000**, *112*, 6787.

(26) Fermann, J. T.; Auerbach, S. M. *J. Phys. Chem. A* **2001**, *105*, 2879.

(27) Czjzek, M.; Jobic, H.; Fitch, A. N.; Vogt, T. *J. Phys. Chem.* **1992**, *96*, 1535.

(28) Gonzales, N. O.; Bell, A. T.; Chakraborty, A. K. *J. Phys. Chem. B* **1997**, *101*, 10058.

(29) Kresse, G.; Furthmuller, J. *Phys. Rev. B* **1996**, *54*, 11169.

(30) Civalleri, B.; D'Arco, P.; Orlando, R.; Saunders, V. R.; Dovesi, R. *Chem. Phys. Lett.* **2001**, *348*, 131.

(31) Shah, R.; Payne, M. C.; Lee, M. H.; Gale, J. D. *Science* **1996**, *271*, 1395.

(32) Sierka, M.; Sauer, J. *J. Comput. Chem.* **2000**, *21*, 1470.

(33) Greatbanks, S. P.; Sherwood, P.; Hillier, I. H. *J. Phys. Chem.* **1994**, *98*, 8134.

(34) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.

(35) Astala, R.; Auerbach, S. M. *J. Am. Chem. Soc.* **2004**, *126*, 1843.

(36) Astala, R.; Auerbach, S. M.; Monson, P. A. *J. Phys. Chem. B* **2004**, *108*, 9208.

(37) Astala, R.; Auerbach, S. M.; Monson, P. A. *Phys. Rev. B* **2005**, *71*, 014112.

(38) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.-Theochem* **1999**, *461*−*462*, 1.

(39) Sauer, J.; Kolmel, C. M.; Hill, J.-R.; Ahlrichs, R. *Chem. Phys. Lett.* **1989**, *164*, 193.

(40) Sierka, M.; Sauer, J. *J. Chem. Phys.* **2000**, *112*, 6983.

(41) Catlow, C. R. A.; Dixon, M.; Mackrodt, W. C. In *Computer Simulations of Solids*; Catlow, C. R. A., Mackrodt, W. C., Eds.; Springer-Verlag: Berlin, 1982; Vol. 166, p 130.

(42) Gale, J. D. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 629.

(43) Aqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.

(44) Chang, Y.-T.; Miller, W. H. *J. Phys. Chem.* **1990**, *94*, 5884.

(45) 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 C.02 ed.; Gaussian, Inc.: Wallingford, CT, 2004.

(46) 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 Development Version; Revision C.01 ed.; Gaussian, Inc.: Wallingford, CT, 2004.

(47) Damin, A.; Bordiga, S.; Zecchina, A.; Lamberti, C. *J. Chem. Phys.* **2002**, *117*, 226.

(48) Damin, A.; Bonino, F.; Ricchiardi, G.; Bordiga, S.; Zecchina, A.; Lamberti, C. *J. Phys. Chem. B* **2002**, *106*, 7524.

(49) Sillar, K.; Purk, P. *J. Mol. Struct.-Theochem.* **2002**, *589*, 281.

(50) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

(51) Vreven, T.; Morokuma, K. *Theor. Chem. Acct.* **2003**, *109*, 125.

(52) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.

(53) Vreven, T.; Frisch, M. J.; Kudin, K. N.; Schlegel, H. B.; Morokuma, K. *Mol. Phys.* **2005**, in press.

(54) Brand, H. V.; Curtiss, L. A.; Iton, L. E. *J. Phys. Chem.* **1992**, *96*, 7725.

(55) Auerbach, S. M.; Jousse, F.; Vercauteren, D. P. Dynamics of Sorbed Molecules in Zeolites. In *Computer Modelling of Microporous and Mesoporous Materials*; Catlow, C. R. A., Van Santen, R. A., Smit, B., Eds.; Academic Press: London, 2004; p 49.

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

CT0501203

JCTC Journal of Chemical Theory and Computation

Accurate QM/MM Free Energy Calculations of Enzyme Reactions: Methylation by Catechol *O*-Methyltransferase

Thomas H. Rod* and Ulf Ryde

Department of Theoretical Chemistry, Chemical Center, Lund University, P.O. Box 124, S-22100 Lund, Sweden

Received April 25, 2005

Abstract: We recently described a method to compute accurate quantum mechanical free energies [Rod, T. H.; Ryde, U. *Phys. Rev. Lett.* **2005**, *94*, 138302]. The method, which we term quantum mechanical thermodynamic cycle perturbation (QTCP), employs a molecular mechanics force field to sample phase space and, subsequently, a thermodynamic cycle to estimate QM/MM free energy changes. Here, we discuss the methodology in detail and test an approach based on a different thermodynamic cycle. We also show that a new way of treating hydrogen link atoms makes the free energy changes converge faster and that extrapolation to higher accuracy can be performed. We finally discuss the quantum mechanical free energy (QM/MM-FE) method in the framework of the QTCP method. All methods considered are applied to the methylation of catecholate catalyzed by catechol *O*-methyltransferase. We compute the free energy barrier for the reaction by computing free energy changes in steps between fixed QM regions along a predetermined reaction pathway. Using the QTCP approach, an extrapolated activation free energy of 69 kJ/mol for the forward reaction and 90 kJ/mol for the reverse reaction are obtained at the level of the B3LYP functional and the 6-311++G(2d,2p) basis set. The value for the forward reaction is in excellent agreement with the experimental value of 75 kJ/mol. Results based on the QM/MM-FE method differ by less than 10 kJ/mol from those values, indicating that QM/MM-FE may be a fairly accurate and cheap alternative to calculate QM/MM free energy changes. Moreover, the results are compared to barriers obtained with a fixed molecular mechanics environment as well as with structures optimized in a vacuum. All the computed free energy barriers are well converged. A major approximation in the current implementation of the QTCP method is that the QM region is fixed. The approximation leads to well-converged free energy barriers, which has been a problem in similar studies.

1. Introduction

The general problem with atomic-scale simulations of enzyme catalysis is that thousands of atoms are involved and chemical bonds are modified. The description of bond breaking or forming calls for fairly accurate quantum mechanical (QM) methods, whereas the huge phase space calls for sampling with a cheaper method like molecular mechanics (MM) or semiempirical QM. The most straightforward way to solve the problem is by sampling phase space using a combined quantum mechanics and molecular mechanics (QM/MM) method where a small subset of atoms are treated by QM and the remaining atoms by MM. This can be pursued either by sampling directly on the Born−Oppenheimer surface or by using a Car-Parrinello approach. The problem is that realistic simulation times are currently limited to a few tens of picoseconds for this approach because of the severe computational load even for small QM systems. This is a rather short simulation time for systems as big as enzymes where many events, besides the reaction catalyzed,

* Corresponding author phone: +46 46 222 4500; e-mail: Thomas.Rod@teokem.lu.se.

10.1021/ct0501102 CCC: $30.25
Published on Web 08/30/2005

occur on time-scales much longer than picoseconds.[1] Various approaches have been proposed and utilized in the literature to solve the problem of computing accurate QM/MM free energies for chemical reactions in solutions[2−12] including enzymatic reactions.[13−23] A basic idea is to use a fast but less accurate method to sample phase space and use this sampling to estimate high-level QM/MM free energies with a modest number of QM/MM calculations. Below we give a summary of methods based on this idea.

In the quantum mechanical free energy (QM-FE) approach by Jorgensen and co-workers,[2−4] a reaction pathway for atoms in the QM region is calculated in a vacuum. Free energies for the interaction between the QM and MM atoms are then calculated along the reaction pathway by performing MM free energy perturbation or thermodynamic integration calculations where electrostatic interactions between the QM and MM atoms are defined via point charge interactions. In the treatment by Jorgensen and co-workers, and later by Kollman and co-workers,[13−15] point charges to represent the QM atoms were derived from calculations in a vacuum, i.e., without an MM region. Jorgensen and co-workers used the method to study organic reactions in solution, and Kollman and co-workers extended the method to that of enzymatic reactions, namely amide hydrolysis in trypsin[13] and methyl transfer by catechol *O*-methyltransferase.[14]

Yang and co-workers applied their QM/MM free energy method (QM/MM-FE) to the enzymes triosephosphate isomerase,[16,17] enolase,[18,17] and 4-oxalocrotonate tautomerase.[19] The QM/MM-FE method is an improvement over the QM-FE method in that a QM/MM optimized reaction pathway and QM energies and point charges derived from QM/MM calculations are used. In this way, polarization of the QM region by the MM region is included. Ishida and Kato employed the same approach to study acylation by serine proteases.[20,21]

An alternative approach is the ab initio QM/MM approach (QM(ai)/MM) by Warshel and co-workers.[5,6,22−24] They sampled phase space by performing molecular dynamics (MD) simulations with a reference potential given by the empirical valence bond (EVB) method.[25] Umbrella sampling ensured that the entire reaction pathway was sampled and made it possible to calculate the potential of mean force (PMF). Free energy changes between the system described by the reference potential and by density functional theory were calculated with free energy perturbation, and in this way a high-level QM/MM PMF can be obtained. The methodology corresponds to using the thermodynamic cycle in Figure 1 (with MM replaced by EVB) and is in principle exact with respect to how the free energy changes are calculated. In practice, the free energy did not converge owing to large fluctuations of the difference between the reference potential and the high-level QM/MM potential, although electrostatic interactions converged. Therefore, Warshel and co-workers used more approximate methods to calculate the free energy difference between the system described by EVB and by a high-level QM/MM method.[6,23]

Like Warshel and co-workers, Wood and co-workers use free energy perturbation in their ab initio/classical free energy perturbation (ABC-FEP) approach, which was used to compute hydration energies of water and the Na^+ and Cl^- ions at different physical conditions.[26−29] In this approach, only solute−solvent interaction energies are perturbed to the QM level. They also studied solute−solvent structural properties as well as water dimer dissociation.[30,31]

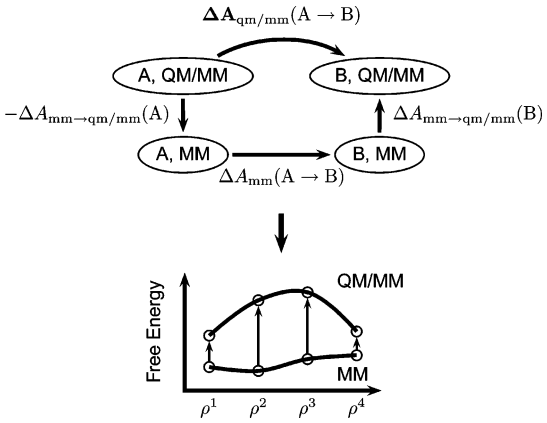


Figure 1. Illustration of the QTCP method, where a thermodynamic cycle is employed to calculate QM/MM free energy changes. The lower figure shows how this can be applied to calculate the free energy barrier along a reaction coordinate defined by ρ. We refer to this particular implementation of the QTCP method by QTCP-U.

Schofield and co-workers and Bandyopadhyay have developed a similar approach termed the molecular mechanics importance based function (MMBIF) method.[7−12] They also use a MM reference potential to sample the phase space and calculate corresponding high-level QM/MM energies for a set of configurations. Based on the two sets of energies, they use a Metropolis-Hastings algorithm to generate a high-level QM/MM canonical ensemble from which QM/MM free energies can be calculated.

Recently, we employed a third approach,[32] which can be considered as a combination of the QM/MM-FE method and the approach by Warshel and co-workers. Like Yang and co-workers as well as Ishida and Kato, we optimized a reaction pathway using QM/MM and selected a number of configurations for the QM region along the reaction pathway. Based on calculated point charges for the QM region, we calculated classical MM-QM interaction free energy changes between subsequent fixed QM configurations along the reaction pathway. Similar to the approach by Warshel and co-workers, we then calculated the MM → QM free energy change for each QM configuration along the reaction pathway, and, in this way, a high-level QM/MM PMF was obtained. With this approach, we obtained a converged PMF for the methyl transfer reaction in catechol *O*-methyltransferase (COMT). Consistent with results by Warshel and co-workers,[22] it shows that the *electrostatic* interaction energies between the QM region and the MM region can be converged to high accuracy. Interactions within the QM region are by definition completely converged in our application since the QM region was fixed in the sampling process.

Here we discuss the approach in a more general framework and study the importance of sampling the phase space of the MM region and ways to limit the computational cost. In

particular, we show that the QM/MM-FE method may be a fair approximation to our approach.

The article is organized as follows. In the next section, we describe the methodology employed, including our treatment of link atoms, and a derivation of the QM-FE and QM/MM-FE approaches. A description of the computational setup follows. Results for the methyl transfer reaction catalyzed by COMT are discussed in section 3. We use two different thermodynamic cycle approaches and demonstrate that free energies can be extrapolated accurately to a different basis set and exchange-correlation functional. The report ends with some concluding remarks in section 4.

2. Method

2.1. Thermodynamic Cycle Approach. The thermodynamic cycle depicted in Figure 1 makes it possible to compute high-level QM/MM free energy changes between two states A and B based on classical sampling. In this approach, the free energy change between A and B described by QM/MM is calculated as the sum of three terms: (1) the negative free energy change between A described by MM and by QM/MM ($-\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{A})$), (2) the free energy change between A and B with both described by the MM potential ($\Delta A_{\mathrm{mm}}(\mathrm{A}\rightarrow\mathrm{B})$), and (3) the free energy change between B described by the MM potential and by QM/MM ($\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{B})$). Hence

$$\Delta A_{\mathrm{qm/mm}}(\mathrm{A}\rightarrow\mathrm{B}) = -\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{A}) + \Delta A_{\mathrm{mm}}(\mathrm{A}\rightarrow\mathrm{B}) + \Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{B}) \quad (1)$$

The methodology is quite general and can be applied to compute free energy changes between different chemical species. For instance, A and B in Figure 1 can be a protonated and deprotonated group in order to compute a pK_a value. The methodology can also be used to compute the free energy along a reaction coordinate as indicated in the lower panel of Figure 1. We term the method the quantum mechanical thermodynamic cycle perturbation (QTCP) method. Because the shape of the thermodynamic cycle in Figure 1 resembles an edged U, we use the acronym QTCP-U when referring specifically to the cycle in Figure 1 (as opposed to the one in Figure 2).

Each of the three terms in the above equation can be calculated by means of free energy perturbation (FEP).[33,34] In the FEP approach, a free energy change, $\Delta A(0\rightarrow 1)$ is calculated as

$$\mathrm{e}^{-\Delta A(0\rightarrow 1)/k_{\mathrm{B}}T} = \langle \mathrm{e}^{-(E_1-E_0)/k_{\mathrm{B}}T}\rangle_0 \quad (2)$$

where k_{B} is Boltzmann's constant and T is the temperature. E_0 and E_1 are the energies of system 0 and 1, respectively, and $\langle ... \rangle_0$ indicates an ensemble average for system 0. Hence, the free energy changes in eq 1 can all be calculated as an ensemble average for the system described by the MM potential, meaning that a QM/MM free energy change can be obtained without sampling the QM/MM potential surface but entirely by sampling the surface of the MM potential.

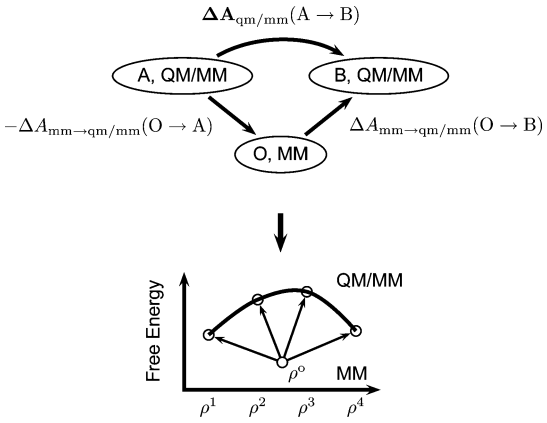


Figure 2. The QTCP method, where a single simulation is used to compute a QM/MM free energy change (QTCP-V). The lower panel shows how this approach can be used to obtain a free energy barrier along a reaction pathway.

Using FEP, the middle term on the right side of eq 1 can be written as

$$\Delta A_{\mathrm{mm}}(\mathrm{A}\rightarrow\mathrm{B}) = -k_{\mathrm{B}}T \ln \langle \mathrm{e}^{-[E^{\mathrm{tot}}_{\mathrm{mm}}(\mathrm{B})-E^{\mathrm{tot}}_{\mathrm{mm}}(\mathrm{A})]/k_{\mathrm{B}}T}\rangle_{\mathrm{mm,A}} \quad (3)$$

and the first and last terms can be written as

$$\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(X) = -k_{\mathrm{B}}T \ln \langle \mathrm{e}^{-[E^{\mathrm{tot}}_{\mathrm{qm/mm}}(X)-E^{\mathrm{tot}}_{\mathrm{mm}}(X)]/k_{\mathrm{B}}T}\rangle_{\mathrm{mm,X}} \quad X = \mathrm{A, B} \quad (4)$$

In the equations above, $E^{\mathrm{tot}}_{\mathrm{qm/mm}}$ and $E^{\mathrm{tot}}_{\mathrm{mm}}$ are the QM/MM and MM potential energies.

The approach above demands an MD simulation per state involved, but the free energy change can in principle be computed from a single simulation. Figure 2 shows how this can be pursued using a single simulation of an intermediate state, O, which we will term an anchor point. In this case, the free energy change, $\Delta A_{\mathrm{qm/mm}}(\mathrm{A}\rightarrow\mathrm{B})$, is calculated as the sum of two terms

$$\Delta A_{\mathrm{qm/mm}}(\mathrm{A}\rightarrow\mathrm{B}) = -\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{O}\rightarrow\mathrm{A}) + \Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{O}\rightarrow\mathrm{B}) \quad (5)$$

where

$$\Delta A_{\mathrm{mm}\rightarrow\mathrm{qm/mm}}(\mathrm{O}\rightarrow X) = -k_{\mathrm{B}}T \ln \langle \mathrm{e}^{-[E^{\mathrm{tot}}_{\mathrm{qm/mm}}(X)-E^{\mathrm{tot}}_{\mathrm{mm}}(X)]/k_{\mathrm{B}}T}\rangle_{\mathrm{mm,O}} \quad (6)$$

Figure 2 shows that the idea also can be applied to compute free energy changes between several points on a reaction pathway. We refer to this particular implementation of the QTCP method by QTCP-V, because the shape of the cycle in Figure 2 resembles a V.

2.2. Definition of Energy Terms. Before discussing the QM-FE and QM/MM-FE methodologies we need to consider the involved potential energy terms in more detail. We denote coordinates of the atoms in the QM and MM region by R_{q} and R_{m}, respectively. We start by defining the MM energy, $E^{\mathrm{tot}}_{\mathrm{mm}}$, in eqs 3, 4, and 6. This term can be separated into three terms

$$E^{\mathrm{tot}}_{\mathrm{mm}} = E^{\mathrm{m}}_{\mathrm{mm}}(R_{\mathrm{m}}) + E^{\mathrm{q/m}}_{\mathrm{mm}}(R_{\mathrm{q}}, R_{\mathrm{m}}) + E^{\mathrm{q}}_{\mathrm{mm}}(R_{\mathrm{q}}) \quad (7)$$

where E^{m}_{mm} describes interactions between MM atoms and $E^{q/m}_{mm}$ describes interactions between the MM and QM regions. The last term, E^{q}_{mm}, describes interactions within the QM region. All these classical terms are defined via a force field, and they can be calculated by standard MM programs such as CHARMM.[35]

To ease the discussion to come, we can further decompose $E^{q/m}_{mm}$ into two terms

$$E^{q/m}_{mm}(R_q, R_m) = E_{ne}(R_q, R_m) + E_{elec}(R_q, R_m) \quad (8)$$

where E_{ne} describes all interactions between the MM region and the QM region besides point charge interactions, which are described by the last term. This term is given by

$$E_{elec}(R_q, R_m) = \sum_{\alpha\in mm}\sum_{\beta\in qm}\frac{Z_\alpha Z_\beta}{|R_\alpha - R_\beta|} \quad (9)$$

where the summations run over MM and QM atoms, and where Z_α and Z_β are point charges of MM and QM atoms, respectively.

In a similar fashion, the total energy, $E^{tot}_{qm/mm}$, can be decomposed into a term that depends solely on the positions of the MM atoms and terms that depend on the positions of both sets of atoms

$$E^{tot}_{qm/mm} = E^{m}_{mm}(R_m) + E_{qm/mm}(R_q, R_m) + E_{corr}(R_q, R_m) \quad (10)$$

where E^{m}_{mm} is the same as in eq 7. E_{corr} is a link atom correction term, which is discussed in the next section.

The term, $E_{qm/mm}$ in eq 10, is composed of a quantum mechanical term, E_{quant}, and a classical interaction term, E_{ne}

$$E_{qm/mm}(R_q, R_m) = E_{quant}(R_q, R_m) + E_{ne}(R_q, R_m) \quad (11)$$

where E_{ne} is the same as in eq 8. In density functional theory notation, the quantum mechanical term is given by

$$E_{quant}(R_q, R_m) = \bar{E}_{vac}(R_q)[n(R_q, R_m)] + \bar{V}_{ext}(R_q, R_m)[n(R_q, R_m)] \quad (12)$$

In this equation, density functionals are indicated by bars, and n denotes the ground-state electron density of the functional $\bar{E}_{vac} + \bar{V}_{ext}$. $\bar{E}_{vac}$ is the density functional for the system in a vacuum, i.e., without external point charges, and $\bar{V}_{ext}$ describes the electrostatic interactions between QM and MM atoms and is given by

$$\bar{V}_{ext}(R_q, R_m)[n] = -\int dr \sum_{\alpha\in mm}\frac{Z_\alpha n(r)}{|R_\alpha - r|} + V_{nuc,\,mm}(R_q, R_m) \quad (13)$$

where the last term defines electrostatic interactions between MM atoms and QM nuclei. Eq 13 is the QM analogue to eq 9. E_{quant} involves polarization of the QM region by the MM region and can be computed by a QM program where external point charges can be included, such as Turbomole.

2.2.1. Link Atom Correction. When calculating E_{quant}, special action is needed when there is a covalent bond between the QM region and the MM region, and several ways to handle such junctions are described in the literature.[36]

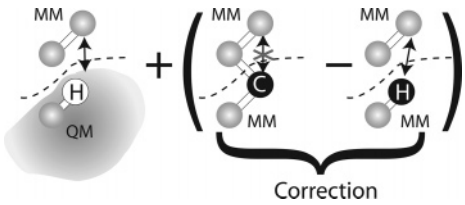


Figure 3. Illustration of the link atom correction employed in the present study. The white ball is a hydrogen link atom, and the black ball represents the corresponding carbon link atom. In the MM system (center) there are no electrostatic interactions between atoms separated by one or two bonds, whereas those interactions exist in the QM/MM system across the MM-QM interface (left). The bond across the interface is removed in the MM calculation to the right, and, hence, the electrostatic interactions across the interface mimic those of the QM/MM calculation.

We use the hydrogen link atom approach, in which the QM region is truncated by hydrogen atoms. The atoms that are converted to hydrogen atoms will be called carbon link atoms. Bond lengths between hydrogen link atoms and adjoining QM atoms are changed such that the relative stretch (or compression) of a given bond from its equilibrium length is preserved compared to that of the bond between the corresponding carbon link atom and QM atom.[37] In our methodology, the coordinates of the link atoms always belong to R_q.

Point charges on MM atoms joining the link atoms are often manipulated to avoid overpolarization of the QM region. In our implementation, we do not manipulate any of the point charges. Instead, we add a correction term, E_{corr}, to eq 11, which is defined as the difference of two classical electrostatic interaction terms

$$E_{corr} = E_{elec} - \tilde{E}_{elec} \quad (14)$$

and which approximately removes unwanted effects from the hydrogen link atoms. E_{elec} is the standard MM electrostatic interaction between the MM region and the QM region (with carbon link atoms) and is defined in eq 9. $\tilde{E}_{elec}$ differs from E_{elec} in that the carbon link atoms are changed to (MM) hydrogen link atoms. This is pursued by deleting the bonds between the link atoms and the MM region and by changing the point charges of the carbon link atoms and reposition them. In that way, $\tilde{E}_{elec}$ models classically the electrostatic interactions, eq 13, which defines the MM-QM electrostatic interactions in E_{quant}.

The above correction modifies only interactions between the MM region and the QM region, whereas interactions entirely within the QM (or MM) region are not changed. Moreover, electrostatic interactions between atoms separated by less than three bonds are approximately removed, and they are scaled if they are between atoms separated by three bonds. This is consistent with the Amber force field employed in the current study. A graphical representation of the correction term is given in Figure 3.

With the correction term, E_{corr}, the free energy between a state described by MM and by QM/MM converges faster because the difference between the MM and QM/MM potential energy surfaces fluctuates less. This is illustrated

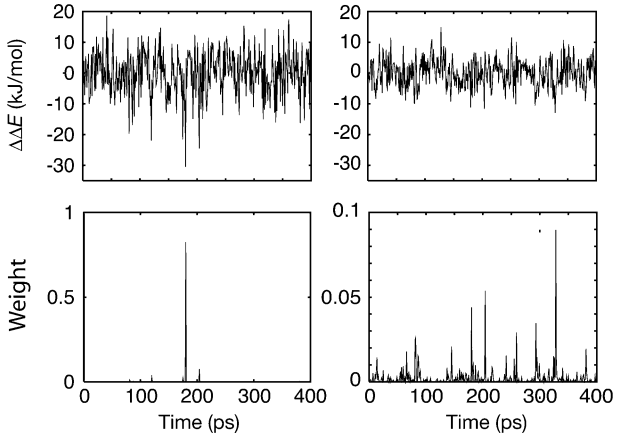


Figure 4. Upper panels show fluctuations of the energy difference between the QM/MM energies and the corresponding MM energies for the simulation with $d_{C-O} = 1.80$ Å. The left panel shows the fluctuations when no link atom correction is employed as opposed to the right panel where a link atom correction is employed. The lower panels show the corresponding contributions to the ensemble average in the free energy perturbation equation.

in Figure 4 where the fluctuations of $(E^{\text{total}}_{\text{qm/mm}} - E^{\text{total}}_{\text{mm}})$ are plotted with and without the correction term included for a simulation of COMT using the setup employed in the current study (section 2.4). The lower panels of Figure 4 show the contributions of the individual configurations to the ensemble averages in the free energy perturbation equations. They demonstrate that many configurations contribute to the free energy if the correction term is employed, whereas the free energy is dominated by a single configuration if the correction is not used.

2.2.2. Free Energy Perturbation Calculations for Fixed QM Regions. In the current application, we use the methodology to compute free energy changes between fixed points along a reaction coordinate as illustrated in Figures 1 and 2. In addition, we fix the entire QM region, such that the free energy depends parametrically of the positions of QM atoms. With that approximation and the definitions in eqs 7−14, eqs 3, 4, and 6 turn into the following formulas

$$\Delta A_{\text{mm}}(R^i_{\text{q}} \rightarrow R^{i+1}_{\text{q}}) = -k_{\text{B}}T \ln\langle e^{-[E^{\text{q/m}}_{\text{mm}}(R^{i+1}_{\text{q}},R_{\text{m}}) - E^{\text{q/m}}_{\text{mm}}(R^{i}_{\text{q}},R_{\text{m}})]/k_{\text{B}}T}\rangle_{\text{mm},R^i_{\text{q}}} + E^{\text{q}}_{\text{mm}}(R^{i+1}_{\text{q}}) - E^{\text{q}}_{\text{mm}}(R^i_{\text{q}}) \quad (15)$$

$$\Delta A_{\text{mm}\rightarrow\text{qm/mm}}(R^i_{\text{q}}) = -k_{\text{B}}T \ln\langle e^{-[E_{\text{quant}}(R^i_{\text{q}},R_{\text{m}}) - \tilde{E}_{\text{elec}}(R^i_{\text{q}},R_{\text{m}})]/k_{\text{B}}T}\rangle_{\text{mm},R^i_{\text{q}}} - E^{\text{q}}_{\text{mm}}(R^i_{\text{q}}) \quad (16)$$

$$\Delta A_{\text{mm}\rightarrow\text{qm/mm}}(R^{\text{o}}_{\text{q}} \rightarrow R^i_{\text{q}}) = -k_{\text{B}}T \ln\langle e^{-[E_{\text{qm/mm}}(R^i_{\text{q}},R_{\text{m}}) - E^{\text{q/m}}_{\text{mm}}(R^{\text{o}}_{\text{q}},R_{\text{m}})]/k_{\text{B}}T}\rangle_{\text{mm},R^{\text{o}}_{\text{q}}} - E^{\text{q}}_{\text{mm}}(R^{\text{o}}_{\text{q}}) \quad (17)$$

where R^i_{q} and R^{i+1}_{q} are two subsequent points on a reaction pathway. Note, that the terms $\Delta E^{\text{q}}_{\text{mm}}$ cancel in the sum eq 1 and therefore simply can be set to zero. This is of great practical advantage in actual simulations, because no bonded parameters need to be defined within the QM region, i.e., for atoms involved in the chemical reaction.

2.3. QM-FE and QM/MM-FE Approaches. In the QM-FE and QM/MM-FE methods, interactions between the MM and QM regions are described classically. To get to that approximation, starting out from the QTCP method, we consider two different configurations of the MM environment, R^0_{m} and R^1_{m}, and write eq 12 as

$$E_{\text{quant}}(R_{\text{q}}, R^{\alpha}_{\text{m}}) = \bar{E}(R_{\text{q}}, R^{\alpha}_{\text{m}})[n(R_{\text{q}}, R^{\alpha}_{\text{m}})]$$
$$= \bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R^{\alpha}_{\text{m}})] + \bar{V}^{\alpha}_{\text{ext}}(R_{\text{q}}, R_{\text{m}})[n(R_{\text{q}}, R^{\alpha}_{\text{m}})] \quad \alpha = 0, 1 \quad (18)$$

It follows from the variational principle[38] that

$$E_{\text{quant}}(R_{\text{q}}, R^1_{\text{m}}) = \bar{E}(R_{\text{q}}, R^1_{\text{m}})[n(R_{\text{q}}, R^1_{\text{m}})]$$
$$= \bar{E}(R_{\text{q}}, R^1_{\text{m}})[n(R_{\text{q}}, R^0_{\text{m}})] + \text{O}[(n(R_{\text{q}}, R^1_{\text{m}}) - n(R_{\text{q}}, R^0_{\text{m}}))^2] \quad (19)$$

Hence, provided that the electron density does not fluctuate too much owing to different polarizations by different MM environments, all the QM/MM calculations in the ensemble average can be approximated with nonself-consistent calculations using a frozen density obtained from a single self-consistent calculation with R^0_{m}.

At the next level of approximation, we represent the density $n(R_{\text{q}}, R^0_{\text{m}})$ by point charges

$$\bar{E}(R_{\text{q}}, R^1_{\text{m}})[n(R_{\text{q}}, R^0_{\text{m}})] = \bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R^0_{\text{m}})] + \bar{V}_{\text{ext}}(R_{\text{q}}, R^1_{\text{m}})[n(R_{\text{q}}, R^0_{\text{m}})]$$
$$\approx \bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R^0_{\text{m}})] + E_{\text{elec}}(R_{\text{q}}, R^1_{\text{m}}) \quad (20)$$

where E_{elec} is given by eq 9. Thus

$$E_{\text{quant}}(R_{\text{q}}, R^1_{\text{m}}) \approx \bar{E}(R_{\text{q}}, R^1_{\text{m}})[n(R_{\text{q}}, R^0_{\text{m}})] \approx \bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R^0_{\text{m}})] + E_{\text{elec}}(R_{\text{q}}, R^1_{\text{m}}) \quad (21)$$

We restrict the discussion to situations where the QM region is fixed. If the point charges used to represent the QM region in the classical simulations are derived from a QM/MM calculation where the coordinates are given by R^0_{m} (so that E_{elec} in eqs 8 and 21 are the same functions), then the above approximation collapses the ensemble average in eq 16 to

$$\Delta A_{\text{mm}\rightarrow\text{qm/mm}}(R^{\text{X}}_{\text{q}}, R_{\text{m}}) \approx \bar{E}_{\text{vac}}(R^{\text{X}}_{\text{q}})[n(R^{\text{X}}_{\text{q}}, R^0_{\text{m}})] - E^{\text{q}}_{\text{mm}}(R^{\text{X}}_{\text{q}}) \quad (22)$$

where R^0_{m} is the configuration of the MM environment in the calculation used to derive the point charges. With this approximation, and using eq 15, the sum in eq 1 can be written as

$$\Delta A_{\text{qm/mm}}(\text{A} \rightarrow \text{B}) = -k_{\text{B}}T \ln\langle e^{-[E^{\text{q/m}}_{\text{mm}}(R^{i+1}_{\text{q}},R_{\text{m}}) - E^{\text{q/m}}_{\text{mm}}(R^{i}_{\text{q}},R_{\text{m}})]/k_{\text{B}}T}\rangle_{\text{mm},R^i_{\text{q}}} + \{\bar{E}_{\text{vac}}(R^{\text{B}}_{\text{q}})[n(R^{\text{B}}_{\text{q}}, R^0_{\text{m}})] - \bar{E}_{\text{vac}}(R^{\text{A}}_{\text{q}})[n(R^{\text{A}}_{\text{q}}, R^0_{\text{m}})]\} \quad (23)$$

The QM terms, $\bar{E}_{\text{vac}}(R_{\text{q}})[n(R_{\text{q}}, R^0_{\text{m}})]$ for A or B, can be computed by obtaining $n(R_{\text{q}}, R^0_{\text{m}})$ from a self-consistent calculation with point charges positioned according to R^0_{m},

i.e., by computing $E_{quant}(R_q, R_m^0) = \bar{E}_{vac}(R_q)[n(R_q, R_m^0)] + \bar{V}_{ext}(R_q, R_m^0)[n(R_q, R_m^0)]$. $\bar{E}_{vac}[n(R_q, R_m^0)]$ can then be computed from a nonself-consistent calculation without point charges.

At another level of approximation, $\bar{E}_{vac}[n(R_q, R_m^0)]$ can be computed with $n(R_q, R_m^0)$ represented by point charges in the calculation of $\bar{V}_{ext}(R_q, R_m^0)[n(R_q, R_m^0)]$, such that

$$\bar{E}_{vac}(R_q)[n(R_q, R_m^0)] = E_{quant}(R_q, R_m^0) - \bar{V}_{ext}(R_q, R_m^0)[n(R_q, R_m^0)]$$

$$\approx E_{quant}(R_q, R_m^0) - E_{elec}(R_q, R_m^0) \quad (24)$$

so that eq 23 is modified to

$$\Delta A_{qm/mm}(A \rightarrow B) = -k_B T \ln\langle e^{-[E_{mm}^{q/m}(R_q^{i+1}, R_m) - E_{mm}^{q/m}(R_q^i, R_m)]/k_B T}\rangle_{mm, R_q^i} + (E_{quant}(R_q^B, R_m^0) - E_{elec}(R_q^B, R_m^0)) - (E_{quant}(R_q^A, R_m^0) - E_{elec}(R_q^A, R_m^0)) \quad (25)$$

This form is employed by Yang and co-workers in their QM/MM-FE method.[16,18,19] In practice, it may be a better approximation since the potential terms $\bar{V}_{ext}(R_q, R_m^1)$ and $\bar{V}_{ext}(R_q, R_m^0)$ are treated at the same level of approximation (eqs 21 and 24) and therefore cancelation of errors may occur.

If the point charges are derived from a calculation in a vacuum, such that $n(R_q, R_m^0) = n_{vac}(R_q)$ and $E_{quant} = \bar{E}_{vac}(R_q)[n_{vac}(R_q)]$, then eqs 21 and 25 become identical, and the formalism corresponds to the QM-FE method employed originally by Jorgensen and co-workers[2,3,4] and later by Kollman and co-workers.[13−15]

2.4. Computational Details. We use the methodologies above to compute free energy changes between fixed QM regions along a predetermined reaction pathway for methylation of catecholate catalyzed by COMT. We emphasize that the topic of this article is not the calculation of reaction pathways but rather how to compute free energies once a reaction pathway is known.

2.4.1. Model System. The crystal structure of COMT (pdb code: 1VID) was used as the starting structure.[39] Hydrogen atoms were positioned with the HBuild routine as implemented in CHARMM.[35] The employed QM system is illustrated in Figure 5 and consists of 44 atoms of which five are hydrogen link atoms.[37] The QM system consists of the catecholate molecule, the Mg^{2+} ion coordinated by catecholate, an $S(CH_3)_3^+$ molecule to model *S*-adenosylmethionine, $HCOO^-$, $HCOO^-$, and $HCONH_2$ to mimic the Mg^{2+} ligands: Asp-141, Asp-169, and Asn-170, respectively. In the crystal structure, a water molecule (HOH-400) also coordinates the Mg^{2+} ion, and that is included in the QM system as well. All atoms were allowed to move in the MD simulations, except the 44 QM atoms.

2.4.2. Reaction Pathway. We calculate the free energy along a reaction pathway adopted from another study.[40] Here it is sufficient to give a brief summary. The distance, d_{C-O}, between the acceptor oxygen of catecholate and the carbon atom in the transferred methyl group was used as reaction coordinate. For 10 different values of the reaction coordinate

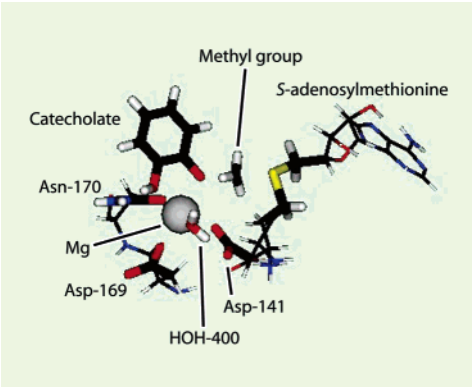


Figure 5. The QM system (bold sticks and sphere) employed in the present study.

with d_{C-O} = 1.47, 1.80, 1.95, 2.00, 2.05, 2.10, 2.13, 2.30, 2.55, and 2.84 Å, the QM region was QM/MM optimized with a constraint on the reaction coordinate and on all the MM atoms. The ComQum program, which in turn is based on the MM software Amber[41] and the QM code Turbomole,[42] was used for these calculations. The transition state was found at 2.13 Å. For each of the points, the RESP protocol[43] as implemented in Amber[41] was used to calculate the point charges in the presence of the MM region using electrostatic potential points obtained by the Merz−Kollman scheme as implemented in Gaussian.[44]

Because a single distance was used as reaction coordinate, it is possible that abrupt changes occur in coordinates orthogonal to the reaction coordinate. This indicates that an improper reaction coordinate is used and that the obtained reaction pathway may be significantly differently than the energy minimum path. It was checked by plotting the progress of other QM degrees of freedom, most noticeably the distance between the transferred methyl group and the donor atom, that this is not the case in the current study, but rather that they progress in a continuous way from the values of the reactant complex to the values of the product complex.

2.4.3. Molecular Dynamics Simulations. Ten molecular dynamics simulations, one for each point along the reaction pathway, were made using the program CHARMM.[35] In these simulations, the QM system was represented by the RESP point charges. The QM region was kept fixed in space during the simulations, and all bond lengths involving hydrogen atoms were constrained by means of the SHAKE algorithm.[45] Atoms of the cofactor and protein that are not part of the QM region were described by the Amber94 force field,[46] and solvents were described explicitly using the TIP3P model.[47] Periodic boundary conditions were employed using an octahedral unit cell and the Particle Mesh Ewald (PME) method[48] with a real-space cutoff of 9 Å to describe the electrostatic interactions. All simulations were performed with a constant volume, and the temperature was kept approximately constant by reassigning the atomic velocities from a Gaussian distribution if the temperature differed by more than 5 K from 298 K.

The 10 simulations were initiated from structures in which the coordinates of the QM system were taken from the QM/MM calculations, whereas the positions of the MM atoms

and the size of the unit cell were taken from the last snapshot of a 200 ps constant pressure (1 atm) and temperature (298 K) simulation for the transition state with $d_{C-O} = 2.13$ Å. The structures were relaxed by 100 steps of steepest-descent minimization with all heavy nonsolvent atoms restrained by a harmonic potential using a mass-weighted force constant of 10 kcal mol^{-1} Å^{-2} amu^{-1}. This was followed by 20 ps of simulation with only the backbone atoms restrained by a harmonic potential using a mass-weighted force constant of 5 kcal mol^{-1} Å^{-2} amu^{-1}. After that, the harmonic restraints were removed. A total of 600 ps of simulations were performed for each of the 10 points defining the reaction pathway.

2.4.4. QM/MM Calculations. For each of the simulations, 400 configurations separated by 1 ps were stored. The configurations were taken from the 200−600 ps time interval of the simulations, and the configurations were used to compute E_{quant} involved in the FEP calculations. In principle, E_{quant} should be computed with periodic boundary conditions employed. However, such calculations are not possible with the QM program employed, and we therefore only include point charges from a single unit cell in the QM calculations as well as in the calculations of $E_{mm}^{q/m}$ in eqs 16 and 17. This is a fair approximation because test calculations for the classical free energy changes defined in eq 3 show that describing the electrostatics in this way rather than with periodic boundary conditions does not change the results significantly.

Density functional theory with the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional[49] was employed for the QM calculations. The resolution-of-the-identity (RI) approximation was used for the Coulomb terms,[50] and the one-electron wave functions were expanded in a 6-31G* basis set with an additional set of diffuse s and p functions located on the oxygen and sulfur atoms. We denote this basis set 6-31(+)G*. The calculations were performed with the program TURBOMOLE.[42]

The term, $\bar{E}_{vac}[n(R_q, R_m^0)]$, used in the QM/MM-FE approximation, eq 23, was calculated by making a self-consistent calculation with point charges included. Then, the point charges were removed, and a single iteration (nonself-consistent calculation) was performed using the wave function from the self-consistent calculation.[51]

3. Results and Discussion

3.1. QTCP-U Approach. The results based on the QTCP-U method (Figure 1) are shown in Figure 6 for the enzyme catalyzed reaction. The barrier for forward reaction is 56 kJ/mol and that of the reverse reaction is 78 kJ/mol. The barrier is well converged with standard errors less than 1 kJ/mol.[52] In addition, a difference of 1.6 kJ/mol is obtained for the forward barrier, when individual classical free energy changes (eq 15) based on forward and reverse transitions are used, cf. Table 1.

The contribution to the free energy barrier from the MM environment is indicated by the dashed line. The steepness of the curve shows that the environment has a significant impact on the reaction and emphasizes the importance of a proper description of the environment. If the environment had been omitted, the reaction would occur almost spontanously.

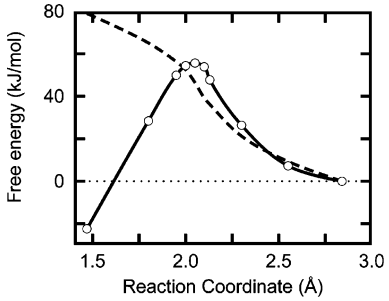


Figure 6. QM/MM free energy barrier at the level of DFT for the methylation of catechol catalyzed by catechol *O*-methyltransferase. The dashed line indicates the contribution to the free energy barrier from the MM environment.

Table 1: Individual FEP Contributions to the Free Energy Barriers Based on the Ensemble Averages in Eqs 15 and 16[a]

			ΔA_{mm}			
d_{C-O}	$\Delta A_{mm\rightarrow qm/mm}$	step	forward	reverse	av	acc
2.84	0.0(0.4)	2.84 → 2.55	9.6(0.2)	9.0(0.2)	9.3	9.3
2.55	−2.1(0.3)	2.55 → 2.30	11.9(0.1)	12.0(0.2)	11.9	21.2
2.30	5.1(0.3)	2.30 → 2.13	14.6(0.2)	15.3(0.2)	15.0	36.2
2.13	11.5(0.5)	2.13 → 2.10	3.1(0.1)	2.9(0.1)	3.0	39.2
2.10	14.7(0.3)	2.10 → 2.05	8.0(0.1)	6.6(0.1)	7.3	46.5
2.05	9.2(0.5)	2.05 → 2.00	5.5(0.1)	6.6(0.1)	6.0	52.5
2.00	1.9(0.7)	2.00 → 1.95	3.5(0.1)	4.1(0.1)	3.8	56.3
1.95	−6.4(0.7)	1.95 → 1.80	9.2(0.1)	8.4(0.2)	8.8	65.1
1.80	−36.8(0.4)	1.80 → 1.47	13.4(0.2)	14.3(0.2)	13.9	79.0
1.47	−101.4(0.4)					

[a] Classical free energy changes are computed for the forward and reverse step, and the averages are listed as well. Accumulated averages (acc) are listed in the last column. All energies are in kJ/mol and reaction coordinates in Å. Statistical standard errors are given in parentheses. Note that $E_{mm}^{q} \equiv 0$, so internal energy changes of the QM system are described exclusively by $\Delta A_{mm\rightarrow qm/mm}$.

Individual contributions from the FEP calculations are listed in Table 1. It can be seen that the barrier is the result of increasingly positive contributions from the classical free energy changes and increasingly negative contributions from the MM → QM/MM calculations. The averages of ΔA_{mm} are used to report the final barrier.

In Figure 7, we investigate how the results vary with the number of data points used in the ensemble averages in eqs 15 and 16. We compare results based on FEP ensemble averages of 20, 50, 100, 200, and 400 data points. It can be seen from the right panel that the barrier height is almost converged already at 20 data points, whereas the left panel shows that additional data points smooth the barrier. However, the transition state at 2.05 Å is not located until 200 data points are used. With 200 data points, the barrier is almost indistinguishable from that using the full data set of 400 points. This shows that it is possible to compute converged free energy barriers with the QTCP approach.

3.2. Extrapolation to Higher Accuracy. The barrier of 56 kJ/mol is lower than the experimental prediction of about 75 kJ/mol.[53] The free energy barrier is converged, but the

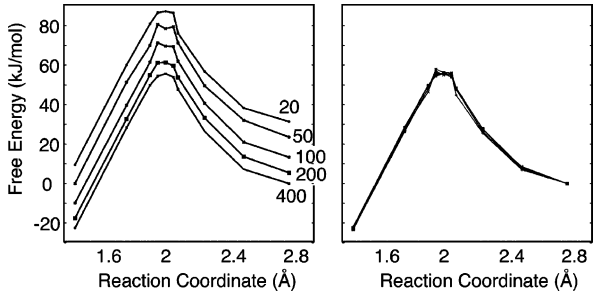


Figure 7. Convergence of free energy barriers with a number of data points utilized. In the left panel, barriers based on 20−400 data points are shifted vertically from one another. In the right panel, the same barriers are superpositioned so that the value at the reactant complex (2.84 Å) is zero.

Table 2: Computed Activation and Reaction Energies (kJ/mol) for Methyl Transfer Using the QM Region Indicated in Figure 5 in a Vacuum[a]

	basis set						
	6-31(+)G*		6-311++G(2d,2p)				
functional	PBE	PBE/RI	PBE	PBE/RI	TPSS/RI	TPSSh	B3LYP
$\Delta E^{\neq}_{fwd}$	12	12	22	22	11	18	27
$\Delta E^{\neq}_{rev}$	111	112	101	101	106	117	115
$\Delta E_{reaction}$	−99	−100	−80	−80	−96	−99	−98

[a] The resolution-of-the-identity (RI) approximation[54] is used in combination with the PBE and TPSS density functionals.

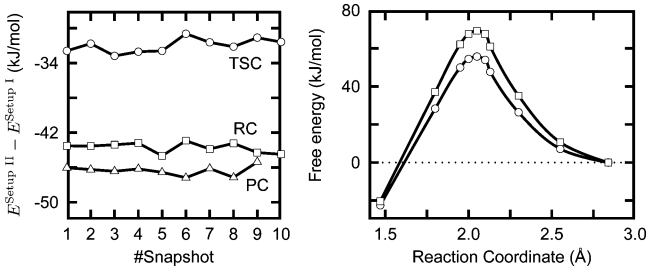


Figure 8. In the left panel, the difference between energies computed with setup II (B3LYP, 6-311++G(2d,2p)) and setup I (PBE, 6-31(+)G*) are plotted for configurations taken from a simulation of the reactant complex (RC), the transition state complex (TSC), and the product complex (PC). A free energy barrier extrapolated from the one based on setup I to one based on setup II (squares) is plotted in the right panel and compared to the one based on setup I (circles).

