JCTC Journal of Chemical Theory and Computation

Hybrid Prefactor Semiclassical Initial Value Series Representation of the Quantum Propagator

Shesheng Zhang and Eli Pollak*

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

Received October 28, 2004

Abstract: One of the central advantages of the Herman Kluk Semiclassical Initial Value Representation (SCIVR) of the quantum propagator is that through its prefactor it approximately conserves unitarity for relatively long times. Its main disadvantage is that the prefactor appearing in the SCIVR propagator is expensive to compute as the dimensionality of the problem increases. When using the SCIVR series method for computation of the numerically exact quantum dynamics, the expense becomes even larger, since each term in the series involves a product of propagators, each with its own prefactor. This expense can be eliminated if one uses prefactor free propagators; however, these do not conserve unitarity as well as the HK propagator. As a compromise, we suggest the use of a hybrid propagator, in which the system variables are treated with the Herman−Kluk prefactor, while the bath variables are treated as prefactor free. Numerical application to a quartic oscillator coupled bilinearly to five harmonic bath oscillators demonstrates the viability of the hybrid method. The results presented are also a first application of the SCIVR series method to a system with six degrees of freedom. Convergence to the numerically exact answer using Monte Carlo sampling is obtained with at most the first two terms in the SCIVR series.

I. Introduction

The semiclassical initial value representation (SCIVR) approximation to the exact quantum propagator[1−3] has been demonstrated to be a practical way of approximately computing quantum mechanical effects through classical molecular dynamics simulations in systems with "many" degrees of freedom.[4−18] The most widely used SCIVR form is that suggested by Herman and Kluk (HK) in 1984.[3] Over a decade ago, Kay demonstrated[5−7] that one may use the HK approximation for the propagator and obtain reasonably good results, when compared to numerically exact quantum computations on small systems. This led to an increasing number of studies of the semiclassical IVR method, including applications to systems with many degrees of freedom.[22,23] Recent reviews may be found in refs 19−21.

A major difficulty though in the application of the HK method is the computation of the prefactor.[12,14,24−27] It is a square root of a determinant of a complex linear combination of matrices and as such must be computed at each time step to ensure continuity of its phase. For a system with N degrees of freedom, one must integrate not only the $2N$ equations of motion in phase space but also the $(2N)^2$ time dependent monodromy matrix elements. Moreover, because the HK prefactor is in the numerator, classical chaotic motion can lead to an exponentially increasing prefactor making the Monte Carlo computation virtually impossible.[25]

One way of overcoming these problems is by using prefactor free propagators,[28,33−35] of which Heller's frozen Gaussian is the first example.[2] However the HK propagator was invented by Herman and Kluk among others to precisely solve an acute problem which appears in the frozen Gaussian propagator. It rapidly loses unitarity.

In recent years, we have developed the SCIVR series method[29−33,36,37] which uses the SCIVR propagator to obtain *exact* quantum propagation. In this formalism, the SCIVR propagator is the zeroth-order term in a perturbation series where the "small" perturbation is a "correction operator"[29] whose explicit form depends on the choice of the zeroth-

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

10.1021/ct0499074 CCC: $30.25
Published on Web 03/31/2005

order SCIVR propagator. The SCIVR series method has been applied successfully to a number of problems, including one-dimensional motion on an anharmonic potential,[31,32] "deep" tunneling through a symmetric Eckart barrier,[36] the interference pattern of a Gaussian wave packet scattered through a two-dimensional double slit potential,[33] and more. In all these studies, only a few terms in the SCIVR series were needed to compute the numerically exact result, demonstrating that one could use Monte Carlo methods to converge real time quantum dynamics computations.

In the SCIVR series method, the computational price of each additional term in the series is heavy, as the n-th term involves a product of n + 1 propagators and n time integrations. One is thus interested in reducing the integration time of the underlying classical dynamics to a minimum. It is at this point, where again the HK prefactor presents a challenge due to the expense involved in its computation. We also note, that thus far the SCIVR series method has been applied at most to systems with two degrees of freedom.

In this paper we present a new method, which is based on a hybrid between the HK propagator and the prefactor free propagator.[33] The method should be especially useful for dissipative systems, that is systems which are bilinearly coupled to harmonic baths. The idea is to use the HK prefactor for the system degrees of freedom but to use the prefactor free method for the bath degrees of freedom. In Section II we derive the hybrid propagator and its associated correction operator by demanding that on the average the contribution from the bath degrees of freedom to the correction operator vanishes.