QM/MM energies depend significantly on the basis set and exchange-correlation functional chosen. This is shown in Table 2, in which the reaction and activation energy for various functionals and basis sets are listed. Energies based on the 6-31(+)G* basis set are not fully converged, and there are also significant changes among the different functionals. On the other hand, energy fluctuations are almost invariant to the basis set and functional employed. This is illustrated in the left panel of Figure 8, where the difference between energies calculated with the original setup (setup I) and a 6-311++G(2d,2p) basis set and the B3LYP hybrid functional (setup II) are plotted for three different simulations. Because the fluctuations are small compared to the shift in average energy along the reaction pathway, it makes sense to extrapolate between the two setups. The result of such an extrapolation is plotted in the right panel in Figure 8.

Using this setup we get a free energy barrier of 69 kJ/mol, which is in excellent agreement with the experimental value of about 75 kJ/mol. The barrier for the reverse reaction is calculated to 90 kJ/mol.

In the extrapolation scheme, a correction term $\Delta E_i^{xpol} = E_i^{setup\ II} - E_i^{setup\ I}$ is added to each point i on the reaction pathway. The strategy which we have chosen and which performs well is one where the same MM environment (i.e. positions and charges of MM atoms) is used to calculate all the ΔE_i^{xpol}. The result shown in Figure 8 are based on energies calculated with the MM environment defined by a configuration taken from the simulation of the transition state, but extrapolation based on calculations with the MM environment turned off, i.e., calculations in a vacuum, performs equally well. The two strategies give almost identical barrier heights; for the forward barrier we compute a barrier of 69 kJ/mol if the TS environment is employed and 71 kJ/mol for the in vacuum corrections. The corresponding barrier heights for the reverse reaction are 90 and 91 kJ/mol, respectively. The agreement between the two strategies validates the extrapolation scheme further.

Cancelation of systematic errors probably occurs when the same MM environment is used to calculate all extrapolated energies along the reaction pathway. In contrast, because the energy difference between setup I and setup II fluctuates somewhat (Figure 8), a rugged free energy barrier is obtained if a different MM environment is used for each point on the reaction pathway. Therefore, we find such a procedure less reliable. A more accurate approach would be to use an average for several energy calculations for each point, but we have not pursued such an approach here.

3.3. CPU Consumption. The total CPU load for the QM/MM calculations using setup I is approximately 8 CPU days on an Intel Xeon 2.4 GHz CPU for the entire free energy barrier based on 4000 QM/MM calculations (∼3 CPU min/calculation). One or two CPU days spend on 10 calculations using setup II should be added if the free energies are extrapolated to higher accuracy (2.5−4 CPU h/calculation).

The fairly small CPU load is due mostly to the RI approach, which saves significant CPU time: a factor of about 6 for the smaller basis set and a factor of about 13 for the larger basis set. With the larger basis set a calculation with the RI approximation takes about 20 CPU min, and, hence, even with the larger basis set, the CPU load does not prohibit convergence of the free energy changes as long as pure density functionals are employed to which the RI approximation applies. We find that the RI approach is sufficiently accurate; the PBE energy changes listed in Table 2 with and without the RI approximation deviate by at most 0.5 kJ/mol from each other.

Finally, the methodology is perfectly suitable for cluster and grid computing since all the single point QM/MM calculations are strictly independent of each other.

3.4. Importance of Sampling. Having established that the environment is important for a proper description of the reaction, it is now interesting to explore the role of sampling. An approach sometimes employed in QM/MM calculations

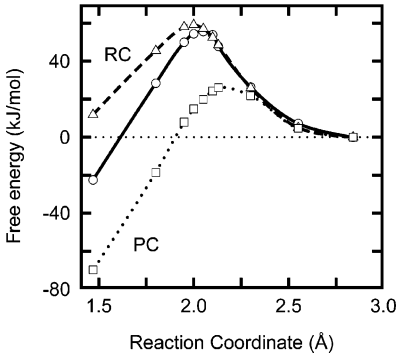


Figure 9. Energy barriers for methyl transfer catalyzed by COMT using a fixed MM enviroment. The curves are averages of 10 configurations taken from a simulation of the reactant complex (RC, dashed) and the product complex (PC, dotted) and compared with the QTCP-U method (solid).

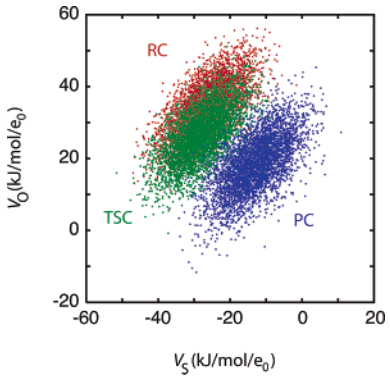


Figure 10. The electrostatic potential caused by the MM environment at the donor atom as a function of that at the acceptor atom in the methyl transfer reaction catalyzed by COMT. The potentials are plotted for a simulation of the reactant (red), transition state (blue), and product (green) complexes.

is to include the environment as a static set of point charges. Figure 9 shows energy barriers computed in this way using the same fixed MM environments (including solvent molecules) for all points on the reaction pathway. Two different configurations for the MM environments are tested. One configuration is taken from a simulation of the reactant complex (dashed curve), and the other is taken from a simulation of the product complex (dotted curve). The former one gives barrier heights of 59 and 47 kJ/mol for the forward and reverse reaction, respectively, and the later one gives barrier heights of 26 and 96 kJ/mol. The difference between the two curves is caused by a change in the polarization of the environment upon going from the reactant complex to the product complex. This can be seen from Figure 10 where the electrostatic potential at the donor atom (V_O) versus that at the acceptor atom (V_S) is plotted for the reactant complex (RC), the transition state complex (TSC), and the product complex (PC). The lack of a significant overlap between the distributions of the product complex and the other complexes also shows that sampling of one complex does not sample important contributions to the free energy of the other complex.[55]

As mentioned previously, omitting the MM environment leads to a very low barrier, if the same QM/MM optimized reaction pathway is employed. A method that has been used extensively in the literature is to omit the MM environment and optimize product, transition state, and the reactant complex in a vacuum. Using this approach with setup II (B3LYP, 6-311+G(2d,2p)) and with a solvent correction, which is based on a continuum model with a dielectric constant of 4, results in a forward barrier of 62 kJ/mol and a barrier of 138 kJ/mol for the reverse reaction. These values should be compared with the extrapolated values of 69 and 90 kJ/mol based on the QTCP-U method (Figure 8). Hence, the forward barrier is fairly well reproduced, whereas the reverse barrier is reproduced less well. The relatively small and large deviations are because the transition state and reactant complexes are similar and polarize the environment in a similar manner as revealed by Figure 10. In contrast, the product and transition state complexes polarize the environment differently in a way that cannot be modeled adequately by the low dielectric continuum model. It should be emphasized that the reaction pathway optimized in a vacuum is different from the QM/MM optimized reaction pathway.

In the calculations in a vacuum, entropy changes and zero-point energies can be estimated from normal mode calculations. Including entropy changes calculated in this way decreases the barrier for forward reaction by 8 kJ/mol and increases that of the reverse reaction by 2 kJ/mol. Entropy changes of the QM region are not included in the results based on the QTCP method, because the QM region is fixed. As a first approximation, the above values can be included. However, the entropy change between the transition-state complex and reactant complex of 8 kJ/mol is probably an overestimation because the orientation of $S(CH_3)_3$ group and the catecholate oxygen in the reactant complex optimized in a vacuum is different from that in the corresponding QM/MM structure. On the other hand, the transition state and product structures are quite similar in the two types of calculations. For these reasons, we do not add these entropy terms to our final results.

3.5. QTCP-V Approach. As described in section 2.1, the actual number of simulations may be reduced if the QTCP-V method (Figure 2) is employed. Here we have tested this approach by computing free energy barriers based on simulations with anchor points at 1.80, 2.05, or 2.55 Å. Using these anchor points, the computed barriers for the forward reaction are found to be 50, 63, or 58 kJ/mol, respectively, and 82, 70 or 61 kJ/mol, respectively, for the reverse reaction. In light of the previous section, these deviations are not surprising since a single simulation does not sample important contributions to all the other points.

We can compute a barrier where each of the three simulations are used only to compute free energy changes in an interval around the point for which the simulation was performed. The three pieces can then be merged together to give the complete free energy barrier. The result from this strategy is plotted in Figure 11 (black line) and compared to the result based on the QTCP-U method in the lower panel. It is seen that the agreement is pretty good. The reaction barrier based on the QTCP-V method is 60 kJ/mol for the

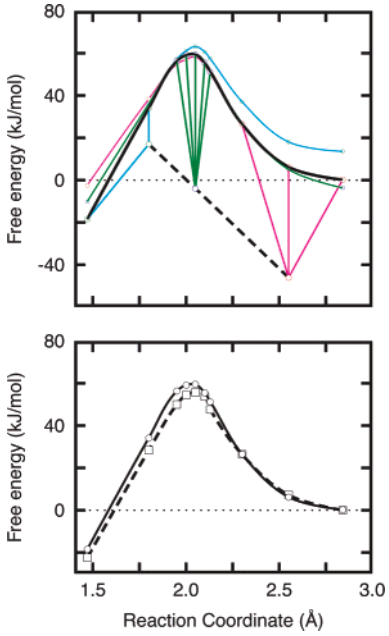


Figure 11. Top: Free energy barriers based on the QTCP-V method using anchor points of $d_{C-O} = 1.80$ Å (cyan), $d_{C-O} = 2.05$ Å (green), or $d_{C-O} = 2.55$ Å (magenta). The bold black curve is a free energy barrier generated by merging pieces from the three other curves together. Bottom: Comparison of the merged black curve from the top panel with the free energy barrier based on the QTCP-U method.

forward reaction and 78 kJ/mol for the reverse reaction, to be compared with 56 and 79 kJ/mol for the QTCP-U method.

The curves based on the 2.55 and 2.05 Å anchor points coincide over a large interval between the two anchor points. In fact, the curve based solely on the 2.55 Å anchor point almost predicts the correct barrier height for forward reaction. This feature is consistent with the scattering plot in Figure 10, which shows that the simulation of the reactant complex (next to the 2.55 Å anchor point) samples important contributions about the transition state complex at 2.05 Å. On the other hand, merging of the curves based on the 2.05 and 1.80 Å anchor points is less well defined. Splicing the two curves either at 1.95 or 1.80 Å gives slightly different results for the reverse barrier height (~4 kJ/mol difference). The solid curve in Figure 11 is based on the average of the two values.

We finally mention that the three curves also can be aligned relative to each other by computing the classical free energy changes between the three anchor points. The single step free energy computations between the anchor points are however associated with great uncertainty. This is revealed by a hysteresis of about 8 kJ/mol for both the 1.80 → 2.05 Å and the 2.05 → 2.55 Å step. Nonetheless, if the averages over the values computed for the forward and reverse transitions are used, almost the same results as that from the splicing of the curves is obtained.

3.6. QM/MM-FE Approach. Figure 12 compares results based on the QM/MM-FE method, eq 23 or 25, with results based on the more exact QTCP-U method. The two QM/MM-FE approaches differ in how the QM energy E_{quant} is approximated (the approximation in eq 24 turns eq 23 into eq 25).

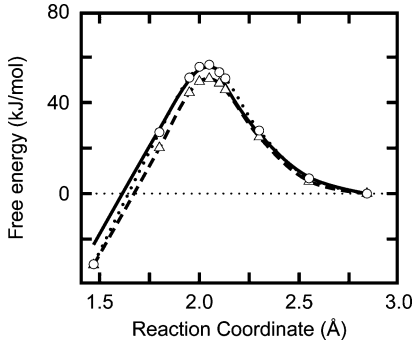


Figure 12. Free energy barrier for the methyl transfer reaction catalyzed by COMT computed by means of the QM/MM-FE approach, eq 23 (circles and dotted line), and compared with the barrier based on the QTCP-U method in Figure 6 (solid line). A free energy barrier based on the QM/MM-FE approach and using the approximation in eq 24 is plotted with the triangles connected with a dashed line.

The results based on eq 23 (dotted line) reproduces the barrier for forward reaction well with a slight overestimation of 1 kJ/mol, whereas the barrier for the reverse reaction is overestimated by 10 kJ/mol. This trend is not surprising, since the point charges employed are derived from a QM/MM calculation where the enzyme is relaxed and the solvent is equilibrated to the reactant complex, and therefore the point charges model the electron density of points close to the reactant complex better than points closer to the product complex.

On the other hand, if eq 25 is used (dashed line), the reaction barrier for forward reaction is underestimated by 5 kJ/mol and that of the reverse reaction by 4 kJ/mol. The similar performance over the entire range might be due to cancellation of errors as discussed in section 2.3.

Nonetheless, with or without the approximation in eq 24 included, the QM/MM-FE approach seems to be a reasonable approximation to the QTCP approach leading to a significant reduction of CPU time consumed.

4. Discussion and Conclusion

We have described a method, QTCP, that makes it possible to compute converged high-level free energy changes. We have used it successfully to compute a free energy barrier for the methyl transfer reaction catalyzed by catechol *O*-methyltransferase. The computations are based on a total of 600 ps simulations per point on a reaction pathway (200 ps equilibration and 400 ps production run) and result in a free energy barrier of 56 kJ/mol using the 6-31(+)G* basis set with the PBE exchange-correlation functional. This value can be extrapolated to 69 kJ/mol using the 6-311++G(2d,-2p) basis set with the B3LYP functional, which is in excellent agreement with an experimental value of about 75 kJ/mol. The results are converged to standard errors to within 1 kJ/mol and with a hysteresis of less than 2 kJ/mol. To obtain such a good convergence we fix the QM region and make a simple link atom correction, which compensates for the improper description of covalent bonds at the interface between the MM and QM region.

An important outcome of our computations is that a significant polarization of the environment occurs as the

reaction progresses. For that reason relaxation of the MM environment is important as obtained through the MD simulations. It is possible that minimization of the total QM/MM energy would give similar results. Electronic polarization of the QM region does not need to be described in great detail. This is illustrated by the fact that energy fluctuations are almost invariant to the exchange-correlation functional and basis set employed. Moreover, using a point charge description of the QM region, as in the QM/MM-FE approach, reproduces the results fairly well.

We have demonstrated that significant CPU time can be saved by using an extrapolation scheme and also that the number of simulations can be reduced. The CPU time can be reduced even further by using the QM/MM-FE approximation.

A major approximation in the current implementation of the QTCP method and in other similar methods is that entropic effects from the internal degrees of freedom of the QM region are ignored or at least considered constant along the reaction pathway. The need for fixing the QM region is caused by the very different potential surfaces for the QM internal degrees of freedom in the QM and MM description of the QM region, which in turn makes the MM → QM FEP calculations difficult to converge, cf. refs 6 and 22. In a first approximation, entropic effects from the QM region can be estimated from normal mode calculations based on optimized QM structures in a vacuum. However, we find such an approach unreliable since the reaction pathway in a vacuum is different from that obtained from the QM/MM reaction pathway. In the current study, this is most noticeable for the forward reaction, since the reaction complex optimized in a vacuum is markedly different from the one optimized by QM/MM calculations. A different approach is to include QM entropic effects by parametrizing the QM region along the reaction pathway and use the parametrized potential surface to describe fluctuations of the QM region.[6,22,56] However, if the parametrization mimics the ab initio QM potential surface well, it should also be possible to obtain converged MM → QM FEP calculations. Indeed, the MM → QM FEP calculations can be considered as a test of a computationally cheaper potential. In this regard, Wood and co-workers find that the FEP calculations can be improved considerably by using better reference potentials. Moreover, the ab initio calculations involved in the MM → QM FEP calculations can be used to optimize the reference potential.[22,26,28,29,31] Warshel and co-workers have also proposed an alternative approach to circumvent the problem of a flexible QM region. The approach is based on the linear response approximation (LRA) and involves sampling of the ab initio potential energy surface in addition to the EVB potential energy surface.[6]

In conclusion, the QTCP and the QM/MM-FE methods seem to be promising methods for computing high-level quantum mechanical free energies. We are currently testing the methods, e.g. for the calculation of proton and electron affinities.

References

(1) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196−1202.

(2) Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974−2975.

(3) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154−163.

(4) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184−189.

(5) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516−17524.

(6) Štrajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333−13343.

(7) Iftimie, R.; Salahub, D.; Wei, D.; Schofield, J. *J. Chem. Phys.* **2000**, *113*, 4852−4862.

(8) Iftimie, R.; Schofield, J. *J. Chem. Phys.* **2001**, *114*, 6763−6773.

(9) Iftimie, R.; Schofield, J. *J. Chem. Phys.* **2001**, *115*, 5891−5902.

(10) Iftimie, R.; Salahub, D.; Schofield, J. *J. Chem. Phys.* **2003**, *119*, 11285−11297.

(11) Iftimie, R.; Schofield, J. *Int. J. Quantum Chem.* **2003**, *91*, 404−413.

(12) Bandyopadhyay, P. *J. Chem. Phys.* **2005**, *122*, 091102.

(13) Stanton, R. V.; Peräkylä, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448−3457.

(14) Kuhn, B.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 2586−2596.

(15) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72−79.

(16) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483−3492.

(17) Zhang, Y.; Liu, H.; Yang, W. Ab Initio QM/MM and Free Energy Calculations of Enzyme Reactions. In *Macromolecules- -Challenges and Applications*; Schlick, T., Gan, H. H., Eds.; Springer-Verlag's Lecture Series in Computational Science and Engineering, Springer: New York, 2002.

(18) Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2000**, *122*, 6560−6570.

(19) Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2003**, *125*, 10384−10393.

(20) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2003**, *125*, 12035−12048.

(21) Ishida, T.; Kato, S. *J. Am. Chem. Soc.* **2004**, *126*, 7111−7118.

(22) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293−2301.

(23) Olsson, M. H. M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025−5039.

(24) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.

(25) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.

(26) Wood, R. H.; Yezdimer, E. M.; Sakane, S.; Barriocanal, J. A.; Doren, D. J. *J. Chem. Phys.* **1999**, *110*, 1329−1337.

(27) Sakane, S.; Yezdimer, E. M.; Liu, W.; Barriocanal, J. A.; Doren, D. J.; Wood, R. H. *J. Chem. Phys.* **2000**, *113*, 2583−2593.

(28) Sakane, S.; Liu, W.; Doren, D. J.; Shock, E. L.; Wood, R. H. *Geochim. Cosmochim. Acta* **2001**, *65*, 4067−4075.

(29) Liu, W.; Sakane, S.; Wood, R. H.; Doren, D. J. *J. Phys. Chem. A* **2002**, *106*, 1409−1418.

(30) Wood, R. H.; Liu, W.; Doren, D. J. *J. Phys. Chem. A* **2002**, *106*, 6689−6693.

(31) Ming, Y.; Lai, G.; Tong, C.; Wood, R. H.; Doren, D. J. *J. Chem. Phys.* **2004**, *121*, 773−777.

(32) Rod, T. H.; Ryde, U. *Phys. Rev. Lett.* **2005**, *94*, 138302.

(33) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(34) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.

(35) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(36) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720−1735.

(37) Ryde, U. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153−164.

(38) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864−B871.

(39) Vidgren, J.; Svensson, L. A.; Liljas, A. *Nature* **1994**, *368*, 354−358.

(40) Rasmussen, T.; Nilsson, K.; Ryde, U. in preparation.

(41) Case, D. A. et al. AMBER 7, University of California, San Francisco, 2002.

(42) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165−169.

(43) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(44) Frisch, M. J. et al. Gaussian 98 (Revision A.1), Gaussian, Inc., Pittsburgh, PA, 11998.

(45) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

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

(48) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

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

(50) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119−124.

(51) In doing this, it should be ensured that the same grid size is employed in the nonself-consistent calculation as in the last iteration of the self-consistent calculation. In Turbomole notation, we used the grid size 'm3' for the self-consistent calculations, meaning that grid size 3 is used for the very last iteration.

(52) Standard errors were computed using a bootstrap procedure.

(53) Schultz, E.; Nissinen, E. *Biochem. Phamacol.* **1989**, *38*, 3953−3956.

(54) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283−289.

(55) We note that the gap between the distributions of the PC and TSC is closed when all simulations are included. This indicates that inclusion of the enzyme and solvent degrees of freedom into the reaction coordinate would not change the barrier height significantly. The lack of a profound hysteresis further supports this point.

(56) Lu, Z.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 89−100.

CT0501102

JCTC Journal of Chemical Theory and Computation

Helix Interactions in Membranes: Lessons from Unrestrained Monte Carlo Simulations

Yana A. Vereshaga, Pavel E. Volynsky, Dmitry E. Nolde, Alexander S. Arseniev, and Roman G. Efremov*

M.M. Shemyakin and Yu.A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Ul. Miklukho-Maklaya, 16/10, Moscow V-437, 117997 GSP, Russia

Received May 5, 2005

Abstract: We describe one of the first attempts at unrestrained modeling of self-association of α-helices in implicit heterogeneous membrane-mimic media. The computational approach is based on the Monte Carlo conformational search for peptides in dihedral angles space. The membrane is approximated by an effective potential. The method is tested in calculations of two hydrophobic segments of human glycophorin A (GpA), known to form membrane-spanning dimers in real lipid bilayers. Our main findings may be summarized as follows. Modeling in vacuo does not adequately describe the behavior of GpA helices, failing to reproduce experimental structural data. The membrane environment stabilizes α-helical conformation of GpA monomers, inducing their transmembrane insertion and facilitating interhelical contacts. The voltage difference across the membrane promotes "head-to-head" orientation of the helices. "Fine-tuning" of the monomers in a complex is shown to be regulated by van der Waals interactions. Detailed exploration of conformational space of the system starting from arbitrary locations of two noninteracting helices reveals only several groups of energetically favorable structures. All of them represent tightly packed transmembrane helical dimers. In overall, they agree reasonably well with mutagenesis data, some of them are close to NMR-derived structures. A possibility of left-handed dimers is discussed. We assume that the observed moderate structural heterogeneity (the existence of several groups of states with close energies) reflects a real equilibrium dynamics of the monomers—at least in membrane mimics used in experimental studies of GpA. The elaborated computational approach is universal and may be employed in studies of a wide class of membrane peptides and proteins.

1. Introduction

Membrane proteins (MP) constitute ∼30% of all proteins encoded by whole genomes.[1] Delineation of the structure−function relationships for MPs represents an intriguing challenge. Apart from fundamental importance, solving the problem would be invaluable in the optimization of these molecules' behavior for pharmaceutical applications. Many of MPs contain several transmembrane (TM) fragments or function as oligomers. It is well-established now that protein−protein interactions in lipid membranes are very important for a large number of crucial cell processes.[2,3] Membrane bound helix associates are the most appropriate objects to study such interactions owing to the relative simplicity of the systems and their stability. In addition, hydrophobic and amphiphilic α-helices represent a dominant structural motif responsible for the binding of MPs to membranes. Thus, TM helix interactions mediate the functional activity of a large number of integral and peripheral MPs: receptors, ion channels, and others.[4] Furthermore, it has recently been established that dimerization of TM α-helices may play a crucial role in the functioning of

* Corresponding author phone: (7-095) 336 20 00; e-mail: efremov@nmr.ru.

10.1021/ct0501250 CCC: $30.25
Published on Web 09/02/2005

receptor tyrosine kinases.[2] Interaction of individual α-helices with membranes and membrane mimics has been extensively studied using both experimental and modeling techniques (see refs 5−7 for reviews). Analysis of atomic resolution structures of MPs revealed a number of common principles of helix packing in membranes.[8,9] On the other hand, molecular mechanisms that drive such association are still poorly understood due to the difficulties of their experimental studies. Therefore, elaboration of independent techniques is especially well-timed. Molecular modeling represents a promising alternative, which considerably extends and complements traditional structural biology tools such as X-ray and NMR spectroscopy.

A suitable system for the development of computational techniques to assess helix interactions in membranes is the hydrophobic segment 69−97 (GpA) of human glycophorin A from erythrocyte membranes.[10] This relatively small system has been extensively studied in experiments. In membranes and detergent micelles GpA forms α-helical homodimers with parallel (head-to-head) orientation. Site-directed mutagenesis[11] has revealed that the LIxxGVxxG-VxxT sequence pattern (where x marks an arbitrary residue) is important for dimerization. A spatial model of the dimer has been elaborated on the basis of these data and the results of computer simulations.[12,13] This has been done for two helices in vacuo using global optimization technique combining molecular dynamics (MD) and simulated annealing. To prevent helix dissociation, a number of intermonomer distance restraints have been employed. The most plausible model of the dimer represents a right-handed helical supercoil. It has been selected in agreement with mutagenesis data and geometrical rules of helix packing in globular proteins. Later, the model has been refined using several interhelical distance restraints obtained by NMR in detergent micelles[14] and by solid-state NMR in lipid bilayers.[15] Despite the atomic resolution of the models, NMR experiments yield only five pairs of restraints on interhelical distances. Because of such a small number of restraints, the question of uniqueness of the model proposed for 3D structure of GpA still remains to be answered. Moreover, there is a wealth of NMR data[16,17] confirming that in many cases there is an equilibrium conformational exchange between different oligomeric states of helices in membrane-mimic environments. Such effects have been also recently reported based on MD simulations of hydrophobic helices in explicit hydrated octane slab.[18]

Noteworthy is that all the aforementioned experiments with GpA have been carried out in membrane mimics, like detergent micelles and lipid vesicles. Even in these artificial environments the spatial structures of the dimer are somewhat different. Furthermore, helix dimerization is mediated by the micellar composition.[19] Real biological membranes are much more complex in terms of physicochemical properties and therefore may strongly affect helix interactions. Nevertheless, the appearance of the 3D model of GpA dimer in membrane mimics (hereafter called "the nativelike structure", although the spatial structure under native conditions is yet to be solved) has stimulated the development of theoretical methods to study helix interactions in MPs. A number of approaches resembling that suggested in ref 12 have since been applied to predict 3D models of a number of α-helical dimers.[20−24] Most of them possess severe limitations. First, the simulations (except in refs 23 and 24) were carried out either in vacuo or in a continuum dielectric with low permeability. Additionally, the monomers were always "rigid", and hence the common occurrence of local distortions in TM helices, like kinks and bends, was not taken into account. Finally, it was a priori proposed that helices adopt a TM orientation. These studies show that the NMR structure of the GpA dimer can be predicted more or less correctly even without media effects—it is sufficient to introduce just a few restraints (helicity, TM orientation, parallel packing). This is because GpA is a well-studied system, and such restraints can be defined before the simulation. In this case the answer already lies (at least partially) in the input—one knows a priori what one wants to get. However, to proceed with the new oligomers, for which the structural experimental data are missing, the membrane effects on structure and/or insertion mechanisms should be taken into account. Thus, the modeling of even the simplest membrane helical oligomers is not straightforward, and the development of new efficient algorithms seems very promising. To have a predictive power, they should not impose any structural restraints and should not be based on a priori knowledge of the mode of membrane binding for the peptides. Previously we elaborated such a computational approach based on unrestrained Monte Carlo (MC) simulations in the presence of heterogeneous implicit membrane (reviewed in ref 7). Applied to a large number of α-helical peptides, this method was efficient in reproducing the main tendencies in peptide−membrane interactions in accord with experimental data.

Here we undertake one of the first attempts at unrestrained ab initio modeling of self-association of two GpA α-helices in implicit membrane. The objective was to check whether the calculations are capable of predicting the dimer structures close to those observed in experiments. In addition to this, because the experimental data on GpA have been obtained in artificial membrane mimics, not in real membranes, it was challenging to explore other conformational possibilities for helix complexes— possibly, depending on the media properties, some of them may be realized under native conditions.

2. Simulation Details

The System. The calculations were performed for two GpA segments with sequences $S^{(69)}$EPEITLIIFGVMAGVIGTIL-LISYGIRR$^{(97)}$. All-atom starting structures of GpA were built in α-helical conformation—previously we have shown[25] that in an implicit membrane the GpA monomer folds in the TM α-helix from the initial random coil. As shown in Figure 1A, 10 and 20 dummy residues were attached to the N-terminus of monomer-1 and introduced between the monomers, respectively. This was done to change the orientation of peptides with respect to each other and to the membrane during the simulation. These "virtual" residues were taken from the standard library of the FANTOM program.[26] They do not contribute to the energy of the system. First the atom of the N-terminal dummy residue was always positioned in the center of the hydrophobic layer with

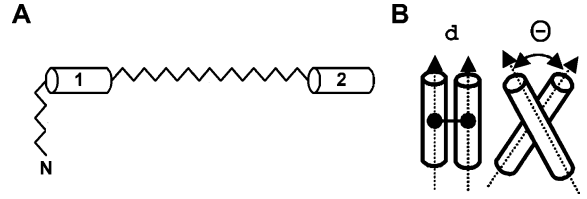


Figure 1. (A) Schematic drawing of the molecular system used in Monte Carlo simulations. Cylinders marked "1" and "2" represent α-helical monomers (segments 69−97) of GpA. Thin broken lines indicate dummy residues. Termini of the system are marked with symbols "N" and "C". (B) Definition of the geometrical parameters for two α-helices in a dimer: Θ and *d* are respectively the angle and the distance between the helical axes.

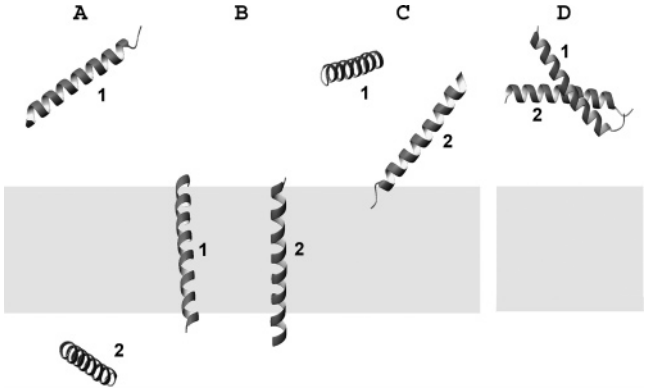


Figure 2. (A−C) The arbitrary chosen starting configurations of two noninteracting α-helices of GpA used in Monte Carlo simulations. (D) The starting structure of the NMR-derived model of the GpA dimer. The peptides are displayed with ribbons. The nonpolar layer of membrane is shown in gray. The monomers are marked "1" and "2".

coordinates (0,0,0). The inclusion of the "dummy" regions was dictated by the necessity of continuity of the protein backbone for simulations in dihedral angles space. Charges were assigned to the peptide atoms at pH 7. Simulations were started from three independent positions of noninteracting α-helices (Figure 2). Atomic coordinates of the NMR-derived model[14] of the dimer (GpA_{NMR}, residues 69−97) were taken from the Brookhaven Protein Data Bank,[27] entry 1afo. MC search for GpA_{NMR} in an implicit membrane was performed with dummy residues introduced as shown in Figure 1A. The starting structure was arbitrarily placed in an aqueous phase (Figure 2D). The lowest-energy state found for GpA_{NMR} in a membrane is referenced as $GpA_{NMR-mem}$.

Simulation Protocol. The peptides' conformational space was explored via MC search in torsion angles space as described elsewhere.[28] The membrane was represented by a "hydrophobic slab" described by an effective solvation potential. This was done using atomic solvation parameters (ASP) for gas−cyclohexane and gas−water transfer, which mimic the hydrophobic membrane core, lipid−water interface, and aqueous solution.[25] All-atom potential energy function for the protein was taken in the following form: $E_{total} = E_{ECEPP/2} + E_{solv} + E_{\Delta\psi}$. The term $E_{ECEPP/2}$ includes van der Waals, torsion, electrostatic, and hydrogen bonding contributions.[29] E_{solv} is the solvation energy

$$E_{solv} = \sum_{i=1}^{N} \Delta\sigma_i \, ASA_i \tag{1}$$

where ASA_i and $\Delta\sigma_i$ are accessible surface area (ASA) and ASP of atom i, respectively, and N is the number of atoms. The values of ASPs were taken from ref 28. Interaction of the protein with both aqueous and membrane environments is given by eq 2, where $\Delta\sigma_i$ depends on the z coordinate of atom i (the axis Z is normal to the membrane plane)

$$\Delta\sigma_i(z) = \begin{cases} \Delta\sigma_i^{mem} - 0.5\cdot(\Delta\sigma_i^{mem} - \Delta\sigma_i^{wat})\cdot e^{(|z|-z_0)/\lambda} & \text{if } |z| < z_0 \\ \Delta\sigma_i^{wat} + 0.5\cdot(\Delta\sigma_i^{mem} - \Delta\sigma_i^{wat})\cdot e^{-(|z|-z_0)/\lambda} & \text{if } |z| \geq z_0 \end{cases} \tag{2}$$

where $\Delta\sigma_i^{mem}$ and $\Delta\sigma_i^{wat}$ are ASP values for the type i atom in aqueous (wat) or nonpolar (mem) environments, respectively; z_0 is a half-width of the membrane (i.e., the hydrophobic layer is restricted by the planes given by the equation $|z| = z_0$); $D = 2z_0$ is the membrane thickness (30 Å); and λ is a characteristic half-width of the water−membrane interface (in this study $\lambda = 1.5$ Å).

Nonbond interactions were truncated with a spherical cutoff of 30 Å. As discussed earlier,[28] electrostatic interactions were treated with distance-dependent dielectric permeability $\epsilon = 4 \times r$. Prior to MC simulations the structures were subjected to 80−150 cycles of conjugate gradients minimization. Then several consecutive MC runs (5×10^3 steps each) with different seed numbers and sampled 5, 3, 2, 1 randomly chosen torsion angles were performed without restraints. At each MC step the structures were minimized via 50−150 conjugate gradients iterations. In each run the initial conformation was the lowest-energy one found in previous runs. In sum, $\sim 2.5 \times 10^4$ MC steps were performed for all systems in one complete MC simulation. To preserve the structure of the GpA_{NMR} model, in the beginning of the MC search (first 3.3×10^3 MC steps), a set of intermonomer distance restraints derived from the NMR structure was applied (Table 1 in the Supporting Information). The later MC stages were performed without restraints. In total, $\sim 4 \times 10^4$ MC steps were done for GpA_{NMR}. One of the resulting low-energy conformers (model GpA_{MC}) was subjected to constant-temperature MC simulation without minimization. This was done using 5×10^5 MC steps at $T = 300$ K.

The effect of TM potential ($\Delta\psi = 300$ mV) was taken into account using a special energy term ($E_{\Delta\psi}$)[7,30,31]

$$E_{\Delta\psi} = (F\Delta\psi/D)\sum_{i=1}^{N} q_i z_i \tag{3}$$

where q_i and z_i are partial charge and coordinate z of atom i, and F is Faraday's constant. For $|z_i| > z_0$, $\psi(z) = \text{const}$. Earlier we have shown[31] that modeling with $\Delta\psi = 100$ mV reproduces the behavior of a signal peptide adequately well as compared to experimental observations made under $\Delta\psi = 30$ mV. So, we assume that the magnitude of $\Delta\psi =$

300 mV used in simulations (later also denoted as "$\Delta\psi \neq 0$") roughly corresponds to $\Delta\psi \sim 100$ mV in real membranes.

The models of both "less hydrophobic" and "more hydrophobic" membranes were constructed as follows. In the former case the membrane's nonpolar part ($|z| \leq 15$ Å) was described by the ASP values obtained on the basis of the gas-octanol free energies of transfer.[28] In the latter case, this region was approximated with ASPs corresponding to a hypothetical solvent that is "more hydrophobic" than cyclohexane. To do this, the gas-cyclohexane values of ASPs were calculated according to the following formula: $\Delta\sigma^{mem'} = a \times \Delta\sigma^{mem}(a - 1) \times \Delta\sigma^{wat}$, where the coefficient a was taken to be equal to 1.2. The modified membrane models were employed in MC simulations analogous to those described above. Other details will be given elsewhere (Efremov et al., manuscript in preparation).

A Hypothetical Left-Handed Model of the GpA Dimer. Such the model (hereafter indicated as GpA_L) was built as follows. The low-energy MC state with parameters ($d \approx 5.7$ Å; $\Theta \approx 58°$) was used as a structural template. d and Θ are the distance and the angle between helical axes of the monomers (Figure 1B). Two ideal α-helices of GpA were fitted into this model over C_α atoms using the least-squares criterion, thus preserving the helix packing inherent in the MC model. The resulting structure was subjected to energy minimization in vacuo using 3×10^2 steepest descent iterations followed by 2×10^4 conjugate gradients steps. Five pairs of NMR distance restraints[14] were employed in the minimization protocol. The calculation and visualization of hydrophobic/hydrophilic properties of GpA were carried out using the molecular hydrophobicity potential (MHP) approach, as described elsewhere.[32] The MHP values were calculated on the solvent-accessible surfaces of α-helical segments Glu72-Ile95 in the models $GpA_{NMR-mem}$, GpA_{MC}, and GpA_L.

Analysis of the Results. Resulting states were analyzed using a set of auxiliary programs specially written for this. Only the low-energy states (in the range $[E_{min}, E_{min}+\Delta E]$, where E_{min} is the minimal energy, $\Delta E = 15$ kcal/mol) were considered. Mutual disposition of the monomers was described in terms of the values d, Θ, and the dimerization interface. The helix axes were calculated using the least-squares fit to coordinates of backbone atoms. Accessible surface areas (ASA) of residues and their secondary structure were assessed using the DSSP program.[33] Residue i was considered to lie on the dimerization interface if the difference between its ASA values in dimer and in monomer exceeds 25 Å^2 (10 Å^2 for glycines). Clustering of low-energy states was done based on the parameters of helix packing (Θ, d) and the composition of dimerization interface.

3. Results

The Choice of the Studied Systems and the Simulation Protocols. To understand to what extent the membrane environment determines the spatial structure of monomers and drives their association MC simulations of two α-helices of GpA were performed in vacuo and in a hydrophobic slab with and without TM voltage. (The term "vacuum" means that the solvation energy was omitted.) Two types of starting configurations were used in each medium, these being the randomly oriented helices and the GpA_{NMR} model (see Methods). Furthemore, in the former case several independent simulations with different random starts were carried out with a view of checking, whether the unrestrained MC search may be used to predict the states that agree with the experimental observations. In the latter case the following questions were addressed: "Whether the experimental structure is stable in membrane mimic?"; "Whether the yielded low-energy states compete in energy with the "nativelike" structures?"; and "How do they bind to the membrane (geometry of insertion)?" Comparison of the results obtained in both cases is required to assess the quality of sampling of the GpA's conformational space: is it sufficient enough to provide solutions that agree with the experimental data. It is important that the low-energy states obtained in simulations of the same type (see below) revealed similar energies, structures, and modes of membrane binding. This gives strong grounds to believe that the essential sampling of the conformational space was reached in MC simulations.

Simulations in Vacuo: Wrong Way to the "Nativelike" Structure of the Dimer. The low-energy states obtained from random starts represent either "hairpin (HP)-helix" structures or antiparallel (head-to-tail) dimers ($TM_{\uparrow\downarrow}$) (not shown). They demonstrate large conformational heterogeneity—HP-helices may be destabilized on different residues, and spatial disposition of the monomers varies in a wide range. Furthermore, interfaces between the monomers drastically differ from those in the GpA_{NMR} structure. A similar situation was seen with the simulations starting from the GpA_{NMR} model. In this case the initial "realistic" structure significantly loses in energy to the conformers found from random starts. In vacuo the "nativelike" model is unstable—the monomers rapidly change their mutual orientation, and the structure converges into the ones obtained from random starts (not shown). Therefore, unrestrained modeling in vacuo cannot be used for adequate simulations of TM helix interactions. Below we present the results obtained in a symmetrical membrane. (The term "symmetrical" means that both sides of the membrane are equivalent in the mathematical sense.)

Two Helices in a Symmetrical Hydrophobic Slab: Predominance of Misfolded Dimers. As seen in Figure 3A, the presence of a hydrophobic layer stabilizes the α-helical structure of both monomers and forces them to adopt TM orientations. Also, in their energetically favorable states the helices pack together and form antiparallel ($|\Theta| > 90°$, $TM_{\uparrow\downarrow}$) or parallel ($|\Theta| < 90°$, $TM_{\uparrow\uparrow}$) dimers. The population of the former ones is considerably higher. Helix packing parameters for some of the low-energy states are shown in Table 1. It can be seen that the states $TM_{\uparrow\downarrow}$ and $TM_{\uparrow\uparrow}$ have low energies of interaction with the environment (E_{solv}): the hydrophobic residues in the central part of helices are exposed to apolar membrane core, while the hydrophilic ones (on the GpA termini) are accessible to water. Also, the overall stability of such dimers is determined by the saturation of their H-bonding potencies (maximal number of residues in α-helix) and by favorable van der Waals interhelical contacts.

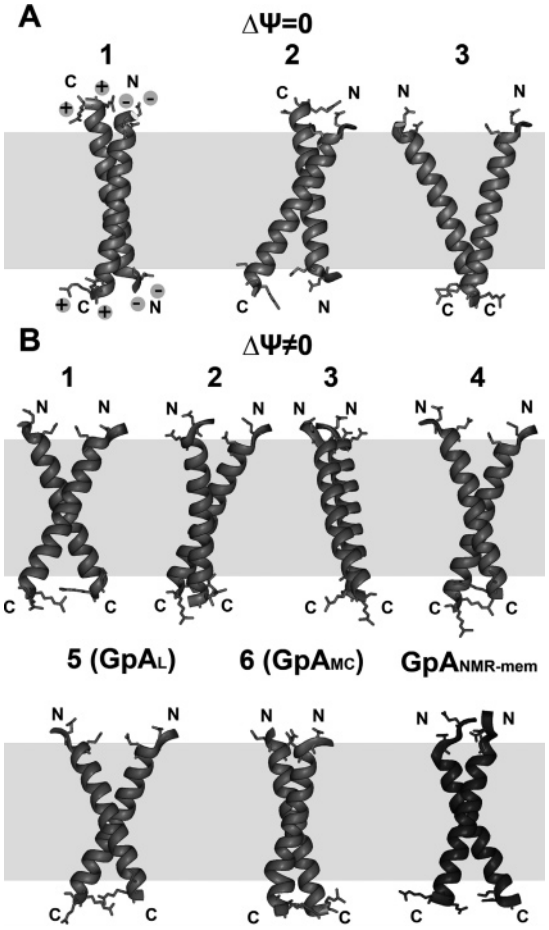


Figure 3. Monte Carlo simulations of two α-helices of GpA in implicit membrane without (A) and with (B) applied TM potential. Groups of the low-energy states are indicated with numbers (as in Table 1). $GpA_{NMR-mem}$ is the lowest-energy state of the NMR-derived model of GpA. The peptide's termini are marked with symbols "N" and "C". Side chains of positively (Arg96, Arg97) and negatively (Glu70, Glu72) charged residues are shown in stick presentation. In the structure A-1 they are also marked with symbols "+" and "−", respectively. The peptides are displayed with ribbons. The nonpolar layer of membrane is shown in gray.

It is remarkable that in all TM dimers occurrence of the motif GxxxG on the helix−helix interface correlates with low values of the E_{vdW} energy term. $TM_{⇅}$ structures reveal smaller values of the electrostatic term (E_{elect}) due to interactions between opposite charges on the monomers' termini. In the calculations starting from the GpA_{NMR} model, the low-energy states ($GpA_{NMR-mem}$) represent TM α-helical dimers, and the monomers almost do not change their packing as compared to the initial structure. Tight packing of parallel helices in $GpA_{NMR-mem}$ (low E_{vdW}) compensates for the unfavorable electrostatic contacts between their termini.

Electrostatic Potential on the Membrane: The Way To Correct Folding of the Dimer? *Unrestrained Simulations from Arbitrary Starts.* The low-energy states obtained from random starts in a membrane with applied TM voltage are characterized as follows. Like in a symmetric ($\Delta\psi = 0$ mV) membrane, the monomers retain well their initial α-helical conformation (Figure 3B). Both α-helices adopt a

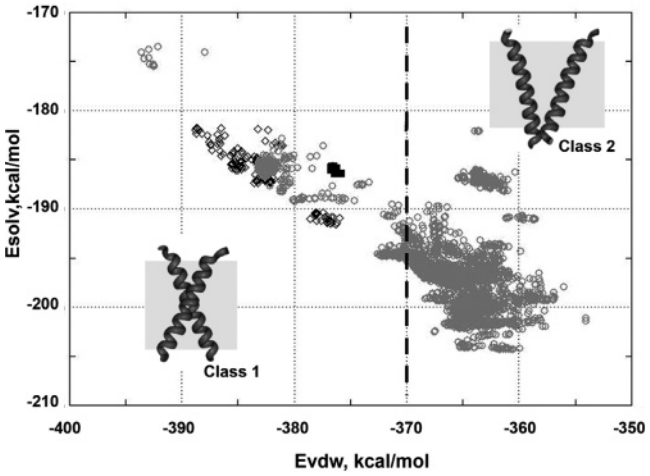


Figure 4. Two classes of the low-energy states found via Monte Carlo simulations in implicit membrane with applied TM potential. Distribution of the dimers over their solvation (E_{solv}) and van der Waals (E_{vdW}) energy terms. Class 1: tightly packed dimers with symmetrical interface; class 2: dimers with helices interacting by the terminal parts. Typical examples of structures from both classes are shown with ribbons. Dashed vertical line indicates an approximate boundary between the two classes. The symbols "◇*" and "■" mark respectively $GpA_{NMR-mem}$ and GpA_{MC} states. For other details see the legend to Figure 3.

TM-orientation, and their hydrophobic parts strongly interact with the medium imitating lipid bilayer—the corresponding impact to the solvation energy of the dimer (E_{solv}) being ~ -35 kcal/mol for the hydrophobic membrane core and ~ -15 kcal/mol for the interfacial region. The peptides' termini reveal highly favorable contacts with water, E_{solv} being ~ -135 kcal/mol. Moreover, in all low-energy states the monomers constitute tightly packed TM complexes, i.e., the formation of protein−protein contacts in a nonpolar environment is an energetically more beneficial process than giving to each of the peptides all the accessible area for contacts with the medium. At the same time, a number of significant differences were observed comparing with the case $\Delta\psi = 0$. For instance, the dipole moments of the α-helices are oriented in the direction of the TM electric field. As a result, $TM_{⇈}$-dimers typical of GpA in membrane are well reproduced in the calculations (Figure 3B). Besides, the presence of $\Delta\psi$ affects the geometry of helix packing: unlike the case $\Delta\psi = 0$, the resulting states agree better with the GpA_{NMR} model (Table 1). Additionally, a lesser degree of conformational heterogeneity of the MC states results in a restricted number of solutions. To define how adequate the dimeric MC structures are, it is necessary to inspect the mutual positions of the helices and to pinpoint the interactions driving the dimerization in the membrane.

With this aim in view, the total energy and its components for both the entire complexes and individual monomers and residues were considered, and the dimerization interfaces were delineated. The results reveal two highly populated classes of conformers (Figure 4). In the first of them the area of intermonomer contacts is located in the central part of the dimer, while in the second class the subunits interact mostly by their termini. States from both classes have

Table 1. Parameters of the Low-Energy States of the GpA Dimer Found via Monte Carlo Simulations in Implicit Membrane with and without Applied Transmembrane Voltage

Group[1]	d [2]	Θ [2]	dimerization interface [3]	E_{total} [4]	$E_{elect.}$	E_{vdW}	E_{solv}
Membrane with $\Delta\psi = 0$							
1 ($TM_{\uparrow\downarrow}$)	9.2 (0.4)	165.3 (2.7)	SEPEITLIIFGVMAGVIGTILLISYGIRR ___E__LI__GV__G__T__LI___I__ R__GY_IL__T___G__VG__I___E___	-584.5 (2.4)	10.3 (1.3)	-373.1 (8.4)	-173.0 (5.4)
2 ($TM_{\uparrow\downarrow}$)	9.8 (0.1)	142.2 (0.4)	_____I__V__G__T__LI__GI_R RR___S___I__I___M___I__T_____	-585.6 (3.6)	13.4 (0.5)	-362.5 (0.6)	-181.9 (0.6)
3 ($TM_{\uparrow\uparrow}$)	8.9 (0.2)	-33.8 (1.3)	_____V___I__I___I__ _____I__IS_____	-577.5 (1.3)	14.1 (0.2)	-357.4 (1.5)	-183.5 (0.5)
Membrane with $\Delta\psi \neq 0$ Class-1 structures							
1 ($TM_{\uparrow\uparrow}$)	10.1 (0.1)	-50.1 (0.5)	SEPEITLIIFGVMAGVIGTILLISYGIRR _____II__VM__V___I_____ _____II__VM__V___I_____	-629.8 (0.5)	18.2 (0.2)	-370.4 (0.3)	-191.8 (0.4)
2 ($TM_{\uparrow\uparrow}$)	9.2 (0.1)	22.1 (1.1)	____I___I__V__GV__T___I__G__ _____I__V__GV__T__LI_____	-632.4 (1.7)	18.9 (0.2)	-382.3 (1.5)	-185.9 (0.6)
3 ($TM_{\uparrow\uparrow}$)	9.0 (0.1)	5.9 (0.2)	____I___I__VM__V___I__I___I__ ___E__LI__GV__G__GT__L___G___	-626.9 (2.7)	18.8 (0.3)	-379.3 (1.9)	-188.0 (1.3)
4 ($TM_{\uparrow\uparrow}$)	6.9 (0.1)	42.6 (0.8)	_____LI__GV_AG__G__LL__Y__R_ ___E___I__GV_AG__G__LL__Y____	-629.3 (1.6)	21.6 (0.4)	-392.3 (1.4)	-174.8 (0.7)
5 ($TM_{\uparrow\uparrow}$)	5.5 (0.2)	60.2 (2.0)	_____I__GV_AG__G__LL__Y____ _____I__GV_AG__G__LL_____	-626.5 (2.9)	19.3 (1.3)	-378.6 (3.0)	-190.2 (1.2)
6 (GPA_{MC}) [5] ($TM_{\uparrow\uparrow}$)	9.1 (0.4)	-25.9 (1.4)	___E__LI__GV__GV__T___I_____ ___E__LI__GV__GV__T___I_____	-628.4 (4.4)	17.4 (0.4)	-372.7 (2.5)	-192.0 (3.7)
$GpA_{NMR\text{-}mem}$ 6 ($TM_{\uparrow\uparrow}$)	7.1 (0.1)	-39.8 (0.3)	___E__LI__GV__GV__T___I_____ ___E__LI__GV__GV__T___I_____	-622.3 (1.0)	20.6 (0.2)	-382.5 (3.7)	-187.0 (2.8)
Class-2 structures [7]							
a ($TM_{\uparrow\uparrow}$)	7.1 (0.9)	-39.9 (4.2)	_____I__GI__ _____T__LI__GI__	-623.1 (2.1)	17.0 (1.5)	-359.7 (2.4)	-196.5 (3.5)
b ($TM_{\uparrow\uparrow}$)	9.5 (0.8)	47.0 (5.9)	__PE__L_____ _____L_____	-624.1 (1.7)	18.1 (0.4)	-361.1 (2.6)	-195.5 (1.3)
c ($TM_{\uparrow\uparrow}$)	10.9 (0.5)	47.8 (4.2)	_____LI__G__G_____ __P___L__F_____	-622.5 (1.4)	19.0 (0.6)	-359.4 (3.9)	-196.5 (2.2)

[1] The group numbers are the same as in Figure 3. The symbols $TM_{\uparrow\downarrow}$ and $TM_{\uparrow\uparrow}$ indicate antiparallel and parallel TM dimers, respectively. [2] d and Θ are respectively the mean distance (in Å) and the mean angle (in deg) between helix axes for a given group of states. Standard deviations are given in brackets. [3] Residues found on the helix−helix interface are indicated with their one-letter code. Gray hatching shows, whether these residues also contribute to dimerization according to mutagenesis data.[11] Sequences of the monomers 1 and 2 are shown either in antiparallel or in parallel orientations—for $TM_{\uparrow\downarrow}$ and $TM_{\uparrow\uparrow}$ states, respectively. [4] E_{total}, E_{elect}, E_{vdW}, and E_{solv} are the mean values of total, electrostatic, van der Waals, and solvation energies, respectively (in kcal/mol). Standard deviations are given in brackets. [5] GpA_{MC} is the calculated model most close to the NMR structure of the dimer.[14] [6] $GpA_{NMR-mem}$ is the lowest-energy state found starting from the NMR-derived model (GpA_{NMR}) of the dimer[14] in implicit membrane. The values Θ and d in $GpA_{NMR-mem}$ and in GpA_{NMR} are very similar. [7] These class-2 states are not shown in Figure 3B.

comparable values of total energy, but the terms E_{solv} and E_{vdW} differ. Thus, the states of class-1 are characterized by a tighter packing of helices (Figure 4). As a result, E_{vdW} for them is ~10 kcal/mol lower than in class-2. The states of class-2 have a smaller area of intermonomer contacts (370 ± 50 Å^2 and 250 ± 70 Å^2 in classes 1 and 2, respectively). Because of this the subunits interact with the membrane to a greater extent, thus resulting in lower values of E_{solv}.

As argued below (see Discussion), the class-1 of states was selected for further inquiry. In total, about 800 such structures were delineated. They have diverse types of helix

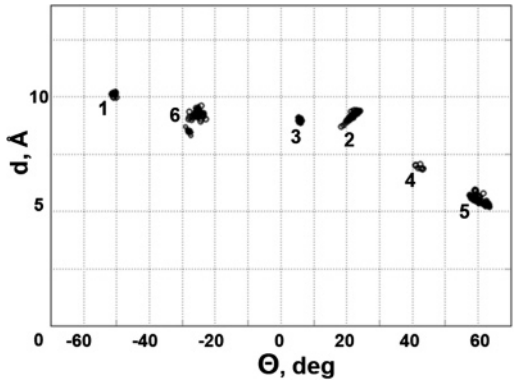


Figure 5. Distribution of the low-energy states of class-1 found via Monte Carlo simulations in implicit membrane with applied TM potential over their helix packing parameters *d* and Θ (distance and angle between helix axes, respectively). Numbering of groups is that as in Table 1.

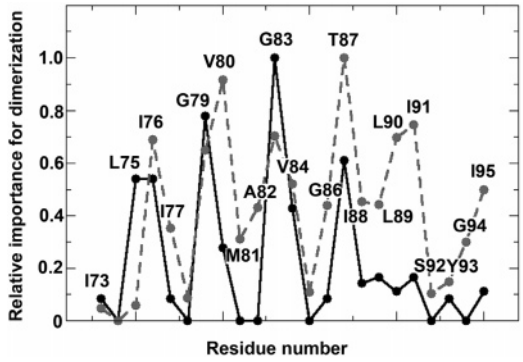


Figure 6. Relative importance of GpA residues in dimerization. Bold line: mutagenesis data, taken from ref 11. Dashed line: frequency of occurrence of residues on the helix−helix interface: results of summary analysis of the groups of low-energy states obtained via Monte Carlo simulations (listed in Table 1). The distributions are normalized on the range [0, 1].

packing (Table 1). Accordingly, the composition of the dimerization interface may vary as well. Nevertheless, in all these structures, most of the residues involved in helix−helix contacts were also found on the interface in experimental studies. The groups of low-energy states represent well-defined and compact clusters which do not overlap (Figure 5), and therefore only a limited number of possible packing geometries should be used for future analysis. No outliers of the clusters were found. As shown in Table 1, a complete conformity with the known motif of dimerization was not observed for the theoretically predicted states, although in some of them six of seven residues were determined correctly. Indeed, summary analysis of dimerization interfaces in all groups listed in Table 1 yields a distribution of residues most frequently involved in helix−helix contacts quite similar to that found via mutagenesis in ref 11 (Figure 6). Putative reasons for the discrepancies observed near the C-terminus are discussed later (see Discussion).

In the considered structures α-helices pack in a symmetrical (or nearly symmetrical) manner with respect to one another (Table 1). Because the α-helices have identical sequences, this result seems quite plausible, although asymmetrical homodimers have also been predicted for some other TM peptides (e.g., ref 18). We should note that the NMR model has been built as a symmetrical dimer as well. Interestingly, the formation of complexes with either negative (groups 1 and 6, Table 1) or positive (groups 2−5) values of the angle Θ is possible, corresponding to right- and left-handed double super-helix, respectively. For both types of structures the largest population of states is observed when $d \sim 9$ Å. This agrees with the data on statistical analysis of high-resolution structures of α-helical complexes in MPs:[8,34] in the majority of the helical pairs $d \sim 9.6$ Å, while those with $d \sim 7.0$ Å are rare. Among the states with $\Theta < 0$ is one (highly populated) group of dimers with packing parameters close to the model GpA_{NMR} (Table 1). The best correspondence is observed for the group-6 (GpA_{MC}). These are the states for which the dimerization interface is entirely symmetrical and agrees well with that found by NMR. To inspect the stability of the resulting dimeric states, one of the found low-energy conformers (GpA_{MC}) was subjected to constant-temperature MC simulation without minimization. It was shown that the GpA_{MC} model was stable during 5×10^5 MC steps: resulting all-atom RMSDs were within 1 Å from the initial structure.

Simulations with the GpA_{NMR} Start. Resulting low-energy states represent TM complexes with the termini accessible to water (Figure 3B, Table 1). The initial structure is well retained—for residues 71−95 the backbone RMSDs with the starting NMR model do not exceed 1.2 Å. The model $GpA_{NMR-mem}$ has the total energy close to that in other groups of states (Table 1). Comparison with one of such structures (GpA_{MC}) reveals that both dimers have similar geometry of membrane binding (Figure 3B) and dimerization interface (Table 1). One exception is provided by residues Leu75—in the model GpA_{MC} their side chains do not form such tight intermonomer contact, as in the experimental structure. Despite a relatively high backbone RMSD as compared to the GpA_{MC} structure (2.5 Å on the region 71−95) and somewhat overestimated distances for a number of helix contacts on the N-terminus (Table 2 in the Supporting Information), in overall, the model GpA_{MC} represents a rather good approximation for the structure observed by NMR in micelles. We should mention that, unlike the GpA_{NMR} model, the helices in the GpA_{MC} states were not restrained to their ideal conformations.

The Role of Hydrophobicity Degree of the Membrane. To estimate the sensitivity of the modeling results to changes of the parameters of the lipid bilayer, calculations analogous to those described above were conducted using "less hydrophobic" and "more hydrophobic" membranes. In the former case the membrane's nonpolar part was approximated with octanol, while in the latter study a hypothetical solvent that is "more hydrophobic" than cyclohexane (in terms of free energy of transfer from water) was employed (see "Methods"). It was shown that the application of such modified membranes leads to unrealistic conformations of the dimer: improper helix packing, helix destabilization, the appearance of nondimeric forms, etc. (data not shown,

manuscript in preparation). This clearly demonstrates the importance of the balance between E_{solv} and other energy terms: underestimation as well as overestimation of the media effects may considerably affect the resulting low-energy states.

Left-Handed α-Helical Dimer. As seen in Table 1, there are several groups of dimers with $\Theta > 0$. Two of them are especially interesting, namely groups 2 and 5. This is because the packing parameters in group-2 are frequently observed in left-handed TM helical pairs,[34] while the densely packed complexes of group-5 bear a resemblance to the left-handed dimers proposed earlier based on mutagenesis data.[11] Monomers in this group have favorable van der Waals contacts and demonstrate symmetrical tight packing of α-helices—mainly due to interactions in their middle parts, via the motif GVxAGxxG. Whether is it possible a dimeric structure with $\Theta > 0$, which also satisfies the constraints, imposed by mutagenesis and NMR results? To answer this question, a hypothetical structure was built based on one of MC conformers from the group-5 (see "Methods"). Detailed analysis of helix packing in structures GpA_L, $GpA_{NMR-mem}$, and GpA_{NMR} reveals that the model GpA_L satisfies reasonably well the two independent sets of NMR distance restraints.[14,15] In the former case, average violations of the restraints are 1.10, 0.86, and 0.74 ± 0.07 Å for GpA_L, $GpA_{NMR-mem}$, and an ensemble of 19 experimental models (GpA_{NMR}). In the latter case, such violations are 0.84, 0.18, and 0.47 ± 0.17 Å, respectively (Table 3 in the Supporting Information). So, the model GpA_L fits better to the solid-state NMR data. Interestingly, MC simulations of the model GpA_{NMR} in an implicit membrane lead to a very good agreement with the restraints measured in lipid vesicles as compared to those obtained in micelles: during MC simulations the "micellar" model converges to the structure observed in lipid vesicles. Therefore, overall, the hypothetical model GpA_L does not contradict NMR results. Analysis of helix−helix contacts in GpA_L demonstrates that the dimerization interface includes residues GXXXG (Table 1). Interestingly, residues Leu75, Ile76, Gly79, Val80, Gly83, and Thr87, the importance of which for helix association was proved by the mutagenesis results, form a helix−helix interface in the model with $\Theta > 0°$. Apart from those, the residues Gly86, Leu89, and Ile91 are also involved in helix association. According to mutagenesis data, replacement of each of the last ones seriously affects dimerization, even though they lie apart from the interface in the model GpA_{NMR}.

Complementarity of Hydrophobic/Hydrophilic Surfaces. Two-dimensional hydrophobicity maps for α-helices in models GpA_{NMR}, GpA_L, and GpA_{MC} are shown in Figure 7. Contour isolines display on their surfaces hydrophobic regions with high positive values of MHP, and the gray-hatched areas indicate dimerization interfaces. It can be seen that the characteristic feature of each α-helix is its strong hydrophobicity and presence of a relatively more polar "Λ-shaped" pattern. Its left and right "arms" are formed by residues Ser92, Thr87, Gly83, Gly79 and Gly94, Gly86, Gly79, respectively. In the right-handed complexes (GpA_{NMR}, GpA_{MC}) the dimerization interface fits to the first polar region (Figure 7A,B), while in the left-handed model (GpA_L)—to the second one (Figure 7C). As seen in Figure 7A,B, in the structure GpA_{MC} the interfacial contact area agrees well with that observed by NMR, although in the theoretical model the contact pattern has a somewhat smaller tilt with respect to the helix axis (and, therefore, smaller value of the angle Θ). In addition, in all the models the interface is composed of parallel hydrophilic and hydrophobic stretches. In the right- and left-handed dimers the last one is formed by residues Ile76, Val80, Val84, Ile91 and Val80, Val84, Leu89, Leu90, respectively. Analysis of the dimerization interface in both right- (GpA_{MC}) and left- (GpA_L) handed MC models shows that the interdigitation of the side chains ("grooves-into-ridges") observed in the NMR structures is well reproduced (Figure 7, bottom). Moreover, in the calculated structures χ_1 rotameric states of residues correspond well to those in the NMR models (data not shown).

4. Discussion

Simulations in Vacuo or in a Hydrophobic Slab? Unrestrained modeling in vacuo clearly demonstrates that media effects are very important for proper description of folding and assembling of TM α-helices. Thus, the simulations from random starts were unable to reproduce the "nativelike" parallel α-helical dimers observed in experiments. In addition, the well-packed GpA_{NMR} structure was unstable in vacuo. The last result points to the fact that the failure of MC search from arbitrarily chosen positions of helices is not related to the problem of insufficient sampling of the conformational space for the system. Instead, the reason is that the absence of heterogeneous polar/apolar membrane environment leads to overestimated electrostatic interactions between charged termini of the peptides. In addition, aliphatic side chains of a large number of residues tend to form hydrophobic contacts with each other but not to be exposed on the surface.

On the contrary, MC search from random starts with noninteracting α-helices reveals that monomers of GpA retain α-helical conformation in the hydrophobic slab, adapting TM orientation and tightly packing together to form stable parallel and antiparallel dimers. In a symmetrical slab only a small fraction of the low-energy states resembles the GpA_{NMR} structure. Modeling in such a membrane does not possess high predictive power because it results in a large number of possible solutions, these being TM complexes with close energies but drastically different packing. Moreover, most of them are misleading, their "head-to-tail" topology not resembling that found in experiments. It is important that, independently of the starting configuration (random starts or GpA_{NMR} model), resulting low-energy states have close values of the total energy. This demonstrates high efficiency of the MC search and makes us confident that the developed protocol provides an essential sampling of states. Similar findings have also been reported in our previous studies of other peptides and proteins (reviewed in ref 7).

As for the surprisingly high population of misfolded ($TM_{\uparrow\downarrow}$) states in a symmetrical membrane, we suppose that it is due to the equivalence of the two membrane sides: since there is no preferential direction for the dipole moments of α-helices, both types of the dimers are energetically favor-

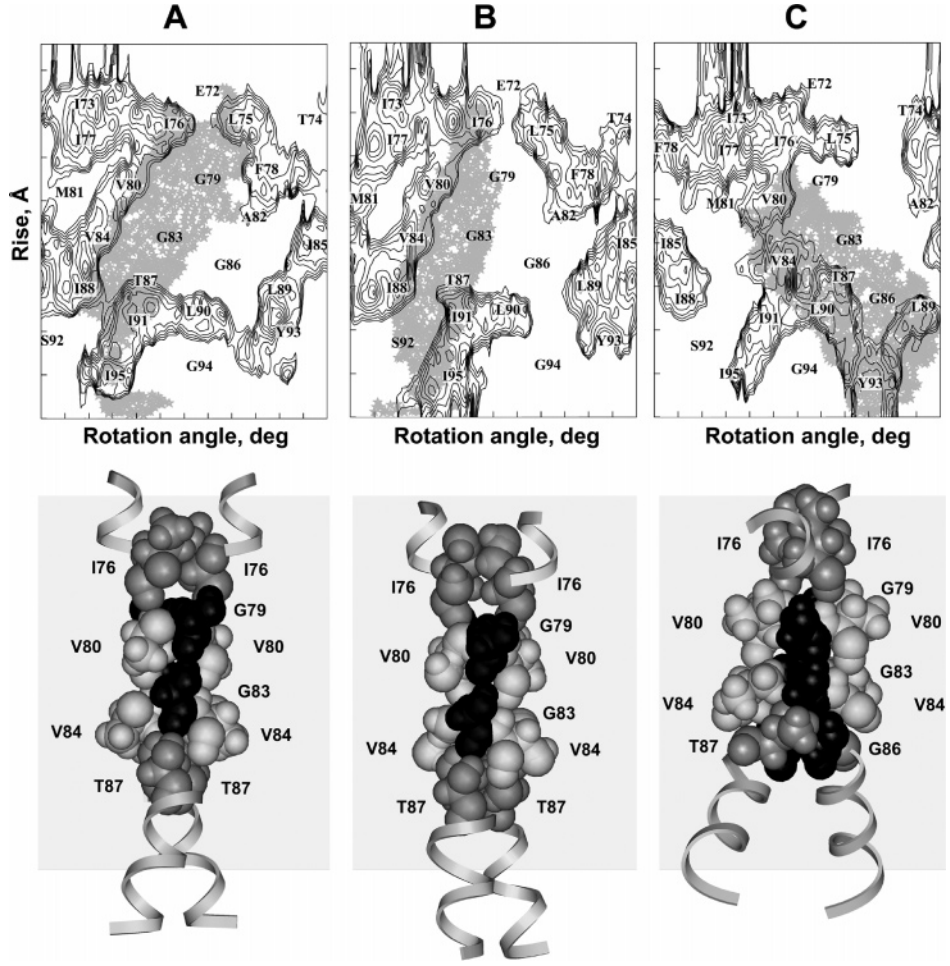


Figure 7. Hydrophobicity and packing of the NMR-derived and calculated models of GpA dimer: GpA_{NMR} (A), GpA_{MC} (B), and GpA_{L} (C). (Top) Hydrophobic properties of α-helices. 2D isopotential map of the molecular hydrophobicity potential (MHP) on the peptide surface calculated as described in ref 31. The value on the *X* axis is the rotation angle about the helix axis; the parameter on *Y*-axis is the distance along the helix axis. MHP is given in octanol−water logP units. Only the hydrophobic areas with MHP > 0.09 are shown. Contour intervals are 0.015. The positions of residues are indicated by letters and numbers. Gray-hatching marks helix−helix interface in the dimer. (Bottom) Molecular models of the dimers. The peptides are shown with ribbons. Residues on the dimerization interface are shown in ball presentation. Glycines are colored in black, valines—in light gray, Ile76 and Thr87—in medium gray. The nonpolar layer of membrane is shown in light gray.

able. In contrast, real biological membranes manifest TM potentials ($\Delta\psi$) present across lipid bilayers. They feed and control many biological processes, ranging from energy conversion and channel gating to protein insertion and translocation. The role of $\Delta\psi$ seems to be especially important for TM helices of GpA because they carry charges of opposite sign on their termini and possess large dipole moments (~90−100 D). It seems reasonable that the presence of $\Delta\psi$ destabilizes the $TM_{\uparrow\downarrow}$-states and sufficiently favors the $TM_{\uparrow\uparrow}$-ones, where the dipole moments of helices are oriented along the electric field. Therefore, to provide a more realistic treatment of helix association, the influence of $\Delta\psi$ should be taken into account.

Correct Assembling of Helices Requires a Hydrophobic Slab and a TM Voltage. In some aspects the results of simulations with applied TM voltage are similar to those in symmetrical membrane: the energetically favorable states represent membrane-spanning and densely packed α-helical dimers. At the same time, occurrence of $\Delta\psi$ provides a far more correct description of helix behavior. In particular, it is important for the "nativelike" association of TM α-helices maintaining their parallel orientation. As mentioned above, the found low-energy states are divided into two large classes (1 and 2, or "X"- and "V"-shaped conformers, respectively). The class-1 of states was selected for further inquiry. Such a choice is determined by the following notions. Primarily, conformers from the class-1 have significantly lower energy of intermolecular interactions as compared to the class-2: -60 ± 15 and -30 ± 5 kcal/mol, respectively. In the former case the contact area comprises 6−10 residues, and in the second one it is much less extended—only 3−6 residues. Second, the class-1 states have quite a similar interface (Table 1, groups 1−6), whereas in the class-2 it varies greatly (see examples in Table 1). Therefore, helix interactions in the class-1 are much more specific. As the oligomeric states of MPs are usually functionally active,[2] the packing of α-helices has to be quite specific. Finally, the average total energies in the groups of class-2 states (Table 1) are somewhat higher than those in the class-1. Taking all the above said into account, we assume that the complexes from class-1 are more

likely to represent the "nativelike" dimeric structures in membrane. However, we should note that the "V"-shaped dimers of TM helices (resembling our states of class-2) have been proposed for some other proteins,[35] and, moreover, the switch between "X"- and "V"-shaped conformations has been suggested as functionally important.[36] But at this stage we have no ideas about the significance and the putative role of such states found for GpA. As discussed below, we cannot also exclude the possibility that the "V"-shaped conformers may be stabilized in particular membrane-mimic media, which are different from those employed so far in the experiments.

Dimerization interfaces in the MC states show reasonable correlation with the mutagenesis data (Figure 6). The discrepancies are mainly observed near the C-terminus—for residues Gly86, Ile88, Leu90, and Ile91. We should note that such an overall comparison may be done only in a qualitative manner for two principal reasons. First, the mutagenesis results are somewhat ambiguous. Thus, as indicated in ref 11, the effects of GpA's residues substitutions with polar residues are not well understood. In addition, the final histogram of the relative degree of disruption of the dimer (Figure 5 in ref 11) was built using an arbitrary chosen scale. Furthermore, similar histogram for the calculated MC states represents just a summation over the ensemble (about 800) of low-energy structures of class-1, whereas the structures of the class-2 were omitted. (However, as seen in Table 1, the dimerization interfaces in the class-2 structures also contain residues that are important for helix association according to mutagenesis data.) Our data may be insufficient to get the real distribution of states, although the essential sampling of the GpA's conformational space was realized (see above). Nevertheless, reasonably good agreement between the experimental and computational results demonstrates eligibility of the theoretical approach. It is principally important that, despite certain conformational heterogeneity, in all of the calculated structures the helices contact each other with the same side, and it is exactly the side revealed in experiments.

Left-Handed α-Helical Dimer—a Putative Alternative to the NMR-Derived Model? The feasibility of dimers with $\Theta > 0°$ has been discussed previously in the mutagenesis work of Lemmon et al.[11] In that study the left-handed models were omitted upon analysis of periodicity in the distribution of residues critical for dimerization. This has been done using Fourier transform processing of the peptide's sequence. On the other hand, such approaches simplify the picture to a huge extent, not taking into account the exact conformations of side chains. In our opinion, this criterion cannot serve as a basis for unambiguous decision about the impossibility of the left-handed α-helical complexes. Interestingly, two distinct families of dimers: right- ($\Theta \sim -50°$) and left- ($\Theta \sim 40°$) handed ones have recently been obtained in MD simulations of GpA in implicit membrane starting from noninteracting TM helices.[24] Although in that study the authors have not considered the membrane insertion of individual helices (like in this work), the results agree well with our findings. This proves that the question concerning the left-handed dimers remains to be answered. In our opinion, the presence of that type of structures raises issue of the possibility of dimeric models alternative to those obtained in the result of NMR data interpretation.[14]

Conformational Heterogeneity of Low-Energy States: A Shortcoming of the Method or an Indicator of Media-Dependent Equilibrium Distribution of Dimers? Several problems related to future applications of the developed computational approach should also be mentioned. The most severe of them is related to the unambiguity of MC solutions: the choice of the "nativelike" structure is not straightforward without a priori knowledge of the helix—helix interface. Although most of the predicted conformations of GpA dimer have quite similar overall patterns of intermonomer contacts, their detailed structures may differ. This depends on a number of factors, like the membrane thickness and hydrophobicity degree, presence of TM voltage, and so forth. For instance, changing ASP values which describe the hydrophobic slab (see above) may lead to incorrect structures of the dimer and even to nondimeric structures. In many cases the objective is not so much the elaboration of an extremely precise spatial structure of a protein oligomer as the establishment of a general mode of its insertion into membrane, along with the delineation of some crucial residues on the dimerization interface. Such a rough model may be employed in the future to rationalize experimental observations and to design new experiments. On the other hand, we assume that the aforementioned problem has a fundamental character. Namely, the moderate structural heterogeneity of predicted GpA dimers (existence of several groups of states with close energies) may reflect a real equilibrium dynamics in membrane-mimic media used in experimental studies. We also suppose that some fraction of left-handed α-helical dimers may present as well. As discussed above, such a conformational heterogeneity may be caused by the fact that the packing of helices is quite sensitive to media effects and the geometry of membrane binding. Such a hypothesis is partially corroborated by the vague results of mutagenesis studies (see discussion in ref 11) as well as by NMR[16,17] and MD[18,24] data that demonstrate media effects on the stability of helical oligomers and provide examples of their multistate equilibrium in membrane mimics. In real biological membranes the situation may be far more complex due to the heterogeneity of their physicochemical characteristics. As a result, one or more predominant conformations of the dimer may occur, although this assumption requires further investigation.

Despite this vagueness, in overall, the modeling results show that under certain conditions, like the presence of $\Delta\psi$, an optimal membrane thickness and hydrophobicity degree, only a limited number of structures are energetically favorable. Some of them are close to those observed by NMR. We assume that the others may also be realized in membranes and artificial membrane mimics with particular properties. In our opinion, future MD simulations in explicit bilayers and micelles starting from these representative conformations will help to gain an additional insight into the equilibrium behavior of helical oligomers. This work is currently in progress in our group. The proposed computational approach therefore represents an indispensable step

toward the development of efficient theoretical methods to study helix interactions in membranes.

Driving Forces for Helix Association. One of the criteria used to assess quality of the resulting models of the dimer is the distribution of hydrophobic/hydrophilic patterns on the surfaces of monomers. Thus, it is known that the helix−helix and helix−lipid interfaces in TM domains of proteins reveal complementarity of such properties.[37,38] To check, whether this is the case with the proposed models of GpA dimers, we applied the MHP approach, which has been previously employed to study a number of membrane-bound α-helices.[32] TM α-helices in MPs usually expose to lipids their nonpolar surfaces, and this is especially true for complexes with a small number of helices, e.g.—for dimers (e.g., ref 39). (In contrast, the stability of MPs with ≥ 4 TM helices is determined to a large extent by interhelical contacts.[40]) Therefore, judging by the MHP maps (Figure 7), one may propose that α-helices of GpA interact via the lengthy hydrophilic stretches on their surfaces. In this case there is a strong complementarity between polarity regions of the monomers, while the most hydrophobic surfaces are exposed to lipids. In principle, according to the proposed criteria, two types of helix packing are possible—via residues forming either a left or a right "arm" of the "Λ-shaped" pattern. Interestingly, both variants may be realized in real and in computational experiments—for both right- (GpA_{NMR}, GpA_{MC}) and left- (GpA_L) handed dimers.

To summarize, the comparison of geometrical and hydrophobic parameters of the predicted α-helical complexes allows the following conclusions to be drawn: (1) Despite minor differences in packing, the majority of structures with $\Theta < 0°$ reveal overall good agreement with NMR and mutagenesis data. (2) In models with the alternative fold ($\Theta > 0°$) the dimerization interface only partially corresponds to that in the GpA_{NMR} model, although these conformers satisfy reasonably well the NMR-derived geometrical restraints. (3) The hydrophobic organization of all types of complexes is quite similar and does not contradict the known principles of packing of TM helices. What drives helix association in the membrane? Analysis of various energy terms on different stages of MC search shows that the initial significant drop of the total energy is caused by the independent insertion of monomers into the membrane. As a result, total E_{solv} decreases by ∼70 kcal/mol. Then the energetically favorable intermonomer van der Waals contacts appear. In all complexes (including GpA_{NMR}) residues on the dimerization interface have low values of E_{vdW}. For example, distribution of E_{vdW} along the sequence in GpA_{MC} and GpA_{NMR} models is quite similar (not shown).

Therefore, the formation of TM dimers is mainly driven by two factors. At the first stage (insertion)—by interaction with membrane, while at the second one (helix association with subsequent "fine-tuning" of the complex)—by van der Waals contacts. Importance of the media effects lies in the fact that the hydrophobic peptides insert one by one into the membrane and adopt a TM orientation. Also, apolar medium considerably promotes their α-helical conformation. TM voltage favors a head-to-head disposition of α-helices, providing a suitable starting point for obtaining the most adequate ("nativelike") final solutions via MC search. This is because the number of possible variants of monomers' association is seriously limited as compared, for example, with two conformationally labile peptides in water or in a vacuum. Hence, at the initial stage the membrane plays a role of a peculiar matrix, which makes the system's components well-prepared for subsequent association. Then the main energy gain is achieved upon creation of favorable intermonomer contacts, although solvation effects still remain important because the optimal docking of helices occurs in cases when residues on the external surface of the dimer have energetically favorable interactions with the medium (see above). These observations agree well with the famous two-stage model of folding of TM domains in proteins.[41]

Several shortcomings are inherent in our method. They are mainly related to the limitations of the implicit membrane model. Thus, a number of important characteristics of biological membranes, like heterogeneity of dielectric properties, chemical composition, and microscopic nature of protein−lipid interactions (e.g., H-bonds), etc. are completely or partially omitted. Also, the influence of the protein on the lipid bilayer is neglected. Finally, the question about treatment of electrostatic interactions in membrane requires a special consideration (see ref 42 for recent review). Thus, during the last several years a number of implicit membrane models based on Generalized Born (GB) theory have appeared.[24,43] The models dealing with solving the Poisson−Boltzmann equation[44] and implementing the Gouy−Chapman term describing counterion-screened electrostatic interactions of a protein with anionic membranes[45] were reported as well. Earlier we also tried various schemes to treat electrostatics but did not get significant improvement of the results (in terms of their consistency with experimental data). Comparison of our results with those obtained with one of the recent GB models[24] shows that both models lead to quite similar conclusions (see above). Interestingly, Im et al.[24] ignored one Glu- on the N-terminus and two Arg+ on the C-terminus of GpA. They mentioned that this is related to the problem of "stabilization of the helical interface". In addition, in their study the start was chosen as two TM helices placed perpendicular to the membrane plane, while we used random starting configurations. One disadvantage of our simplified electrostatic screening model becomes apparent in studies of binding of positively charged proteins to membranes composed of anionic lipids. Thus, application of the Gouy−Chapman theory allowed better description of interactions of cardiotoxins from snake venom with negatively charged membranes,[45] although in zwitterionic bilayers and micelles (like here) the results are similar.

On the other hand, as for any theoretical model based on empirical parametrization (including ours, naturally), the only criterion of validity is its accord with experiments. Our membrane model was proved to be adequate for a large number of peptides and proteins with different fold (α-helical, *β*-structural) and mode of binding (TM and peripheral). These items were reviewed in ref 7. Being computationally efficient, the proposed technique permits exploration of a number of alternative scenarios that are too costly to be tested experimentally (e.g., comparative analysis of binding

for wild-type and mutant proteins). Finally, the method may provide good starting points for subsequent simulations in full-atom lipid bilayers and micelles. This computational approach is currently being tested on other TM α-helical complexes. We are also studying the sensitivity of the simulation results to the thickness and hydrophobicity degree of the membrane, as well as to some other factors (Efremov et al., manuscript in preparation). The approach will be refined in the future with the appearance of new experimental structural information concerning the nature of protein−protein interactions in membranes.

Abbreviations used: GpA, hydrophobic segment 69−97 of the human glycophorin A; MP, membrane protein; TM, transmembrane; MD, molecular dynamics; MC, Monte Carlo; NMR, nuclear magnetic resonance; DPC, dodecylphosphocholine; MHP, molecular hydrophobicity potential; RMSD, root-mean-square deviation; GpA_{NMR}, NMR-derived model of the GpA dimer; $GpA_{NMR-mem}$, the lowest-energy conformer found for GpA_{NMR} in membrane; GpA_{MC}, the calculated model most close to the NMR structure of the dimer; GpA_L, left-handed model of the GpA dimer.

Acknowledgment. This work was supported by the Program RAS MCB, by the Russian Foundation for Basic Research (Grant 04-04-48875-a), by the Russian Ministry of Education and Science (Lead 05, Living systems: Integrated project 3/001). R.G.E. is grateful to the Russian Science Support Foundation for the grant awarded. Access to computational facilities of the Joint Supercomputer Center (Moscow) is gratefully acknowledged.

Supporting Information Available: Distance restraints used in MC simulations of the GpA_{NMR} model in implicit membrane and interatomic distances in models calculated via MC simulations in implicit membrane and in models obtained by NMR spectroscopy (Tables 1−3). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Wallin, E.; von Heijne, G. *Protein Sci.* **1998**, *7*, 1029−1038.

(2) Arkin, I. T. *Biochim. Biophys. Acta* **2002**, *1565*, 347−363.

(3) Harder, T. *Philos. Trans. R. Soc. London B. Biol. Sci.* **2003**, *358*, 863−868.

(4) Ubarretxena-Belandia, I.; Engelman, D. M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 370−376.

(5) Forrest, L. R.; Sansom, M. S. *Curr. Opin. Struct. Biol.* **2000**, *10*, 174−181.

(6) Liang, J. *Curr. Opin. Chem. Biol.* **2002**, *6*, 878−884.

(7) Efremov, R. G.; Nolde, D. E.; Konshina, A. G.; Syrtcev, N. P.; Arseniev, A. S. *Curr. Med. Chem.* **2004**, *11*, 2421−2442.

(8) Eilers, M.; Patel, A. B.; Liu, W.; Smith, S. O. *Biophys. J.* **2002**, *82*, 2720−2736.

(9) DeGrado, W. F.; Gratkowski, H.; Lear, J. D. *Protein Sci.* **2003**, *12*, 647−665.

(10) Furthmayr, H.; Marchesi, V. T. *Biochemistry* **1976**, *15*, 1137−1144.

(11) Lemmon, M. A.; Flangan, J. M.; Treutlein, H. R.; Zhang, J.; Engelman, D. M. *Biochemistry* **1992**, *31*, 12719−12725.

(12) Treutlein, H. R.; Lemmon, M. A.; Engelman, D. M.; Brünger, A. T. *Biochemistry* **1992**, *31*, 12726−12733.

(13) Adams, P. D.; Engelman, D. M.; Brünger, A. T. *Proteins: Struct., Funct., Genet.* **1996**, *26*, 257−261.

(14) MacKenzie, K. R.; Prestegard, J. H.; Engelman, D. M. *Science* **1997**, *276*, 131−133.

(15) Smith, S. O.; Song, D.; Shekar, S.; Groesbeek, M.; Ziliox, M.; Aimoto, S. *Biochemistry* **2001**, *40*, 6553−6558.

(16) Orekhov, V. Y.; Abdulaeva, G. V.; Musina, L. Y.; Arseniev, A. S. *Eur. J. Biochem.* **1992**, *210*, 223−229.

(17) Gratkowski, H.; Dai, Q.; Wand, A. J.; DeGrado, W. F.; Lear, J. D. *Biophys. J.* **2002**, *83*, 1613−1619.

(18) Stockner, T.; Ash, W. L.; MacCallum, J. L.; Tieleman, D. P. *Biophys. J.* **2004**, *87*, 1650−1656.

(19) Fisher, L. E.; Engelman, D. M.; Sturgis, J. N. *Biophys J.* **2003**, *85*, 3097−3105.

(20) Pappu, R. V.; Marshall, G. R.; Ponder, J. W. *Nat. Struct. Biol.* **1999**, *6*, 50−55.

(21) Smith, S. O.; Smith, C.; Shekar, S.; Peersen, O.; Ziliox, M.; Aimoto, S. *Biochemistry* **2002**, *41*, 9321−9332.

(22) Dobbs, H.; Orlandini, E.; Bonaccini, R.; Seno, F. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 342−349.

(23) Ducarme, P.; Thomas, A.; Brasseur, R. *Biochim. Biophys. Acta* **2000**, *1509*, 148−154.

(24) Im, W.; Feig, M.; Brooks, C. L., III *Biophys. J.* **2003**, *85*, 2900−2918.

(25) Efremov, R. G.; Volynsky, P. E.; Nolde, D. E.; Arseniev, A. S. *Theor. Chem. Acc.* **2001**, *106*, 48−54.

(26) von Freyberg, B.; Braun, W. *J. Comput.Chem.* **1991**, *12*, 1065−1076.

(27) Berman, H. M.; Bhat, T. N.; Bourne, P. E.; Feng, Z.; Gilliland, G.; Weissig, H.; Westbrook, J. *Nat. Struct. Biol.* **2000**, *7*, 957−959.

(28) Efremov, R. G.; Nolde, D. E.; Vergoten, G.; Arseniev, A. S. *Biophys. J.* **1999**, *76*, 2448−2459.

(29) Némethy, G.; Pottle, M. S.; Scheraga H. A. *J. Phys. Chem.* **1983**, *87*, 1883−1887.

(30) Roux, B. *Biophys. J.* **1997**, *73*, 2980−2989.

(31) Efremov, R. G.; Volynsky, P. E.; Nolde, D. E.; van Dalen, A.; de Kruijff, B.; Arseniev, A. S. *FEBS Lett.* **2002**, *526*, 97−100.

(32) Efremov, R. G.; Vergoten, G. *J. Phys. Chem.* **1995**, *99*, 10658−10666.

(33) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

(34) Bowie, J. U. *Nat. Struct. Biol.* **1997**, *4*, 915−917.

(35) Kairys, V.; Gilson, M. K.; Luy, B. *Eur. J. Biochem.* **2004**, *271*, 2086−2092.

(36) Fleishman, S. J.; Schlessinger, J.; Ben-Tal, N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15937−15940.

(37) Rees, D. C.; DeAntonio, L.; Eisenberg, D. *Science* **1989**, *245*, 510−513.

(38) Efremov, R. G.; Vergoten, G. *J. Prot. Chem.* **1996**, *15*, 63−76.

(39) Popot, J. L.; Engelman, D. M. *Annu. Rev. Biochem.* **2000**, *69*, 881−922.

(40) Efremov, R. G.; Vergoten, G.; Arseniev, A. S. *Theor. Chem. Acc.* **1999**, *101*, 73−76.

(41) Popot, J. L.; Engelman, D. M. *Biochemistry* **1990**, *29*, 4031−4037.

(42) Tobias, D. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 253−261.

(43) Spassov, V. Z.; Yan, L.; Szalma, S. *J. Phys. Chem. B* **2002**, *106*, 8726−8738.

(44) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244−1253.

(45) Lazaridis, T. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 518−527.

CT0501250

JCTC Journal of Chemical Theory and Computation

Quantum Mechanics/Molecular Mechanics Calculations of the Vanadium Dependent Chloroperoxidase

Joslyn Yudenfreund Kravitz,[†] Vincent L. Pecoraro,[†] and Heather A. Carlson*[,†,‡]

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055, and Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, Michigan 48109-1065

Received May 13, 2005

Abstract: Large quantum mechanics/molecular mechanics (QM/MM) calculations are used to probe the resting and initial protonated states of the vanadium dependent chloroperoxidase from the pathogenic fungus *Curvularia inaequalis.* QSite was used to model 433 residues and 24 structural waters with molecular mechanics, while 8 active-site residues and the vanadate cofactor (161 atoms) were represented at the B3LYP/lacvp* level of theory. Our previous study of small model systems implied that the resting state of the enzyme contains a trigonal bipyramidal vanadate with one hydroxyl group in the equatorial plane and another in the axial position. This study uses a much larger model of the biological system at a higher level of theory to identify the location of the equatorial hydroxo group with respect to the enzyme active site. We also identify a second resting-state configuration with an axial water and three equatorial oxo moieties that is nearly isoenergetic with the previously identified state. We propose that the resting state is a hybrid of these two configurations, stabilized by the long-range electrostatic field of the protein environment. The first step in catalysis is believed to be protonation of the vanadate. Our previous small models indicated that there were two protonated configurations, but this study shows that the configuration containing an axial water and one hydroxo group in the equatorial plane is significantly lower in energy than any other configuration. Additionally, we can now assign an important role for lysine 353 in the catalytic cycle. Based on our calculations and other model studies, we provide an updated catalytic cycle for vanadium dependent haloperoxidase activity. Further, we demonstrate the importance of system set up. In particular, maintaining the proper electrostatic field at the active site is crucial for identifying the correct minima in a truncated protein model.

Introduction

Recent computational studies have resulted in new proposals for the structure[1] and catalytic reactivity[2] of the active site of the vanadium dependent haloperoxidases. These unusual enzymes catalyze the two electron conversion of halide ions to the corresponding hypohalous acids using hydrogen peroxide as an oxidant:[3]

$$H_2O_2 + H^+ + X^- \rightarrow H_2O + \text{“HOX”} \qquad (1)$$

Chloride, bromide, or iodide can be used as substrate for a vanadium dependent chloroperoxidase (VCPO), but only bromide and iodide can be oxidized by vanadium dependent bromoperoxidases (VBPO). The identity of “HOX” is dependent upon both the pH of the reaction and the halide involved. Studies indicate that VCPO oxidizes a chloride ion

* Corresponding author e-mail: carlsonh@umich.edu.
[†] Department of Chemistry.
[‡] Department of Medicinal Chemistry, College of Pharmacy.

10.1021/ct050132o CCC: $30.25
Published on Web 09/20/2005

to HOCl,[4] whereas the products of bromide oxidation are likely a thermodynamic distribution of Br_3^-, HOBr, and Br_2.[5]

Unlike heme-based peroxidases, which require the formation of highly oxidized intermediates, the vanadium does not undergo redox cycling during catalysis. It is thus believed that the vanadium ion plays the role of a strong Lewis acid which activates the peroxide.[6] Kinetic studies indicated that protonation of the bound peroxo group is a crucial step in the heterolytic cleavage of the O−O bond. These observations formed the basis for a proposal that the protonated oxygen is then transferred to the halide according to an oxo-transfer mechanism.[7] Subsequent computational studies now suggest that the nonprotonated peroxo oxygen is the atom transferred to the substrate.[2]

Upon oxidation, the halide can be added to an organic substrate if one is present.[8] In some cases, this addition is stereoselective.[9] This reaction is thought to be the origin of many halogenated species in the environment.[10] In fact, Butler and co-workers recently carried out the first experiments which established the role of these enzymes in the biosynthesis of brominated metabolites from marine red algae.[11] At high pH and in the absence of organic substrate, the oxidized halogen can react with a second equivalent of peroxide to produce singlet oxygen.[12] Additionally, VBPOs can oxidize thioethers to the corresponding sulfoxide[13] using a mechanism similar to that of halide oxidation.[14]

Peroxo−vanadium complexes are good functional models of the VCPO and have been shown to oxidize organic[15] and inorganic compounds,[16] including alcohols[17] and sulfides.[14] They are also capable of hydroxylating hydrocarbons[18] and epoxidizing alkenes.[19] However, attempts to synthesize structural models of the VCPO active site have not been as successful, and a synthetic complex reproducing the exact coordination environment of vanadium in VCPO has not yet been obtained. It has been proposed that such a complex has not been isolated because a five-coordinate imidazole complex is unstable unless sequestered in a protein active site.[20]

The only crystal structure of a VCPO that has been solved is the enzyme from the fungus *Curvularia inaequalis*.[21] In the native state, the vanadate cofactor (VO_4^{3-}) is bound to the protein through a single coordinate covalent bond from the V to the N_ϵ of His496 (Figure 1). The high negative charge of the cofactor appears to be offset by a number of protonated amino acids in the active site which donate hydrogen bonds to the oxygen atoms of the cofactor (Lys353, Arg360, and Arg490). Additional hydrogen bonds are donated to vanadate's equatorial oxygens by the side chain of Ser402 and the backbone amide of Gly403. His404 may participate in a hydrogen bond to or from an axial hydroxo group.

Mutagenesis studies[22] in which each of the three basic active-site residues and His496 were replaced with alanine indicated that His496 and Lys353 are the two most important residues for activity. Replacement of Arg490 or Arg360 leads to a lesser, but still significant, reduction in activity. Additional mutants which eliminate almost all activity are His404, which is involved in a hydrogen bond with the axial oxygen moiety, and Asp292, which helps orients Arg490 through a strong salt bridge.[23]

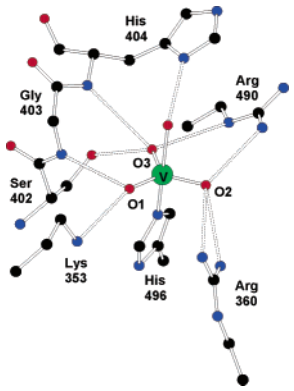


Figure 1. Active site configuration observed in the crystal structure of the VCPO from *C. inaequalis*.[1] The hydrogen-bonding interactions shown above are implied from distances between the heavy atoms. The equatorial oxygens of the vanadate are labeled with the notation used throughout our discussions.

The crystal structure of this enzyme was interpreted to contain a vanadate unit in a trigonal bipyramidal structure with a hydroxide and His496 in the axial positions and three oxo moieties in the equatorial plane. However, due to inherent limitations in resolution, the crystallographic study did not reveal hydrogen positions, and the distinction between oxide and hydroxide moieties was inferred by slight differences in ligand−metal bond distances which were less than the uncertainty in the atomic positions.

Computational studies of this enzyme complement and correlate well to available experimental techniques. Michael Bühl and co-workers have carried out an extensive series of DFT calculations in which ^{51}V NMR chemical shift values were calculated.[24−27] Valeria Conte and co-workers have carried out calculations for small models of each catalytic intermediate. These calculations, done on vanadate models of fewer than 10 atoms, provided some insight into the order and location of peroxide[28] and bromide binding[29] and reactivity.[30] More recent studies, carried out by De Gioia and co-workers,[2] involved slightly larger models of the active site, incorporating methylamine and imidazole moieties to represent Lys353 and His496. This study examined models of each catalytic intermediate in the cycle.