In Section III we apply this methodology and compare it with the HK based SCIVR series method for a particle moving in a quartic potential but coupled bilinearly to a bath of five harmonic oscillators. We find that the unitarity of the hybrid propagator is substantially improved as compared to the prefactor free propagator. This is also the first example of the application of the SCIVR series method to a problem with six degrees of freedom. The results presented show convergence using only the first two terms of the SCIVR series. We end the paper in Section IV with a Discussion of the results and future applications.

II. A Hybrid SCIVR Propagator

A. The SCIVR Series Method. Briefly, we review first the SCIVR series formalism. The exact quantum propagator $\hat{K}(t)$ obeys the equation of motion

$$i\hbar\frac{\partial}{\partial t}\hat{K}(t) = \hat{H}\hat{K}(t),\ \hat{K}(0) = \hat{I} \tag{2.1}$$

while the SCIVR propagator $\hat{K}_0(t)$ equation of motion has the general form:[29]

$$i\hbar\frac{\partial}{\partial t}\hat{K}_0(t) = \hat{H}\hat{K}_0(t) + \hat{C}(t),\ \hat{K}_0(0) = \hat{I} \tag{2.2}$$

Here, $\hat{H}$ is the Hamiltonian operator and $\hat{C}(t)$ is termed the "correction operator". The formal solution for the SCIVR propagator is then readily seen to be[32]

$$\hat{K}_0(t - t_i) = \hat{K}(t - t_i) - \frac{i}{\hbar}\int_{t_i}^{t} dt'\hat{K}(t - t')\hat{C}(t') \tag{2.3}$$

We may now represent the exact propagator in terms of a series, in which the j-th element is of the order of $\hat{C}(t)^j$:

$$\hat{K}(t - t_i) = \sum_{j=0}^{\infty}\hat{K}_j(t - t_i) \tag{2.4}$$

Inserting the series expansion into the formal solution 2.3 one finds the recursion relation:

$$\hat{K}_{j+1}(t - t_i) = \frac{i}{\hbar}\int_{t_i}^{t} dt'\hat{K}_j(t - t')\hat{C}(t') \tag{2.5}$$

This recursion relation together with the known form of the correction operator provides a series representation of the exact propagator. Past experience has shown that this series typically converges rapidly. In the next subsection we shall derive the hybrid SCIVR propagator and its associated correction operator.

B. The Hybrid SCIVR Propagator. We assume that we are dealing with a system with N_s degrees of freedom and a bath with $N_b = N - N_s$ degrees of freedom. The classical phase space (with mass weighted coordinates and momenta) may be separated into the system phase space $\mathbf{p}_s = (p_1, p_2, .., p_{N_s})$, $\mathbf{q}_s = (q_1, q_2, .., q_{N_s})$, and bath phase space $\mathbf{p}_b = (p_{N_{s+1}}, p_{N_{s+2}}, ..., p_N)$, $\mathbf{q}_b = (q_{N_{s+1}}, q_{N_{s+2}}, ..., q_N)$. We will further use the notation $\mathbf{Y}_s = (\mathbf{p}_s, \mathbf{q}_s)$, $\mathbf{Y}_b = (\mathbf{p}_b, \mathbf{q}_b)$, $\mathbf{Y} = (\mathbf{Y}_s, \mathbf{Y}_b)$. We assume that the Hamiltonian may be subdivided into two parts, a system dependent part only and then all the rest

$$\hat{H}(\mathbf{Y}) = \hat{H}_s(\mathbf{Y}_s) + \hat{H}_{b,s}(\mathbf{Y}) = \hat{T}_s + \hat{V}_s(\mathbf{Y}_s) + \hat{T}_b + \hat{V}_{b,s}(\mathbf{Y}) \tag{2.6}$$

where $\hat{T}$ and $\hat{V}$ are the kinetic and potential operators, respectively. The hybrid SCIVR propagator will be defined rather generally by introducing a function $f(\mathbf{Y}, t)$ such that initially $f(\mathbf{Y}, 0) = 1$:

$$\hat{K}_0(t) = \int_{-\infty}^{\infty}\frac{d\mathbf{Y}}{(2\pi\hbar)^N}R(\mathbf{Y}_s, t)f(\mathbf{Y}, t)e^{(i/\hbar)S_s(\mathbf{Y}_s,t)}|g(\mathbf{Y}, t)\rangle\langle g(\mathbf{Y}, 0)| \tag{2.7}$$