The conclusions from these studies agree well with our recent studies of small models to represent the active site of both the VCPOs and VBPOs. These calculations revealed that the resting state of the enzyme included an anionic, doubly protonated vanadate.[1] In this species, the axial position and one equatorial position were protonated, while the other two equatorial positions contained oxo moieties. This is slightly different than the proposed active site based on the crystal structure of the VCPO where it was suggested that three oxo groups were found in the equatorial plane.[31] Based on the atomic arrangement of the peroxide-bound crystal structure, it was proposed that the axial hydroxo group is protonated and then leaves as water.[21] This is consistent with our calculations which showed that the energy difference between protonation of either the axial hydroxo or one

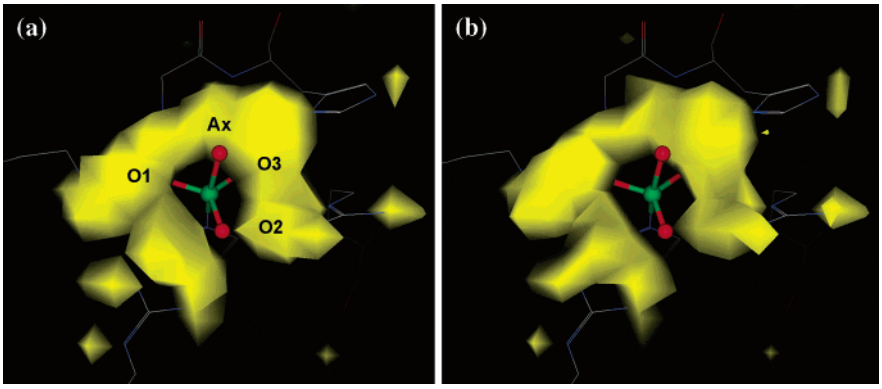


Figure 2. An electrostatic isosurface in the active site calculated for (a) the full chloroperoxidase and (b) the truncated protein model. The isosurfaces pictured here represent only the protein's contribution to the electric field (vanadate not included in the calculation). The isosurface is shown at 8 kT for the full protein and 9 kT for the truncated protein.

equatorial oxo moieties was negligible, indicating that either could be the site of the protonation that begins the catalytic cycle. From our previous model calculations, it is clear that the structure and stability of the vanadate cofactor is highly dependent upon the hydrogen-bonding partners available to the various oxygen ligands. To characterize the active site of the enzyme fully, a larger system with a complete active site is needed.

This paper presents a hybrid quantum mechanics/molecular mechanics (QM/MM) study of the resting and first protonation states of the VCPO from the fungus *C. inaequalis*. This computational method treats part of the protein quantum mechanically (the vanadate cofactor and selected surrounding residues), while the remainder of the protein is treated using molecular mechanics. In this way, a high level electronic calculation of the active site can be carried out in the presence of the actual protein environment. In our application of this powerful new technique, we were able to use a quantum mechanical region of 161 atoms (with 42 additional interface atoms), making this one of the largest, high-accuracy QM/MM calculations to date. Our application also highlights the need for careful attention to long-range, electrostatic interactions provided by the protein environment.

Computational Methods

Friesner and co-workers have developed QSite,[32−34] a QM/MM program well suited to study metalloenzymes. They have shown that the binding energy of O_2 to hemerythrin[35] and the activation barrier in cytochrome P450cam hydrogen abstraction[36] could be calculated quite accurately using this program. Unless noted, all of our calculations were carried using QSite version 2.5, revision 20.[37] QM/MM calculations were completed with the default parameters in QSite, including the OPLS 1999 force field[38] and a distance-dependent dielectric constant. We used the B3LYP[39,40] functional with the lacvp* basis set which uses the Los Alamos ECP including the outermost core orbitals[41] for vanadium and 6-31G*[42−47] for all other atoms. In all calculations, backbone atoms were held fixed. Any residues more than 20 Å from the vanadium were also held fixed (including protons).

Protein Preparation. Using the program Maestro, the atoms of the side chains of six residues that were unresolved in the crystal structure (PDB code 1VNI)[21] were added. Prolines 361 and 401 were converted to alanines because of software limitations in the QM/MM interface region (backbone atoms were constrained to the positions in the crystal structure, so this change is minimal). Protons were added and oriented to maximize the hydrogen-bonding network within the protein using the program MOE.[48]

While QSite allows a maximum of 8000 atoms and bonds in the system, the chloroperoxidase is comprised of over 9000 atoms. To deal with the size limitation, the protein had to be truncated by removing sections that were distant from the vanadate cofactor. We cannot overemphasize how important it is to maintain the electrostatic field in the active site when creating a truncated model. It was important to choose distant residues that would have less effect on the electrostatics at the active site. The chloroperoxidase has an overall charge of −26, and careful choices were made to minimize any disruption of the electric field at the active site. Thus, the electric field of the entire protein and that of several potential truncation models were calculated using the Poisson−Boltzmann routine in the program MOE. The field of the protein alone was calculated to determine the environment with which the vanadate cofactor will interact. It was necessary to identify which truncated model best reproduced the field of the whole protein. The parameters used in the calculation were as follows: the dielectric constant of the interior of the protein was set to 20; the dielectric constant for the exterior of the protein was set to 80; offset was 2 Å; the counterion and solvent radii were 2.5 and 1.4 Å, respectively; the grid spacing was 1 Å and the grid extended 20 Å beyond the protein; the salt (NaCl) and solute (protein) concentrations were 0.15 M and 0.001 M, respectively. The appropriate truncated system was chosen based on its ability to most closely reproduce the field of the whole protein at the active site (Figure 2). Though it was not possible to maintain the magnitude of the field when eliminating so many charges, it was possible to maintain the topography of the field. This produces the same gradients, so the relative effect on the wave functions will be the same. Protonation biases will also be as similar as possible.

The truncation model which yielded the best result was based on an approximate 27-Å cutoff from the vanadate. To maintain the electrostatic characteristic of the protein envi-

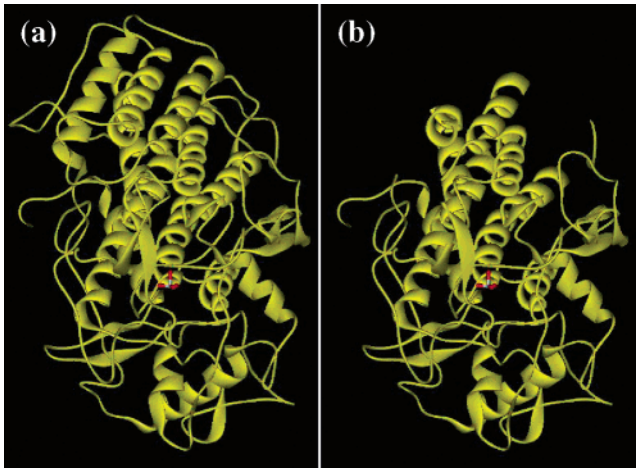


Figure 3. (a) The complete chloroperoxidase and (b) the truncated model used in this study. Vanadate is shown in ball-and-stick representation, roughly at the center of both structures.

ronment, the charges of the residues at the cutoff distance were considered when choosing which residues to remove. For instance, neutral residues were cut preferentially over charged ones. In cases where a charged residue near the cutoff was part of a salt bridge, both residues were removed. At a distance greater than 25 Å, a salt bridge has an electrostatic effect similar to a neutral dipole. All chains that were cut were capped with *N*-methylamide (NME) or acetyl (ACE) residues as appropriate (Figure 3). A list of the residues which were included in the truncated model of the protein is available in the Supporting Information. Surface water molecules were removed, but structural waters in the interior were retained. This procedure resulted in a careful pruning of the protein to contain 7018 atoms with an overall charge of −14 and an appropriate electric field gradient in the active site.

The positions of all protons were minimized using Impact version 2.5 revision 20.[49]

Definition of Quantum Mechanics Region. As in our previous study, our investigation proceeded by building up the QM region in a systematic fashion starting from a dianionic vanadate. In all calculations, the vanadate group and the side chain of His496 were included in the QM region. The initial calculations used a QM region that also included Lys353, Arg360, and Arg490. This QM region had a total charge of +1 and contained 113 lacvp* atoms with 28 additional interfacial atoms (four residue calculations). A second set of calculations included the same four residues plus the side chains of Asp292 and His404 in the QM region (six residue calculations). The QM region was net neutral in charge and contained 138 lacvp* atoms with 36 interface atoms. A final set of calculations for the anionic resting state of the protein was carried out in which a proton was added to the system in all possible locations (full resting state calculations). These calculations also included the backbone of His404 and all of Gly403 in the QM region. The first protonation step in the catalytic cycle was generated by again adding a proton to all appropriate locations in the active site (first protonation state calculations). The full resting state and first protonation state calculations contained 160 and 161 lacvp*atoms in the QM region, respectively, with 42 interfacial atoms (a total of 202 and 203 atoms in the largest calculations). The QM region had a net charge of +1 for the full resting state and +2 for the first protonation state (Figure 4).

In the four residue calculations, the heavy atoms of all QM side chains were frozen, but the vanadate atoms and all QM protons were allowed to move. For all other calculations, vanadate and all heavy atoms of the side chains in the QM region were allowed to move, except His496. All protons in the QM region were allowed to move, except 1HZ and 3HZ of Lys353, 2HH1 of Arg360, and HE2 of His404. The positions of these four protons had to be restrained due to interactions at the QM/MM interface. For Lys353 and His496, only the side chain was included in the QM region. The arginines required inclusion of the entire residue, backbone and side chain. Sample input files are available in the Supporting Information.

Results

Four Residue Calculations. Vanadate was modeled in the dianionic state with an axial hydroxo and three equatorial oxo moieties surrounded by three charged residues for these initial calculations. Vanadate, Lys353, Arg360, Arg490, and His496 were all included in the QM region (Figure 4a). The

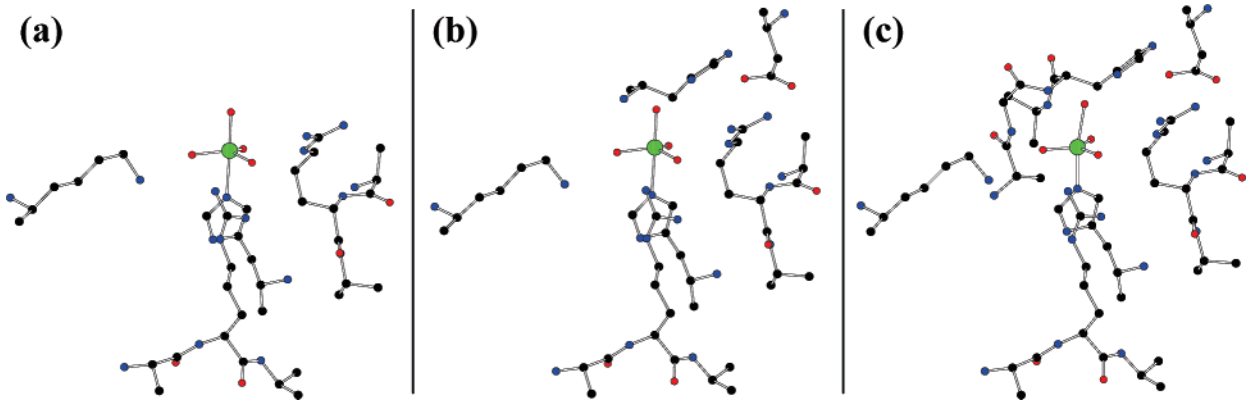


Figure 4. Residues in the QM region for (a) four residue, (b) six residue, and (c) full resting state and first protonated state calculations. For clarity, protons and the interfacial atoms have been omitted. It should be noted that a significant portion of the neighboring backbone atoms are required when including arginines in the QM region with QSite.

Table 1: For the Four Residue Calculations, the Energy Difference of $V(OH)O_3$ Surrounded by Three Cationic Neighbors versus the Transfer of a Proton from the Noted Residue to the Nearest Equatorial Oxygen on Vanadate[d]

deprotonated side chain	ΔE (kcal/mol)
3 cationic residues + $V(OH)O_3$	0
Lys353[a]	−2.8
Arg360[b]	5.3
Arg490η[b]	8.4
Arg490ϵ[c]	14.9

[a] Proton transferred to O1. [b] Proton transferred to O2. [c] Proton transferred to O3. [d] Negative values indicate that the proton transfer is favorable.

goal was to determine the protonation state of vanadate and the hydrogen-bonding scaffold. Our previous calculations of model systems indicated that the equatorial oxygens can abstract protons from neighboring cationic residues. Each of the three charged residues could donate a proton to the vanadate. There are two possible positions from which Arg490 could donate a proton: the terminal $\eta 2$ nitrogen (denoted 490η) or the ϵ nitrogen (denoted 490ϵ). Table 1 presents a series of calculations to examine proton placement. For each residue, we calculated the energy for the charged residue + $V(OH)O_3$ versus transferring the proton to an equatorial oxygen to produce a neutral residue + $V(OH)_2O_2$. Only the transfer of a proton from Lys353 to vanadate lowered the energy of the system. Combinations of transferring a second proton, in addition to the one from Lys353 to O1, were also examined. Upon minimization, a second proton on O2 or O3 simply transferred back to the appropriate arginine nitrogen.

Six Residue Calculations. As a second series of calculations, all four residues from the previous calculations, plus the side chains of Asp292 and His404, were included in the QM region (Figure 4b). This QM region had an overall neutral charge. The protonation states calculated in the four residue system were minimized in this larger, more complete system. The calculations were initiated with partial minimization where the heavy atoms of the side chains were held fixed, and then, the partial minima were refined with a second minimization where the side-chain heavy atoms were also unrestrained. At the point of freeing the side chains, it became necessary to constrain the positions of certain atoms at the QM/MM interface and His496. The ring atoms of His496 were constrained to the crystal structure positions, and the terminal protons of Lys353, proton 2HH1 from Arg360, and proton HE2 from His404 were constrained to remain fixed in positions taken from the minimum structure in which a proton from Lys353 has been transferred to the vanadate. Only two lysine protons were frozen, even in cases where there were three protons on the lysine, because all protons which interact directly with the vanadate were allowed to move freely in all minimizations.

In this series of calculations, transferring a proton from Lys353 to the vanadate is even more favorable by well over 7 kcal/mol. All other structures in which one or two protons are donated to vanadate are over 9 kcal/mol higher in energy than the global minimum. Contrary to the four residue calculations, structures where O2 or O3 are protonated, in addition to O1 being protonated by Lys353, were stable minima but high in energy (Table 2).

Table 2: For the Six Residue Calculations, Energy Difference of Transferring Protons from the Three Cationic Residues to Vanadate's Equatorial Oxygens

deprotonated side chain	ΔE (kcal/mol)
3 cationic residues + $V(OH)O_3$	0
Lys353[a]	−7.7
Arg360[b]	6.0
Arg490η[b]	2.8
Arg490ϵ[c]	1.9
Lys353 + Arg360[d]	14.1
Lys353 +Arg490η[d]	5.0
Lys353 + Arg490ϵ[e]	6.2

[a] Proton transferred to O1. [b] Proton transferred to O2. [c] Proton transferred to O3. [d] Protons transferred to O1 and O2. [e] Protons transferred to O1 and O3.

Full Resting State Calculations. In the calculations presented above, we have established that the dianionic form of the vanadate is not preferred in the active site. Instead, the anionic state is formed preferentially by abstracting a proton from Lys353. However, the small model calculations indicated that the residues in the active site were fully protonated in the resting state. Therefore, the next set of calculations included an additional proton in the active site, making the overall charge of the active site +1 (+1 from each Arg and Lys and −1 each from Asp and the anionic, doubly protonated vanadate). For completeness, we also examined structures in which the vanadate was singly protonated and the added proton was located on His404. To include His404 properly in the calculation, the entirety of His404 and Gly403 were required in the QM region (Figure 4c). Since the structure in which Lys353 donated a proton to vanadate was the previous global minimum, the coordinates for that structure served as the basis for this set of calculations, with the positions of the protons in the active site altered as needed. Four calculations were carried out in which Lys353 was in the free base form and O1 was protonated. In these calculations, the additional proton was added to O2, O3, the axial hydroxide, or His404. Five calculations were carried out based on the minimum with Lys353 protonated and the vanadate in the dianionic form. In these calculations, the additional proton was added to O1, O2, O3, the axial hydroxide, or His404.

Figure 5 shows the two lowest-energy structures in which Lys353 and O2 were protonated (Table 3b) or Lys353 was protonated and the axial position contained a water moiety (Table 3d). The difference in energy between these two structures was only a third of a kcal/mol, and both were ∼6 kcal/mol lower in energy than the next most favorable structure (Table 3c). The structures in which O1 and either Lys353 or His404 were protonated were the highest-energy structures (Table 3 (parts a and g, respectively)). In the first case, the proton on O1 was oriented away from Lys353 to minimize the repulsion between the two protons. In the second case, the proton of the axial hydroxo group was oriented away from His404 for the same reason. Adding a proton to O2 or O3 while O1 was protonated and Lys353 was in the free-base form resulted in the transfer of the O1

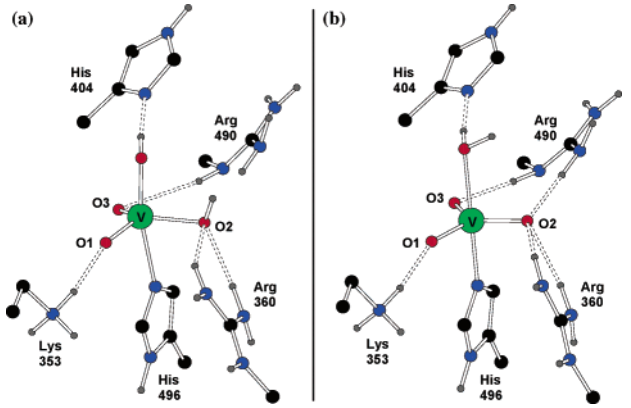


Figure 5. The two lowest-energy minima from the full resting state calculations. For clarity, many of the atoms in the QM region are not shown. (a) Minimum with axial and equatorial hydroxo groups and two equatorial oxo groups. (b) Minimum with an axial water and three equatorial oxo groups. In (a), a hydrogen bond exists from Arg490 to O2 as it does in (b), but it is not drawn in (a) so that the reader can better understand that the hydrogen on O2 points up toward the axial hydroxo. In (a), the axial O−V−O2−H atoms are nearly coplanar as are the equivalent O2−V−axial O−H atoms in (b).

Table 3: Relative Energies for Each Stable Minimum from the Full Resting State Calculations

structures with Lys353 protonated	ΔE (kcal/mol)
(a) O1 protonated	26.4
(b) O2 protonated	0
(c) O3 protonated	6.0
(d) axial water	0.3
(e) His404 protonated	18.1
structures with Lys353 deprotonated	**ΔE (kcal/mol)**
(f) O1 protonated and axial water	22.6
(g) O1 and His404 protonated	31.8

proton to Lys353 to produce structures similar to those in Table 3b,c. Protonation of the axial position does not result in transfer of the O1 proton to Lys353, but the structure is relatively high in energy (Table 3f). Schematic pictures and QM coordinates of each minimum listed in Table 3 may be found in the Supporting Information.

First Protonation State Calculations. The first step in the catalytic cycle is the protonation of one vanadate oxygen group.[7] It is not known which site is protonated in this first step. To determine this, we protonated each of the two minimum energy structures of the full resting state (Figure 5) in every possible location. These calculations were identical to the previous set of calculations except that the charge of the QM region was increased to +2 by the addition of the proton (Figure 4c).

The global minimum was the structure with an axial water and a protonated O3 (Table 4c and Figure 6). The next lowest-energy structure was 7 kcal/mol higher in energy (Table 4h), and O2 and His404 were protonated. All other structures are higher in energy by 20 kcal/mol or more (Table 4). Schematic pictures and QM coordinates of each minimum listed in Table 4 may be found in the Supporting Information.

Table 4: Relative Energy (in kcal/mol) for Each Stable Minimum from the First Protonation State Calculations

structures with axial water	ΔE vs axial water and O3 protonated
(a) O1 protonated	32.9
(b) O2 protonated	26.9
(c) O3 protonated	0
(d) His404 protonated	19.7
structures with axial hydroxide	**ΔE vs axial water and O3 protonated**
(e) O1 and O2 protonated	41.0
(f) O2 doubly protonated	44.0
(g) O2 and O3 protonated	37.9
(h) O2 and His404 protonated	7.1

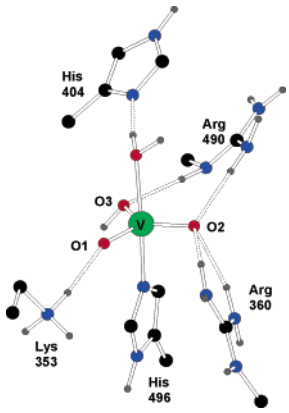


Figure 6. QM region of the global minimum from the first protonated state calculations. For clarity, not all the atoms in the QM region are shown.

Discussion

Enzymology,[11,50−52] synthetic model chemistry,[7,53−57] and small model calculations[1,2,25,58] have provided a wealth of information on the catalytic activity of the VCPOs. This study is the first application of QM/MM calculations to VCPOs which allows us to add additional atomic details to the understanding of the influence of the protein on the catalysis. We used this technique to describe more accurately the protonation states and interactions of the vanadate and surrounding residues of the active site.

The system setup was very carefully done to ensure that the electrostatic influence of the whole protein was unaltered in the truncated model. The truncated model provides the same topography of the potential surface, ensuring that wave functions would be influenced in nearly the same manner in the full and truncated systems. It should be stressed that similar comparisons should be done in any QM/MM calculation when the MM region must be truncated.

The QM region was built up in a systematic fashion to examine as many protonation states as possible, yet focus the most expensive calculations on the most likely states. The six residue calculations clearly showed that Lys353 is the only possible proton donor in the active site. This is logical because lysines have lower pK_as than arginines. The only other possibility was a protonated His404 playing the role of an acid, but the largest calculations all

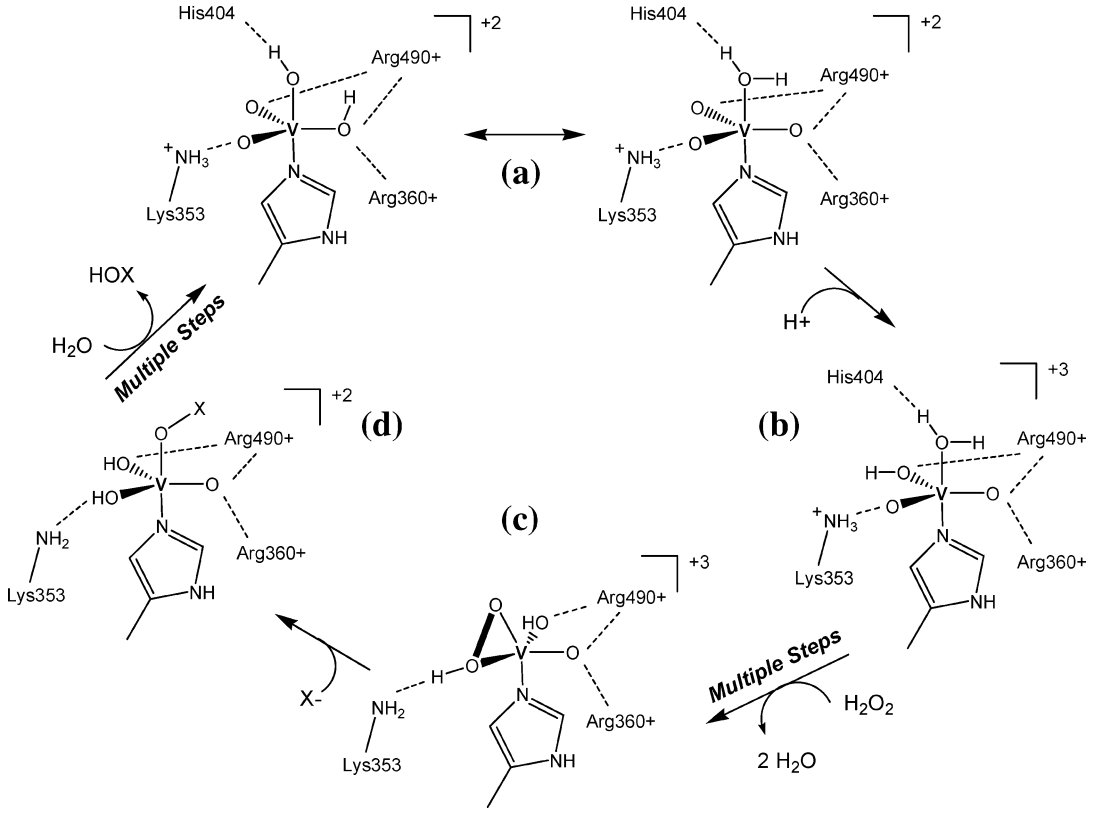


Figure 7. Proposed catalytic cycle based on the current and previous calculations. Panels (a) and (b) represent the minima calculated in this QM/MM study. Panels (c) and (d) represent minima calculated in other studies[2] combined with insights into the active site from our QM/MM calculations. Panel (a) is the hybrid of the two resting state configurations. Panel (b) is the first protonated state. Panel (c) represents the hydroperoxide-bound structure. Panel (d) represents the state in which the halide has been oxidized and is still bound to the vanadate.

show that His404 is not protonated in the resting or first protonated state.

Previously published small model calculations[1,2] suggest that the equatorial plane of the anionic vanadate should be surrounded by active-site residues which are all fully protonated. An additional proton was added to the calculation to provide the full resting state. Models with Lys353 in both the protonated and deprotonated states were examined. This resulted in nine calculations with the same large QM region in which we protonated the vanadate at each of the oxygens, including the axial position, and also at His404. In the systems with Lys353 neutral and O1 protonated, only two systems produced stable minima, but they were very high in energy. (Adding the proton to either O2 or O3 resulted in the O1 proton transferring back to Lys353.) We found that the protonation of Lys353 was overwhelmingly preferred over the double protonation of the equatorial plane of the vanadate. The results also show that when Lys353 is fully protonated, O1 is no longer the preferred position for the equatorial hydroxo group as might be expected. The most favorable equatorial position for protonation is at O2, which accepts hydrogen bonds from both arginines. The added proton on O2 is oriented up toward the axial hydroxo (Figure 5a). It is also highly favorable to protonate the axial position, resulting in an axial water. The added proton to create the axial water is oriented over O2 (Figure 5b). Both minima in Figure 5 show that the four atoms (O2−V−axial O−added proton) lay in a near-planar arrangement. When the two structures are overlaid, the positions of the added proton on the axial water vs O2 are only 2.2 Å apart. The proton can most likely tunnel between these two nearby locations. Because the two structures are nearly isoenergetic and easily interconverted, we propose a hybrid resting state, shown in Figure 7a.

Our previous small model calculations indicated that the sole structure of the resting state included two hydroxo groups, one in the axial position and one in the equatorial plane, likely at O1.[1,2] The more complete systems presented here refine the previous proposal on two accounts. First, this study shows that the resting state is comprised of not one but two structures. Second, they indicate that it is not the O1 position which is protonated but rather the O2 position. This may be surprising; after all, O2 is complemented by two arginines, while O1 and O3 only interact with one charged residue each. The electrostatic field of two Arg side chains should make O2 the least favorable site to protonate, but the reader is referred to Figure 2. When the full electrostatic field of the entire protein is calculated, one can see that the positive potential is actually weakest in the region of O2 and the axial hydroxo. It is reasonable for the proton to occupy the gap in the field between O2 and the axial position.

The strongest positive field is located at O1, and there is a very strong hydrogen bond between Lys353 and O1. Mutagenesis studies have shown that Lys353 is more important for catalysis than the arginines in the active site.[22]

The tight hydrogen bond from Lys353 to vanadate, coupled with the information from the mutagenesis studies, prompted us to propose an interesting catalytic role for Lys353 (more detail below).

Since the small molecule experiments of Colpas et al. showed that the first step in catalysis is the protonation of the vanadate,[59] we examined the first protonation state of the enzyme as well. Our previous small model calculations indicated that there were two structures in equilibrium in the protonated state. One included three hydroxo groups: one in the axial position and two in the equatorial plane. Our current, more complete calculations show that such a configuration is a stable minimum but much too high in energy to be relevant. The second minimum in the small model calculations contained a water moiety in the axial position and one hydroxo group in the equatorial plane. This QM/MM study shows that the later minimum is the lowest in energy, and O3 is the favorable position for the added proton in the first protonated state. Again, the long-range electrostatics can explain why O3 is the most favorable site for protonation. The position between O2 and the axial hydroxo is already occupied, and the highest positive potential in the active site is located at O1. When protonation occurs, O3 is the most favorable site to accept the proton. Upon protonation of O3, the hybrid form of the resting state appears to become biased toward the axial water configuration.

While this study has only quantified the resting state and first protonation state of the enzyme, it has laid the groundwork for the confirmation of the other intermediates proposed in this catalytic cycle, and we can propose a modified catalytic cycle (Figure 7). The calculations of the full resting state imply that it is a hybrid of the two lowest-energy minima (Figure 7a). In the first protonation state, O3 is protonated, and the hydrogen tunneling between the axial and O2 positions is localized on the axial position, creating a water molecule (Figure 7b).

In the next steps of the catalytic cycle, it is known that the incoming hydrogen peroxide displaces the axial water and one equatorial oxygen moiety (Figure 7c). The lowest unoccupied molecular orbital (LUMO) of the first protonation state shows that any electron density contributed to this orbital by the attacking peroxide would weaken and lengthen the bonds between the vanadium and the equatorial oxo groups (Figure 8). A longer, weaker oxo bond would be more basic, making it very likely that O1 abstracts a proton from Lys353. We propose that O1 is the second oxygen moiety which is displaced (quite possibly leaving as a water molecule after a second protonation by the hydrogen peroxide). Previous computational studies of model systems indicate that the equatorial hydroperoxide oxygen is one which is protonated.[2] This group would favorably donate a hydrogen bond to the resulting free-base form of Lys353 (as shown in Figure 7c). Furthermore, both computational studies and the crystal structure of the peroxide-bound form[21] indicate a strong hydrogen bond between Lys353 and the equatorially bound peroxide oxygen.

The halide can then attack the pseudoaxial hydroperoxide oxygen (Figure 7d), as indicated by the study by Zampella et al.[2] A protein-bound water molecule or the O3 equatorial hydroxide could then protonate the axial OX^-, making it a better leaving group. The HOX species is displaced. Transfer of a proton from the former equatorial hydroperoxide oxygen back to Lys353 reforms the resting state (Figure 7a).

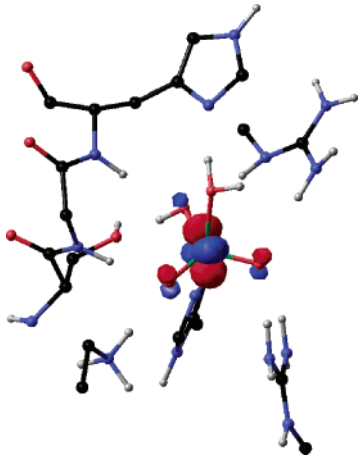


Figure 8. The LUMO of the first protonation state is shown with a cutoff of −0.100. The LUMO is made up of an antibonding orbital between vanadium and primarily the two oxo groups, with minor contribution from the hydroxo group. We found that the HOMO is simply the lone pair on His404 that accepts a hydrogen bond from the axial water (data not shown).

The crystal structure of the peroxide bound form[21] was interpreted as having the hydroxo group located at O2, not O3 as we are proposing. The crystal structure was not of high enough resolution to identify positions of hydrogens, and the uncertainty in atomic positions makes it very speculative to differentiate between an oxo and hydroxo bond length. Questions about the catalytic steps and protonation states of the next intermediates could be answered by carrying out QM/MM calculations of the later stages in the catalytic cycle. There are multiple steps between parts b and c of Figure 7 as well as between parts c and d of Figure 7. Computational studies are one of a few ways to gain insights into such fleeting and reactive states.

Conclusions

We have used QM/MM calculations to refine the resting state and first protonation state of the vanadium dependent chloroperoxidase. We have shown how the LUMO of the first protonated state and the strong hydrogen bond between Lys353 and O1 support a modification of the catalytic cycle.

Our previous small model calculations predicted that an anionic vanadate unit with one hydroxo group in the equatorial position and a second in the equatorial plane would be most favorable. These calculations support that prediction and add new information by indicating O2 as the likely position of the equatorial hydroxo group in the resting state. The calculations also reveal that an anionic vanadate with an axial water, which was not stable in the small model calculations, is actually very stable in the protein environment and is nearly isoenergetic with the previously proposed resting state. Our current calculations suggest that the resting state is actually a hybrid of these two minima, with the

configuration containing an axial water being more important for generating the first protonated state. Our previous small model calculations indicated that there were two structures that were nearly isoenergetic upon protonation of the active site. However, these QM/MM calculations reveal only one low-energy minimum with an axial water and an equatorial hydroxo group at O3. The influence of long-range electrostatics in this system appears to significantly influence the most favorable positions for protonation. These findings highlight the importance of incorporating the protein environment into quantum mechanical models of protein active sites.

Acknowledgment. The authors appreciate the generosity of Prof. William L. Jorgensen and Dr. D. C. Lim for their donation of the ChemEdit software which was used to make several figures in this paper. We thank the following agencies for support of this work: Beckman Young Investigator Program (H.A.C.), NIH GM42703 (V.L.P.), and NIH GM008270 Michigan Molecular Biophysics Training Grant (J.Y.K.). We would also like to thank Dale Braden and Jeff Saunders of Schrödinger, Inc. for helpful conversations and Dr. Anna Bowman for her help with the figures.

Supporting Information Available: A list of residues included in the truncated protein model; sample input files for QM/MM calculations; and schematic pictures, energies, and QM coordinates for all minima from the full resting state and first protonated state calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Zampella, G.; Kravitz, J. Y.; Webster, C. E.; Fantucci, P.; Hall, M. B.; Carlson, H. A.; Pecoraro, V. L.; De Gioia, L. *Inorg. Chem.* **2004**, *43*, 4127−4136.

(2) Zampella, G.; Fantucci, P.; Pecoraro, V. L.; De Gioia, L. *J. Am. Chem. Soc.* **2005**, *127*, 953−960.

(3) Everett, R. R.; Butler, A. *Inorg. Chem.* **1989**, *28*, 393−395.

(4) van Schijndel, J. W. P. M.; Barnett, P.; Roelse, J.; Vollenbroek, E. G. M.; Wever, R. *Eur. J. Biochem.* **1994**, *225*, 151−157.

(5) Butler, A. *Coord. Chem. Rev.* **1999**, *187*, 17−35.

(6) de Boer, E.; Tromp, M. G. M.; Plat, H.; Krenn, G. E.; Wever, R. *Biochim. Biophys. Acta* **1986**, *872*, 104−115.

(7) Colpas, G. J.; Hamstra, B. J.; Kampf, J. W.; Pecoraro, V. L. *J. Am. Chem. Soc.* **1996**, *118*, 3469−3478.

(8) de la Rosa, R. I.; Clague, M. J.; Butler, A. *J. Am. Chem. Soc.* **1992**, *114*, 760−761.

(9) Martinez, J. S.; Carrol, G. L.; Tschirret-Guth, R. A.; Altenhoff, G.; Little, R. D.; Butler, A. *J. Am. Chem. Soc.* **2001**, *123*, 3289−3294.

(10) Wever, R.; Kustin, K. In *Advances in Inorganic Chemistry*; Sykes, A. G., Ed.; Academic Press: New York, 1990; Vol. 35, pp 103−137.

(11) Carter-Franklin, J. N.; Butler, A. *J. Am. Chem. Soc.* **2004**, *126*, 15060−15066.

(12) Everett, R. R.; Soedjak, H. S.; Butler, A. *J. Biol. Chem.* **1990**, *265*, 15671−15679.

(13) ten Brink, H. B.; Tuynman, A.; Dekker, H. L.; Hemrika, W.; Izumi, Y.; Oshiro, T.; Schoemaker, H. E.; Wever, R. *Inorg. Chem.* **1998**, *37*, 6780−6784.

(14) Smith, T. S., II; Pecoraro, V. L. *Inorg. Chem.* **2002**, *41*, 6754−6760.

(15) Conte, V.; Di Furia, F.; Modena, G. *J. Org. Chem.* **1988**, *53*, 1165−1669.

(16) Butler, A.; Clague, M. J.; Meister, G. E. *Chem. Rev.* **1994**, *94*, 625−638.

(17) Bortolini, O.; Di Furia, F.; Scrimin, P.; Modena, G. *Nouv. J. Chim.* **1985**, *9*, 147−150.

(18) Mimoun, H.; Saussine, L.; Daire, E.; Postel, M.; Fischer, J.; Weiss, R. *J. Am. Chem. Soc.* **1983**, *105*, 3101−3110.

(19) Sheldon, R. A.; Kochi, J. K. *Metal-catalyzed oxidations of organic compounds: mechanistic principles and synthetic methodology including biochemical processes*; Academic Press: New York, 1981.

(20) Ligtenbarg, A. G. J. In *Chemistry*; University of Groningen: Groningen, Holland, 2001.

(21) Messerschmidt, A.; Prade, L.; Wever, R. *Biol. Chem.* **1997**, *378*, 309−315.

(22) Hemrika, W.; Renirie, R.; Macedo-Ribeiro, S.; Messerschmidt, A.; Wever, R. *J. Biol. Chem.* **1999**, *274*, 23820−23827.

(23) Macedo-Ribeiro, S.; Hemrika, W.; Renirie, R.; Wever, R.; Messerschmidt, A. *J. Biol. Inorg. Chem.* **1999**, *4*, 545−559.

(24) Buhl, M. *Angew. Chem., Int. Ed* **1998**, *37*, 142−144.

(25) Buhl, M. *J. Comput. Chem.* **1999**, *20*, 1254−1261.

(26) Buhl, M.; Hamprecht, F. A. *J. Comput. Chem.* **1998**, *19*, 113−122.

(27) Buhl, M.; Parrinello, M. *Chem. Eur. J.* **2001**, *7*, 4487−4494.

(28) Conte, V.; Di Furia, F.; Moro, S. *J. Mol. Catal. A* **1997**, *117*, 139−149.

(29) Bortolini, O.; Carraro, M.; Conte, V.; Moro, S. *Eur. J. Biochem.* **2003**, 42−46.

(30) Conte, V.; Di Furia, F.; Moro, S.; Rabbolini, S. *J. Mol. Catal. A* **1996**, *113*, 175−184.

(31) Messerschmidt, A.; Wever, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 392−396.

(32) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468−1494.

(33) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442−1457.

(34) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *Chem. Phys. Lett.* **2000**, *321*, 113−120.

(35) Wirstam, M.; Lippard, S. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2003**, *125*, 3980−3987.

(36) Guallar, V.; Baik, M.-H.; Lippard, S. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6998−7002.

(37) QSite 2.5 revision 20, Schrodinger, LLC, New York City, 2003.

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

(39) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(40) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(41) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299−310.

(42) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724−728.

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

(44) Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 4233−4234.

(45) Binkley, J. S.; Pople, J. A. *J. Chem. Phys.* **1977**, *66*, 879−880.

(46) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654−3665.

(47) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(48) Molecular Operating Environment, Chemical Computing Group, Inc., Montreal, Quebec, Canada, 2002.

(49) Impact 2.5 revision 20, Schrodinger, LLC, New York City, 2003.

(50) de Boer, E.; Wever, R. *J. Biol. Chem.* **1988**, *263*, 12326−12332.

(51) Carter, J. N.; Beatty, K. E.; Simpson, M. T.; Butler, A. *J. Inorg. Biochem.* **2002**, *91*, 59−69.

(52) Casny, M.; Rehder, D.; Schmidt, H.; Vilter, H.; Conte, V. *J. Inorg. Biochem.* **2000**, *80*, 157−160.

(53) Colpas, G. J.; Hamstra, B. J.; Kampf, J. W.; Pecoraro, V. L. *J. Am. Chem. Soc.* **1994**, *116*, 3627−3628.

(54) Colpas, G. J.; Hamstra, B. J.; Kampf, J. W.; Pecoraro, V. L. *Inorg. Chem.* **1994**, *33*, 4669−4675.

(55) Andersson, M.; Conte, V.; Di Furia, F.; Moro, S. *Tetrahedron Lett.* **1995**, *36*, 2675−2678.

(56) Bortolini, O.; Carraro, M.; Conte, V.; Moro, S. *Eur. J. Biochem.* **1999**, 1489−1495.

(57) Clague, M. J.; Keder, N. L.; Butler, A. *Inorg. Chem.* **1993**, *32*, 4754−4761.

(58) Conte, V.; Bortolini, O.; Carraro, S.; Moro, S. *J. Inorg. Biochem.* **2000**, *80*, 41−49.

(59) Colpas, G. J.; Hamstra, B. J.; Kampf, J. W.; Pecoraro, V. L. *J. Am. Chem. Soc.* **1996**, *118*, 3469−3478.

CT050132O

JCTC Journal of Chemical Theory and Computation

A Semiempirical Quantum Model for Hydrogen-Bonded Nucleic Acid Base Pairs

Timothy J. Giese, Edward C. Sherer, Christopher J. Cramer, and Darrin M. York*

Department of Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455-0431

Received April 15, 2005

Abstract: An exploratory semiempirical Hamiltonian ($PM3_{BP}$) is developed to model hydrogen bonding in nucleic acid base pairs. The $PM3_{BP}$ Hamiltonian is a novel reparametrization of the PM3 Hamiltonian designed to reproduce experimental base pair dimer enthalpies and high-level density-functional results. The parametrization utilized a suite of integrated nonlinear optimization algorithms interfaced with a d-orbital semiempirical program. Results are compared with experimental values and with benchmark density-functional (*m*PWPW91/MIDI!) calculations for hydrogen-bonded nucleic acid dimers and trimers. The $PM3_{BP}$ Hamiltonian is demonstrated to outperform the AM1, PM3, MNDO, and MNDO/H Hamiltonians for dimer and trimer structures and interaction enthalpies and is shown to reproduce experimental dimer interaction enthalpies that rival density-functional results for an over 3 orders of magnitude reduction in computational cost. The tradeoff between a high accuracy gain for hydrogen bonding at the expense of sacrificing some generality is discussed. These results provide insight into the limits of conventional semiempirical forms for accurate modeling of biological interactions.

1. Introduction

The accurate calculation of the electronic structure and associated properties of biomolecules remains an important challenge in computational biochemistry.[1] Biological processes are often mediated by a delicate balance of subtle and highly specific molecular interactions that allow the myriad of cellular events to proceed under physiological conditions. It is a goal of applied quantum chemistry to provide accurate, robust methods to model these interactions that include specific binding and recognition events as well as complex catalytic reaction mechanisms.[2−10] Unfortunately, for many biological applications, accurate ab initio methods are thwarted by the computational cost associated with the inherently large system size, broad temporal domain, or high degree of phase-space sampling required by the problem. A pragmatic alternative is to take recourse into empirical or semiempirical quantum methods that are able to provide accuracy that often surpasses low-level ab initio methods[11] for a fraction of the computational cost.

Semiempirical quantum methods have traditionally not been considered to be of sufficient accuracy for biological chemistry, largely because their development has focused on more general ground-state thermochemical applications.[12,13] Because of their immense computational advantage, there has been a recent resurgence in interest to develop new semiempirical quantum models[14,15,17,18] specifically designed to provide high accuracy for biological reactions[19] and that can be used with linear-scaling electronic structure[20,21] and implicit solvent methods[22,23] as well as hybrid quantum mechanical/molecular mechanical (QM/MM) simulations.[24,25]

The interaction of nucleic acid bases in DNA and RNA structures plays an integral role in macromolecular structure and function.[26,27] Nucleic acid bases can interact via specific hydrogen-bonding arrangements and aromatic base stacking.[28] These interactions have been an area of intense investigation both experimentally and with electronic structure methods.[29] Hydrogen-bonding interactions between nucleic acid base pairs is vital to the integrity of duplex DNA and responsible for the transfer of genetic information. An accurate description of nucleic acid base pairs requires a proper description of the dipole moments and delocalization

* Author to whom correspondence should be addressed. E-mail: york@chem.umn.edu

10.1021/ct050102l CCC: $30.25
Published on Web 09/13/2005

of π bonds of the individual bases, and of intermolecular hydrogen bonding.[30] These features are not adequately reproduced by any of the standard semiempirical models.[31−33]

In this paper, an exploratory $PM3_{BP}$ Hamiltonian is developed specifically for hydrogen bonding in nucleic acid base pairs. The purpose of this paper is to explore the parametrizational limits of existing Hamiltonian forms in adequately modeling biologically relevant interactions. This is a key step toward the development of simple quantum Hamiltonian models that provide accuracy comparable to the highest feasible ab initio methods for biomolecules and, therefore, can be readily extended to linear-scaling quantum calculations[34−36] or hybrid QM/MM simulations.[37,38] Achievement of this goal would represent a major advance in the modeling of important biological reactions. With careful parametrization of the semiempirical $PM3_{BP}$ Hamiltonian, accuracy comparable to density-functional theory results are obtained with over 3 orders of magnitude less computational cost. The results presented here demonstrate promise for the future development of extremely fast quantum models especially designed for biological systems.

2. Background

The formalism for the electronic part of the MNDO,[14,39,40] AM1,[41] PM3,[12,42] and MNDO/H[43] Hamiltonians is based on the neglect of the diatomic differential overlap (NDDO) approximation and is identical for all the methods (see ref 44 for an overview). The four Hamiltonians differ only in the way core−core repulsions are treated. In the MNDO method, the repulsion between two nuclear cores (A and B) is calculated as

$$E_N^{MNDO}(A,B) = Z'_A Z'_B \langle s_A s_A | s_B s_B \rangle \, (1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) \quad (1)$$

where Z'_A and Z'_B are the effective nuclear charges (nuclear charge minus number of core electrons), $\langle s_A s_A | s_B s_B \rangle$ is a Coulomb repulsion integral between an *s*-symmetry orbital centered on A and an *s*-symmetry orbital centered on B, and α_A and α_B are parameters in the exponential term that account for decreased screening of the nucleus by the electrons at small interatomic distances. For O−H and N−H bonds, a modified form of the screening term is used

$$E_N^{MNDO}(A,H) = Z'_A Z'_H \langle s_A s_A | s_H s_H \rangle \, (1 + R_{AH} e^{-\alpha_A R_{AH}} + e^{-\alpha_H R_{AH}}) \quad (2)$$

For many intermolecular interactions, particularly hydrogen bonds, the MNDO model is problematic and often incorrectly predicts essentially unbound hydrogen-bonded complexes. The PM3 and AM1 models include a set of Gaussian core−core terms that alleviate excessive repulsion at close range and offer significant improvement for intermolecular interactions. The modified core−core term takes the form

$$E_N^{AM1/PM3}(A,B) = E_N^{MNDO}(A,B) + \frac{Z'_A Z'_B}{R_{AB}} \left(\sum_k a_{kA} e^{-b_{kA}(R_{AB} - c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB} - c_{kB})^2} \right) \quad (3)$$

These terms considerably improve the description of hydrogen bonds; although, they are, in general, still considerably underbound. Alternatively, one could substitute the Gaussian core−core terms by other functions[45,46] or introduce new functional forms to the Hamiltonians.[14,18] A promising approach is to design new semiempirical methods based on density-functional theory, such as the SCC-DFTB method.[47]

The MNDO/H Hamiltonian is a modification of the MNDO Hamiltonian, where nuclear repulsion in bonds of the type A···H taking part in hydrogen bonds A···H−D (A, D = N, O, F) takes the form

$$E_N^{MNDO/H}(A,H) = Z'_A Z'_H \langle s_A s_A | s_H s_H \rangle \, (1 + e^{-\alpha R_{AH}^2}) \quad (4)$$

where α was proposed[43] to equal 2.0 Å^{-2}. As part of the MNDO/H modification to MNDO, the user must choose which pairs A···H take part in the formation of hydrogen bonds. We have chosen to use the default settings as implemented in the MNDO97 program;[48] that is, the minimum and maximum A···H distances were chosen as 1.1 and 5.0 Å, respectively, and a minimum A···H−D angle of 90 degrees was selected.

Other successful Hamiltonian forms of note, although not directly compared against here, include the PDDG/PM3 and PDDG/MNDO Hamiltonians, which employ pairwise distance-dependent Gaussian core−core terms,[49,50] the AM1/d model for molybdenum with bond-specific (i.e., pairwise) core−core exponential repulsion terms,[51] a redefinition of core−core terms for hydrogen-bonded systems,[45,46] the use of bond-based corrections for improving heats of formation,[52] and potential energy scaling procedures.[53]

3. Methods

This section describes the methods used to develop the semiempirical $PM3_{BP}$ model that is subsequently analyzed and tested. The first subsection describes the quantum data set used as the reference data to fit the semiempirical $PM3_{BP}$ parameters for nucleic acid base pairs. The second subsection describes the details of the parametrization procedure itself.

3.1. Quantum Dataset for Nucleic Acid Base Pairs. The quantum reference data set employed here to parametrize the new semiempirical method has been described in detail elsewhere[54] and is briefly summarized here. Geometries were optimized with the Kohn−Sham density-functional theory (DFT) method using the *m*PWPW91 exchange-correlation functional[55,56] with the MIDI! basis set.[57] Stationary points were verified to be minima through standard frequency calculations (positive Hessian eigenvalues for all vibrational modes) that were also used (unscaled) to calculate zero-point and thermal contributions to the gas-phase enthalpy at 298.15 K and 1 atm. Basis set superposition errors were corrected using the procedure of Xantheas.[58] All ab initio calculations were performed with the Gaussian 98 suite of programs.[59] The quantum reference data[54] calculated at this level will henceforth be designated as "*m*PWPW" in the tables and text. The interaction enthalpies obtained from the quantum data set were previously demonstrated[54] to compare favorably with available experimental[60,61] values for AT, GC, UU, CC, TT, and AU pairs and also with computations[62] for a large

number of other base pairs carried out with larger basis sets and more complete levels of electronic structure theory and were, thus, deemed to be an appropriate and convenient test set against which to parametrize the semiempirical model. Formally, the contributions to the experimental interaction enthalpies require sampling of all relevant conformations. In the present work, calculated enthalpies are based on the single, lowest-energy ab initio configuration, as in other work,[54] or on a Boltzmann-weighted average.

3.2. Semiempirical Parametrization Procedure. This section describes the $PM3_{BP}$ parametrization procedure for nucleic acid base pairs based on the density-functional quantum data set described in the previous section. The first step is to construct an appropriate $\chi^2(\boldsymbol{\lambda})$ merit function that measures the goodness of fit of a set of molecular properties, calculated with a set (vector) of semiempirical parameters $\boldsymbol{\lambda}$, with the corresponding reference values. The second step is to use nonlinear optimization methods to find a suitable set of parameters by minimization of the $\chi^2(\boldsymbol{\lambda})$ merit function.

3.2.1. Construction of the $\chi^2(\boldsymbol{\lambda})$ Merit Function. The form of the $\chi^2(\boldsymbol{\lambda})$ merit function used in this work is given by

$$\chi^2(\boldsymbol{\lambda}) = \sum_{i}^{\text{mol}} \sum_{\alpha}^{\text{prop}} w_{i\alpha}[Y_{i\alpha}^{\text{PM3}_{\text{BP}}}(\boldsymbol{\lambda}) - Y_{i\alpha}^{\text{REF}}]^2 \quad (5)$$

$$w_{i\alpha} = (\sigma_{i\alpha}{}^2)^{-1} \quad (6)$$

where the first sum with index i in eq 5 runs over molecules (or complexes) and the second sum with index α runs over properties of the molecule (or complex). The argument λ represents a trial set of $PM3_{BP}$ parameters that are the variational degrees of freedom, $Y_{i\alpha}^{\text{PM3}_{\text{BP}}}(\boldsymbol{\lambda})$ is the value of the property α for molecule (complex) i calculated with the trial parameter set λ, $Y_{i\alpha}^{\text{REF}}$ is the corresponding reference value (taken either from the experiment or calculated with DFT), and $w_{i\alpha}$ is the associated least-squares weight in the fitting. The weights $w_{i\alpha}$ are proportional to the inverse square of the $\sigma_{i\alpha}$ values in eq 6. The $\sigma_{i\alpha}$ values have the same units as the molecular property to which they are associated and control the sensitivity of the merit function to deviations of that property from the reference value. The properties contained in the χ^2 merit function, the number of reference data, and the σ weight values for each property are summarized in Table 1.

For the semiempirical calculations, a modified version of the MNDO97[48] program was used. The properties considered include relative energies, optimized bond lengths, angles, torsions, and dipole moments for neutral species. Each structure of the data set is fully optimized at the semiempirical level for a given set of parameters before calculating these properties and constructing the $\chi^2(\boldsymbol{\lambda})$ function. Previous work in the development of specific reaction parameter Hamiltonians did not perform geometry optimization but, instead, performed single-point calculations at a stationary point and penalized the norm of the gradient in the $\chi^2(\boldsymbol{\lambda})$ function. This procedure works well for very simple molecules and reactions with only small allowable variations in a few semiempirical parameters.[63] This is not a productive strategy in the present case. The large number of degrees of freedom make it extremely difficult to lock down the norm of the gradient to a sufficient degree so that they accurately reflect the energies and geometries associated with the fully optimized geometries, especially when kcal/mol accuracy is the primary goal.

Table 1. Reference Data Contained in the χ^2 Merit Function[a]

description	N	σ	unit
bond lengths	999	0.005	Å
bond angles	1518	2.000	deg
bond torsion angles	1075	5.000	deg
dipole moments (*m*PWPW)	36	0.040	D
dipole moments (exptl)	5	0.040	D
H-bond distances	65	0.020	Å
H-bond angles	65	0.040	deg
intermolecular heavy atom distances	65	0.020	Å
dimerization enthalpies (*m*PWPW)	18	0.500	kcal/mol
dimerization enthalpies[b] (exptl)	11	0.500	kcal/mol
conformationally relative dimerization enthalpies[c] (*m*PWPW)	30	0.250	kcal/mol

[a] The terms exptl and *m*PWPW refer to available experimental and DFT data, respectively. If not explicitly specified, the reference data refers to the use of DFT reference data. N is the number of reference data points for the given property, and σ is the associated weight within the χ^2 definition. [b] A dimerization enthalpy is defined as the difference in enthalpy between a dimer and isolated monomers. Only six unique experimental values were used in the parametrization but appear more than once in the χ^2 definition since the experimental numbers are not orientationally distinguishable. [c] The conformationally relative dimerization enthalpies are defined here as the difference in dimerization enthalpies between base-pair arrangements, for example, the difference in dimerization enthalpy between AT_{WC} and AT_{RWC}.

3.2.2. Nonlinear Optimization of the $\chi^2(\boldsymbol{\lambda})$ Merit Function. Semiempirical parameters were obtained by optimization of the $\chi^2(\boldsymbol{\lambda})$ merit function of eq 5 with respect to the set of semiempirical parameters $\boldsymbol{\lambda}$ for H, N, and O atoms (parameters for C atoms were held fixed to the PM3 values). For this purpose, a suite of integrated nonlinear optimization methods for semiempirical parameter development has been used. The details of the integrated suite are forthcoming;[64] a brief overview of the algorithms is provided here.

Three nonlinear optimization methods working in concert were applied in the present work: (1) genetic algorithm, (2) Monte Carlo simulated annealing, and (3) direction set minimization methods. Genetic algorithms[65,66] have been demonstrated elsewhere to be useful in semiempirical parameter optimization.[51,63,67,68] The implementation of the genetic algorithm was loosely based on the description by Goldberg[65] and tailored for the several issues encountered in the semiempirical optimization application. A new method of partitioning subsets of the population (referred to as "tribes") into different local minima on the $\chi^2(\boldsymbol{\lambda})$ surface called *fitness-weighted eigenvector niching* was employed. The fitness of members was determined from a Gaussian distribution of the χ^2 values of population members within a "tribe" with the Gaussian width proportional to the χ^2 variance. The full details of the method are described elsewhere.[64]

Table 2. Parameters in the PM3 and $PM3_{BP}$ Hamiltonians[a]

parameter	H	C	N	O
U_{ss} (eV)	−12.755 118 80	−47.270 320 00	−48.794 933 85	−87.387 093 24
	−13.073 321 00	*−47.270 320 00*	*−49.335 672 00*	*−86.993 002 00*
U_{pp} (eV)		−36.266 918 00	−46.579 459 03	−71.702 685 70
		−36.266 918 00	*−47.509 736 00*	*−71.879 580 00*
β_s (eV)	−4.878 234 60	−11.910 015 00	−14.338 846 65	−46.877 410 23
	−5.626 512 00	*−11.910 015 00*	*−14.062 521 00*	*−45.202 651 00*
β_p (eV)		−9.802 755 00	−19.308 628 53	−24.742 325 18
		−9.802 755 00	*−20.043 848 00*	*−24.752 515 00*
α (eV)	3.356 386 00	2.707 807 00	2.830 545 00	3.217 102 00
	3.356 386 00	*2.707 807 00*	*2.830 545 00*	*3.217 102 00*
H_{sp} (eV)		2.290 980 00	1.136 713 00	0.593 883 00
		2.290 980 00	*1.136 713 00*	*0.593 883 00*
G_{ss} (eV)	15.023 337 45	11.200 708 00	12.415 221 41	15.261 643 45
	14.794 208 00	*11.200 708 00*	*11.904 787 00*	*15.755 760 00*
G_{pp} (eV)		10.796 292 00	13.966 114 12	13.659 300 75
		10.796 292 00	*11.754 672 00*	*13.654 016 00*
G_{sp} (eV)		10.265 027 00	7.345 402 26	10.413 326 25
		10.265 027 00	*7.348 565 00*	*10.621 160 00*
G_{p2} (eV)		9.042 566 00	10.410 219 25	12.420 510 17
		9.042 566 00	*10.807 277 00*	*12.406 095 00*
ζ_s (Å^{-1})	0.967 807 00	1.565 085 00	2.028 094 00	3.796 544 00
	0.967 807 00	*1.565 085 00*	*2.028 094 00*	*3.796 544 00*
ζ_p (Å^{-1})		1.842 345 00	2.313 728 00	2.389 402 00
		1.842 345 00	*2.313 728 00*	*2.389 402 00*
a_1 (unitless)	1.121 725 94	0.050 107 00	1.501 671 53	−1.131 176 77
	1.128 750 00	*0.050 107 00*	*1.501 674 00*	*−1.131 128 00*
b_1 (Å^{-2})	5.095 167 07	6.003 165 00	5.903 991 75	6.009 998 15
	5.096 282 00	*6.003 165 00*	*5.901 148 00*	*6.002 477 00*
c_1 (Å)	1.536 937 00	1.642 214 00	1.710 426 69	1.607 311 00
	1.537 465 00	*1.642 214 00*	*1.710 740 00*	*1.607 311 00*
a_2 (unitless)	−1.064 925 25	0.050 733 00	−1.515 716 18	1.130 989 09
	−1.060 329 00	*0.050 733 00*	*−1.505 772 00*	*1.137 891 00*
b_2 (Å^{-2})	6.023 153 66	6.002 979 00	5.975 794 98	5.872 165 45
	6.003 788 00	*6.002 979 00*	*6.004 658 00*	*5.950 512 00*
c_2 (Å)	1.571 307 32	0.892 488 00	1.710 935 13	1.603 474 21
	1.570 189 00	*0.892 488 00*	*1.716 149 00*	*1.598 395 00*

[a] Standard notation for parameters taken from refs 42 and 70. The original PM3 parameters are shown in italics immediately below the $PM3_{BP}$ values. Note: the parameters for C were held fixed to the standard PM3 values.

Several genetic algorithm runs were performed, with the number of generations ranging from 50 to 200 using a population of 64−128 members. The final population from the genetic algorithm optimization was then passed to a Monte Carlo simulated annealing procedure. The Monte Carlo procedure[69] used multidimensional simplex moves and variable exponentially decaying annealing schedules to explore the local region of parameter space around the final population provided by the genetic algorithm. The resulting parameters were then passed to a quadratically convergent direction set optimization method[69] to arrive at the final optimized $PM3_{BP}$ parameter set (Table 2). Recently, these methods have been extended and improved to make the parametrization more (although not completely) automated and robust, a detailed description of which is forthcoming.[64]

4. Results and Discussion

This section presents results and compares the performance of semiempirical Hamiltonian models with respect to experimental and density-functional calculations for nucleic acid base dimers and trimers. A detailed comparison and extended discussion of nucleic acid base monomer geometries and dipole moments are provided in the Supporting Information. The semiempirical methods include the new $PM3_{BP}$ method of the present work and the conventional semiempirical AM1,[41] PM3,[12,42] MNDO,[39,40] and MNDO/H[43] Hamiltonian models. The error metrics (error = calculated − experimental/reference value) shown in the tables are the maximum (signed) error (MAXE), root-mean-square error (RMSE), mean unsigned error (MUE), and mean signed error (MSE).

4.1. Relationship between the PM3 and $PM3_{BP}$ Parameters. Overall, the $PM3_{BP}$ parameters do not change dramatically from the PM3 parameters that were the starting point for optimization (Table 2). Note that the parameters for carbon in the $PM3_{BP}$ method were held fixed to the PM3 values, as were the H_{sp} parameters for each atom. For hydrogen, the greatest change occurs for the β_s and G_{ss} parameters that were shifted from the PM3 values in the positive direction by 0.75 and 0.21 eV, respectively. Similarly

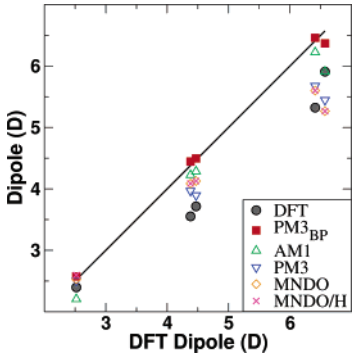


Figure 1. Regression of semiempirical and DFT *m*PWPW[54] dipole moments for nucleic acid bases with B3LYP/cc-pVTZ[71] (*x*-axis reference) values. A linear fit for each method produces intercept (*b*), slope (*m*), and correlation coefficient (*c*) values of DFT: $b = 0.253$ D, $m = 0.809$, $c = 0.982$; $PM3_{BP}$: $b = 0.081$ D, $m = 0.981$, $c = 1.000$; AM1: $b = 0.116$ D, $m = 0.918$, $c = 0.996$; PM3: $b = 0.289$ D, $m = 0.818$, $c = 0.988$; MNDO: $b = 0.349$ D, $m = 0.807$, $c = 0.985$; and MNDO/H: $b = 0.367$, $m = 0.805$, $c = 0.985$.

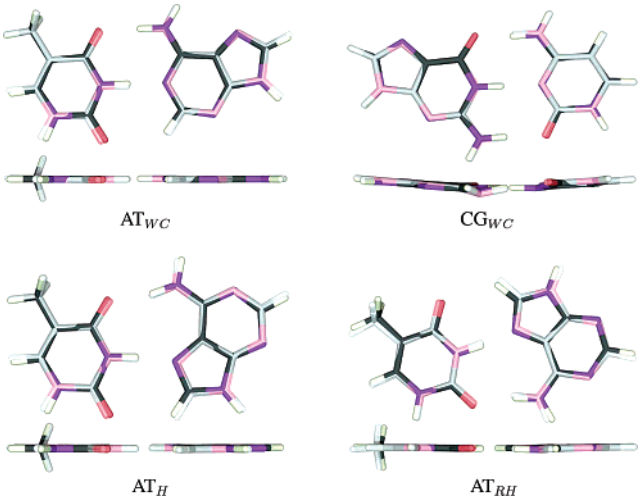


Figure 2. Superimposed root-mean-squared fit of $PM3_{BP}$ (lighter colors) geometry optimized structures to DFT *m*PWPW[54] structures (darker colors) for AT_{WC} (upper left), CG_{WC} (upper right), AT_H (lower right), and AT_{RH} (lower left). Each pane shows the face-on view (upper) and side view (lower).

for nitrogen and oxygen, the β_s and G_{ss} parameters also exhibited significant change; however, the β_s parameter was shifted toward slightly more negative values, whereas the G_{ss} parameter was shifted by +0.51 eV for oxygen and −0.49 eV for nitrogen. For nitrogen, the β_p and G_{pp} parameters exhibited even more pronounced change from the PM3 values than the β_s and G_{ss} parameters (+0.74 and 2.21 eV, respectively). Aside from these parameters, the $PM3_{BP}$ parameters deviated from the PM3 values by typically only a few percent or less.

4.2. Nucleic Acid Base Monomers. In this section, a comparison of the internal geometry and dipole moments for cytosine, guanine, adenine, thymine, and uracil nucleotide bases is briefly summarized. An extended discussion can be found in the Supporting Information. All of the semiempirical methods perform reasonably well for the internal geometries of the base monomers with respect to the *m*PWPW results of Sherer et al.,[54] with the $PM3_{BP}$ method performing best overall.

The semiempirical dipole moments for the nucleic acid bases are compared with the density-functional calculations of Sherer et al.,[54] and the high basis-set level (B3LYP/cc-pVTZ) calculations of Li et al.[71] are illustrated in Figure 1. The $PM3_{BP}$ method performs the best with respect to the high basis DFT results, with a RMSE of 0.101 D. The MNDO/H method has the largest RMSE (0.717 D) of the semiempirical methods. It is of interest to note that the DFT dipole moments calculated with the smaller basis set (*m*PWPW) have a RMSE (0.761 D), with respect to the higher basis-set (B3LYP/cc-pVTZ) values, that is larger than any of the semiempirical RMSE values.

4.3. Nucleic Acid Base Dimers. The main focus of this paper is on the structure and binding enthalpy of hydrogen-bonded nucleic acid base pairs. A host of standard and nonstandard base-pairing interactions were considered, including Watson−Crick (WC), reverse Watson−Crick (RWC), Hoogstein (H), reverse Hoogstein (RH), and mismatched base pairs. Figure 2 illustrates various base pair geometries (*m*PWPW geometries) for a representative subset of hydrogen-bonded dimers ($PM3_{BP}$ geometries are RMS overlaid onto the DFT structures of Sherer et al.;[54] figures of the DFT structures and their relation to the nomenclature are found within the appendix of ref 54). Subsection 4.3.1 provides a brief summary comparison of the semiempirical results for gas-phase intermolecular hydrogen-bonding and binding enthalpies with recent density-functional calculations,[54] a full discussion of which is provided in the Supporting Information. Subsection 4.3.2 compares density-functional and semiempirical hydrogen-bond lengths and dimerization enthalpies to experimental values. Subsection 4.3.3 examines an adiabatic binding potential energy curve for a representative hydrogen-bonded base pair and addresses potential problems associated with the use of Gaussian core−core functions.

4.3.1. Comparison with Density-Functional Calculations. This subsection provides a brief summary of the comparison results for the set of 31 hydrogen-bonded nucleic acid base dimers calculated with the semiempirical and density-functional (*m*PWPW) quantum models. An extended discussion and presentation of data is provided in the Supporting Information and has also been discussed, in part, elsewhere.[54]

The MNDO Hamiltonian does not predict stable hydrogen bonds, MNDO/H forms hydrogen-bond lengths that are too short, AM1 predicts hydrogen-bond lengths that are too long, and PM3 and $PM3_{BP}$ predict hydrogen-bond lengths that agree the most closely with the *m*PWPW values of any of the other semiempirical models considered. Results for the hydrogen-bond angles are qualitatively similar in that errors are most significant for the MNDO and AM1 methods. A comparison of semiempirical and *m*PWPW dimerization enthalpies shows that MNDO is critically underbound; AM1 and PM3 are significantly underbound by over 5 kcal/mol on average, whereas MNDO/H predicts dimers that are

Table 3. Comparison of Semiempirical and DFT Binding Enthalpies with Experimental Values for Nucleic Acid Base Dimers[a]

molecule	exptl	*m*PWPW	$PM3_{BP}$	AM1	PM3	MNDO	MNDO/H
CG_{WC}	−21.0	−22.4	−21.4	−13.8	−11.8	−3.9	−29.2
AT_{WC}		−11.3	−12.4	−4.9	−5.8	−0.6	−17.7
AT_{RWC}		−10.6	−12.5	−4.7	−5.9	−0.4	−17.1
AT_{H}		−11.2	−13.6	−4.9	−6.8	−0.9	−17.4
AT_{RH}		−10.7	−13.7	−5.0	−6.9	−1.0	−17.2
AT*	−13.0	−11.1	−13.5	−4.9	−6.7	−0.8	−17.4
CC	−16.0	−17.0	−18.8	−7.5	−9.4	−3.4	−26.6
TT_1		−8.6	−8.7	−5.9	−4.6	−0.5	−13.4
TT_2		−9.4	−8.4	−6.0	−4.4	−0.4	−13.8
TT_3		−7.9	−8.9	−5.9	−4.7	0.2	−13.0
TT*	−9.0	−9.2	−8.7	−5.9	−4.6	−0.4	−13.6
UU_1		−8.3	−8.7	−6.0	−4.5	−1.8	−13.4
UU_2		−9.4	−8.7	−6.0	−4.3	−0.3	−13.8
UU_3		−7.5	−8.8	−5.9	−4.6	−0.2	−13.0
UU*	−9.5	−9.2	−8.8	−6.0	−4.5	−1.6	−13.6
AU_{WC}	−14.5	−11.4	−12.6	−4.9	−5.8	−0.4	−17.8
MSE		**0.5**	**0.1**	**6.7**	**6.7**	**12.1**	**−5.9**
MUE		**1.3**	**1.1**	**6.7**	**6.7**	**12.1**	**5.9**
RMSE		**1.6**	**1.4**	**7.1**	**6.9**	**12.5**	**6.4**
MAXE		**3.1**	**−2.8**	**9.6**	**9.2**	**17.1**	**−10.6**

[a] Comparison of binding enthalpies (kcal/mol) for nucleic acid base dimers from semiempirical ($PM3_{BP}$, AM1, PM3, MNDO, and MNDO/H) and DFT *m*PWPW[54] (*m*PWPW) calculations with experimental values[60,61] (exptl). An asterisk indicates that the value used in comparison to the experimental value is a Boltzmann-weighted average of several structures (individually listed immediately above the averaged result) at 298.15 K. Summarized at the bottom are the error metrics (bold) for the semiempirical and DFT (*m*PWPW) values with corresponding experimental results.

overbound by over 6 kcal/mol on average relative to the *m*PWPW results. The $PM3_{BP}$ dimerization enthalpies, on the other hand, are in close agreement with the *m*PWPW values with a RMSE or 1.3 kcal/mol.

4.3.2. Comparison with Experimental Values. Table 3 compares the calculated binding enthalpies with experimental values.[60,61] In some instances, a direct comparison cannot be made since the experimental values can often not distinguish between different binding orientations. In these instances, the computed value was used to compare with the experiment result from a Boltzmann-weighted average of the available minima at 298.15 K. Note that experimental measurements may contain fractions of very different binding motifs, such as stacked base interactions, which are not accounted for within the scope of the currently selected geometries. Consequently, the "MSE" and "RMSE" values reported in Table 3 must be regarded as approximate. Nonetheless, the DFT values appear to slightly underestimate the binding enthalpy (the MSE is 0.5 kcal/mol), and the overall RMSE is 1.6 kcal/mol. The MSE and RMSE values for AM1 (6.7 and 7.1 kcal/mol, respectively) and PM3 (6.7 and 7.9 kcal/mol, respectively) relative to the experimental binding enthalpies are slightly larger than the corresponding MSE and RMSE values relative to the DFT results. The MNDO/H method is still overbound with respect to experimental values, with a MSE value of −5.9 kcal/mol and RMSE of 6.4 kcal/mol. This underscores the inadequacy of these methods for biological applications where hydrogen bonding is involved. The $PM3_{BP}$ method performs best relative to the experimental binding enthalpies (Figure 3) with a MSE of only 0.1 kcal/mol and a RMSE of 1.4 kcal/mol. The largest errors occur for the CC dimer ($PM3_{BP}$ error of −2.8 kcal/mol) and the AU_{WC} dimer ($PM3_{BP}$ error of 1.9 kcal/mol). The error trends of the DFT (*m*PWPW) values with respect to those of the experiment have the same sign for these dimers (−1.0 and 3.1 kcal/mol for CC and AU_{WC}, respectively).

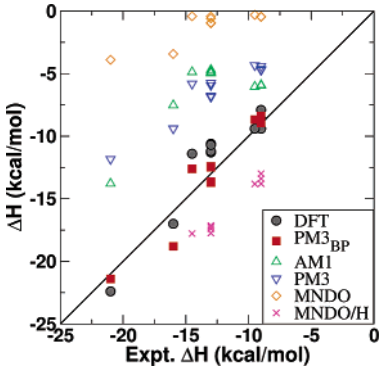


Figure 3. Regression of semiempirical and DFT *m*PWPW[54] binding enthalpies for nucleic acid base dimers with experimental[60,61] (*x*-axis reference) values. A linear fit for each method produces intercept (*b*), slope (*m*), and correlation coefficient (*c*) values of DFT: $b = 2.009$ kcal/mol, $m = 1.086$, $c = 0.940$; $PM3_{BP}$: $b = 1.510$ kcal/mol, $m = 1.117$, $c = 0.965$; AM1: $b = -0.006$ kcal/mol, $m = 0.495$, $c = 0.695$; PM3: $b = 1.211$ kcal/mol, $m = 0.598$, $c = 0.944$; MNDO: $b = 1.211$ kcal/mol, $m = 0.598$, $c = 0.945$; and MNDO/H: $b = 2.865$ kcal/mol, $m = 0.308$, $c = 0.850$.

The distance between the heavy atoms acting as hydrogen-bond acceptors and donors have been resolved for the AU_{WC} and CG_{WC} base pairs in sodium adenylyl-3′-5′-uridine (ApU) and sodium guanylyl-3′,5′-cytidine nonahydrate (GpC) crys-

Table 4. Comparison of Experimental, Semiempirical, and DFT Hydrogen-Bond Heavy-Atom Acceptor−Donor Separations for AU and CG Base Pairs[a]

	AU_{WC}		CG_{WC}		
method	$N_1\cdots H-N_3$	$N_6-H\cdots O_4$	$N_4-H\cdots O_6$	$N_3\cdots H-N_1$	$O_2\cdots H-N_2$
exptl	2.82	2.95	2.91	2.95	2.86
*m*PWPW	2.73 (−0.09)	2.84 (−0.11)	2.71 (−0.20)	2.83 (−0.12)	2.87 (0.01)
$PM3_{BP}$	2.77 (−0.05)	2.77 (−0.18)	2.76 (−0.15)	2.78 (−0.17)	2.80 (−0.06)
AM1	3.05 (0.23)	3.11 (0.16)	3.06 (0.15)	3.05 (0.10)	3.10 (0.24)
PM3	2.81 (0.01)	2.82 (−0.13)	2.82 (−0.09)	2.80 (−0.15)	2.85 (−0.01)
MNDO	4.22 (1.40)	5.10 (2.15)	4.22 (1.31)	4.12 (1.17)	4.09 (1.23)
MNDO/H	2.55 (−0.27)	2.54 (−0.41)	2.52 (−0.39)	2.55 (−0.40)	2.60 (−0.26)

[a] Comparison of semiempirical ($PM3_{BP}$, AM1, PM3, MNDO, and MNDO/H) and DFT *m*PWPW[54] (*m*PWPW) calculated hydrogen-bond heavy-atom acceptor−donor separations (Å) with experimental (exptl) X-ray crystal structure analysis of ApU and GpC.[26,72,73] Errors relative to experimental values are indicated in parentheses.

tals using X-ray diffraction.[26,72,73] Table 4 compares the experimental acceptor−donor distances with *m*PWPW and semiempirical methods. Reasonable agreement with experimental values is found between DFT and $PM3_{BP}$ with errors ranging from 0.01 to 0.2 Å. Both *m*PWPW and $PM3_{BP}$ are in reasonable agreement with crystallographic data with MUE values of 0.11 and 0.13 Å, respectively. PM3 agrees best with experimental values with a MUE error of 0.08 Å; however, an examination of the base pair geometries shows considerably artificial nonplanarity. The MNDO/H method (using the suggested α parameter[43] of 2.0 $Å^{-2}$) predicts acceptor−donor separations that are systematically too short by 0.29 Å.

4.3.3. Dimer Potential Energy Curve. The use of core−core functions can lead to artificial stationary points in potential energy surfaces. In this section, the adiabatic binding energy potential energy curve for a hydrogen-bonded base pair is explored to address this issue. Figure 4 displays the adiabatic binding potential energy curve for AT_{WC} (defined here as the center of mass separation between ADE and THY) for the semiempirical methods and *m*PWPW. The monomer geometries and relative orientation with respect to one another were taken from the DFT-optimized AT_{WC} structure (Figure 2).

None of the semiempirical methods show artificial stationary points in the potential energy curve. The minimum energy separations and relative binding energies are qualitatively similar to the dimer hydrogen bond lengths and dimer binding enthalpies summarized in Section 4.3 and extensively discussed in the Supporting Information. At first glance, it appears that the AM1 method has the best qualitative *shape* of the potential energy curve when compared to DFT, although severely underbound, whereas the PM3 and $PM3_{BP}$ methods have a spuriously steep potential well near the minimum. A careful comparison of these three potential energy curves beyond 6 Å reveals that their long-range attractive tails are nearly parallel, suggesting that the steep potential well near the minimum observed in the PM3 and $PM3_{BP}$ methods are due to core−core functions. A better potential energy curve might be obtained with core−core functions with smaller Gaussian exponents or having based the parametrization off of AM1 as opposed to PM3. For the design of new-generation semiempirical methods for QM/MM methods, it is likely best to avoid completely the use of off-center Gaussian core−core terms.

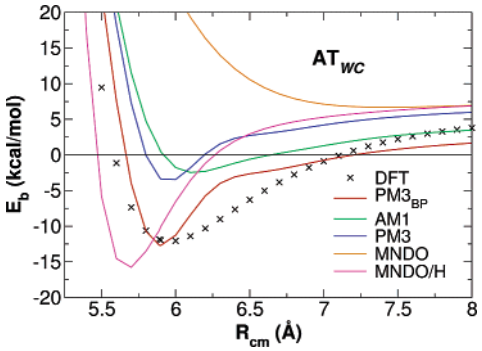


Figure 4. Binding energy of AT_{WC} as a function of rigid base-pair separation. The potential energy curve is defined as the center of mass separation between each monomer along the vector joining their center of mass. The monomer geometries and relative orientation with respect to each other were held fixed during the scan to those determined from DFT optimization of the energy minimum (see Figure 2). The DFT binding energies are counterpoise corrected, but since nonstationary points are involved, zero-point energy corrections and thermal corrections to the enthalpy were not included. The zero of energy is defined as the energy of the two isolated monomers. Since the monomers are constrained to those found in the AT_{WC} DFT optimized structure, the binding energy asymptotically reaches a positive value at large center of mass separations. The semiempirical geometries used were the same as used in the ab initio single-point calculations.

4.4. Nucleic Acid Base Trimers. The elementary next step in evaluating the limits of semiempirical methods in studying nucleic acid base interactions is to examine trimer interactions. This is an interesting test for the $PM3_{BP}$ method, since no trimer data were used in the parametrization procedure. Base pair trimers (sometimes referred to as "triplexes") have been studied in the past with Hartree−Fock[74], DFT (B3LYP[5] and *m*PWPW),[54] and MP2[75,76] methods. In addition, Yanson et al.[61] have reported trimerization enthalpies by analysis of mass spectral peak intensities in multicomponent mixtures.

Table 5 compares values for the nucleic acid base trimer binding enthalpies calculated with DFT,[54] MP2,[76] $PM3_{BP}$, AM1, PM3, MNDO, and MNDO/H. The naming of the trimers and illustrations of the DFT trimer geometries are

Table 5. Comparison of the Semiempirical and DFT Binding Enthalpies with Experimental Values for Nucleic Acid Base Trimers[a]

molecule	exptl	*m*PWPW	$PM3_{BP}$	AM1	PM3	MNDO	MNDO/H
CCC_1		−14.0	−16.8	−24.1	−8.5	−8.5	−17.4
CCC_2		−28.8	−33.4	−25.5	−17.1	−8.3	−40.2
CCC_4	−33 (−38) ± 4	−22.0	−28.9	−14.8	−13.5	−4.4	−29.1
UUA_1		−21.0	−25.2	−8.8	−12.0	−1.4	−33.9
UUA_2		−21.4	−25.1	−8.8	−11.9	−1.2	−34.1
UUA_3		−17.0	−20.5	−11.5	−10.6	−1.3	−21.9
UUA_4		−17.4	−20.6	−12.2	−11.9	−2.2	−25.0
UUA*	−27 (−29) ± 4	−21.3	−25.2	−12.0	−11.9	−1.8	−34.0
UUU_1		−8.5	−13.1	−10.0	−7.0	−2.6	−19.9
UUU_2		−11.3	−14.6	−10.3	−8.4	−0.7	−18.3
UUU*	−20 (−22) ± 4	−11.3	−14.5	−10.2	−8.3	−2.5	−19.8
UUT	−23 (−25) ± 4	−7.1	−12.7	−9.9	−6.6	−0.5	−18.6
TAT	−23.8 (MP2)	−21.3	−24.9	−8.9	−11.9	−1.0	−34.1
CGC^+	−65.2 (MP2)	−68.3	−65.2	−39.9	−46.0	−23.2	−79.8
GGG		−36.8	−33.3	−28.5	−19.6	−8.7	−40.3

[a] Comparison of binding enthalpies (kcal/mol) for nucleic acid base trimers (triplexes) from semiempirical ($PM3_{BP}$, AM1, PM3, MNDO, and MNDO/H) and DFT *m*PWPW[54] (*m*PWPW) calculations with experimental values[60,61] (exptl) or MP2/6-31G(d) calculations[76] (MP2). The MP2 results[76] involve geometry optimization, zero-point vibrational energy correction, and thermal contributions at the HF/6-31G(d) level followed by BSSE correction and MP2/6-31G(d) calculation with all d polarization functions using an exponent of 0.25. The naming convention of the molecules follows from Sherer et al.[54] Of special note is the difference between CCC_1, CCC_2, and CCC_4: CCC_1 and CCC_2 are unmethylated cytosine triplex structures, whereas CCC_4 is a (N_1,N_4)-dimethylcytosine triplex structure. An asterisk indicates that the value is a Boltzmann-weighted average of several structures (individually listed immediately above the averaged result) at 298.15 K.

presented in ref 54. As was done with the comparisons of dimer binding enthalpies with experimental values, Boltzmann-weighted averages of the computed energies from the available geometries are used to compare to available experimental trimer enthalpy results. The different experimental values reported in Table 5 result from different choices of the dimer equilibrium constants in the analysis of the spectral intensity ratios.[61]

The AM1, PM3, and MNDO semiempirical methods considerably underestimate the binding enthalpy for all trimers when compared to experimental values, MP2, or the Boltzmann-averaged DFT data. The smallest error of all of the conventional (AM1, PM3, MNDO) semiempirical methods is with AM1, although closer inspection reveals an incorrect rank order of some of the binding enthalpies relative to that of the DFT values. The MNDO/H method, in general, is considerably overbound, but not in all cases, such as the CCC_4 trimer. The DFT values are underbound relative to those of the experiment and are typically just outside the lower bound of the experimental error. $PM3_{BP}$ is also underbound relative to experiment, but more bound than the DFT values by roughly 3 kcal/mol, which is consistent with the behavior observed with the dimers (see Supporting Information). As pointed out previously,[54] the experimentally determined binding enthalpy of (N_1,N_4)-dimethylcytosine (CCC_4) is unjustifiably overbound; a binding enthalpy of −33 kcal/mol only seems plausible if hydrogen-bonding sites are freed by demethylation of the monomers, the lowest enthalpy of which is the CCC_2 structure. In fact, the $PM3_{BP}$ binding enthalpy of the unmethylated structure exactly reproduces a binding enthalpy of −33 kcal/mol.

For structures where experimental binding enthalpies are unavailable (TAT and CGC^+), a comparison is made to counterpoise corrected MP2 enthalpies with zero-point vibrational energy and thermal corrections to the energy.[76] In both cases, $PM3_{BP}$ agrees better with the MP2 enthalpies (−23.8 and −65.2 kcal/mol for TAT and CGC^+, respectively) than the *m*PWPW[54] enthalpies. Although the DFT values are in good agreement with the MP2 enthalpies (errors of 2.5 and 3.1 kcal/mol for TAT and CGC^+, respectively), $PM3_{BP}$ agrees exceptionally well (errors of 1.1 and 0.0 kcal/mol for TAT and CGC^+, respectively).

4.5. Transferability to Molecules not in the Parametrization Set. Although it is the purpose of the present work to focus on hydrogen bonding in nucleic acid bases, it is instructive to test and compare the $PM3_{BP}$ method with a more general set of hydrogen-bonded complexes. Toward this end, a test set of molecules was considered in order to compare the ability of the $PM3_{BP}$ method and other semiempirical methods to model intermolecular hydrogen bonding between neutral molecules[31] and some biologically relevant ions. A summary of the error metric results for the dimerization enthalpies, dipole moments, and hydrogen-bond lengths relative to the *m*PWPW values is provided in Table 6, the complete set of data being provided in the Supporting Information. Overall, the $PM3_{BP}$ method makes a considerable improvement relative to the other semiempirical methods for all of these properties. For example, the RMSE values for the dimerization enthalpy, dipole moment, and hydrogen-bond distances are 1.58 kcal/mol, 0.63 D, and 0.23 Å, respectively, for $PM3_{BP}$, whereas the next lowest RMSE values from any of the other semiempirical methods are 1.91 kcal/mol (MNDO/H), 0.83 D (PM3), and 0.26 Å (PM3), respectively. These results suggest that the strategy outlined here of careful, specific reparametrization, using some consistency constraints (such as fixing the C parameters and allowing a relatively small deviation from the more general PM3 parameter values) can assist in maintaining a significant level of robustness and transferability for the properties included in the parametrization procedure.

Table 6. Comparison of Semiempirical Dimerization Enthalpy, Dipole Moment, and Hydrogen-Bond Length Error Metrics Relative to DFT for Hydrogen-Bonded Complexes Not Contained in the Parameterization[a]

property	metric	$PM3_{BP}$	AM1	PM3	MNDO/H
ΔH (kcal/mol)	MSE	0.34	0.81	1.61	−0.93
	MUE	1.11	2.37	2.04	1.63
	RMSE	1.58	3.27	2.87	1.91
	MAXE	4.47	10.45	8.65	−5.26
μ (D)	MSE	0.17	−0.36	−0.25	0.01
	MUE	0.40	0.63	0.61	0.69
	RMSE	0.63	0.90	0.83	0.91
	MAXE	1.91	−2.78	−2.43	1.93
H···X (Å)	MSE	−0.04	0.39	0.05	−0.24
	MUE	0.19	0.39	0.19	0.32
	RMSE	0.23	0.49	0.26	0.37
	MAXE	0.76	1.27	0.72	0.90

[a] The DFT reference values were obtained at the B3LYP/6-311++G(3df,2p)//B3YLP/6-31++G(d,p) level of theory with zero-point energy and thermal corrections to the enthalpy derived from a frequency analysis at the B3YLP/6-31++G(d,p) level of theory. The dimerization enthalpy (ΔH) error metrics involved the comparison of 37 dimers. The dimerization enthalpy is defined as the difference in enthalpy between the dimer and the isolated monomers. From these 37 dimers, there were 43 hydrogen-bond lengths. The hydrogen-bond lengths were measured from the hydrogen position to the heavy atom to which it is hydrogen-bonding, denoted as H···X. The dipole moment (μ) statistics involved a total of 45 *neutral* dimers and monomers. A complete comparison of the data is tabulated in the Supporting Information.

5. Conclusion

The present paper reports an exploratory semiempirical Hamiltonian ($PM3_{BP}$) for modeling hydrogen-bonded nucleic acid bases that significantly outperforms the AM1, PM3, MNDO, and MNDO/H Hamiltonians and accurately reproduces nucleic acid base pair interaction enthalpies and optimized geometries when compared to experimental and *m*PWPW calculations. The $PM3_{BP}$ model was applied to hydrogen-bonded nucleic acid base trimers not contained in the parametrization set and found to agree much better with prior, higher-level calculations than the other tested semiempirical Hamiltonians.

Overparametrization of semiempirical methods to focused chemical problems can distort the physical nature of the model away from general applicability, which calls their very usefulness into question and limits their general predictive capability. On the other hand, the very broadly parametrized semiempirical models are not of sufficient quantitative accuracy to be useful in biological applications without additional ad hoc corrections. This suggests that the forms of current semiempirical models might be reaching their inherent limits. Consequently, further progress needs to be made in the development of new semiempirical methods[17] with treatments for those phenomena not described well with current Hamiltonian forms, such as dispersive attraction and proper polarization to electric fields while using a small basis set.

Nonetheless, the present work takes a significant step forward in testing the ability of the common semiempirical Hamiltonian forms to accommodate and reliably reproduce hydrogen-bonded nucleic acid base interactions. None of the tested Hamiltonians contain a term to properly account for the long-range dispersion effects that play an important role in stabilizing base-stacking interactions (or condensed phase simulations, in general), and further refinement of the methods to include such terms is likely to lead to more accurate and robust semiempirical models for biomolecular interactions.

Acknowledgment. D.Y. is grateful for financial support provided by the National Institutes of Health (Grant GM62248) and the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement number DAAD19-01-2-0014. C.J.C. thanks the National Science Foundation for support (CHE02-03346). Computational resources were provided by the Minnesota Supercomputing Institute.

Supporting Information Available: The Supporting Information contains an extended discussion and comparison of nucleic acid base monomer bond lengths, angles, torsion angles, and dipole moments predicted at the semiempirical and DFT levels of theory. Nucleic acid base dimer hydrogen-bond lengths, hydrogen-bond angles, and dimerization binding enthalpies are tabulated and compared in the extended discussion. Dimerization enthalpies, hydrogen-bond lengths, and the dipole moments of the 37 hydrogen-bonded complexes not considered in the parametrization are also tabulated. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Friesner, R. A.; Beachy, M. D. *Curr. Opin. Struct. Biol.* **1998**, *8*, 257−262.

(2) Alkorta, I.; Elguero, J. *J. Phys. Chem. B* **2003**, *107* (22), 5306−5310.

(3) Brandl, M.; Meyer, M.; Sühnel, J. *J. Phys. Chem. A* **2000**, *104* (47), 11177−11187.

(4) Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100* (1), 143−167.

(5) Peters, M.; Rozas, I.; Alkorta, I.; Elguero, J. *J. Phys. Chem. B* **2003**, *107* (1), 323−330.

(6) Sivanesan, D.; Babu, K.; Gadre, S. R.; Subramanian, V.; Ramasami, T. *J. Phys. Chem. A* **2000**, *104* (46), 10887−10894.

(7) Williams, N. G.; Williams, L. D.; Shaw, B. R. *J. Am. Chem. Soc.* **1989**, *111* (18), 7205−7209.

(8) Zhanpeisov, N. U.; Šponer, J.; Leszczynski, J. *J. Phys. Chem. A* **1998**, *102* (50), 10374−10379.

(9) Kawahara, S.; Uchimaru, T.; Sekine, M. *THEOCHEM*, **2000**, *530*, 109−117.

(10) Bosch, D.; Campillo, M.; Pardo, L. *J. Comput. Chem.* **2003**, *24* (6), 682−691.

(11) For example, minimal basis set Hartree−Fock methods.

(12) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.

(13) Stewart, J. J. P. *J. Mol. Model.* **2004**, *10*, 6−12.

(14) Thiel, W. In *Adv. Chem. Phys.*; Prigogine, I., Rice, S. A., Eds.; John Wiley and Sons: New York, 1996; Volume 93, pp 703−757.

(15) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495−506.

(16) Clark, T. *THEOCHEM* **2000**, *530*, 1−10.

(17) Winget, P.; Selçuki, C.; Horn, A.; Martin, B.; Clark, T. *Theor. Chem. Acc.* **2003**, *110* (4), 254−266.

(18) Thiel, W. Semiempirical Theories. In *Handbook of Molecular Physics and Quantum Chemistry*; Wilson, S., Ed.; John Wiley and Sons: Chicester, U. K., 2003; Volume 2, pp 487−502.

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

(20) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085−1123.

(21) Yang, W.; Pérez-Jordá, J. M. In *Encyclopedia of Computational Chemistry*; von Schleyer, P. R., Ed.; John Wiley and Sons: New York, 1998; pp 1496−1513.

(22) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(23) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99* (8), 2161−2200.

(24) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(25) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.

(26) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(27) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.

(28) Maki, A.; Brownewell, F. E.; Liu, D.; Kool, E. T. *Nucleic Acids Res.* **2003**, *31* (3), 1059−1066.

(29) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2002**, *61*, 3−31.

(30) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.

(31) Jurema, M. W.; Shields, G. C. *J. Comput. Chem.* **1993**, *14* (1), 89−109.

(32) Lively, T. N.; Jurema, M. W.; Shields, G. C. *Int. J. Quantum Chem.* **1994**, *52* (21), 95−107.

(33) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzkík, P.; Vondrášek, J. *J. Comput. Chem.* **1997**, *18* (9), 1136−1150.

(34) Khandogin, J.; York, D. M. *J. Phys. Chem. B* **2002**, *106*, 7693−7703.

(35) Khandogin, J.; Musier-Forsyth, K.; York, D. M. *J. Mol. Biol.* **2003**, *330*, 993−1004.

(36) Khandogin, J.; York, D. M. *Proteins* **2004**, *56*, 724−737.

(37) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2003**, *125*, 7178−7179.

(38) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 7504−7513.

(39) Dewar, M. J.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99* (15), 4899−4907.

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

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

(42) Stewart, J. J. P. *Rev. Comput. Chem.* **1990**, *1*, 45−81.

(43) Burstein, K. Y.; Isaev, A. N. *Theor. Chim. Acta* **1984**, *64* (5), 397−401.

(44) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons: Chichester, England, 2004.

(45) Bernal-Uruchurtu, M.; Ruiz-López, M. *Chem. Phys. Lett.* **2000**, *330*, 118−124.

(46) Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C. C.; Millot, M. F. R.-L. *J. Comput. Chem.* **2000**, *21*, 572−581.

(47) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. Status Solidi B* **2000**, *217*, 357−376.

(48) Thiel, W. *MNDO97*, version 5.0; University of Zurich: Zurich, Switzerland, 1998.

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

(50) Tubert-Brohman, I.; Guimaraes, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2003**, *25*, 138−150.

(51) Voityuk, A. A.; Rösch, N. *J. Phys. Chem. A* **2000**, *104*, 4089−4094.

(52) Long, D. A.; Anderson, J. B. *Chem. Phys. Lett.* **2005**, *402*, 524−528.

(53) Brauer, B.; Chabanb, G. M.; Gerbe, R. B. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2543−2556.

(54) Sherer, E. C.; York, D. M.; Cramer, C. J. *J. Comput. Chem.* **2003**, *24*, 57−67.

(55) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108* (2), 664−675.

(56) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54* (23), 16533−16539.

(57) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1996**, *93* (5), 281−301.

(58) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104* (21), 8821−8824.

(59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; 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.9; Gaussian, Inc.: Pittsburgh, PA, 1998.

(60) Sukhodub, L. F.; Yanson, I. K. *Nature* **1976**, *264* (5583), 245−247.

(61) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149−1170.

(62) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.

(63) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231−236.

(64) Giese, T. J.; York, D. M. Nonlinear parameter optimization algorithms. Manuscript in preparation.

(65) Goldberg, D. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Reading, MA, 1989.

(66) Coley, D. A. *An Introduction to Genetic Algorithms for Scientists and Engineers*; World Scientific: River Edge, NJ, 1999.

(67) Cundari, T. R.; Deng, J.; Fu, W. *Int. J. Quantum Chem.* **2000**, *77*, 421−432.

(68) Hutter, M. C.; Reimers, J. R.; Hush, N. S. *J. Phys. Chem. B* **1998**, *102*, 8080−8090.

(69) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, W. P. *Numerical Recipes in Fortran*, 2nd ed.; Cambridge University Press: Cambridge, U. K., 1992.

(70) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391−404.

(71) Li, J.; Xing, J.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *111* (3), 885−892.

(72) Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Kim, J.; Rich, A. *J. Mol. Biol.* **1976**, *104* (1), 109−144.

(73) Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104* (1), 145−167.

(74) Pundlik, S. S.; Gadre, S. R. *J. Phys. Chem. B* **1997**, *101* (46), 9657−9662.

(75) Poltev, V. I.; Shulyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *3* (4), 739−765.

(76) Sponer, J.; Burda, J. V.; Mejzlik, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613.