Here, $R(\mathbf{Y}_s, t)$ is taken to be the Herman−Kluk preexponential factor, defined though on the phase space $\{\mathbf{Y}_s\}$ only. It does depend on the full phase space, in the sense that the trajectories used to generate it are trajectories in the full phase space of the combined system and bath. Its initial time value is unity $R(\mathbf{Y}_s, 0) = 1$. The classical action $S(\mathbf{Y}_s, t)$ is defined on the system phase space $\{\mathbf{Y}_s\}$ as

$$S_s(\mathbf{Y}_s, t) = \int_0^t dt'[\mathbf{p}_s(t')\dot{\mathbf{q}}_s(t') - H_s(\mathbf{Y}_s(t'))] \tag{2.8}$$

however it too depends on the full phase space in the sense that the trajectories used to evaluate the system action are propagated in the full phase space of the system and the bath. The coordinate representation of the coherent state in eq 2.7 takes the form

$$\langle \mathbf{x}|g(\mathbf{p}, \mathbf{q}, t)\rangle = \left(\frac{\det(\mathbf{\Gamma})}{\pi^N}\right)^{1/4} \exp\left[-(1/2)[\mathbf{x}-\mathbf{q}_t]^T\mathbf{\Gamma}[\mathbf{x}-\mathbf{q}_t] + \frac{i}{\hbar}\mathbf{p}_t\cdot(\mathbf{x}-\mathbf{q}_t)\right] \quad (2.9)$$

and $\mathbf{\Gamma}$ in this paper will be defined as a *diagonal* time independent width parameter matrix so that it may be divided into two submatrices $\mathbf{\Gamma}_s$ for the system degrees of freedom and $\mathbf{\Gamma}_b$ for the bath degrees of freedom. With these definitions one notes that at time $t = 0$, the generalized hybrid SCIVR propagator of eq 2.7 is the identity matrix. The time dependence in the coherent state is obtained from Hamilton's equations of motion for the classical limit of the quantum Hamiltonian operator.

By expressly taking the time derivative of the hybrid SCIVR propagator and using Hamilton's equations of motion one derives the following expression for the correction operator

$$\hat{C}(t) = \int_{-\infty}^{\infty}\frac{d\mathbf{Y}}{(2\pi\hbar)^N}[\Delta V_s(\hat{\mathbf{q}}_s, t) + \Delta V_{b,s}(\hat{\mathbf{q}}, t)]R(\mathbf{Y}_s, t)\, f(\mathbf{Y}, t)\, e^{(i/\hbar)S_s(\mathbf{Y}_s,t)}|g(\mathbf{Y}, t)\rangle\langle g(\mathbf{Y}, 0)| \quad (2.10)$$

where the potential difference operators are

$$\Delta V_s(\hat{\mathbf{q}}_s, t) = i\hbar\frac{\dot{R}}{R} + V_s(\mathbf{Y}_s, t) + \nabla_s V_s[\mathbf{q}_s(t)]\cdot(\hat{\mathbf{q}}_s - \mathbf{q}_s(t)) - V_s(\hat{\mathbf{q}}_s) - \frac{\hbar^2}{2}Tr[\mathbf{\Gamma}_s] + \frac{\hbar^2}{2}(\hat{\mathbf{q}}_s - q_s(t))^T\mathbf{\Gamma}_s\mathbf{\Gamma}_s(\hat{\mathbf{q}}_s - \mathbf{q}_s(t)) \quad (2.11)$$

$$\Delta V_{b,s}(\hat{\mathbf{q}}, t) = i\hbar\frac{\dot{f}}{f} + \frac{1}{2}\mathbf{p}_b(t)^T\cdot\mathbf{p}_b(t) + \nabla V_{b,s}[\mathbf{q}(t)]\cdot(\hat{\mathbf{q}} - \mathbf{q}(t)) - V_{b,s}(\hat{\mathbf{q}}) - \frac{\hbar^2}{2}Tr[\mathbf{\Gamma}_b] + \frac{\hbar^2}{2}(\hat{\mathbf{q}}_b - \mathbf{q}_b(t))^T\mathbf{\Gamma}_b\mathbf{\Gamma}_b(\hat{\mathbf{q}}_b - \mathbf{q}_b(t)) \quad (2.12)$$

∇_s is the gradient operator defined on the subspace of the system coordinates, and ∇ is defined on the full space of the system and the bath coordinates.