CT050102L

JCTC Journal of Chemical Theory and Computation

Effects of Calcium Binding on Structure and Autolysis Regulation in Trypsins. A Molecular Dynamics Investigation

Elena Papaleo, Piercarlo Fantucci, and Luca De Gioia*

Department of Biotechnology and Biosciences, University of Milano-Bicocca, P.za della Scienza 2, 20126 Milan, Italy

Received April 6, 2005

Abstract: The calcium ion was proposed to be involved in protein structure stabilization against thermal and proteolytic degradation, such as autolysis phenomena, in trypsin-like serine proteases. However, molecular details related to the role played by the metal ion are still largely unknown. Several molecular dynamics simulations of 6 ns have been used to investigate the dynamic behavior of bovine and salmon trypsins in calcium-bound and calcium-free forms, with the aim of evaluating the role of the calcium ion in trypsin three-dimensional structure and autoproteolysis propensity. It turned out that the calcium-free trypsins are characterized by a more flexible structure, revealing structure−function relationships connecting Ca^{2+} binding and autoproteolysis propensity. In particular, the removal of Ca^{2+} not only increases the flexibility of regions around its binding site, in the N-terminal domain, but also leads to channeling of the fluctuations to remote sites in the C-terminal domain, possibly involving the interdomain loop. Two primary autolysis sites are strongly influenced by calcium binding (R117 and K188) in bovine trypsin, whereas Ca^{2+} plays a less crucial role in salmon trypsin.

Introduction

Trypsin, a member of the serine protease family that is highly specific for cleavage at lysine and arginine residues, plays an important role in many biological processes. The large number of studies carried out on trypsins in their native form, as well as in the presence of a variety of ligands, allowed an understanding of fundamental issues related to substrate binding, specificity, and fine details of the catalytic mechanism.[1] Trypsins and several other eukaryotic serine proteases bind a Ca^{2+} ion, which is known to play an important role as a regulator of physiological functions and in the preservation of structural integrity in many proteins.[2] In particular, the calcium ion in trypsin-like serine proteases has been proposed to be involved in the stabilization against thermal and proteolytic degradation, such as autolysis phenomena.[3,4] The elucidation of factors regulating autolysis propensity is particularly relevant since the inability of trypsin to self-degradate has been linked to human hereditary pancreatitis, a genetic disorder due to inappropriate activation of trypsin within the pancreas.[5]

Even though several structural details related to Ca^{2+} binding to trypsins are known,[1] the detailed role played by Ca^{2+} in structure stabilization and autolysis is not yet fully understood. The calcium-binding loop runs from residue 69 to 80 (numeration of bovine trypsin, BT) and connects two antiparallel β strands in the N-terminal domain. The loop structure is maintained by hydrogen bonds formed with other parts of the protein. In addition, several water molecules form bridges between the loop and other protein regions.[3] The role played by Ca^{2+} in trypsin structure has been investigated using several spectroscopic techniques,[6−10] leading to the conclusion that the protein assumes a less-compact structure when the metal ion is removed. This behavior has been correlated to modifications in the environment of a conserved tryptophan residue (W141),[9] even though FT-IR spectroscopy did not show detectable perturbation of the protein secondary structure.[10]

The analysis of trypsin structures obtained by X-ray diffraction (mammalian and salmon trypsins) revealed that

* Corresponding author. Phone: +39 02 64483463. E-mail: luca.degioia@unimib.it.

10.1021/ct050092o CCC: $30.25
Published on Web 10/05/2005

the calcium-binding loop interacts with a protein portion usually described as the autolysis loop (residues 143−155, BT numeration).[3] This loop owes its name to its similarity with chymotrypsins, but it is still unclear if it constitutes a conserved primary autolysis site in trypsins. In fact, a well-known autoproteolytic site in the BT autolysis loop (K145) is not conserved in salmon trypsin (ST), where the closest putative autolysis site (K154) seems to be too buried and rigid to be a favorable autolysis site.[3] Moreover, several other autolysis sites have been characterized in mammalian trypsins.[11,12] The peptide bonds between R117−V118 and K145−S146 are simultaneously cleaved, leading to an enzyme form that retains activity.[12] Cleavage can also takes place at K60−S61[12] and K188−D189,[11] leading to enzyme inactivation. Finally, it was also shown that the autolysis of wild-type rat trypsin begins with cleavage at R117, and K60 is another relevant autolysis site.[13,14] Notably, the peptide segment between K60 and R117, which contains the so-called interdomain loop that connects the two globular protein domains, is part of the longest peptide portion in trypsin that does not include disulfide bridges.[1] It has been suggested that this region, which includes the Ca^{2+}-binding loop, may function as a built-in target for autolysis: the cleavage at any end of the peptide segment K60−R117 is followed by cleavages at several trypsin-sensitive sites between the two ends.[13] It was also proposed that Ca^{2+} could stabilize an autolysis-resistant conformation of the K60−R117 segment.[13,15] Interestingly, ST and other fish trypsins[3,16−18] are characterized by autolysis rates that are less dependent on the presence of calcium ions than those of mammalian trypsins. However, it has been noted that this difference does not necessarily indicate that Ca^{2+} has a lower stabilizing effect in ST, since slower autolysis rates could be due to the presence of fewer exposed autolysis sites.[3] As for the other factors that can affect autolysis propensity, Fontana et al.[19] have suggested that protein portions characterized by large flexibility (evaluated on the basis of B-factor values) are particularly susceptible to proteolytic attack, whereas Novotny and Bruccoleri[20] concluded that solvent accessibility is more important than flexibility in affecting proteolysis.

Molecular dynamics (MD) is a powerful tool used to understand the structure and functional features of a protein in atomic detail, and it can lead to significant insights into the atomic machinery underlying protein function.[21−23] In fact, MD simulations on trypsins from organisms adapted to different environmental conditions have been previously reported, revealing details about the relationship between protein dynamics and functions.[24,25] However, the role played by Ca^{2+} on the dynamic properties of trypsins has never been investigated.

With the aim of clarifying the role of calcium on trypsin structure and autoproteolysis propensity, we have carried out multiple 6 ns MD simulations of ST and BT proteins, in their native holo (calcium-bound) and apo (calcium-free) forms. Analogous approaches were applied successfully to investigate the role of calcium in α-lactalbumin and c-type lysozyme,[26] parvalbumin,[27] and calmodulin.[28,29]

Sequence and structure characteristics of ST and BT relevant to the proper analysis of results from the MD simulations are presented in the first part of the manuscript, whereas the results and analysis of MD simulations are discussed in the second section.

Methods

To collect known structures and sequences of mammalian and fish trypsins, the Protein Data Bank (PDB)[30] and the nonredundant sequence database were searched with Blast[31] using the amino acid sequences of BT and ST as probes. Multiple sequence alignments were generated with ClustalW[32] using default parameters.

MD simulations were performed using the GROMACS simulation software package,[33] implemented on a parallel architecture. The X-ray structures of native bovine (BT) and atlantic salmon (ST) trypsins (PDB entries 3PTB[34] and 2TBS,[35] respectively) were used as starting points for the computational investigation.

Initial structures for the apo forms have been obtained according to the following procedure: the Ca^{2+} ion has been removed from the PDB file, the apo-protein structure was optimized by molecular mechanics (1000 steepest descent cycles followed by 10 000 conjugate gradient steps), and the resulting structure was submitted for MD simulations.

MD simulations of the holo and apo forms have been carried out according to the following protocol. Protein structures, including crystallographic water molecules, were soaked in a dodecahedral box of 4523 (BT simulations) or 5244 (ST simulations) SPC water molecules.[36] The minimum distance between the solute (protein) and the box edges was set to 0.5 nm. The ionization state of charged residues was set to be consistent with neutral pH: Lys and Arg residues were positively charged, whereas Asp and Glu were negatively charged. The protonation state of the histidine residues was predicted using GROMACS tools and confirmed by visual inspection of the molecular environment of each histidine. To neutralize the overall charge of the system, a number of water molecules equal to the norm of the protein charge, and located within 4 Å of the protein surface, were replaced by Cl^- (BT) or Na^+ (ST) ions.

Initially, solvent molecules were relaxed by molecular mechanics (steepest descent method, 1000 steps). The optimization step was followed by 20 ps MD at 300 K (time step 1 fs) while restraining protein atomic positions using a harmonic potential. MD simulations were performed in the NPT ensemble at 283 and 310 K, applying periodic boundary conditions and using an external thermal bath[37] with a coupling constant τ of 0.1 ps for the protein and 0.002 ps for nonprotein groups. Pressure was kept constant (1 atm) by modifying the box dimensions. The time constant for the pressure coupling was set to 1 ps.[37] The LINCS algorithm[38] was used to constrain the lengths of bonds involving hydrogen atoms. Electrostatic interactions were calculated using the particle-mesh Ewald summation scheme,[39] with a cutoff of 0.12 nm for the separation of the direct and reciprocal space sum cutoffs. van der Waals and Coulomb interactions were truncated at 1.0 nm. The nonbonded pair list was updated every 10 steps, and conformations were stored every 2 ps. The time step was set to 2 fs. The coupling

constant of the external bath was set to 0.001 ps for both protein and nonprotein elements.

To improve the conformational sampling, two or three simulations, obtained initializing the dynamic run with different Maxwellian distributions of initial velocities, were carried out for each protein system. The rmsd (root-mean-square deviation), which is a crucial parameter in the evaluation of the stability of MD trajectories, was evaluated for main-chain atoms using as a reference the starting structure of the MD simulations.

MD trajectories characterized by different initial velocities have been merged to compute rmsd matrices, in which the rmsd values computed comparing all pairs of frames are organized in a two-dimensional map, which allows an evaluation of the resampling of similar substructures. The rmsd matrices representing merged trajectories have been processed using the Jarvis−Patrick method,[40] to extract information on possible clusters of conformations. The average structures of the clusters, defined as the protein structure with the smallest average distance to the other conformation belonging to the same cluster, have also been calculated.[33]

The secondary structure content has been calculated using the DSSP program.[41] The solvent accessibility degree has been evaluated with the NACCESS program[42] on structures collected from the trajectories every 20 ps. The root-mean-square fluctuation (rmsf) has been calculated on main-chain atoms at a residue base, using as a reference the average structure from the trajectories.

The visual analysis of protein structures was carried out using INSIGHT II tools (Biosym Technologies/Molecular Simulations, San Diego, CA) and visual molecular dynamics.[43]

In the following, the numbering of residues is referred to BT if not otherwise specified. The N- (16−135) and C-terminal (136−245) domains of trypsin are labeled as N and C, respectively. The β strands are numbered from β-1 to β-13, while the α helices are numbered from α-1 to α-4. BT and ST sequences are composed of 223 and 222 total amino acids, respectively.

Results and Discussion

Analysis of Trypsin Structures and Sequences. Structures of mammalian and fish trypsins collected in the PDB (59 entries; March 30, 2005) were initially analyzed to characterize the environment of the calcium ion. In addition, the BT sequence was aligned with trypsins from different organisms (Figure 1) to evaluate the conservation of calcium-binding residues. It turned out that the side-chain oxygen atoms of E70, E80, and E77, which are strictly conserved in all trypsins; the backbone oxygen atoms of N72 and V75 (less strictly conserved); and a water molecule define the Ca^{2+} coordination environment in the analyzed trypsins. The molecular environment of the calcium-binding loop (69−80) comprises the loop at the N-terminal end (16−28), the loop between Nβ-1 and Nβ-2 (37−39), part of the interdomain loop between Nβ-6 and Cβ-1 (110−133), and the autolysis loop (144−154) between Cβ-1 and Cβ-2 (Figure 2a). The conformation of the calcium-binding loop in BT and ST is extremely similar, as is the overall fold of the proteins. The main difference in the backbone conformation is localized in correspondence to the autolysis loop (143−155), which assumes significantly different conformations in the two proteins. In fact, the loop conformation observed in ST has been proposed to be stabilized by the ion pairs D150−K154, E21−H71, and K74−D153, which are typical of fish trypsins (Figure 1).[3]

Six disulfide bonds are present in ST and BT (Figure 1, Figure 2b): C22−C157 (connecting the N-terminal end to the Cβ-2 strand), C42−C58 (connecting Nβ-2 to the loop Nβ-3/Nβ-4), C128−C232 (connecting loop Nβ-6/Cβ-1 to Cβ-2), C136−C201 (connecting Cβ-1 to Cβ-4), C168−C182 (connecting Cβ-2 to Cβ-3), and C191−C220 (connecting the loop Cβ-3/Cβ-4 to the loop Cβ-5/Cβ-6). All disulfide bridges are conserved in mammalian and fish trypsins, with the exception of human trypsin II. Notably, only two disulfide bridges are localized in the N-terminal domain, and one of them connects the N- to the C-terminal domain.

Trypsins are proteolytic enzymes specific for arginine and lysine residues. In BT, there are 16 putative autoproteolytic sites (K60, R66, K87, K107, K109, R117, K145, K156, K159, K169, K188, K204, K222, K224, K230, and K239), while in ST, the putative autolytic sites are 12 (K23, K60, R62, R66, K74, R87, R90, K107, K110, K154, K188, and K230). Only a few sites, conserved in most of the sequences, are common to the two proteins (K60, R66, K/R87, K107, K188, and K230) (Figure 1). As mentioned in the Introduction, K60, K145, K188, and R117 were characterized as primary autolysis sites in trypsins.[15,16] It is relevant to observe that R117 is typical of mammalian trypsins, whereas K145 is present only in three trypsins (from bovine, porcine, and Antarctic fish; Figure 1). Most of the lysine and arginine residues are located in loop regions or in proximity of the ends of β strands or α helices, with the exception of K/R87 (in Nβ-5), BT K159 (in Cβ-2), BT K169 (in α-1), and BT K239 (in α-2). Interestingly, two of these residues, K/R87 and K159, are surrounded by other trypsin-sensitive sites, which are localized in loop regions and might be more susceptible to proteolysis.

The solvent accessibility degree of the putative autoproteolytic sites, as well as their structural features, were also analyzed. The side chains of putative autolysis sites are generally well-exposed to the solvent (solvent accessibility degree larger than 30%), with the exception of R66, BT K107 and K156. The solvent accessibility degrees for the autolysis sites that have been characterized experimentally are 38.9% for BT K60, 39.9% for ST K60, 40.6% for BT R117, 60.7% for BT K145, 57.7% for BT K188, and 54.1% for ST K188. Notably, the solvent accessibility degree for K154, which has been proposed to functionally correspond to the BT K145 autolysis site,[3] is as large as 42.2%, in apparent disagreement with the hypothesis that the low propensity to autoproteolysis of this site could be due to solvent inaccessibility.[3] In fact, ST K154 is involved in a salt-bridge cluster with E21, H71, and D150, and therefore, its side chain could not be available to properly interact with the specificity pocket of trypsins. In a similar way, K74, which is located in the calcium-binding loop, can form an ion pair with D153.[3]

```
                                                                C22                 C42
Q8AV11Oncorhynchus_keta          ------------------------IVGGYECKAYSQPHQVSLNSGYHFCG 26
P35032Salmo_SalarII              -----------AAFAT----EDDKIVGGYECKAYSQPHQVSLNSGYHFCG 35
P35031Salmo_salarI               MISLVFVLLIGAAFAT----EDDKIVGGYECKAYSQTHQVSLNSGYHFCG 46
Q91515Fugu_rubripes              --------LIAAAYAAPI-DEDDKIVGGYECRKNSVAYQVSLNSGYHFCG 41
Q92099Paranotothenia_magellani   MRSLVFVLLIGAAFAT----EEDKIVGGKECSPYSQPHQVSLNSGYHFCG 46
Q9W7Q7Paralichthys_olivaceus     MRSLVFVLLIGAAFAM----EDDKIVGGYECTPYSQPHQVSLNSGYHFCG 46
Q9W7Q6Paralichthys_olivaceusII   ---LVFILLIGAAFAT----EDDKIVGGYECTPYSQPHQVSLNSGYHFCG 43
O93266Pleuronectes_americanus    MRSLVFVLLIGAAFAL----EDDKIVGGYECTPHSQAHQVSLNSGYHFCG 46
Q6RI79Tautogolabrus_adspersus    MRSLVFVLLIGAAFAI----DDDKIVGGYECTPHSQPHTVSLQSGYHFCG 46
P16049Gadus_morhuaI              MKSLIFVLLLGAVFAE-----EDKIVGGYECTKHSQAHQVSLNSGYHFCG 45
Q91041Gadus_morhuaII             MKSLIFVLLLGAVFAE-----EDKIVGGYECTRHSQAHQVSLNSGYHFCG 45
Q98TG9Engraulis_japonicusII      MRSLVFLVLLGAAFAE-----DDKIVGGYECQPYSQPHQVSLNSGYHFCG 45
Q98TH0Engraulis_japonicusI       MRPLVFLVLLGAAFAE-----DDKIVGGYECQAHSQPHTVSLNSGYHFCG 45
Q7SX90Brachydanio_rerio          MRSLVFLVLLGAAFALD----DDKIVGGYECQPNSQPWQASLNSGYHFCG 46
Q8QGW3Anguilla_japonica          MRSLVFILLLGVAVALD----DDKIVGGYECEPHSQPWQASLNAGYHFCG 46
Q7T1R8Pangasius_hypophthalmus    MRSLVLLLLVGACFALE----DDKIVGGYECTPYSQPWQVSLNVGYHFCG 46
Q6R670Oreochromis_aureus         MKYFILLALFAAAYAAPIED--DKIIGGYECAKNSVPYMVSLNIGYHFCG 48
Q6R671Oreochromis_niloticus      MKYFILLALFAAAYAAPIED--DKIIGGYECAKNSVPYMVSLNIGYHFCG 48
P35033Salmo_salarIII             ---------FAVAFAAPIDDEDDKIVGGYECRKNSASYQASLQSGYHFCG 41
P07477Homo_sapiensI              MNPLLILTFVAAALAAPFDDDD-KIVGGYNCEENSVPYQVSLNSGYHFCG 49
P07478Homo_sapiensII             MNLLLILTFVAAAVAAPFDDDD-KIVGGYICEENSVPYQVSLNSGYHFCG 49
P06872Canis_familiarisII         MNPLLILAFLGAAVATPTDDDD-KIVGGYTCEENSVPYQVSLNAGYHFCG 49
P00763Rattus_norvegicusII        MRALLFLALVGAAVAFPVDDDD-KIVGGYTCQENSVPYQVSLNSGYHFCG 49
Q792Z1Mus_musculusI              MSTLLFLALVGAAVAFPVDDDD-KIVGGYTCRENSVPYQVSLNSGYHFCG 49
P07146Mus_musculusII             MSALLILALVGAAVAFPVDDDD-KIVGGYTCRESSVPYQVSLNAGYHFCG 49
P00762Rattus_norvegicusI         MSALLILALVGAAVAFPLEDDD-KIVGGYTCPEHSVPYQVSLNSGYHFCG 49
P00760Bos_taurusI                ---FIFLALLGAAVAFPVDDDD-KIVGGYTCGANTVPYQVSLNSGYHFCG 46
P06871Canis_familiarisI          MKTFIFLALLGATVAFPIDDDD-KIVGGYTCSRNSVPYQVSLNSGYHFCG 49
P00761Sus_scrofaI                ---------------FPTDDDD-KIVGGYTCAANSIPYQVSLNSGSHFCG 34
P08426Rattus_norvegicusIII       MKALIFLAFLGAAVALPLDDDDDKIVGGYTCQKNSLPYQVSLNAGYHFCG 50
P70059Xenopus_laevisII           MKFLVILVLLGAAVAF---EDDDKIVGGFTCAKNAVPYQVSLNAGYHFGG 47
P19799Xenopus_laevisI            MKFLLLCVLLGAAAAF---DDD-KIIGGATCAKSSVPYIVSLNSGYHFCG 46
                                                     *:**  *   : .  .**: * ** *
```

```
                                          ;    C58 K60          Ca2+ b.loop
Q8AV11Oncorhynchus_keta          GSLVNENWVVSAAHCYKS--R-EVRLGEHNIKVTEGSEQFISSSRVIRHP 73
P35032Salmo_SalarII              GSLVNENWVVSAAHCYQS--RVEVRLGEHNIQVTEGSEQFISSSRVIRHP 83
P35031Salmo_salarI               GSLVNENWVVSAAHCYKS--RVEVRLGEHNIKVTEGSEQFISSSRVIRHP 94
Q91515Fugu_rubripes              GSLVNENWVVSAAHCYKS--RVVVRLGEHNIRANEGTEQFISSSRVIRHP 89
Q92099Paranotothenia_magellani   -SLVNENWVVSAAHCYKS--RVEVRMGEHHIRVTEGKEQFISSSRVIRHP 93
Q9W7Q7Paralichthys_olivaceus     GSLVNENWVVSAAHCYKS--RVEVRMGEHHIKINEGTEQFISSERVIRHP 94
Q9W7Q6Paralichthys_olivaceusII   GSLVNENWVVSAAHCYKS--RVEVRIGEHNLRVYEETEQFISSSRVIRHP 91
O93266Pleuronectes_americanus    GSLVNENWVVSAAHCYKS--RVEVRMGEHKIRVNEGTEQFVSSSRVIRHP 94
Q6RI79Tautogolabrus_adspersus    GSLVNEDWVVSAAHCYKS--RIQVRLGEHHIRVNEGTEQIISSSRVIRHP 94
P16049Gadus_morhuaI              GSLVSKDWVVSAAHCYKS--VLRVRLGEHHIRVNEGTEQYISSSSVIRHP 93
Q91041Gadus_morhuaII             GSLVSKDWVVSAAHCYKS--VLRVRLGEHHIRVNEGTEQFISSSSVIRHP 93
Q98TG9Engraulis_japonicusII      GSLVSDSWVVSAAHCYKS--RVEVRMGEHHIGMTEGNEQFIDSSRVIRHP 93
Q98TH0Engraulis_japonicusI       GSLVNENWVVSAAHCYKS--RVEVRLGEHHIGQNENTEQFIDSSRVIRHP 93
Q7SX90Brachydanio_rerio          GSLVSEYWVVSAAHCYKS--RVEVRLGEHNIVINEGTEQFITSEKVIRNP 94
Q8QGW3Anguilla_japonica          GSLVNENWVVSAAHCYKSPSRLEVRLGEHHIGLNEGTEQFIGASHVIRNP 96
Q7T1R8Pangasius_hypophthalmus    GSLINQNWVVSAAHCYQS--RIEVRLGEHNIQINEGTEQFISSSRVIRHP 94
Q6R670Oreochromis_aureus         GSLISSTWAVSAAHCYQS--SIQLRLGEHNIAVNEGTEQFISSSRVIRHQ 96
Q6R671Oreochromis_niloticus      GSLISSTWAVSAAHCYQS--SIQLRLGEHNIAVNEGTEQFISSSRVIRHQ 96
P35033Salmo_salarIII             GSLISSTWVVSAAHCYKS--RIQVRLGEHNIAVNEGTEQFIDSVKVIMHP 89
P07477Homo_sapiensI              GSLINEQWVVSAGHCYKS--RIQVRLGEHNIEVLEGNEQFINAAKIIRHP 97
P07478Homo_sapiensII             GSLISEQWVVSAGHCYKS--RIQVRLGEHNIEVLEGNEQFINAAKIIRHP 97
P06872Canis_familiarisII         GSLISDQWVVSAAHCYKS--RIQVRLGEYNIDVLEGNEQFINSAKVIRHP 97
P00763Rattus_norvegicusII        GSLINDQWVVSAAHCYKS--RIQVRLGEHNINVLEGNEQFVNAAKIIKHP 97
Q792Z1Mus_musculusI              GSLINDQWVVSAAHCYKS--RIQVRLGEHNINVLEGNEQFIDAANIIKHP 97
P07146Mus_musculusII             GSLINDQWVVSAAHCYKY--RIQVRLGEHNINVLEGNEQFVDSAKIIRHP 97
P00762Rattus_norvegicusI         GSLINDQWVVSAAHCYKS--RIQVRLGEHNINVLEGDEQFINAAKIIKHP 97
P00760Bos_taurusI                GSLINSQWVVSAAHCYKS--GIQVRLGEDNINVVEGNEQFISASKSIVHP 94
P06871Canis_familiarisI          GSLINSQWVVSAAHCYKS--RIQVRLGEYNIAVSEGGEQFINAAKIIRHP 97
P00761Sus_scrofaI                GSLINSQWVVSAAHCYKS--RIQVRLGEHNIDVLEGNEQFINAAKIITHP 82
P08426Rattus_norvegicusIII       GSLINSQWVVSAAHCYKS--RIQVRLGEHNIDVVEGGEQFIDAAKIIRHP 98
P70059Xenopus_laevisII           -SLINSQWVVSAAHCYKS--RIQVRLGEHNIALNEGTEQFIDSQKVIKHP 94
P19799Xenopus_laevisI            GSLITNQWVVSAAHCYKA--SIQVRLGEHNIALSEGTEQFISSSKVIRHS 94
                                  **:.. *.***.***:      :*:** ::   *  ** : :   * :
```

```
                                          R117      C128   C136
Q8AV11Oncorhynchus_keta                   NYSSYNIDNDIMLIKLSKPATLNTYVQPVALPSSCAPAGTMCTVSGWGNT 123
P35032Salmo_SalarII                       NYSSYNIDNDIMLIKLSKPATLNTYVQPVALPTSCAPAGTMCTVSGWGNT 133
P35031Salmo_salarI                        NYSSYNIDNDIMLIKLSKPATLNTYVQPVALPTSCAPAGTMCTVSGWGNT 144
Q91515Fugu_rubripes                       NYSSYNIDNDIMLIKLSKPATLNQYVQPVALPSSCAAAGTMCKVSGWGNT 139
Q92099Paranotothenia_magellani            NYSSYNIDNDIMLIKLSKPATLNQYVQAVALPSSCAPAGTMCTVSGWGST 143
Q9W7Q7Paralichthys_olivaceus              NYSSYNINNDIMLIKLREPATLNQYVQPVALPTSCAPAGTMCTVSGWGNT 144
Q9W7Q6Paralichthys_olivaceusII            NYSSYNINNDIMLIKLSEPATLNQYVQPVALPTSCAPAGTMCTVSGWGDT 141
O93266Pleuronectes_americanus             NYDSWNIDNDIMLIKLSKPATLNQYVKTVALPSSCAPAGTMCKVSGWGNT 144
Q6RI79Tautogolabrus_adspersus             RYSSYNIDNDIMLIKLSQSATLNQYVKTVALPTSCAPAGTMCKVAGWGNT 144
P16049Gadus_morhuaI                       NYSSYNINNDIMLIKLTKPATLNQYVHAVALPTECAADATMCTVSGWGNT 143
Q91041Gadus_morhuaII                      NYSSYNIDNDIMLIKLTEPATLNQYVHAVALPTECAADATMCTVSGWGNT 143
Q98TG9Engraulis_japonicusII               QYDSYNIDNDIMLIKLSKPATLNQYVQTVALPSSCAPAGTMCLVSGWGNT 143
Q98TH0Engraulis_japonicusI                QYSSYNIDNDVMLIKLSTPATLNQYVQPVALPSRCASAGTMCLVAGWGNT 143
Q7SX90Brachydanio_rerio                   NYDSWDLDSDIMLIKLSKPATLNKYVQPVALPNGCAADGTMCRVSGWGNT 144
Q8QGW3Anguilla_japonica                   NYNSWDLDSDIMLIKLSKPATLNSYVQPVALPTRCAPAGTMCRVTGWGNT 146
Q7T1R8Pangasius_hypophthalmus             NYNSWTIDNDIMLIKLSQSASVNNYVQPVALPSSCPPAGTWCIVSGWGNT 144
Q6R670Oreochromis_aureus                  SYNSYTLDNDIMLIKLSQPATLNSYVKTVSLPSGCAGAGTSCLISGWGNT 146
Q6R671Oreochromis_niloticus               SYNSYTLDNDIMLIKLSQPATLNSYVKTVSLPSGCAGAGTSCLISGWGNT 146
P35033Salmo_salarIII                      SYNSRNLDNDIMLIKLSKPASLNSYVSTVALPSSCASSGTRCLVSGWGNL 139
P07477Homo_sapiensI                       QYDRKTLNNDIMLIKLSSRAVINARVSTISLPTAPPATGTKCLISGWGNT 147
P07478Homo_sapiensII                      KYNSRTLDNDILLIKLSSPAVINSRVSAISLPTAPPAAGTESLISGWGNT 147
P06872Canis_familiarisII                  NYNSWILDNDIMLIKLSSPAVLNARVATISLPRACAAPGTQCLISGWGNT 147
P00763Rattus_norvegicusII                 NFDRKTLNNDIMLIKLSSPVKLNARVATVALPSSCAPAGTQCLISGWGNT 147
Q792Z1Mus_musculusI                       KFKKKTLDNDIMLIKLSSPVTLNARVATVALPSSCAAAGTQCLISGWGNT 147
P07146Mus_musculusII                      NYNSWTLDNDIMLIKLASPVTLNARVASVPLPSSCAPAGTQCLISGWGNT 147
P00762Rattus_norvegicusI                  NYSSWTLNNDIMLIKLSSPVKLNARVAPVALPSACAPAGTQCLISGWGNT 147
P00760Bos_taurusI                         SYNSNTLNNDIMLIKLKSAASLNSRVASISLPTSCASAGTQCLISGWGNT 144
P06871Canis_familiarisI                   RYNANTIDNDIMLIKLSSPATLNSRVSAIALPKSCPAAGTQCLISGWGNT 147
P00761Sus_scrofaI                         NFNGNTLDNDIMLIKLSSPATLNSRVATVSLPRSCAAAGTECLISGWGNT 132
P08426Rattus_norvegicusIII                SYNANTFDNDIMLIKLNSPATLNSRVSTVSLPRSCGSSGTKCLVSGWGNT 148
P70059Xenopus_laevisII                    NYNSRNLDNDIMLIKLSTTARLSANIQSVPLPSACASAGTNCLISGWGNT 144
P19799Xenopus_laevisI                     GYNSYTLDNDIMLIKLSSPASLNAAVNTVPLPSGCSAAGTSCLISGWGNT 144
                                           :.   ::.*::****   . :.  : .:.**      .* . ::***.

                                          K145   K154 C157       C168          C182  K188 C191
Q8AV11Oncorhynchus_keta                   MSSTADK-NKLQCLNIPILSYSDCNNSYPGMITNAMFCAGYLEGGKDSCQ 172
P35032Salmo_SalarII                       MSSTADK-NKLQCLNIPILSYSDCNNSYPGMITNAMFCAGYLEGGKDSCQ 182
P35031Salmo_salarI                        MSSTADS-NKLQCLNIPILSYSDCNNSYPGMITNAMFCAGYLEGGKDSCQ 193
Q91515Fugu_rubripes                       MSSTADR-NKLQCLNIPILSDRDCENSYPGMITDAMFCAGYLEGGKDSCQ 188
Q92099Paranotothenia_magellani            QSSSADG-NKLQCLNIPILSDRDCDNSYPGMITDAMFCAGYLQGGKDSCQ 192
Q9W7Q7Paralichthys_olivaceus              MSSTANR-DMLQCLDLPILSDRDCENSYPGMITPAMFCAGYLEGGKDSCQ 193
Q9W7Q6Paralichthys_olivaceusII            MSSTDS--SRLQCLDLPILSERDCENSYPGMITNAMFCAGYLEGGKDSCQ 189
O93266Pleuronectes_americanus             MSSADNG-DLLQCLDIPILSFSDCNNAYPGMITDSMFCAGYLEGGKDSCQ 193
Q6RI79Tautogolabrus_adspersus             MSSSADG-DKLQCLNIPILSERDCDNSYPGMITDAMFCAGYLEGGKDSCQ 193
P16049Gadus_morhuaI                       MSSVADG-DKLQCLSLPILSHADCANSYPGMITQSMFCAGYLEGGKDSCQ 192
Q91041Gadus_morhuaII                      MSSVDDG-DKLQCLNLPILSHADCANSYPGMITQSMFCAGYLEGGKDSCQ 192
Q98TG9Engraulis_japonicusII               MS-NVSG-DKLQCLQIPILSDRDCKNSYPGMITESMFCAGYLEGGKDSCQ 191
Q98TH0Engraulis_japonicusI                MS-NVSG-DKLQCLQIPILSDRDCDNSYPGMITDAMFCAGYLEGGKDSCQ 191
Q7SX90Brachydanio_rerio                   MSSTADS-NKLQCLEIPILSDRDCNNSYPGMVTDTMFCAGYLEGGKDSCQ 193
Q8QGW3Anguilla_japonica                   MNPAVSG-DKLQCLEIPILSESDCSNSYPGMITSTMFCAGYLEGGKDSCQ 195
Q7T1R8Pangasius_hypophthalmus             MSSTADR-NKLQCVEVPILSDEDCNNSYPGMITKAMFCAGFLEGGKDSCQ 193
Q6R670Oreochromis_aureus                  STSGSNYPDRLMCLNAPILSDTDCRNSYPGEITNNMFCAGFLEGGKDSCQ 196
Q6R671Oreochromis_niloticus               STSGSNYPDRLMCLNAPILSDTDCRNSYPGEITNNMFCAGFLEGGKDSCQ 196
P35033Salmo_salarIII                      SGSSSNYPDTLRCLDLPILSSSSCNSAYPGQITSNMFCAGFMEGGKDSCQ 189
P07477Homo_sapiensI                       ASSGADYPDELQCLDAPVLSQAKCEASYPGKITSNMFCVGFLEGGKDSCQ 197
P07478Homo_sapiensII                      LSSGADYPDELQCLDAPVLSQAECEASYPGKITNNMFCVGFLEGGKDSCQ 197
P06872Canis_familiarisII                  LSSGTNYPELLQCLDAPILTQAQCEASYPGQITENMICAGFLEGGKDSCQ 197
P00763Rattus_norvegicusII                 LSSGVNEPDLLQCLDAPLLPQADCEASYPGKITDNMVCVGFLEGGKDSCQ 197
Q792Z1Mus_musculusI                       LSSGVNNPDLLQCLDAPLLPQADCEASYPGKITKNMICVGFLEGGKDSCQ 197
P07146Mus_musculusII                      LSNGVNNPDLLQCVDAPVLPQADCEASYPGDITNNMICVGFLEGGKDSCQ 197
P00762Rattus_norvegicusI                  LSNGVNNPDLLQCVDAPVLSQADCEAAYPGEITSSMICVGFLEGGKDSCQ 197
P00760Bos_taurusI                         KSSGTSYPDVLKCLKAPILSDSSCKSAYPGQITSNMFCAGYLEGGKDSCQ 194
P06871Canis_familiarisI                   QSIGQNYPDVLQCLKAPILSDSVCRNAYPGQISSNMMCLGYMEGGKDSCQ 197
P00761Sus_scrofaI                         KSSGSSYPSLLQCLKAPVLSDSSCKSSYPGQITGNMICVGFLEGGKDSCQ 182
P08426Rattus_norvegicusIII                LSSGTNYPSLLQCLDAPVLSDSSCKSSYPGKITSNMFCLGFLEGGKDSCQ 198
P70059Xenopus_laevisII                    LSSGTNYPDLLQCLNAPILTDSQCSNSYPGEITKNMFCAGFLAGGKDSCQ 194
P19799Xenopus_laevisI                     LSNGSNYPDLLQCLNAPILTNAQCNSAYPGEITANMICVGYMEGGKDSCQ 194
                                               .  . * *:. *:*.   *  :*** ::  *.* *:: *******
```

```
                                  C201          C220          C232
Q8AV11Oncorhynchus_keta           GDSGGPVVCNGELQGVVSWGYGCAEPGNPGVYAKVCIFNDWLTSTMATY- 221
P35032Salmo_SalarII               GDSGGPVVCNGELQGVVSWGYGCAEPGNPGVYAKVCIFNDWLTSTMATY- 231
P35031Salmo_salarI                GDSGGPVVCNGELQGVVSWGYGCAEPGNPGVYAKVCIFNDWLTSTMASY- 242
Q91515Fugu_rubripes               GDSGGPVVCNNELQGVVSWGYGCAERDHPGVYAKVCLFNDWLESTMASY- 237
Q92099Paranotothenia_magellani    GDSGGPVVCNGELQGVVSWGYGCAERDHPGVYAKVCLN-DWLETSMANY- 240
Q9W7Q7Paralichthys_olivaceus      GDSGGPVVCNGELQGVVSWGYGCAERDHPGVYARVCIFIDWLETTMASY- 242
Q9W7Q6Paralichthys_olivaceusII    GDSGGPVVCNGQLQGVVSWGYGCAQRDHPGVYAKVCIFIDWLERTMSSY- 238
O93266Pleuronectes_americanus     GDSGGPVICNGELQGVVSWGYGCAERGNPGVYAKVCLFNDWLESTMASY- 242
Q6RI79Tautogolabrus_adspersus     GDSGGPVVCNGQLQGVVSWGYGCAERDHPGVYAKVCLFNDWLESTMASN- 242
P16049Gadus_morhuaI               GDSGGPVVCNGVLQGVVSWGYGCAERDHPGVYAKVCVLSGWVRDTMANY- 241
Q91041Gadus_morhuaII              GDSGGPVVCNGVLQGVVSWGYGCAERDNPGVYAKVCVLSGWVRDTMASY- 241
Q98TG9Engraulis_japonicusII       GDSGGPVVCNGELQGIVSWGYGCAERDHPGVYAKVCLFNDWIDSTMAQYN 241
Q98TH0Engraulis_japonicusI        GDSGGPVVCNGELQGVVSWGYGCAERDHPGVYAKVCIFTDWLQSTMASN- 240
Q7SX90Brachydanio_rerio           GDSGGPVVCNGELHGIVSWGYGCAEKNHPGVYGKVCMFSQWIADTMRNN- 242
Q8QGW3Anguilla_japonica           GDSGGPVVCNGELQGVVSWGYGCAEQNHPGVYPKVCMFSDWLRTTMAST- 244
Q7T1R8Pangasius_hypophthalmus     GDSGGPVVCNGELQGIVSWGYGCAEKNHPGVYTKVCIFTDWIAQTIASN- 242
Q6R670Oreochromis_aureus          GDSGGPVVCNGQLQGIVSWGYGCAQRNRPGVYTKVCNYNSWISNTMANN- 245
Q6R671Oreochromis_niloticus       GDSGGPVVCNGQLQGIVSWGYGCAQRDRPGVYTKVCNYNSWISNTMANN- 245
P35033Salmo_salarIII              GDSGGPVVCNGQLQGVVSWGYGCAQRNKPGVYTKVCNYRSWISSTMSSN- 238
P07477Homo_sapiensI               GDSGGPVVCNGQLQGVVSWGDGCAQKNKPGVYTKVYNYVKWIKNTIAANS 247
P07478Homo_sapiensII              GDSGGPVVSNGELQGIVSWGYGCAQKNRPGVYTKVYNYVDWIKDTIAANS 247
P06872Canis_familiarisII          GDSGGPVVCNGELQGIVSWGYGCAQKNKPGVYTKVCNFVDWIQSTIAANS 247
P00763Rattus_norvegicusII         GDSGGPVVCNGELQGIVSWGYGCALPDNPGVYTKVCNYVDWIQDTIAAN- 246
Q792Z1Mus_musculusI               GDSGGPVVCNGQLQGIVSWGYGCAQKDNPGVYTKVCNYVDWIQNTIAAN- 246
P07146Mus_musculusII              GDSGGPVVCNGELQGIVSWGYGCAQPDAPGVYTKVCNYVDWIQNTIADN- 246
P00762Rattus_norvegicusI          GDSGGPVVCNGQLQGIVSWGYGCALPDNPGVYTKVCNFVGWIQDTIAAN- 246
P00760Bos_taurusI                 GDSGGPVVCSGKLQGIVSWGSGCAQKNKPGVYTKVCNYVSWIKQTIASN- 243
P06871Canis_familiarisI           GDSGGPVVCNGELQGVVSWGAGCAQKGKPGVSPKVCKYVSWIQQTIAAN- 246
P00761Sus_scrofaI                 GDSGGPVVCNGQLQGIVSWGYGCAQKNKPGVYTKVCNYVNWIQQTIAAN- 231
P08426Rattus_norvegicusIII        GDSGGPVVCNGQLQGVVSWGYGCAQKGKPGVYTKVCNYVNWIQQTVAAN- 247
P70059Xenopus_laevisII            GDSGGPVVCNGQLQGVVSWGYGCAQRNYPGVYTKVCNFVTWIQSTISSN- 243
P19799Xenopus_laevisI             GDSGGPVVCNGQLQGVVSWGYGCAMRNYPGVYTKVCNYNAWIQNTIAAN- 243
                                  *******:... *:*:**** ***
```

Figure 1. Multiple sequence alignment of mammalian and fish trypsins. Sequences of mammalian and fish trypsins are indicated with their SwissProt code. The identical (*) and similar residues (: or .) are highlighted. The calcium-binding loop, the conserved cysteine residues, and the putative autolytic sites are highlighted in gray and boldfaced.

Molecular Dynamics. ST and BT were chosen as representative structures for fish and mammalian trypsins, respectively, since they have been thoroughly characterized experimentally.[1,3,16] Multiple 6 ns MD simulations of calcium-bound (holoBT) and calcium-free (apoBT) bovine trypsins, as well as calcium-bound (holoST) and calcium-free (apoST) atlantic salmon trypsins, were carried out at both 283 and 310 K, as outlined in the Methods. The choice of two different temperatures was motivated by the observation that BT and ST belong to mesophilic and psycrophilic organisms, respectively. However, results related to the effects of Ca^{2+} on the structure and dynamic properties discussed in the present contribution are essentially independent from the chosen temperature values (not shown). Therefore, only results for MD runs obtained at 310 K will be presented and discussed.

Several criteria have been used to evaluate convergence of the MD runs. Indeed, it is well-known that multiple trajectories help to identify recurring features and to avoid artifacts arising from the simulation procedure.[44] Therefore, to efficiently sample the conformational space, two or three independent MD simulations were carried out starting from the same atomic coordinates but using different initial velocities randomly taken from a Maxwellian distribution compatible with the simulation temperature. In the following, MD trajectories collected for the same system but characterized by different initial velocities are labeled as simX (as a subscript), where X runs from 1 to the number of MD simulations carried out for a particular system.

The rmsd value calculated for the main-chain atoms of ST and BT (Figure 3) reached a stable value after about 1.5 ns in all simulations, with the exception of $apoBT_{sim2}$ and $holoST_{sim1}$, which reach convergence after 2.7 ns. As a consequence, the analyses of MD trajectories have been carried out using only the last 4.5 or 3.3 ns.

The interatomic distance between the atoms of the coordinating residues (70, 72, 75, 77, and 80) and Ca^{2+}, as well as the protein gyration radius and total and potential energies of the system, are characterized by negligible fluctuations throughout the simulations, indicating stable trajectories (data not shown).

The removal of Ca^{2+} did not significantly affect the secondary structure content in BT and ST, possibly because of the presence of six disulfide bridges that stabilize the N-terminal (C42−C58) and C-terminal domains (C128−C232, C136−C201, C168−C182, and C191−C220) and cross-link the N- and the C-terminal domains (C22−C157). However, it was generally possible to observe a slight reduction of the average number of residues characterized by regular secondary structure in trajectories of the apo forms (Table 1). Since the N-terminal domain presents a fewer number of disulfide bridges, the secondary structure content was also evaluated separately for the two domains. The behavior of C- and N-terminal domains is similar, the only

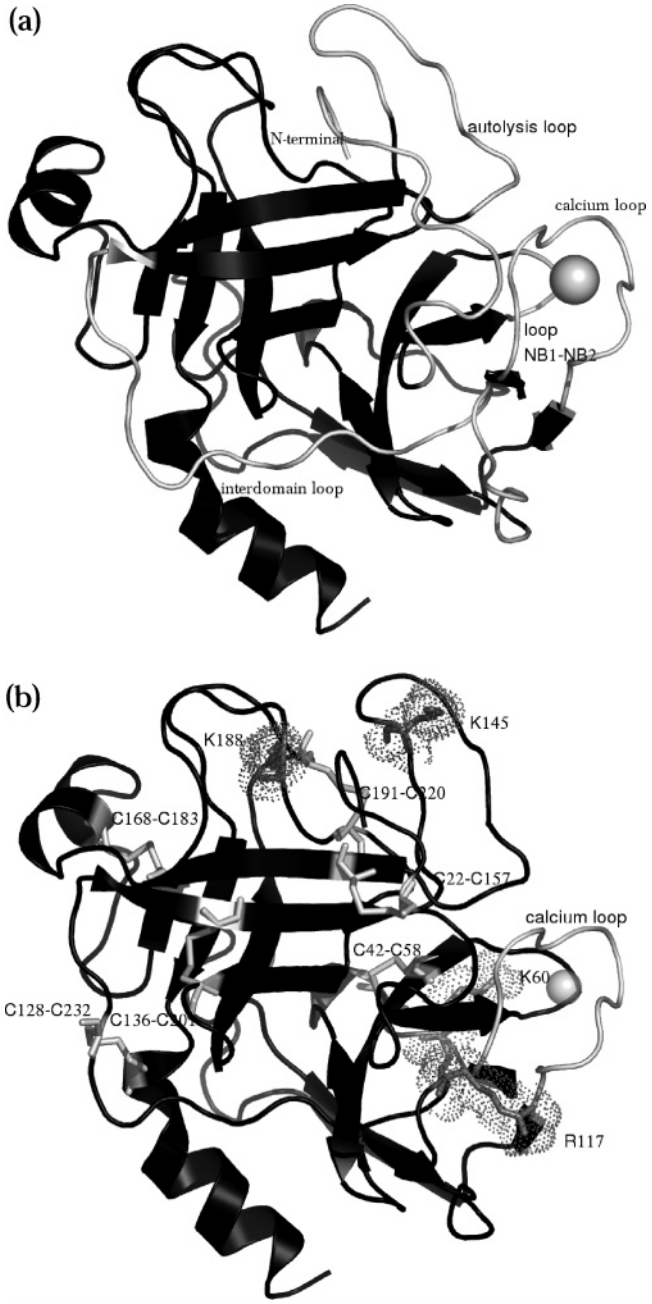


Figure 2. 3D structure of BT trypsin. (a) The calcium ion and the loop regions which surround the calcium-binding loop are shown in gray and light gray, respectively. (b) The calcium ion, the six conserved disulfide bridges, and the putative autolysis sites are shown as gray spheres, light gray sticks, and light gray sticks and dots, respectively. The secondary structure elements are shown as ribbons.

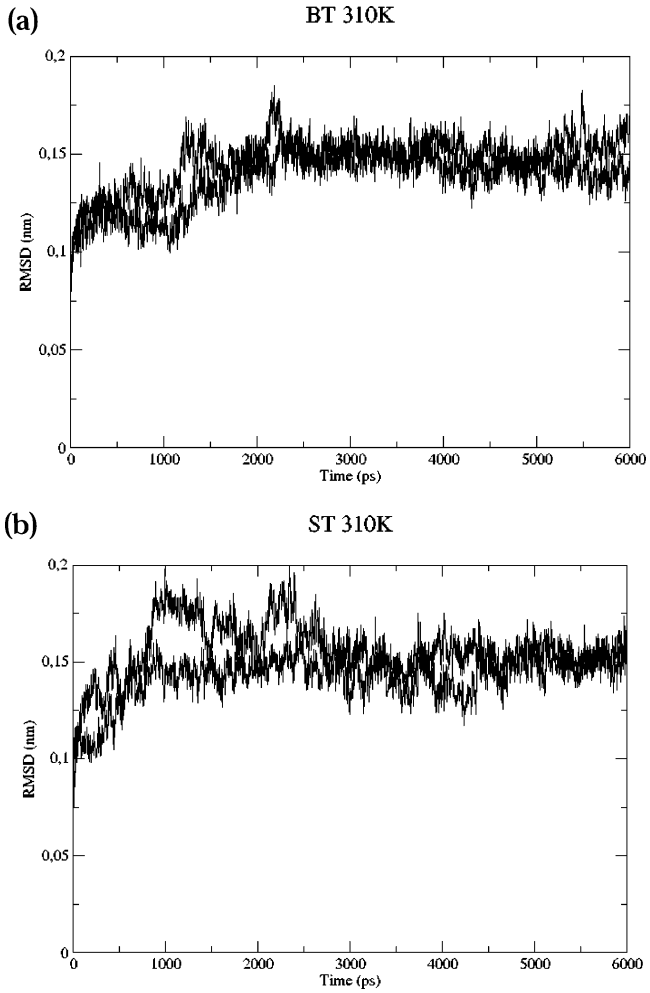


Figure 3. rmsd of main-chain atoms as a function of time. (a) rmsd as a function of time of BT at 310 K for the holo (black line) and apo forms (gray line). (b) rmsd as a function of time of ST at 310 K for the holo (black line) and apo forms (gray line). For the sake of clarity, only values relative to the simulation sim1 are shown.

appreciable difference being that the N-terminal domain in the apo forms is generally characterized by a lower number of residues in β-strand conformations (not shown).