The function f has yet to be determined. As in our previous derivation of prefactor free propagators,[33] we will define it by demanding that on the average, at any point and time in the full phase space $\{Y\}$, the part of the correction operator associated with the bath will vanish. That is we define f by demanding

$$\langle g(\mathbf{Y}, t)|\Delta\hat{V}_{b,s}(\hat{\mathbf{q}}, t)|g(\mathbf{Y}, t)\rangle = 0 \quad (2.13)$$

From eq 2.12 one then finds that the solution for f is

$$f(\mathbf{Y}, t) = \exp\left(\frac{i}{\hbar}S_{b,s}(\mathbf{Y}, t)\right) \quad (2.14)$$

and the bath-system action $S_{b,s}$ is found to be

$$S_{b,s}(\mathbf{Y}, t) = \int_0^t dt'\left(\frac{1}{2}\mathbf{p}_b(t')^T\cdot\mathbf{p}_b(t') - \tilde{V}_{b,s}(\mathbf{q}(t')) - \frac{\hbar^2}{4}Tr[\mathbf{\Gamma}_b]\right) \quad (2.15)$$

The coherent state bath-system averaged potential is defined as

$$\tilde{V}_{b,s}(\mathbf{q}(t)) = \left(\frac{\det[\mathbf{\Gamma}_b]}{\pi^N}\right)^{1/2}\int_{-\infty}^{\infty}d\mathbf{x}e^{-[\mathbf{x}-\mathbf{q}(t)]^T\mathbf{\Gamma}[\mathbf{x}-\mathbf{q}(t)]}V_{b,s}(\mathbf{x}) \quad (2.16)$$

Having defined the function f we can now insert it back into the original definition of the SCIVR propagator 2.7 to get the central formal results of this paper. The zeroth-order hybrid propagator is

$$\hat{K}_0(t) = \int_{-\infty}^{\infty}\frac{d\mathbf{Y}}{(2\pi\hbar)^N}R(\mathbf{Y}_s, t)e^{(i/\hbar)[S_s(\mathbf{Y}_s,t)+S_{b,s}(\mathbf{Y},t)]}|g(\mathbf{Y}, t)\rangle\langle g(\mathbf{Y}, 0)| \quad (2.17)$$

and the associated correction operator takes the form

$$\hat{C}(t) = \int_{-\infty}^{\infty}\frac{d\mathbf{Y}}{(2\pi\hbar)^N}R(\mathbf{Y}_s, t)e^{(i/\hbar)[S_s(\mathbf{Y}_s,t)+S_{b,s}(\mathbf{Y},t)]}\Delta V(\hat{\mathbf{q}}, t)|g(\mathbf{Y}, t)\rangle\langle g(\mathbf{Y}, 0)| \quad (2.18)$$

The potential difference operator is

$$\Delta V(\hat{\mathbf{q}}, t) = \Delta V_s(\hat{\mathbf{q}}_s, t) + \Delta V_{b,s}(\hat{\mathbf{q}}, t) \quad (2.19)$$

$$\Delta V_s(\hat{\mathbf{q}}_s, t) = i\hbar\frac{\dot{R}}{R} + V_s(\mathbf{Y}_s, t) + \nabla_s V[\mathbf{q}(t)]\cdot(\hat{\mathbf{q}}_s - \mathbf{q}_s(t)) - V_s(\hat{\mathbf{q}}_s) - \frac{\hbar^2}{2}Tr[\mathbf{\Gamma}_s] + \frac{\hbar^2}{2}(\hat{\mathbf{q}}_s - \mathbf{q}_s(t))^T\mathbf{\Gamma}_s\mathbf{\Gamma}_s(\hat{\mathbf{q}}_s - \mathbf{q}_s(t)) \quad (2.20)$$

$$\Delta V_{b,s}(\hat{\mathbf{q}}, t) = \tilde{V}_{b,s}(\mathbf{q}(t)) + \nabla V_{b,s}[\mathbf{q}(t)]\cdot(\hat{\mathbf{q}} - \mathbf{q}(t)) - V_{b,s}(\hat{\mathbf{q}}) - \frac{\hbar^2}{4}Tr[\mathbf{\Gamma}_b] + \frac{\hbar^2}{2}(\hat{\mathbf{q}}_b - \mathbf{q}_b(t))^T\mathbf{\Gamma}_b\mathbf{\Gamma}_b(\hat{\mathbf{q}}_b - \mathbf{q}_b(t)) \quad (2.21)$$