The evaluation of conformational sampling in MD trajectories is crucial to avoid a misleading interpretation of MD simulations.[44,45] A simple but efficient index of sampling is the extension of resampling in the phase space. The system revisiting similar structures is a good indication of the convergence of the simulation on structures that represent the conformational properties of the system. Therefore, multiple trajectories, obtained by different Maxwellian distributions of initial velocities, have been merged and analyzed. In particular, the data from rmsd matrices, obtained as outlined in the Methods, were used to carry out a cluster analysis and obtain further insight in the configurations visited by the system. Only one cluster has been identified in the MD trajectories for the holo forms, indicating that simulations initialized with different Maxwellian distributions converge to similar conformational basins. On the other hand, all the MD trajectories for the apo forms presented two different clusters, indicating a greater conformational freedom of calcium-free structures. However, when the secondary structure content of the average structures obtained from the cluster analysis (see Method) was analyzed, the only appreciable difference observed comparing the apo and holo forms was a greater disorder for the Nβ-2 and Nβ-5 strands in the apo forms, in agreement with the above observations about the secondary structure content of the N-terminal domain. The average 3D structures obtained from the conformational clustering were analyzed also to evaluate the structural features of some key regions of the protein: the

Table 1. Average Number of Residues in a Given Secondary Structure of BT and ST at 310 K in the Holo and Apo Forms, According to the DSSP Software[41] [a]

	BT				ST			
	$holo_{sim1}$	apo_{sim1}	$holo_{sim2}$	apo_{sim2}	$holo_{sim1}$	apo_{sim1}	$holo_{sim2}$	apo_{sim2}
structure	121.9 (5.7)	120.8 (5.2)	118.2 (5.3)	117.2 (5.3)	125.8 (5.5)	118.1 (5.2)	118.0 (5.9)	114.2 (7.0)
coil	58.6 (3.3)	58.8 (3.5)	60.9 (3.4)	60.0 (3.4)	55.6 (3.1)	61.3 (3.3)	61.0 (3.6)	63.1 (4.1)
β sheet	71.4 (3.3)	70.5 (3.4)	69.4 (3.7)	69.6 (3.6)	70.0 (3.2)	70.8 (3.4)	68.2 (3.3)	65.3 (4.1)
β bridge	6.6 (2.3)	5.0 (1.8)	5.0 (1.9)	7.2 (2.3)	6.4 (1.9)	4.8 (1.9)	6.2 (1.7)	6.2 (2.2)
bend	38.9 (4.1)	38.5 (3.8)	41.8 (3.8)	38.3 (4.1)	38.7 (4.4)	40.3 (4.2)	38.1 (4.4)	40.9 (4.8)
turn	26.1 (3.7)	26.2 (4.0)	25.4 (3.7)	26.5 (3.6)	30.1 (4.2)	23.0 (4.0)	27.1 (4.7)	24.9 (4.6)
helix	17.8 (2.0)	18.0 (2.1)	16.4 (1.5)	16.9 (1.5)	19.2 (1.7)	20.4 (1.2)	16.3 (1.3)	17.7(1.9)
3_{10} helix	3.6 (3.7)	3.9 (4.1)	5.2 (2.2)	4.5 (2.4)	1.9 (4.2)	2.4 (4.0)	4.8 (2.6)	3.8 (2.9)

[a] For the sake of clarity, only values relative to the simulations sim1 and sim2 are shown. Standard deviations are given in parentheses.

catalytic triad (H57, D102, and S195), the specificity pocket (D189, G216, and V226), and the hydrophobic pair I73−W141. The structural properties of the catalytic triad and of the specificity pocket are always very similar in the apo and holo forms. The same holds true for the hydrophobic pair I73−W141, which, according to experimental data,[9,10] might be implicated in the transmission of the structural modifications observed upon calcium removal. In fact, the interatomic distances between the C atoms of I73 and W141 are nearly constant during all MD simulations. This indicates that no relevant changes in the interaction between W141 and I73 take place upon calcium removal.

The rmsd per residue from the average structure (rmsf) was calculated to better evaluate the role of Ca^{2+} on protein flexibility (Figure 4).[47] The overall profiles of the rmsf deviations are quite similar for the apo and holo forms, as well as for ST and BT: the rmsf peaks are located in corresponding positions, whereas the intensity of the fluctuations is generally larger in the apo-form simulations. Regions characterized by a regular secondary structure show small fluctuations during the simulations, whereas pronounced fluctuations are observed for some loop regions.

To better highlight regions characterized by different flexibility in the apo and holo forms, the rmsf profile for each holo simulation was subtracted from the corresponding profile obtained for the apo form (rmsf-diff; Figure 4); values of rmsf-diff lower than zero indicate regions where the fluctuations are larger for the apo form. All proteins present a larger number of residues with rmsf-diff values lower than zero, confirming the previous observations indicating a larger flexibility for the apo forms. In particular, BT and ST show 127 and 135 residues (out of 223 and 222 total residues), respectively, with an rmsf-diff value lower than zero. BT and ST trajectories are characterized by similar rmsf-diff profiles, and most importantly, the lack of Ca^{2+} has effects not only on the N-terminal domain, where the calcium-binding site is localized, but also on the C-terminal domain.

Protein portions characterized by rmsf-diff values lower than −0.05 nm (greater flexibility in the apo form) or higher than 0.05 nm (greater flexibility in the holo form) have been highlighted on the average three-dimensional structures of BT (Figure 5). In the N-terminal domain, the regions characterized by enhanced fluctuations in the apo forms are the calcium-binding loop between Nβ-4 and Nβ-5 (in particular, the region 72−80), the loop at the N end (in particular the region 23−28), and part of the interdomain loop, between Nβ-6 and Cβ-1 (in particular, the regions 93−103 and 110−119) (Figure 5). Notably, all the latter regions surround the calcium-binding loop. Fluctuation intensity changes significantly in three regions of the C-terminal domain upon Ca^{2+} removal (Figure 5): residue 141 in ST and the pair of interacting loops between Cβ-3 and Cβ-4 (in particular, the region 184−191) and between Cβ-5 and Cβ-6

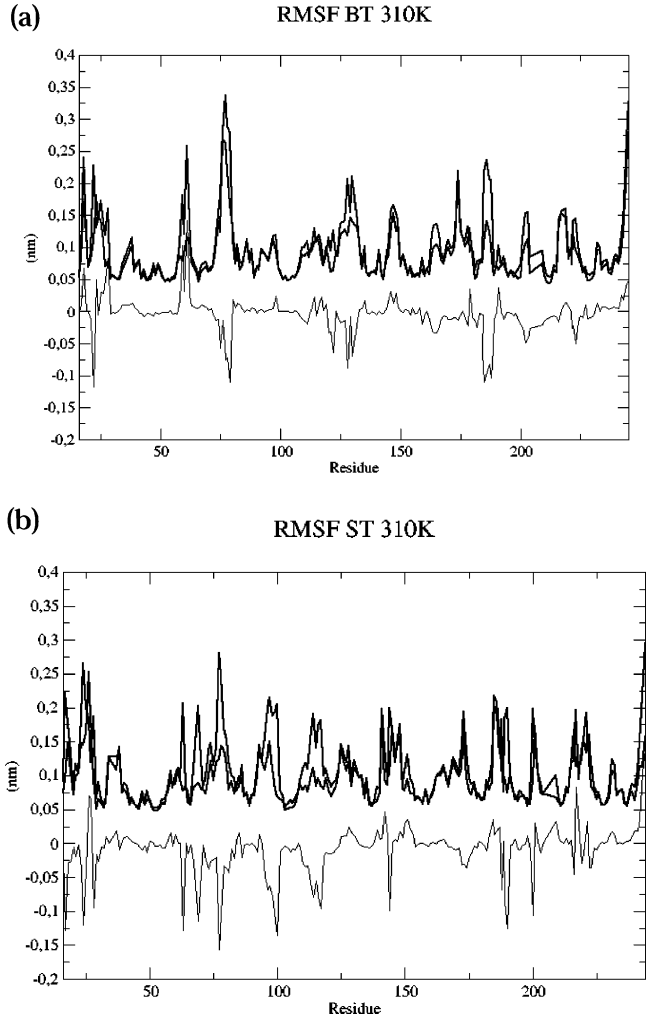


Figure 4. rmsf and rmsf-diff as a function of residue numbers. (a) rmsf values of each residue of the holo (black line) and apo forms (gray line) of BT at 310 K. (b) rmsf values of each residue of the holo (black line) and apo forms (gray line) of ST at 310 K. The rmsf-diff is indicated by a thick black line.

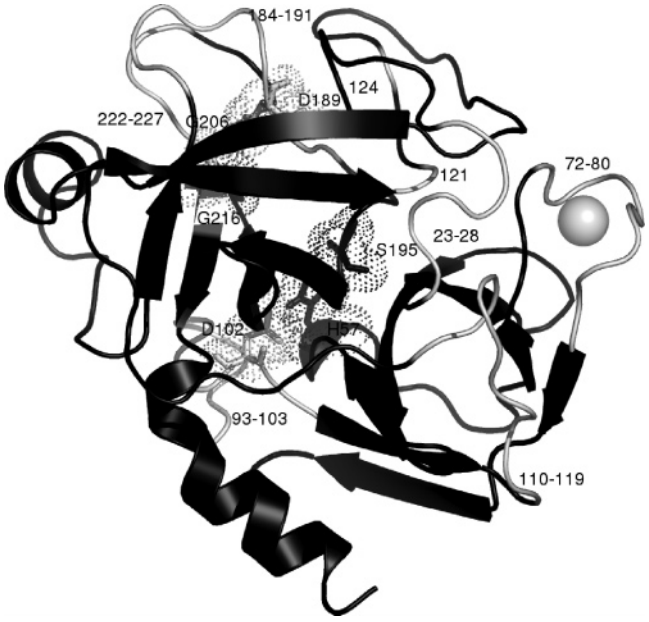


Figure 5. Regions characterized by greater fluctuation in the apo-trypsin structure. The calcium ion and the regions characterized by rmsf values greater in apo BT than those in holo BT are indicated in gray and light gray, respectively. The secondary structure elements are shown as ribbons. The catalytic triad (S195, H57, and D102) and the specificity pocket amino acids (G216, G206, and D189) are shown as sticks and dots.

(in particular, the region 222−227). The two latter loops are located in proximity of the substrate specificity pocket. Therefore, the removal of Ca^{2+} not only increases the flexibility of regions around its binding site in the N-terminal domain but also leads to channeling of the fluctuations to remote sites in the C-terminal domain. In particular, the analysis of the rmsf data in light of the three-dimensional structure of BT and ST suggests that the channeling could involve the interdomain loop, which is a long disordered region that connects the two protein domains.

The loop between Nβ-1 and Nβ-2, where the primary autolysis site K60 is located, is the only region surrounding the calcium-binding loop whose flexibility is not influenced by calcium removal, as a result of the presence of a disulfide bridge (C42−C58) that locally constrains the structure. The behavior of the autolysis loop, where BT K145 and ST K154 are located, can be explained by similar considerations. In fact, three disulfide bridges are located around this region: C22−C157, C136−C201, and C168−C192. Among the primary sites of autolysis in BT, R117, which is located in the interdomain loop, and K188 are the only residues characterized by significantly larger flexibility in the apo form. Notably, in ST, R117 is not conserved and the flexibility of K188 does not increase significantly upon Ca^{2+} removal. The latter observation is in good agreement with the experimental data indicating weak dependence of the autolysis rate on Ca^{2+} binding in ST.[16−18] The reliability of the rmsf analysis has been validated, splitting the MD trajectories (after equilibration) and computing the rmsf values for R117 and K188 separately for every trajectory portion (Table 2).

Table 2. rmsf Values for R117 and K188[a]

	R117	K188
holo $BT_{merged\ sim}$	0.101	0.064
holo BT_{sim1}	0.083	0.053
holo $BT_{sim1\ (first\ half)}$	0.057	0.048
holo $BT_{sim1\ (second\ half)}$	0.090	0.053
holo BT_{sim2}	0.117	0.069
holo $BT_{sim2\ (first\ half)}$	0.126	0.070
holo $BT_{sim2\ (second\ half)}$	0.055	0.063
holo BT_{sim3}	0.075	0.059
holo $BT_{sim3\ (first\ half)}$	0.094	0.061
holo $BT_{sim3\ (second\ half)}$	0.069	0.055
apo $BT_{merged\ sim}$	0.130	0.115
apo BT_{sim1}	0.154	0.107
apo $BT_{sim1\ (first\ half)}$	0.148	0.080
apo $BT_{sim1\ (second\ half)}$	0.123	0.090
apo BT_{sim2}	0.107	0.112
apo $BT_{sim2\ (first\ half)}$	0.096	0.090
apo $BT_{sim2\ (second\ half)}$	0.144	0.089
holo $ST_{merged\ sim}$		0.099
holo ST_{sim1}		0.088
holo $ST_{sim1(first\ half)}$		0.102
holo $ST_{sim1\ (second\ half)}$		0.071
holo ST_{sim2}		0.065
holo $ST_{sim2\ (first\ half)}$		0.064
holo $ST_{sim2\ (second\ half)}$		0.071
apo $ST_{merged\ sim}$		0.095
apo ST_{sim1}		0.051
apo $ST_{sim1\ (first\ half)}$		0.050
apo $ST_{sim1\ (second\ half)}$		0.051
apo ST_{sim2}		0.093
apo $ST_{sim2\ (first\ half)}$		0.089
apo $ST_{sim2\ (second\ half)}$		0.096

[a] The rmsf values for the merged trajectories are highlighted in bold. To evaluate the reliability of the rmsf analysis, single trajectrories have also been split and rmsf values have been computed for every trajectory portion. Values in nm. R177 values for ST are not reported because this amino acid is not conserved in fish trypsins.

As for other putative autoproteolysis sites, only ST K23, ST K74 (in the calcium-binding loop), BT K222, and BT K224 are localized in regions characterized by relevant rmsf differences between the apo and holo forms.

Since the susceptibility to autoproteolysis has been proposed to be related to both flexibility[19] and solvent accessibility of the autolysis sites,[20] the solvent-accessible area of putative autolysis sites has been monitored during MD simulations. The analysis of trajectories of the holo forms reveals that some putative autolysis sites located in loop regions (ST R62, ST K74, BT K109, ST K110, BT K204, BT K222, BT K224) present large solvent accessibility values when compared to primary autolysis sites (Table 3). In addition, the primary autolysis site BT R117, which is characterized by relatively high flexibility (Figure 4), presents low values of solvent accessibility.

The comparison of solvent accessibility values for primary autolysis sites in holo and apo simulations (Table 3) reveals that the removal of Ca^{2+} causes only small effects on some putative autolysis sites (K23, R62, R66, K110, and K154), whereas all other sites are essentially not affected by Ca^{2+} removal. In particular, solvent accessibility for BT K145 and ST K154 are not significantly different in the apo and holo

Table 3. Average Solvent Accessibility Degree (%) of Lysine and Arginine Side Chains for Putative Autolysis Sites in Holo and Apo BT and ST, According to the NACCESS Program[42] [a]

	BT		ST	
	holo	apo	holo	apo
K23	- - -	- - -	89.2 (10.0)	74.0 (17.7)
K60	**54.9 (8.5)**	**52.3 (8.1)**	**51.9 (6.9)**	**46.0 (7.6)**
R62	- - -	- - -	71.1 (10.5)	65.4 (7.6)
R66	21.9 (9.1)	17.9 (5.1)	27.2 (7.9)	19.0 (7.3)
K74	- - -	- - -	72.9 (5.8)	73.8 (6.1)
K87	64.2 (9.7)	63.0 (9.3)	63.0 (12.7)	61.3 (7.9)
R90	- - -	- - -	38.1 (7.7)	39.0 (6.1)
K107	26.5 (7.0)	27.5 (7.9)	31.8 (6.9)	26.7 (4.6)
K109	66.4 (6.2)	69.6 (6.5)	- - -	- - -
K110	- - -	- - -	81.0 (8.6)	72.3 (10.8)
R117	**20.3 (14.6)**	**14.7 (6.0)**	- - -	- - -
K145	**63.4 (6.1)**	**63.2 (5.8)**	- - -	- - -
K154	- - -	- - -	43.9 (6.1)	38.8 (7.2)
K156	26.6 (3.8)	25.8 (3.8)	- - -	- - -
K159	51.6 (5.8)	54.2 (5.7)	- - -	- - -
K169	48.8 (8.3)	52.0 (11.3)	- - -	- - -
K188	**40.5 (8.7)**	**51.5 (10.3)**	**45.0 (7.7)**	**52.2 (9.5)**
K204	56.0 (7.7)	55.4 (8.1)	- - -	- - -
K222	82.2 (7.8)	81.6 (8.2)	- - -	- - -
K224	40.1 (7.5)	37.1 (7.6)	- - -	- - -
K230	9.6 (3.0)	15.8 (7.5)	16.1 (6.0)	16.8 (4.9)
K239	72.0 (8.2)	70.8 (8.3)	- - -	- - -

[a] Standard deviations are given in parentheses. The primary autolysis sites are highlighted in bold, and - - - indicates that, in the corresponding position, lysine or arginine residues are missing.

Table 4. Average Solvent Accessibility Degree of Side Chains of W141 and Hydrophobic Core Residues (%) of BT and ST at 310 K in the Holo and Apo Forms, According to the NACCESS Program[a]

	BT		ST	
	holo	apo	holo	apo
W141	2.2 (1.3)	1.7 (1.3)	1.0 (0.9)	1.4 (1.0)
47	3.0 (2.6)	2.2 (2.2)	2.5 (3.0)	2.1 (2.1)
238	2.3 (2.7)	1.0 (1.8)	0.6 (1.0)	0.2 (0.4)
242	25.7 (6.3)	28.0 (5.8)	9.5 (4.4)	14.2 (3.8)
63	4.9 (3.9)	4.3 (3.5)	1.9 (1.7)	0.04 (0.2)
85	0.7 (1.8)	0.3 (1.3)	0.5 (1.1)	0.7 (2.0)
88	4.2 (6.7)	2.4 (4.4)	2.0 (2.7)	2.7 (4.1)
138	0.2 (0.5)	0.1 (0.2)	0.5 (0.7)	0.4 (0.9)
160	0.1 (0.5)	0.5 (1.3)	1.2 (1.4)	0.2 (0.3)
183	0.2 (0.4)	0.1 (0.3)	0.1 (0.3)	0.03 (0.2)

[a] Standard deviations are given in parentheses.

forms. This may be due to the presence of two stable ion pairs (K154−D151 and H71−E21, in ST) in their proximity. On the other hand, relevant differences in solvent accessibility are evident for BT R117 and K188 when comparing apo and holo trajectories.

Experimental data[6−10] indicate that trypsin assumes a more compact structure upon calcium coordination. In particular, the environment of W141 is affected by metal binding. Analyses of MD trajectories indicate that the solvent accessibility degree for W141 and amino acids forming the three trypsin core regions[48] (47, 238, 242; 63, 85, 88; 138, 160, 183) remains nearly unchanged upon calcium removal (Table 4). The only significant difference is observed for residue 242 (Ile in BT and Met in ST), which is more solvent-exposed and, consequently, less involved in hydrophobic interactions in the apo forms.

Conclusions

The present investigation has probed some effects of Ca^{2+} binding to trypsins, correlating conformational and dynamic properties to relevant functional features of the protein.

Several putative autoproteolytic sites (lysine and arginine residues) are present in ST and BT, and some of the sites are characterized by suitable features for proteolytic cleavage in terms of location in the loop regions, lack of ion pairs or disulfide bridges in their proximity, and solvent accessibility. Most of the potential trypsin-sensitive sites are located in the K60−R117 stretch, which does not include disulfide bridges and which was previously proposed as a built-in target for autolysis.[13]

The removal of Ca^{2+} affects the structural and dynamic properties of specific regions of trypsin. The effects caused by calcium removal on the three-dimensional structure are more pronounced in the N-terminal, where a decrease in the number of residues in β-strand conformations is observed, than in the C-terminal domain. In fact, the N-terminal domain presents a low number of disulfide bridges. Ca^{2+} removal increases the flexibility of regions around its binding site but also leads to channeling of the fluctuations to sites located in the C-terminal domain. This observation is in agreement with other general mechanisms by which the signal induced by metal binding is transmitted to remote regions in the 3D structure.[26−29] In particular, the long disordered interdomain loop, which connects the two globular trypsin domains, is proposed to be involved in the transmission of signals correlated to Ca^{2+} binding.

It was previously observed that the autoproteolysis rate at R117 in rat trypsin is strongly decreased upon Ca^{2+} binding, leading to the suggestion that metal binding can stabilize an autolysis-resistant conformation of the protein.[13] The present results reveal that R117 (in the interdomain loop) and K188 (in the C-terminal domain) are the only primary autolysis sites which are strongly influenced by Ca^{2+} binding in BT, disclosing the molecular relationship connecting Ca^{2+} binding to autoproteolysis propensity in mammalian trypsins. This conclusion is in nice agreement with a site-directed mutagenesis experiment, which indicated that R117 replacement in rat trypsin leads to proteins almost completely resistant to autolysis.[13,15] Remarkably, R117 is not conserved in fish trypsins and the flexibility of K188 is not significantly affected by Ca^{2+} removal, in agreement with the observed weak dependence of autolysis propensity on Ca^{2+} binding in fish trypsins.[16−18]

Acknowledgment. The authors thank L. Riccardi and R. Gonella Diaza for fruitful discussions and comments.

References

(1) Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **2002**, *102*, 4501−4524.

(2) Burdette, S. C.; Lippard, S. J. Meeting of the minds: metalloneurochemistry. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3605−3610.

(3) Smalas, A. O.; Heimstad, E. S.; Hordvik, A.; Willassen, N. P.; Male, R. Cold adaptation of enzymes: structural comparison between salmon and bovine trypsins. *Proteins* **1994**, *20*, 149−166.

(4) Read, R. J.; James, M. N. Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. *J. Mol. Biol.* **1988**, *200*, 523−551.

(5) Toth, M. S.; Graf, L.; Toth, M. Trypsinogen stabilization by mutation Arg 117→His: a unifying pathomechanism for hereditary pancreatitis? *Biochem. Biophys. Res. Commun.* **1999**, *264*, 505−508.

(6) Sipos, T.; Merkel, J. R. An effect of calcium ions on the activity, heat stability and structure of trypsin. *Biochemistry* **1970**, *9*, 2766−2775.

(7) Gabel, D.; Kasche, V. Autolysis of beta-trypsin. Influence of calcium ions and heat. *Acta Chem. Scand.* **1973**, *27*, 1971−1981.

(8) Bulaj, G.; Otlewski, J. Denaturation of free and complexed bovine trypsinogen with the calcium ion, dipeptide Ile-Val and basic pancreatic trypsin inhibitor. *Eur. J. Biochem.* **1994**, *223*, 939−946.

(9) Bulaj, G.; Otlewski, J. Ligand-induced changes in the conformational stability of bovine trypsinogen and their implications for the protein function. *J. Mol. Biol.* **1995**, *247*, 701−716.

(10) Buono, R. A.; Prestrelski, S. J.; Liebman, M. N.; Byler, D. M. Infrared spectroscopic studies of calcium binding to inhibited beta-trypsins. *Biochim. Biophys. Acta* **1994**, *204*, 124−128.

(11) Smith, R. L.; Shaw, E. Pseudotrypsin. A modified bovine trypsin produced by limited autodigestion. *J. Biol. Chem.* **1969**, *244*, 4704−4712.

(12) Maroux, S.; Desnuelle, P. On some autolyzed derivatives of bovine trypsin. *Biochim. Biophys. Acta* **1969**, *181*, 59−72.

(13) Varallyay, E.; Pal, G.; Patthy, A.; Szilagyi, L.; Graf, L. Two mutations in rat trypsin confer resistance against autolysis. *Biochem. Biophys. Res. Commun.* **1998**, *243*, 56−60.

(14) Kaslik, G.; Patthy, A.; Balint, M.; Graf, L. Trypsin complexed with alpha 1-proteinase inhibitor has an increased structural flexibility. *FEBS Lett.* **1995**, *370*, 179−183.

(15) Li, X. F.; Nie, X.; Tang, J. G. Anti-autolysis of trypsin by modification of autolytic site Arg117. *Biochem. Biophys. Res. Commun.* **1998**, *250*, 235−239.

(16) Outzen, H.; Berglund, G. I.; Smalas, A. O.; Willassen, N. P. Temperature and pH sensitivity of trypsins from Atlantic Salmon in comparison with bovine and porcine trypsin. *Comp. Biochem. Physiol. B* **1996**, *115*, 33−45.

(17) Hjelmeland, K.; Raa, J. Characteristics of two trypsin type isozymes isolated from the artic fish capelin. *Comp. Biochem. Physiol. B* **1982**, *71*, 557−562.

(18) Reeck, G. R.; Neurath, H. Pancreatic trypsinogen from the African lungfish. *Biochemistry* **1972**, *11*, 503−510.

(19) Fontana, A.; Fassina, G.; Vita, C.; Dal Zoppo, D.; Zamai, M.; Zambonin, M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **1986**, *25*, 1847−1951.

(20) Novotny, J.; Bruccoleri, R. E. Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Lett.* **1987**, *211*, 185−189.

(21) Norberg, J.; Nilsson, L. Advances in biomolecular simulations: methodology and recent applications. *Q. Rev. Biophys.* **2003**, *36*, 257−306.

(22) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular simulations: recent developments in force fields, simulations of enzyme-catalysis, protein−ligand, protein−protein, and protein−nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−243.

(23) Hansson, T.; Oostenbrink, C.; van Gusteren, W. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190−196.

(24) Heimstad, E. S.; Hansen, L. K.; Smalas, A. O. Comparative molecular dynamics simulation studies of salmon and bovine trypsins in aqueous solution. *Protein Eng.* **1995**, *8*, 379−388.

(25) Brandsdal, B. O.; Heimstad, E. S.; Sylte, I.; Smalas, A. O. Comparative molecular dynamics of mesophilic and psycrophilic protein homologues studied by 1.2 ns simulations. *J. Biomol. Struct. Dyn.* **1999**, *17*, 493−506.

(26) Iyer, L. K.; Qasba, P. K. Molecular dynamics simulation of α-lactalbumin and calcium binding c-type lysozyme. *Protein Eng.* **1999**, *12*, 129−139.

(27) Cates, S. M.; Teodoro, M. L.; Phillips, G. N. Molecular mechanism of calcium and magnesium binding to parvalbumin. *Biophys. J.* **2002**, *82*, 1133−1146.

(28) Yang, C.; Jas, G. S.; Kuczera, K. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *Biochim. Biophys. Acta* **2004**, *1697*, 289−300.

(29) Komeiji, Y.; Ueno, Y.; Uebayasi, M. Molecular dynamics simulations revealed Ca(2+)-dependent conformational change of calmodulin. *FEBS Lett.* **2002**, *521*, 133−139.

(30) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissing, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(31) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI−BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.

(32) Higgins, D.; Thompson, J.; Gibson, T.; Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalities and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.

(33) van der Spoel, D.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Hess, B.; Feenstra, K. A.; Lindhal, E.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual version 3.1.1*, * University of Groningen: Groningen, The Netherlands, 2002. www.gromacs.org.

(34) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr., Sect. B.* **1983**, *39*, 480.

(35) Smalas, A. O.; Hordvik, A.; Hansen, L. K.; Hough, E.; Jynge, K. Crystallization and preliminary X-ray crystallographic studies of benzamidine-inhibited trypsin from the North Atlantic Salmon (Salmo salar). *J. Mol. Biol.* **1990**, *214*, 355−358.

(36) Berendsen, H. J. C.; Postma, J. P. M.; van Gusteren, W. F.; Hermans, J. *Interaction models for water in relation to protein hydratation*; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.

(37) Berendsen, H. J. C.; Postma, J. P. M.; van Gusteren, W. F.; Di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(38) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraahje, J. G. E. M. LINCS: a linear constraint solver for molecular interactions. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(39) Darden, T. A.; York, D. M.; Pedersen, L. G. Particle mesh Ewald. An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(40) Jarvis, R. A.; Patrick, E. A. Clustering using a similarity measure based on shared nearest neighbors. *IEE Trans. Comput.* **1973**, *11*, 1025−1034.

(41) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.

(42) Hubbard, S. J.; Thortnon, J. M. *NACCESS*; 1993.

(43) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.

(44) Tsai, J.; Levitt, M.; Baker, D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J. Mol. Biol.* **1999**, *291*, 215−225.

(45) Hess, B. Convergence of sampling in protein simulations. *Phys. Rev. E.* **2002**, *65*, 031910/1−031910/10.

(46) Smith, L. J.; Daura, X.; van Gusteren, W. F. Assesing equilibration and convergence in biomolecular simulations. *Proteins* **2002**, *48*, 487−496.

(47) Note that the analysis of single trajectories led to similar results, with some differences in the intensity of the peaks in the rmsf plots.

(48) Leiros, H. K.; Willassen, N. P.; Smalas, A. O. Structural comparison of psycrophilic and mesophilic trypsins. Elucidating the molecular basis of cold-adaptation. *Eur. J. Biochem.* **2000**, *267*, 1039−1049.

CT050092O

JCTC Journal of Chemical Theory and Computation

Theoretical Group 14 Chemistry. 4. Cyclotriplumbanes: Relativistic and Substituents Effects

Rainer Koch,* Torsten Bruhn, and Manfred Weidenbruch

Institut für Reine und Angewandte Chemie, Universität Oldenburg, D-26111 Oldenburg, Germany

Received June 6, 2005

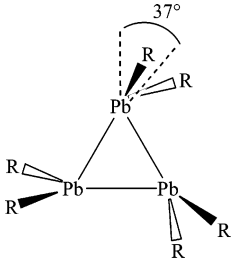
Abstract: We report a study on the first newly synthesized homonuclear lead ring system, the cyclotriplumbane Pb_3R_6. Its geometrical features can be best reproduced using perturbation theory (MP2) together with the Stuttgart-Dresden basis set and ECP for lead. The experimentally observed tilting of the groups R in the cyclotriplumbanes is attributed to the bonding situation: the lead−lead contacts, formed from weak interactions of plumbylene lone pairs with empty p orbitals of neighboring lead atoms, try to maximize overlap. Surprisingly and in contrast to the inert pair effect, hybridization of the former plumbylene lone pair orbitals in the cyclotriplumbane Pb_3R_6 is observed, depending on the substituent. Hybrid orbitals with a 6s orbital contribution of only about 40% are found. Hydrogen atoms and methyl groups promote this effect, while for phenyl substitution the expected 6s lone pair orbital is identified as the bond-forming orbital.

Introduction

The chemistry of the stable heavy elements of the periodic system is not only influenced by group tendencies but also by increasingly important relativistic effects.[2] One of the main consequences is the energetic lowering of 6s electrons (in case of lead) whose availability for chemical bonding is therefore limited.[3]

While homonuclear double bonds between silicon, germanium, and tin atoms are well-known in the literature for some time,[4] it has been possible only a few years ago to isolate the first representative of diplumbenes, compounds with a short Pb═Pb double bonds.[5] However, these bonds with distances of 2.90−3.05 Å are longer than those of a typical Pb−Pb single bond, for instance in hexaphenyldiplumbane with 2.844 Å.[6] In agreement with theoretical calculations on the parent compound $H_2Pb{=}PbH_2$,[7−9] the experimental bond lengths and the pronounced trans-bending of the substituents from the Pb−Pb vector indicate that the double bond in diplumbenes is formed by dual donor−acceptor interactions of the occupied $6s^2$ orbitals with the empty 6p orbitals of the neighboring lead atom.

Recently, we were able to obtain cyclotriplumbane **1** as the first homocyclic compound formed from lead atoms.[10] Lead−lead distances of averaged 3.18 Å together with a substituent twisting of 37° out of their ideal positions indicate that bonding in the three-membered ring is dominated by donor−acceptor interactions and not by overlapping sp-hybridized orbitals. However, the structural parameters of **1** are only in limited accordance with an early theoretical study on the parent system Pb_3H_6.[11] Despite the sterically undemanding hydrogen atoms, a longer Pb−Pb distance of 3.231 Å and trans-bent angles of 51° were determined. The employment of the outdated Hartree−Fock level of theory, a relatively small basis set for the geometry optimizations, and the restriction to hydrogen as substituent does not allow reliable conclusions.



1, R = 2,4,6-$Et_3C_6H_2$

* Corresponding author fax: +49-441-798-3329; e-mail: rainer.koch@uni-oldenburg.de.

10.1021/ct050144+ CCC: $30.25
Published on Web 10/12/2005

We think it is timely to obtain more information on the interplay of relativistic effects and substituent influence and have therefore embarked on such a study on cyclotriplumbane with different substituents and report herein the results.

Computational Details

In the present study all calculations at the Hartree−Fock (HF),[12] density functional theory (DFT, employing the hybrid functional B3LYP),[13,14] many-body perturbation theory methods (second-order Møller−Plesset (MP2),[15,16] and coupled-cluster theory including single, double, and triple excitations (CCSD(T))[17−19] were performed with Gaussian 98,[20] together with Pople's 6-31G(d)[21−23] basis set for C and H and the Stuttgart/Dresden (SDD) basis set and effective core potentials (ECP)[24,25] for Pb (denoted as "I"). For comparison, calculations with the LANL2DZ[26] basis set and pseudopotential for Pb ("II") and a triple-ζ ("TZ") base (cc-pVTZ-PP for lead[27] and 6-311G(2d,2p)[28,29] for other atoms) were also performed. Both the SDD and the LANL2DZ basis sets employed for lead are of double-ζ quality, but the former basis set is generally considered superior over the latter because polarization functions are included. The CCSD(T) calculations used the relatively small basis set I and the triple-ζ ("TZ") base. The nature of all optimized structures was confirmed as minima or transition state by calculating harmonic frequencies. Population analyses were performed with NBO[30,31] as implemented in Gaussian 98 and atoms-in-molecules investigations[32,33] with AIM2000[34] employing the MP2 wave function.

Results and Discussion

In a first step, Pb_3R_6 with R═H, Me, and phenyl were optimized at various levels of theory (see below and Table 1 for details). At the Hartree−Fock level of theory, Pb−Pb bond lengths are largely overestimated independent of the employed basis set/ECP for lead, compared to either the experimental data or coupled-cluster calculations which serve as reference data for the smaller groups where no experimental data are available. Hartree−Fock is not a sufficient level of theory to describe the geometry of cyclotriplumbanes! The general trend is an elongation of the Pb−Pb bond distances when increasing the steric demand of the substituents. DFT calculations with B3LYP differ only slightly from the MP2 results for R = H, but larger substituents result in a Pb−Pb bond length overestimation: for instance, in hexaphenylcyclotriplumbane, the distances are calculated to be 0.1 Å longer compared to the MP2 data. The latter calculations give the shortest Pb−Pb distances for R = Ph (3.216 Å) and the best agreement with the experimental result which surprisingly shows even shorter contacts despite bulkier groups (3.185 Å). The comparison between SDD and LANL2DZ shows that both basis sets give similar results. In combination with perturbation theory, SDD gives slightly shorter distances, while the other two approaches yield shorter contacts with LANL2DZ. However, SDD is generally considered to be superior due to the inclusion of polarization functions and is therefore to be preferred.

Although both theory and basis set should be systematically better than the other levels of theory, the "TZ" basis set in combination with perturbation and coupled cluster theory yields bond lengths that are significantly shorter (about 3.0 Å) than all other calculated and experimental data. One should therefore use these results with care as weak lead−lead interactions are rare and may be overestimated in this basis set. Also, these "TZ" calculations are very sensitive to the proper choice of the ECP. For instance use of the SDD ECP gives distances of 3.26 Å, much too long.

Table 1. Selected Geometrical Properties of Pb_3R_6 and **1** with Different Theoretical Methods Respectively from Experimental Results[a]

structure	method	Pb−Pb bond length	trans-bent angle
Pb_3H_6	HF/I	3.270 Å	49.9°
	HF/II	3.269 Å	50.8°
	B3LYP/I	3.165 Å	50.9°
	B3LYP/II	3.142 Å	51.7°
	MP2/I	3.132 Å	50.3°
	MP2/II	3.140 Å	51.6°
	MP2/TZ	2.972 Å	51.1°
	CCSD(T)/I	3.161 Å	51.4°
	CCSD(T)/TZ	3.026 Å	51.1°
$Pb_3(CH_3)_6$	HF/I	3.336 Å	44.5°
	HF/II	3.352 Å	45.1°
	B3LYP/I	3.215 Å	43.5°
	B3LYP/II	3.182 Å	42.6°
	MP2/I	3.161 Å	43.8°
	MP2/II	3.192 Å	44.2°
	MP2/TZ	2.986 Å	44.8°
Pb_3Ph_6	HF/I	3.579 Å	49.2°
	B3LYP/I	3.310 Å	47.1°
	MP2/I	3.216 Å	50.2°
$Pb_3(2,4,6\text{-}Me_3C_6H_2)_6$	B3LYP/I	3.364 Å	35.0°
	B3LYP/I	3.185 Å (constrained)	31.2°
1	exptl	3.185 Å	37°

[a] Basis set for C and H is 6-31G(d), basis set and ECP for Pb is I: SDD and II: LANL2DZ.

The twisting of the substituents is estimated to be about 50° for R = H and Ph, while it is slightly smaller with the sterically more demanding methyl group (∼ 44°). There is a discrepancy to the experimental value of 37° which is observed with the much larger 2,4,6-$Et_3C_6H_2$ substituent. B3LYP optimization (MP2 is not feasible) of $Pb_3(2,4,6\text{-}Me_3C_6H_2)_6$ results in a reduction of the trans-bent angles to 35°, now in excellent agreement with the experiment, indicating that a phenyl group is by no means an adequate geometry model for the bulky triethylphenyl substituent![1,35] To summarize, the MP2/I calculations give the best agreement with the experimental structure with B3LYP, and in particular HF methods show significant deficiencies in the prediction of Pb−Pb distances.

The reported structures with long lead−lead distances and our population analyses clearly indicate that these "bonds" are not classical single bonds but rather weak interactions between three plumbylene units. To obtain information on how weak these contacts really are, bond dissociation energies are determined. However, one has to keep in mind that these values include the release of ring strain of the three-

Table 2. Bond Dissociation Energies (kJ/mol) for the Reaction $Pb_3R_6 \rightarrow 3\ PbR_2$ at Different Levels of Theory

structure	B3LYP/I	MP2/I	CCSD(T)/I	CCSD(T)/TZ
Pb_3H_6	129.7	133.8	121.6	185.4
Pb_3Me_6	114.1	132.8		
Pb_3Ph_6	73.1	150.7		

Table 3. Bond Dissociation Energies (kJ/mol) for the Reaction $E_3H_6 \rightarrow 3\ EH_2$ at the Perturbation Theory Level

structure	MP2/I
C_3H_6[a]	1151[b]
Si_3H_6	582[b]
Ge_3H_6	555[b]
Sn_3H_6	449
Pb_3H_6	134

[a] CH_2 calculated as triplet. [b] The 6-31G(d) basis set was used for the central atom.

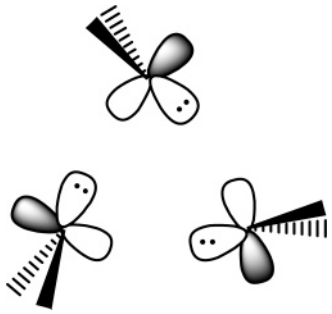


Figure 1. Schematic interactions of the lead−lead bonds in cyclotriplumbanes (reproduced from ref 11).

membered rings as well and therefore do not directly correspond to the energy of three single bonds.

One can clearly see from Table 2 that independent of the employed level of theory, the bond dissociation energy for cyclotriplumbanes is almost constant at 45 kJ/mol per bond/interaction. The exceptions are the DFT-optimized hexaphenylcyclotriplumbane with too long and hence too weak Pb−Pb contacts and the CCSD(T)/TZ-derived values with too short and hence too strong Pb−Pb bonds. The comparison with the lighter homologues confirms the weak nature of the lead−lead interactions: a C−C bond is almost nine times stronger and even cyclotristannane possesses three times stronger bonds (Table 3).

To explain why the substituents are twisted out of their "ideal" positions one has to look on the orientation of the orbitals forming the Pb−Pb contacts in cyclotriplumbane. Plumbylenes usually exist in a singlet state with the lone pair formed predominantly from the 6s orbital. This is confirmed by NBO calculations of different MP2-optimized PbR_2 structures (contribution of more than 85%, independent of R). As expected for a singlet carbene analogue, the lone pair and the unoccupied p orbital are about orthogonal to each other (92° in PbH_2 and $Pb(CH_3)_2$ and 95° in the phenyl-substituted plumbylene). To maximize overlap between doubly occupied $6s^2$ and empty 6p plumbylene orbitals in cyclotriplumbane, the PbR_2 units have to arrange in a tilted manner as shown in Figure 1. This kind of bonding bears a similarity to the double bonds formed by heavier homologues

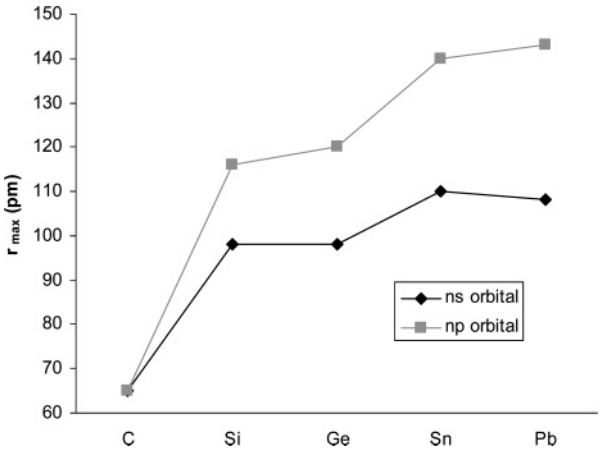


Figure 2. The calculated sizes of the valence *n*s and *n*p orbitals of group 14 elements.[36]

of group 14 elements where also the same interactions and trans-bent angles exist. The smaller trans-bent angles in **1** and in $Pb_3(2,4,6\text{-}Me_3C_6H_2)_6$ result from the interplay between maximum overlap and the increased steric demand of the substituents.

The predominance of singlet carbene analogues can be contributed to the relativistic contraction of the 6s electrons.[2,3] It is known that with the increasing atomic number the difference in the radial extensions of the s and p orbitals of group 14 elements also increases (Figure 2).[36] The splitting of *n*p orbitals due to spin−orbit coupling is not taken into account in the figure because although the resulting Pb $p_{1/2}$ and the $p_{3/2}$ orbitals differ in size, they are still significantly larger than the 6s orbital.[2] The described radii difference renders hybridization of s and p orbitals highly unfavorable with the higher homologues ("inert pair effect"). Another study suggests a dependency of size of the orbital radii and hence the hybridization on the metal partial charge.[37]

From the literature it is known that the trans-bent angle in double bond systems can be correlated with the singlet−triplet splitting of the corresponding carbene analogues.[38] A large gap corresponds to a large displacement angle. We calculated the singlet−triplet gaps of the according plumbylenes and found that the gap rises from 132 kJ/mol for PbH_2 to 171 kJ/mol for $PbPh_2$. The trans-bent angles however do not follow this trend but decrease from hydrogen to methyl and rise again to phenyl. Therefore, the energy gap cannot be correlated with the trans-bent angle in case of the cyclic structures.

Atoms-in molecule population analysis performed on cyclotriplumbanes confirm the nature of interactions between the lead atoms. Relevant bond- and ring-critical points could be located, thereby describing the weak donor−acceptor interactions as bonding. Contour plots of the electron density and the Laplacian of the electronic charge density (charge concentrations, $\nabla^2\rho(r) < 0$) reveal areas of increased electron density where the donor orbitals are located but do not show significant differences for the employed substituents; Figure 3 illustrates the situation in Pb_3H_6.

As mentioned above, our NBO analyses of the cyclotriplumbanes show that the lead−lead bonds are formed by weak interactions between an occupied orbital and an empty

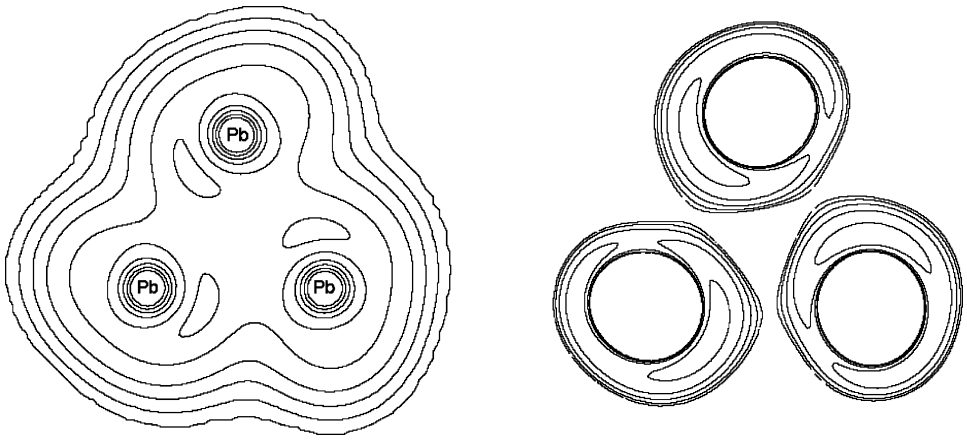


Figure 3. Contour plots of the electron density ρ (left) and the Laplacian of the electronic charge density $\nabla^2\rho(r)$ (right) of cyclotriplumbane.

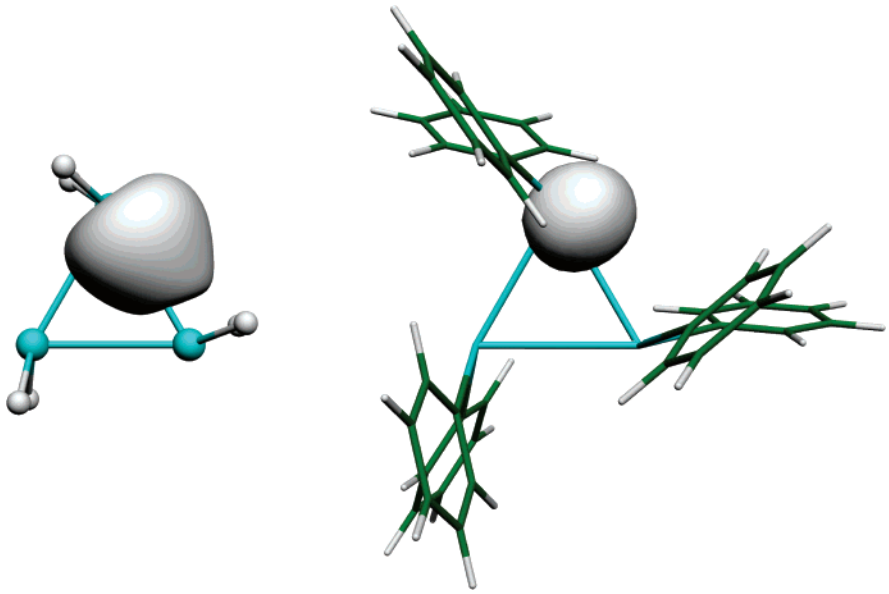


Figure 4. Influence of substitution on hybridization of bond-forming orbital: a lead sp hybrid orbital in Pb_3H_6 (left) and an almost pure s orbital in Pb_3Ph_6 (right). The spherical form of the latter indicates the absence of a p contribution.

p orbital of a neighboring Pb atom with the occupied orbital being the former lone pair of the plumbylene unit. Interestingly, there are differences in the composition of these occupied orbitals in the trimeric structure which depend on the substituent R. The s orbital contribution in the plumbylenes is above 85%, it is more or less the same for hexaphenylcyclotriplumbane (81%), resulting in doubly occupied, almost pure s orbitals. However, the smaller hydrogen atoms and the methyl group lead to a reduction of the s/p ratio in the cyclotriplumbanes, hence hybridization of the "inert" 6s and 6p orbitals takes place: the formerly pure s lone pair possesses now only around 40% s character, and the Pb−Pb bond-forming occupied orbitals are now formally between sp and sp^2 hybrid orbitals (Table 4). Figure 4 illustrates the different occupied orbitals. This is somewhat surprising considering the textbook knowledge of the inert pair effect where no hybridization should take place.

Analysis of the frontier orbitals (Table 4) does not give further indication why the phenyl-substituted cyclotriplumbane hybridizes and the other two do not. The relevant plumbylene orbitals have very similar shapes, and the only notable feature is the lowering of the HOMO/LUMO gap in Pb_3Ph_6.

Table 4. NLMO Analysis of Hybridization and Frontier Orbital Energies (in a.u.) in Cyclotriplumbanes and Plumbylenes

species	Pb−R s/p ratio	"lone pair" s/p ratio[a]	HOMO/LUMO	H/L gap
PbH_2	8/92	85/15	−0.318/0.002	0.320
Pb_3H_6	29/71	43/57	−0.308/0.005	0.313
$Pb(CH_3)_2$	6/94	88/12	−0.288/0.022	0.310
$Pb_3(CH_3)_6$	31/69	38/62	−0.269/0.035	0.304
$PbPh_2$	7/93	87/13	−0.302/0.013	0.315
Pb_3Ph_6	9/91	81/19	−0.279/0.008	0.287

[a] s/p contributions in % of the former lone pair, forming the Pb−Pb bond in the trimer.

Conclusions

It is of utmost importance to employ a reasonably high level of theory together with a reliable basis set/ECP combination

in order to reproduce both geometry and bonding situation of plumbylenes and the trimeric cyclotriplumbane Pb_3R_6. Substitution has a pronounced influence on the geometry, so that MP2 calculations are necessary to get the required good agreement with the experimental structure. Also, phenyl groups are not adequate for the reproduction of the trans-bent angles in **1**. The bulkier trimethylphenyl substituents give the correct tilting of the groups R in the cyclotriplumbanes which is attributed to the bonding situation: the lead−lead contacts are formed from weak interactions of plumbylene lone pairs with empty p orbitals of neighboring lead atoms, and in order to maximize overlap, the plumbylene units have to twist from their "ideal" positions. The trans-bending cannot be correlated with the singlet−triplet splitting as in double bond systems.

Another striking observation is the hybridization of the former lone pair orbitals in the cyclotriplumbane Pb_3R_6. Depending on the substituent, hybrid orbitals with a 6s orbital contribution of only about 40% are found. Hydrogen atoms and methyl groups promote this effect, in contrast to the inert pair effect, while for phenyl substitution the expected 6s lone pair orbital is identified as the bond-forming orbital.

Acknowledgment. This work was supported by generous allocation of computer time at the Norddeutscher Verbund für Hoch- und Höchstleistungsrechnen (HLRN), Hannover, Berlin, and at the Hochschulrechenzentrum, Universität Oldenburg. Financial support by the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

(1) Part 3: Koch, R.; Bruhn, T.; Weidenbruch, M. *J. Mol. Struct. (THEOCHEM)* **2005**, *714*, 109−115.

(2) For a review see Kutzelnigg, W. *Angew. Chem.* **1984**, *96*, 262−286, and references therein.

(3) Pyykkö, P. *Chem. Rev.* **1988**, *88*, 563−594.

(4) For recent reviews, see: (a) Power, P. P. *Chem. Rev.* **1999**, *99* 3463−3503. (b) Weidenbruch, M. *J. Organomet. Chem.* **2002**, *646*, 36−52. (c) Weidenbruch, M. *Organometallics* **2003**, *22*, 4348−4360.

(5) (a) Stürmann, M.; Saak, W.; Marsmann, H.; Weidenbruch, M. *Angew. Chem.* **1999**, *11*, 145−147; *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 187−189. (b) Stürmann, M.; Saak, W.; Weidenbruch, M.; Klinkhammer, K. W. *Eur. J. Inorg. Chem.* **1999**, 579−582. (c) Klinkhammer, K. W. *Polyhedron* **2002**, *21*, 587−598.

(6) Preut, H.; Huber, F. Z. *Anorg. Allg. Chem.* **1976**, *419*, 92−96.

(7) (a) Trinquier, G.; Malrieu, F. *J. Am. Chem. Soc.* **1987**, *109*, 5303−5315. (b) Trinquier, G. *J. Am. Chem. Soc.* **1990**, *112*, 2130−2137.

(8) Jacobsen, H.; Ziegler, T. *J. Am. Chem. Soc.* **1994**, *116*, 3667−3679.

(9) Klinkhammer, K. W.; Fässler, T. F.; Grützmacher, H. *Angew. Chem.* **1998**, *110*, 114−116; *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 124−126.

(10) Stabenow, F.; Saak, W.; Marsmann, H.; Weidenbruch, M. *J. Am. Chem. Soc.* **2003**, *125*, 10172−10173.

(11) Nagase, S. *Polyhedron* **1991**, *10*, 1299−1309.

(12) Jensen, F. *Introduction to Computational Chemistry*, Wiley: Chichester, 1998.

(13) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(14) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Mater.* **1988**, *37*, 785−789.

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

(16) Head-Gordon, M.; Pople, J. A.; Frisch, M. *J. Chem. Phys. Lett.* **1988**, *153*, 503−506.

(17) Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35−89.

(18) Purvis, G. D., III; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910−1918.

(19) Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F., III *J. Chem. Phys.* **1988**, *89*, 7382−7387.

(20) *Gaussian 98, Revision A.11.3*; 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.; Pomeli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Rega, N.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortziz, 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, A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzales, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 2002.

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

(22) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(23) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654−3665.

(24) Kuechle, W.; Dolg, M.; Stoll, H.; Preuss, H. *Mol. Phys.* **1991**, *74*, 1245−1263.

(25) Dolg, M.; Kuechle, W.; Stoll, H.; Preuss, H.; Schwerdtfeger, P. *Mol. Phys.* **1991**, *74*, 1265−1285.

(26) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299−310.

(27) Basis set: Peterson, K. A. *J. Chem. Phys.* **2003**, *119*, 11099−11112; pseudopotential: Metz, B.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2000**, *113*, 2563−2569. Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 02/25/04, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352, U.S.A., and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multiprogram laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RLO 1830. Contact Karen Schuchardt for further information.

(28) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650−654.

(29) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265−3269.

(30) *NBO Version 3.1*; Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold F.

(31) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735−746.

(32) Bader, R. F. W. *Acc. Chem. Res.* **1985**, *18*, 9.

(33) Bader, R. F. W. *Atoms in Molecules: a Quantum Theory*; Clarendon Press: Oxford, 1990.

(34) Biegler-König, F.; Schönbohm, J.; Bayles, D. *J. Comput. Chem.* **2001**, *22*, 545.

(35) Koch, R.; Bruhn, T.; Weidenbruch, M. *J. Mol. Struct. (THEOCHEM)* **2004**, *680*, 91−97.

(36) Karni, M.; Apeloig, Y.; Kapp, J.; Schleyer, P. v. R. Theoretical aspects of compounds containing Si, Ge, Sn and Pb. In *Chemistry of Organic Silicon Compounds*, Rappoport, Z., Apeloig, Y., Eds.; Wiley: Chichester, 2001; Vol. 3, p 1. Data taken from Nagase, S. In *The Transition State*; Fueno, T., Ed.; Gordon and Breach: London, 1999; pp 147−161.

(37) Kaupp, M.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1993**, *115*, 1061−1073.

(38) Chen, W.-C.; Su, M.-D.; Chu, S.-Y. *Organometallics* **2001**, *20*, 564−567.

CT050144+

JCTC Journal of Chemical Theory and Computation

PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique

Kenneth W. Borrelli, Andreas Vitalis, Raul Alcantara, and Victor Guallar*

Department of Biochemistry, Washington University School of Medicine, St Louis, Missouri 63108

Received July 26, 2005

Abstract: Combining protein structure prediction algorithms and Metropolis Monte Carlo techniques, we provide a novel method to explore all-atom energy landscapes. The core of the technique is based on a steered localized perturbation followed by side-chain sampling as well as minimization cycles. The algorithm and its application to ligand diffusion are presented here. Ligand exit pathways are successfully modeled for different systems containing ligands of various sizes: carbon monoxide in myoglobin, camphor in cytochrome P450cam, and palmitic acid in the intestinal fatty-acid-binding protein. These initial applications reveal the potential of this new technique in mapping millisecond-time-scale processes. The computational cost associated with the exploration is significantly less than that of conventional MD simulations.

1. Introduction

Obtaining a molecular-level understanding of protein biochemical and biophysical processes is a complex task requiring the characterization of long-time conformational rearrangements. Molecular dynamics methods, in which the evolution of the system is projected as a series of snapshots resulting from the integration of classical equations of motion, have been largely used with great success to model conformational rearrangement in biomacromolecules.[1−6] These simulations, however, are sensitive to the size of the system and to the total time of propagation, requiring significant computational resources to reach times on the order of hundreds of nanoseconds. Different approximations have recently been introduced to modify conventional molecular dynamic methods in order to capture large-scale conformational rearrangements: simplified electrostatics models,[7] multiple replica dynamics,[8−10] stochastic path approaches,[11] Monte Carlo/molecular dynamics combinations,[12] targeted constraint dynamics,[13] modified potential landscape,[14] and so forth. These methods offer a valuable approach to model long-time dynamics, although often limited to the constraint of providing initial and final states and to the employment (for some methods) of large biasing potentials. The modeling of long-time dynamics still remains a challenge.[15,16] Of particular interest is the elucidation of protein pathways, describing a biophysical or biochemical event, for which only the initial state is known. For this purpose, various approaches often referred to as energy landscape exploration techniques have been recently developed.[16] While using different sampling methods, the overall idea is to project a set of local minima describing the protein conformational changes. Among others, sampling methods include eigenvector following[17] and backbone dihedral pivoting.[18,19] Various simpler mechanical approaches, including normal-mode analysis,[20] essential dynamics analysis,[21] and the Gaussian network model,[22] are widely used to characterize long-range conformational changes and collective protein dynamics. These mechanical methods offer valuable information on conformational changes, requiring only the initial state, but at the cost of a loss in detailed atomic resolution. The combination of these approaches with atomic detailed landscape exploration methods, like the one proposed here, could certainly constitute a new generation of landscape exploration and dynamical tools.

Parallel to the development of the above techniques, there has been a significant effort toward the development of theoretical methods for protein structure prediction based on the use of rotamer libraries.[23−26] Recent work by Hellinga et al. has proven the possibilities of theoretical guided protein engineering when using protein structure prediction algo-

* Corresponding author. E-mail: guallarv@biochem.wustl.edu.

10.1021/ct0501811 CCC: $30.25
Published on Web 10/12/2005

rithms.[27,28] Jacobson et al. have recently developed PLOP, a program for protein modeling using all-atom energy functions and specialized sampling algorithms for side-chain and loop prediction.[29,30] The program uses an implicit surface-generalized Born (SGB) continuum solvent model[31] and the recently updated version of the OPLS-AA force field parametrization (OPLS2001). The latter was demonstrated[32] to provide significant improvements in side-chain prediction as the fit to quantum chemical data in experimentally relevant regions of phase space was improved. The sampling algorithms include the use of highly detailed rotamer state libraries for side-chain conformational searching, hierarchical screening methods based on steric overlap, and approximate electrostatics to rapidly eliminate obviously incorrect conformations. With these technological advances, rigorous sampling of side chains can be done at a rate of one side chain per second on a regular desktop (throughout the paper, CPU estimates refer to a 2.8 GHz Pentium IV PC).

Using these specialized protein structure sampling algorithms, we have developed a novel approach to explore the protein landscape dynamics. The program has been named PELE: Protein Energy Landscape Exploration. The method uses a localized steered perturbation coupled to side-chain prediction and minimization algorithms. The procedure includes the possibility to perform single-point mixed quantum mechanics/molecular mechanics (QM/MM) calculations[33] to update the charges of complex ligands or to obtain quick estimates of a biochemical reaction.

We have initially applied this method to the study of ligand diffusion on several enzymes. In all the cases studied, myoglobin, cytochrome P450, and the intestinal fatty-acid-binding protein, we have obtained entry/exit pathways for one representative ligand for each system: CO, camphor, and palmitate, respectively. All diffusive pathways are consistent with experimental data. The computational cost varies with the system, from 1 day for myoglobin to 4 days on a single CPU for P450 and the fatty-acid-binding systems. In this first paper, we present the methodology and results on camphor migration on P450cam. To demonstrate the potential of the method, a summary of the results on myoglobin and the intestinal fatty-acid-binding protein are also given; more extensive results on these two systems will be presented elsewhere.

Because of the abundant experimental information, camphor migration on P450cam is a well-defined test case. A recent study trapped two different synthetic molecular wires, indicating the importance of the motion of the F and G helix in the substrate entry pathway[34] (Figure 1). Furthermore, simulations using random steered molecular dynamics have also been performed on this system.[35] This method, however, required a large force pushing away the ligand from the heme center, resulting in multiple paths for ligand exit. Our results are consistent with the synthetic molecular wires' crystallographic results and reveal a detailed atomic description of the protein response along the entire migration pathway.

2. Methods

We have developed PELE, a new approach to map the energy landscape for long-time conformational rearrangement in proteins. This procedure incorporates Monte Carlo moves, rotamer library side-chain optimizations, truncated Newton minimizations, and Metropolis acceptance tests. The method is used to generate and propagate changes in a system by generating a series of structurally similar local minima that are then combined into a trajectory.

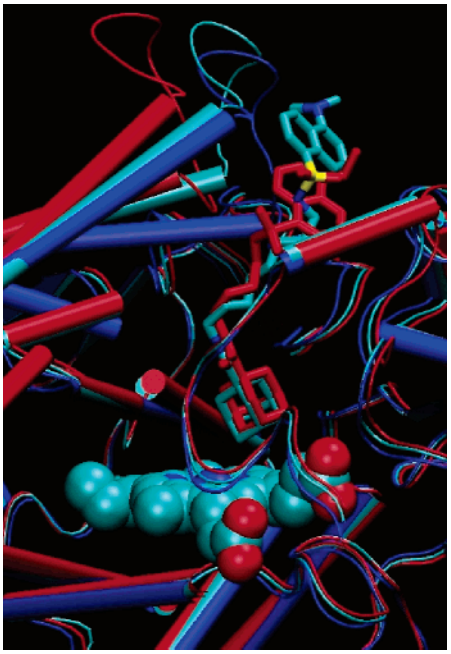


Figure 1. Local superposition of the crystal structures for P450cam with camphor (PDB: 1PCH; dark blue), the short D-4d wire ligand (PDB: 1RF9; red), and the longer 8D−d wire ligand. (PDB: 1RE9; light blue). A clear opening of the F/G helixes and the loop connecting them is observed for the wire ligands.

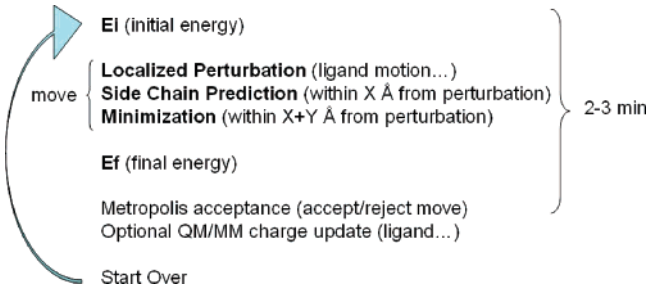


Figure 2. Sampling procedure used in PELE.

Sampling Procedure. *Local Perturbation*. The procedure employed (Figure 2) begins with the generation of a local perturbation of the ligand. Many ligands can be treated as rigid bodies so only three rotational and three translational degrees of freedom are required. Flexible ligands, such as palmitate, cannot be adequately described as a single rigid unit. Thus, the perturbation of these flexible ligands includes additional degrees of freedom from the dihedral angles of rotatable bonds, while the bond distances and bending angles are kept fixed. At this stage, the user has to provide a list of rotatable bonds for the ligand as well as OPLS-AA style parameters (at least atom types). Next, a series of filters are applied to determine if there is any steric contact between the ligand and the backbone of the protein and, in the case of flexible ligands, between the ligand and itself. If any such

contacts are found, the perturbation is rejected. A pair of nonbonded atoms were determined to create a steric clash if the distance between them was less than 70% of the sum of the van der Waals radii of the two atoms. A set of 400 perturbations are generated, and the one with the best scoring is selected. The scoring function used is an OPLS-AA force field in which only the ligand and the backbone of the protein were considered. The solvation effects were neglected as were the nonbonded interactions with residues more than 15 Å from the ligand in its position at the start of the step.

Side-Chain Sampling. The algorithm proceeds by optimally arranging all side chains local to the atoms perturbed in the first step with a rotamer library side-chain optimization at a rotamer resolution of 10°. The side-chain algorithm[32] uses steric filtering and a clustering method to reduce the number of rotamers to be minimized (side-chain minimization only). The typical time required for the side-chain sampling of 20−30 residues, corresponding to a sampling distance of 8 Å from a medium-size ligand, is in the 20−40 s range. Future versions will speed up the side-chain sampling procedure by adapting the side-chain rotamer resolution along the landscape exploration; side chains experiencing less conformational changes will decrease their rotamer resolution.

Minimization. The last step in every move involves the minimization of a region including, at least, all residues local to the atoms involved in steps 1 and 2. The truncated Newton minimization algorithm uses a multiscale protocol, 1−2 orders of magnitude more efficient than conventional approaches.[29] The minimization is intended to generate a backbone response to the initial local perturbation and to possible side-chain rearrangement in the first steps. The overall approach is designed on the basis of the assumption that side chains will act as protein sensors, responding to ligand motion, protein−protein interactions, biochemical activity, and so forth. The backbone will then follow the side-chain response.

The *X* and *Y* distances for the local side chain and minimizations (see Figure 2) are measured in Å and are chosen by the user. They determine the local areas for side chains and minimization. All residues with an atom within the *X* (or *Y*) distances from any ligand atom are included in the list of local residues.

These three steps compose a *move* which is accepted (defining a new minimum) or rejected on the basis of a Metropolis criterion for a given temperature

$$\Delta V < 0$$
$$\exp(-\Delta V/K_B T) < R \qquad (1)$$

that is, by a decrease of the potential surface, $\Delta V < 0$, or by satisfying the second criterion where K_B is the Boltzmann constant, T the temperature chosen for the simulation, and R is a random number with a [0, 1] range. The conformational changes are propagated to the nonlocal environment by means of the diffusion of the local perturbed region and by introducing a long-range step similar to the local step but involving a longer sampling radius. In the results presented here, the local step was confined to residues within 8 Å of the perturbation. The long-range step was conducted every 10 moves and extended out to 25 Å from the ligand. The overall time for each move depends on the chosen side chain and minimization region, and on ligand size. For the default values, 8 Å for side chains and 10 Å for minimization, a complete iteration required 2−3 min of CPU time for the tested ligands.

To perform efficient sampling of rare events, the direction of the perturbation is kept constant for a set number of steps making up a steering cycle. The translation is described by where $\vec{r}$ is the translational vector, λ is a variable describing

$$\vec{r} = (\lambda + R\lambda)\vec{r} \qquad (2)$$

the translational range, and R is, again, a random number. Thus, translations are always forced within the $[\lambda, 2\lambda]$ interval, with λ typical values between 0.05 and 0.2 Å (user-defined variable). $\vec{r}$ is previously aligned with the normalized steering vector using a Gaussian function with a tightness criterion, which ensures that the perturbation will be within 10−30° of the steered direction. A steered cycle of 20 indicates that after a period of 20 moves the steered direction might be randomly updated. The update takes place if the direction was an unsuccessful direction search (with a user-defined number of accepted steps or overall displacement). The Metropolis acceptance/rejection criterion for each move ensures that maintaining a perturbation direction does not result in a large energy increase. The result of this procedure is a series of local minima which are structurally highly correlated. However, large structural changes can be attained from one end of the trajectory to the other. In this way, the atomic-scale details of possible mechanisms can be explored.

The algorithm can be used to attain different goals by adjusting the steered and acceptance parameters. An equilibration procedure with a very short steered cycle and low Metropolis temperature can be used to produce local energy minimization and exploration. Such an equilibration procedure is necessary prior to any landscape exploration. A setup with a longer steered cycle and higher temperature can be used to explore possible pathways. Initially, the exploration runs are carried out at high Metropolis temperatures (1000 °K, for example), defining one or various ligand reaction coordinates. As emphasized by Mousseau et al., the Metropolis temperature does not correspond to a real thermal bath, the effective temperature being significantly lower.[36] The ligand reaction coordinates are reduced to an increase or decrease of a distance between the ligand center of mass and a target(s) (a set of *x*, *y*, and *z* coordinates). Various parallel trajectories at a lower Metropolis temperature are then started at the initial point. The information from the different nodes is shared using the MPI communications protocol. Whenever any trajectory is significantly farther along the reaction coordinate than any of the other trajectories, the trailing trajectory is abandoned and restarted from the position of the leading trajectory. We should emphasize that the higher-temperature exploration runs only determine the ligand target, a criterion by which the different nodes spawn their geometries. They do not introduce any further bias in the protein dynamics. Thus, trajectories produced at any given temperature are entirely produced from the starting structure at the given temperature, with no memory of any

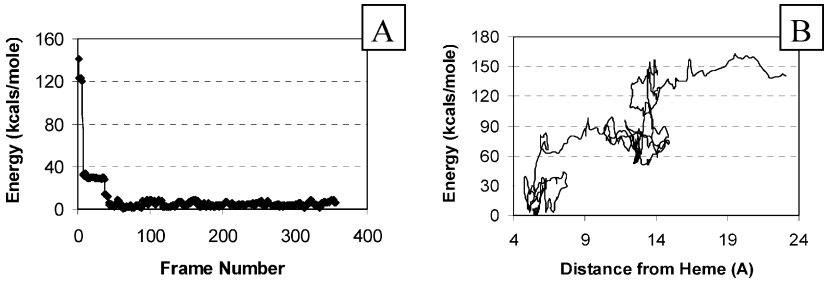


Figure 3. (A) Representative energy profile of an equilibration run. A stable minimum is found after about 50 steps and is maintained for another 300 steps without significant change. (B) Initial exploration run for P450cam run at 2000 K.

previous production at higher temperatures. Obviously, lower temperatures require longer production runs (or more processors). Using this parallel procedure, we have been able to run ligand escape simulations at room temperature for each of the systems studied.

Following recent studies[37,38] where we combined protein structure prediction algorithms with QM/MM techniques, the landscape exploration algorithm has been coupled with Qsite, a QM/MM program.[39] These techniques contain the elements necessary to properly describe the potential energy surfaces relevant to biochemical processes.[33] We have used single-point QM/MM calculations to obtain the initial atomic charges for the ligand and to account for the charge polarization of the ligand along its migration in the protein frame (using electrostatic potential charges). Single-point calculations are performed only after a large ligand rearrangement (every ~50 steps or longer), requiring on the order of 3−15 min of CPU time (3−4 min for camphor, ~15 for the fatty acid). To soften discontinuities in the potential surface, any new set of charges are scaled along six consecutive accepted sampling steps. We should emphasize that the overall charge is conserved. As seen in Figure 2, the charge update is always performed after the Metropolis test. Furthermore, an additional minimization follows any charge update; the initial and final energies for any Metropolis evaluation correspond to minimized structures with the same charge set. The QM region for all systems was reduced to the ligand. For the P450 and myoglobin systems, an additional initial QM calculation, including the heme group and the proximal residue in the QM region, provided the set of charges (fixed along the simulation) for the heme group. QM calculations are carried out using the B3LYP functional in combination with the 6-31G* basis set.

All classical calculations are done with an OPLS-AA description of the ligand and protein atoms and with an implicit SGB solvent model. As stated above, the atomic charges for the ligand (and the heme group in P450 and myoglobin) are obtained from QM/MM single-point calculations. The continuum solvation model has been extensively parametrized by Levy and co-workers to provide accurate predictions for a wide range of small molecule solvation free energies.[40] This parametrization includes the use of a new nonpolar term, which more accurately reproduces nonpolar solvation free energies for molecules of different shapes (e.g., cyclic and branched alkanes) as well as providing first-shell corrections to the generalized Born electrostatics. Finally, Friesner et al. have reparametrized key terms in order to improve conformational side-chain prediction as compared to experimental data.[31] A particular focus has been the ability to predict the formation of surface salt bridges, which are sensitive to the continuum solvation parameters for ionizable side chains. The unmodified SGB model overpredicts the formation of surface salt bridges by a substantial margin; reparametrization of the effective screened pair interaction term for these structures leads to much more accurate results. These improved Born models have been shown to give accurate forces and relative energies when compared with Poisson−Boltzmann and explicit solvent models.[10,41,42] In particular, reparametrized Born models perform well with sampling methods such as the one described in this work; the use of SGB with molecular dynamics requires a careful selection of a friction degree of freedom.[43−45]

PELE has been merged with PLOP and will be available free of charge for academic purposes from the UCSF site (http://francisco.compbio.ucsf.edu/~jacobson/plop_manual/plop_license.htm). To run the QM/MM options, it will require the installation of Qsite.[39]

3. Results and Discussion

Results focus mainly on camphor migration in cytochrome P450cam. The existence of crystallographic direct evidence of the ligand entry/escape channel makes P450cam an ideal test system. A summary of our results on CO and palmitate fatty acid diffusion on myoglobin and intestinal fatty-acid-binding protein, respectively, are also shown. These results indicate the possibilities of our methodology in treating small (CO) as well as complex (palmitate) ligands. More extensive results on these two systems will be presented elsewhere.

3.1. Ligand Migration on P450cam. The first step before launching an exploration run is to obtain an equilibrated local minimum. Such a structure is obtained by adjusting the local perturbation so that it always performs a random update of the direction search. Additionally, the steered cycle is reduced to <5. Thus, there is no successful search direction, reducing the landscape exploration to a local refinement of a given initial conformation. Panel A in Figure 3 introduces the energy profile for the camphor equilibration run. Similar energy profiles are obtained for each of the other systems. As seen in Figure 3, a stable plateau is obtained after 50 iterations. These equilibration runs required about 10 CPU-hours on a single processor.

After the equilibration run, we performed a quick landscape exploration by using high Metropolis temperatures.

At this stage, we usually keep the steered cycle at ~20 and adapt the localized perturbation amplitude and the temperature until an acceptance rate of >50% is reached. The initial exploration is intended to find different possible pathways. From the analysis of the different pathways, we select the optimal reaction coordinate to run a multiprocessor production run at lower Metropolis temperatures. This initial exploration is performed on a single processor and has an associated computational cost of 10−20 CPU-hours. For the P450cam system, we have performed 10 different exploration paths with Metropolis temperatures ranging from 500 to 2000 K. All paths, even at 2000 °K, resulted in camphor leaving the active site in a path closely following the wire ligands shown in Figure 1. No alternate pathway was observed, and no other protein cavities, distant from the exit pathway, were visited. In the random steered molecular dynamics studies,[35] multiple pathways were obtained. One of them closely corresponds to the pathway observed by us (in agreement with the crystallographic wire). However, the large force introduced, necessary to observe the camphor migration, apparently affects the potential landscape and results in alternate pathways. In other cases, myoglobin, for example, this procedure does predict ligand diffusion to a diverse set of experimentally observed inner cavities (details follow).

Panel B in Figure 3 displays an overlay of a few exit snapshots along one of the 10 different trajectories and the energy associated with the ligand motion. As a result of the large temperature acceptance associated with the exploration run, the energy profile is significantly large. Once the pathway(s) of interest have been identified, reaction coordinates can be developed that track the progress along these pathways. For the case of camphor, only a single pathway was observed, and hence, a simple reaction coordinate was implemented, which tracks the distance between the camphor center of mass and the iron of the heme. Using the parallel version of the code, trajectories were conducted using 6−10 processors at Metropolis temperatures of 750, 500, 400, and 250 K. The difference between the energy profiles of the last two temperatures were reduced only to 1−2 kcal/mol. As described in the Methods section, each of these trajectories is independent. By running multiple temperature trajectories, one can follow the convergence of the potential energy surface, as seen in Figure 4. Additionally, a future objective is to build the density of states by means of an energy histogram[46−48] along different bids of the reaction coordinate by means of multiple temperature trajectories. Each of these refinement runs required 2−4 days of CPU time (for each processor).

An update of the atomic ligand charges by means of single-point QM/MM calculations did not introduce any significant change in the ligand escape pathway and its potential surface. The atomic charge standard deviations were in the 0.02−0.06 range, with no abrupt change in atomic charges. Interestingly, the oxygen atom on camphor is the heavy atom with less charge fluctuation (with a 0.026 standard deviation). One would expect these fluctuations to be even less important when employing a polarizable force field for the protein atoms. Future studies need to address the importance of ligand polarization on a larger diversity of ligand and ligation

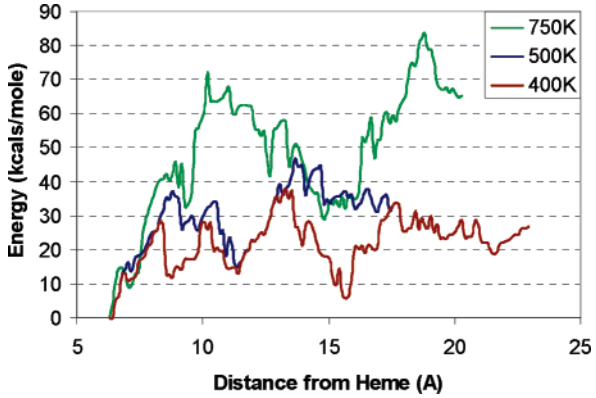


Figure 4. Energy profile associated with camphor exiting P450cam at three different temperatures of coordinated runs. Each successively lower temperature reduced the barriers associated with camphor leaving. At 400 K, energy barriers of ~35 kcal/mol were found (250 K gives essentially the same energy profile as 400 K).

pathways. A QM/MM description of the ligand atomic charges, for example, has recently been shown to considerably improve docking simulations.[49]

As seen in Figure 5, the exit pathway passes through the F−G loop, displacing both the F and G helix by a small amount, and is gated by the movements of several phenylalanine residues (F87, F98, and F193). Panel E in Figure 5 indicates a superposition of several camphor snapshots with the crystallographic wire along the exit pathway. The same panel also indicates the overall motion of the F and G helixes. Panels A−D highlight the molecular details of the phenylalanine motion along the camphor migration. The highest energy minimum, located at a camphor−heme distance of about 13 Å, is shown in panel B. At the present time, the landscape exploration algorithm only samples local minima. A work in progress is applying double-ended methods[50−52] for the characterization of the energy barriers. The ultimate goal is to obtain a network and a transition-state matrix characterizing the dynamics of the system in a fashion similar to recent studies on protein folding.[10,53−55] The highest energy minimum, shown in panel B of Figure 5, corresponds to the point in the trajectory where the camphor must pass between the two phenylalanine residues (F87 and F98). This large amplitude motion of the phenylalanine residues is also observed in the molecular dynamic studies by Wade et al.[35] Furthermore, the important role of these residues agrees with recent observations of the Batista's group at Yale University (personal communication). Their studies have observed a large coupling of these phenylalanine vibrations with the abstraction of a hydrogen atom from the camphor substrate by the putative active species, the oxy-ferryl heme species also known as compound I. Thus, the ligand diffusion (and possibly the enzymatic activity) seems to be mainly gated by the movements of these phenylalanine residues.

3.2. Ligand Migration on Other Systems: Myoglobin and Intestinal Fatty-Acid-Binding Protein. *Myoglobin.* The study of CO diffusion on globular proteins was the simplest and fastest application, requiring only temperatures of 300−400 K on a single processor, for producing successful migration pathways. In 20 exploration trajectories in the

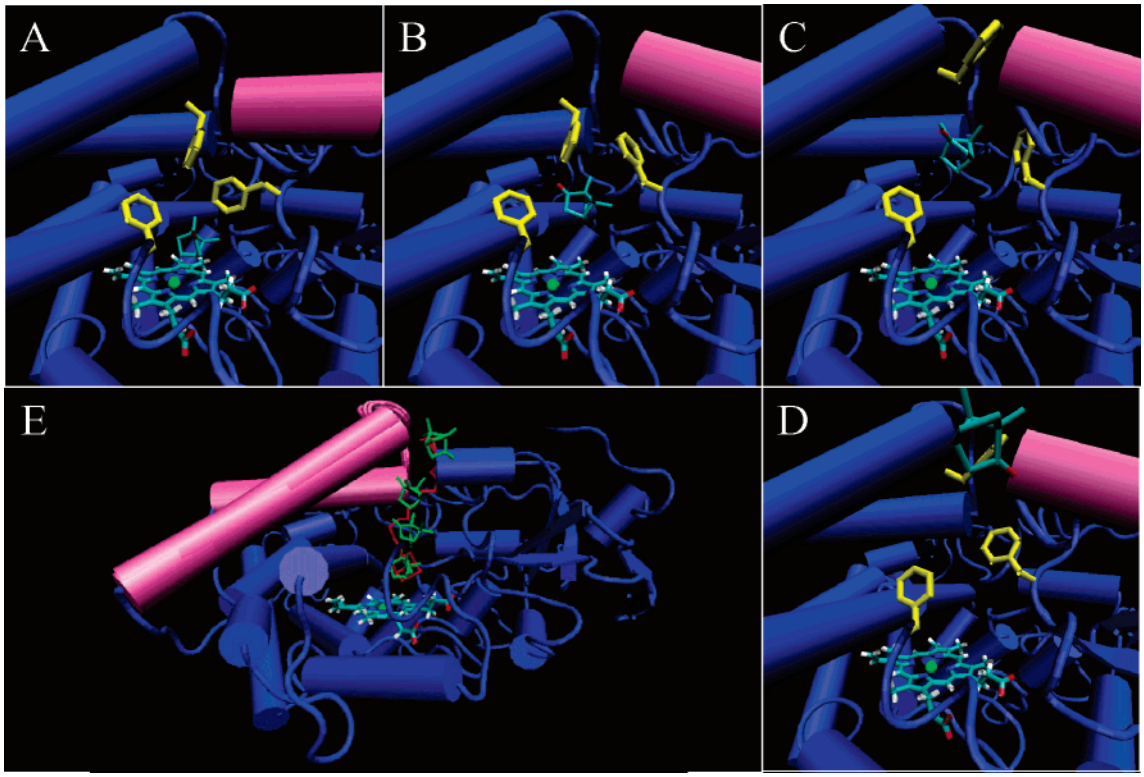


Figure 5. Exit pathway of camphor from P450cam as identified by PELE at room temperature (300 K). Panels A−D show four points along the exit pathway. The three phenylalanine residues experiencing large conformational changes, F87, F98, and F193 are shown in yellow. Panel B represents the structure at the largest energy barrier. Panel E displays an overall view of the pathway and is shown in green. The path identified by molecular wire studies is shown in red. Helices F and G as well as the F−G loop are shown in pink.

myoglobin system, 90% of the escape pathways involved pathways in the heme distal site. In most of the trajectories, the ligand visited different inner cavities (known as xenon cavities) before exiting the protein. The trace of the CO ligand is shown with a sphere representation for the C and O atoms in Figure 6. Such cavities have been clearly characterized previously.[56−60] The ligand exit, however, involved mainly the cavity connecting the distal site with the CD loop. Half of those pathways involve the opening of the distal histidine, a mechanism largely discussed in the literature.[61] The other half of the migration pathways involved the path shown in Figure 6 with an orange arrow. This migration pathway was observed in early calculations, using a time-dependent Hartree approximation, by Elber and Karplus.[59] This cavity contains three of the six phenylalanine residues present in myoglobin (shown with a white stick representation in Figure 6). This result is consistent with mutational studies by Scott et al. where mutation of these phenylalanine residues (and residues next to them) largely affected the entry and escape kinetics; mutation of all residues in the vicinity of the other cavities (Xe1, Xe4, etc.) did not show any significant effect on the ligand kinetics (see, for example, Figure 5 in Scott et al.[62]). Our results indicate that the CO stacking on top of the phenylalanine side chain participates actively in the migration pathway. Protein fluctuations in the CD loop region are essential for the escape of the ligand. The importance of the CD region on the globin heme-binding family has been corroborated by Fernandez-Alberti et al.[63] Using Gaussian network model analysis, the authors showed the existence of conserved low modes involving the CD helixes and the loop connecting them.

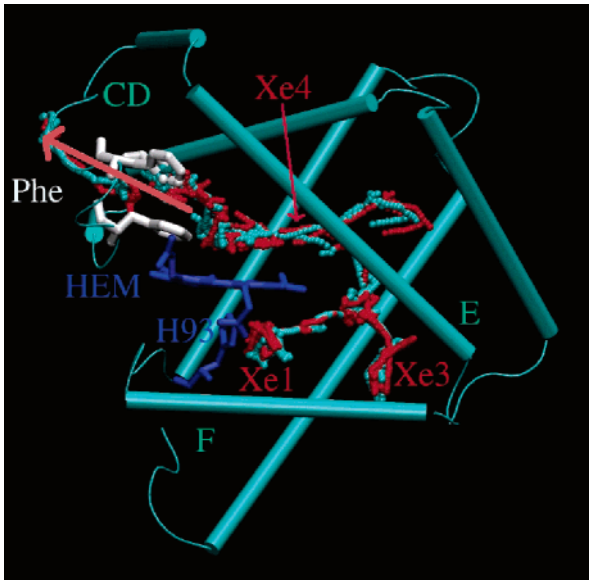


Figure 6. Carbon monoxide migration in myoglobin. The protein is shown in a cartoon representation with helices E and F and the loop CD being labeled. CO migration is shown in a sphere representation (red for oxygen and cyan for carbon), with some xenon cavities labeled in red. Phe 33, 43, and 46 are shown in white. The heme group and the proximal histidine, H93, are shown in dark blue. The exit pathway is underlined with a pink arrow.

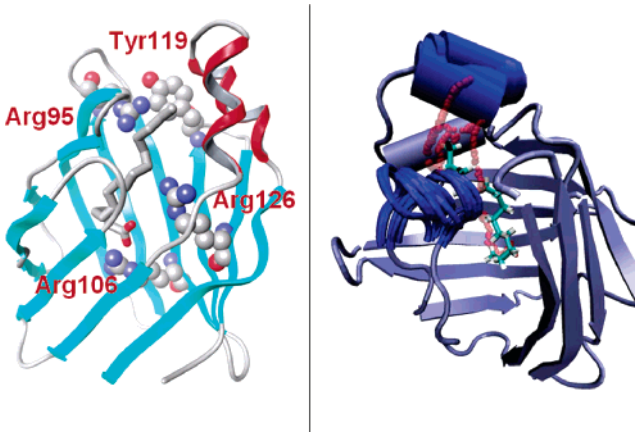


Figure 7. (Left) Important residues along the palmitate escape pathway. (Right) Main protein conformational changes obtained in our simulation along the exit pathway involving mainly helix αII and turns βC−βD and βE−βF.

Intestinal Fatty-Acid-Binding Protein. Palmitate diffusion in the intestinal fatty-acid-binding protein has been associated with the overall motion of the helix αII and turns between βC−βD and βE−βF.[64,65] Covalent pH-dependent cross-linking of these two secondary structure elements trapped palmitate. Figure 7 (right panel) indicates the main protein conformational changes associated with this process and a series of snapshots where several conformations of the fatty acid ligand are captured along the deligation pathway. The (long-range) protein conformational changes obtained in our simulation involve mainly helix αII and turns βC−βD and βE−βF, in great agreement with the experimental observations.[64,65] Our results not only confirm those experimental results but represent the first theoretical simulation where the escape of the palmitate ligand is observed. In particular, several polar residues in the interior of the protein appear to be crucial. These residues, shown in the left panel in Figure 7, induce a 180° turn on the fatty acid from its bound crystal structure, inducing an orientation where the carboxylic head of the ligand faces the solution. In particular, Arg126 appears to be a crucial residue in the proper alignment of the ligand toward the active site and in its initial rotation in the escape pathway. Mutational studies on Arg126 confirm the importance of this residue in the binding event.[66] The complex deligation pathway observed in our simulations is consistent with the experimental observation suggesting that lipid association is more complex than that described within a single equilibrium event.[65] Obtaining an exit pathway for this system involved 4 days of a single CPU time. This result evidences the possibilities of the method in sampling the energy landscape of a large ligand with several rotatable bonds.

4. Conclusions

We have introduced a novel landscape exploration technique based on Metropolis sampling and protein structure prediction algorithms. Ligand escape pathways in myoglobin, P450, and fatty-acid-binding protein have been presented and are consistent with experimental and theoretical data. Those processes, expanding from the microsecond to the millisecond time scale, require large protein conformational changes, which were captured by our methodology. Our approach, therefore, signifies an innovative methodology, capable of describing slow deligation events at a detailed atomic level. The landscape exploration output could be used as an input for several free-energy methodologies (umbrella sampling, etc). Present studies are focusing on obtaining such a description using only small modifications of the algorithms described herein. An accurate free-energy study, however, will necessarily involve a more rigorous solvation model. A mixed explicit/SGB solvation scheme has shown promising results.[67] Additionally, the algorithm proposed here cannot account for a detailed solvent−protein (or solvent−ligand) interaction, which might be essential in some particular protein and ligand dynamics. Further research in mixed solvation models is necessary to address this point. Current studies are also investigating the possibilities of landscape exploration into a wider scope of applications, including the protein response to a biochemical process as well as protein folding.

Acknowledgment. Financial support has been provided by startup package from Washington University School of Medicine. We are very grateful to Prof. Nathan Baker and Prof. Rohit Pappu for helpful discussions.

References

(1) Warshel, A.; Levitt, M. A. *J. Mol. Biol.* **1976**, *103*, 227−249.

(2) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631−639.

(3) Tidor, B. *Curr. Biol.* **1997**, *7*, R525−R527.

(4) Kleinekathoefer, U.; Isralewitz, B.; Dittrich, M.; Schulten, K. *Biophys. J.* **2004**, *86*, 246A−246A.

(5) Zhou, R. H.; Huang, X. H.; Margulis, C. J.; Berne, B. J. *Science* **2004**, *305*, 1605−1609.

(6) Zhou, H. X.; Wlodek, S. T.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9280−9283.

(7) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.

(8) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925−932.

(9) Nymeyer, H.; Gnanakaran, S.; Garcia, A. E. In *Numerical Computer Methods, Pt D*; Academic Press Inc.: San Diego, CA, 2004; Vol. 383, pp 119 ff.

(10) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801−6806.

(11) Elber, R.; Ghosh, A.; Cardenas, A. *Acc. Chem. Res.* **2002**, *35*, 396−403.

(12) Zhou, R. H.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931−14936.

(13) Schlick, T. *Biophys. J.* **2003**, *85*, 1−4.

(14) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919−11929.

(15) Elber, R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151−156.

(16) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, MA, 2003.

(17) Wei, G. H.; Mousseau, N.; Derreumaux, P. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 464−474.

(18) Madras, N.; Sokal, A. D. *J. Stat. Phys.* **1988**, *50*, 109−186.

(19) da Silva, R. A.; Degreve, L.; Caliri, A. *Biophys. J.* **2004**, *87*, 1567−1577.

(20) Brooks, B.; Karplus, M. *PNAS* **1983**, *80*, 6571−6575.

(21) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.

(22) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173−181.

(23) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 389−408.

(24) Al-Lazikani, B.; Jung, J.; Xiang, Z. X.; Honig, B. *Curr. Opin. Chem. Biol.* **2001**, *5*, 51−56.

(25) Xiang, Z. X.; Honig, B. *J. Mol. Biol.* **2001**, *311*, 421−430.

(26) Dunbrack, R. L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431−440.

(27) Dwyer, M. A.; Looger, L. L.; Hellinga, H. W. *Science* **2004**, *304*, 1967−1971.

(28) Looger, L. L.; Dwyer, M. A.; Smith, J. J.; Hellinga, H. W. *Nature* **2003**, *423*, 185−190.

(29) Jacobson, M. P.; Friesner, R. A.; Xiang, Z. X.; Honig, B. *J. Mol. Biol.* **2002**, *320*, 597−608.

(30) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Honig, B.; Friesner, R. A. *Proteins* **2004**, *55*, 351−367.

(31) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.

(32) Jacobson, M. P.; Kaminski, G. A.; Friesner, R. A.; Rapp, C. S. *J. Phys. Chem. B* **2002**, *106*, 11673−11680.

(33) Friesner, R. A.; Guallar, V. *Ann. Rev. Phys. Chem.* **2005**, *56*.

(34) Hays, A.-M. A.; Dunn, A. R.; Chiu, R.; Gray, H. B.; Stout, C. D.; Goodin, D. B. *J. Mol. Biol.* **2004**, *344*, 455−469.

(35) Ludemann, S. K.; Lounnas, V.; Wade, R. C. *J. Mol. Biol.* **2000**, *303*, 797−811.

(36) Wei, G. H.; Mousseau, N.; Derreumaux, P. *J. Chem. Phys.* **2002**, *117*, 11379−11387.

(37) Guallar, V.; Borrelli, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3954−3959.

(38) Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. *J. Mol. Biol.* **2004**, *337*, 227−239.

(39) *Qsite*; Schrödinger, Inc.: Portland, Oregon, 2001.

(40) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517−529.

(41) Alexey Onufriev, D. A. C.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297−1304.

(42) Wagoner, J.; Baker, N. A. *J. Comput. Chem.* **2004**, *25*, 1623−1629.

(43) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6760−6764.

(44) Zhu, J.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2005**, *109*, 3008−3022.

(45) Shen, T. Y.; Wong, C. F.; McCammon, J. A. *Biopolymers* **2003**, *70*, 252−259.

(46) McDonald, I. R.; Singer, K. *Discuss. Faraday Soc.* **1967**, *43*, 40−49.

(47) Mastny, E. A.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 14109−6.

(48) Wang, F. G.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.

(49) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. A. *J. Comput. Chem.* **2005**, *26*, 915−931.

(50) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082−2094.

(51) Henkelman, G.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9978−9985.

(52) Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580−5601.

(53) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, *109*, 6722−6731.

(54) Ozkan, S. B.; Dill, K. A.; Bahar, I. *Protein Sci.* **2002**, *11*, 1958−1970.

(55) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. *J. Phys. Chem. B* **2004**, *108*, 6582−6594.

(56) Tilton, R. F. J.; Kuntz, I. D. J.; Petsko, G. A. *Biochemistry* **1974**, *19*, 9, 2849−2857.

(57) Teeter, M. M. *Protein Sci.* **2004**, *13*, 313−318.

(58) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205−208.

(59) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.

(60) Bossa, C.; Anselmi, M.; Roccatano, D.; Amadei, A.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2004**, *86*, 414a−414a.

(61) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690−7697.

(62) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177−5188.

(63) Maguid, S.; Fernandez-Alberti, S.; Ferrelli, L.; Echave, J. *Biophys. J.* **2005**, *89*, 3−13.

(64) Hodsdon, M. E.; Cistola, D. P. *Biochemistry* **1997**, *36*, 2278−2290.

(65) Ory, J.; Kane, C. D.; Simpson, M. A.; Banaszak, L. J.; Bernlohr, D. A. *J. Biol. Chem.* **1997**, *272*, 9793−9801.

(66) Richieri, G. V.; Ogata, R. T.; Kleinfeld, A. M. *Mol. Cell Biochem.* **1999**, *192*, 77−85.

(67) Yu, Z. Y.; Jacobson, M. P.; Josovitz, J.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* 2004, ***108***, 6643−6654.

CT0501811