To complete the formulation we write down here the explicit expression for the Herman−Kluk prefactor

$$R(\mathbf{Y}_s, t) = \left(\det\left[\frac{1}{2}\left(\frac{\partial\mathbf{q}_t}{\partial\mathbf{q}} + \frac{\partial\mathbf{p}_t}{\partial\mathbf{p}} - i\hbar\mathbf{\Gamma}_s\frac{\partial\mathbf{q}_t}{\partial\mathbf{p}} + \frac{i}{\hbar}\mathbf{\Gamma}_s^{-1}\frac{\partial\mathbf{p}_t}{\partial\mathbf{q}}\right)\right]\right)^{1/2} \quad (2.22)$$

where the various terms in the brackets are the four matrices for the system variables only.

We note, that in the limit that there are no bath variables, that is $N = N_s$ then the hybrid propagator and associated correction operator are identical to the Herman−Kluk propagator and correction operator. Conversely, if all degrees of freedom are considered as the bath degrees of freedom, that is $N = N_b$, then the hybrid propagator and correction operator reduces to the prefactor free frozen Gaussian propagator of Heller and its associated correction operator.

III. Application to a Dissipative System

A. Preliminaries. As a test of the hybrid method, we studied a dissipative system of a quartic double well potential ($N_s = 1$)

$$V_s(q) = -\frac{1}{2}q^2\left(1 - \frac{q^2}{2q_a^2}\right) \quad (3.1)$$

coupled bilinearly to a bath of harmonic oscillators

$$V_{b,s}(\mathbf{q}) = \sum_{j=1}^{N_b}\frac{1}{2}\left(\omega_j x_j - \frac{c_j}{\omega_j}q\right)^2 \quad (3.2)$$

The frequencies and coupling coefficients of the harmonic bath are chosen so as to mimic a cutoff bath with cutoff frequency ω_c and friction coefficient η so that the spectral density is

$$J(\omega) = \frac{\pi}{2}\sum_{j=1}^{N_b}\frac{c_j^2}{\omega_j}\delta(\omega - \omega_j) = \eta\omega \exp^{-\omega/\omega_c} \qquad (3.3)$$

The discretization for a finite number of oscillators is carried out as suggested by Gelabert et al.[13] For the specific case of the exponential cutoff bath, this means that

$$\omega_j = -\omega_c \log\left[1 - \frac{j}{N_b}\left(1 - \exp\left(-\frac{\omega_m}{\omega_c}\right)\right)\right] \qquad (3.4)$$

$$c_j = \omega_j \sqrt{\frac{2\eta\omega_c}{\pi N_b}\left[1 - \exp\left(-\frac{\omega_m}{\omega_c}\right)\right]} \qquad (3.5)$$

where ω_m is the maximal bath frequency.

To simplify the numerical computation, we chose factorized initial conditions for the system and the bath. Accordingly, the initial condition for the system variable is that of a Gaussian wave packet

$$\langle q|\Psi\rangle = \left(\frac{\gamma}{\pi}\right)^{1/4} \exp\left[-\frac{\gamma}{2}(q - q_0)^2\right] \qquad (3.6)$$

The bath variables were distributed canonically with the bath Hamiltonian:

$$\hat{H}_b = \sum_{j=1}^{N_b}\frac{1}{2}(\hat{p}_j^2 + \omega_j^2\hat{x}_j^2) \qquad (3.7)$$

The actual parameters chosen in the numerical simulations are $q_0 = q_a = 2$, $\omega_c = \omega/2$, $\omega_m = 4\omega_c$, $\gamma = 10$, $\Gamma_s = 2.25$, $N_b = 5$, $T = 0$, $\Gamma_{b,j} = (\omega_j/\hbar)$, and $\omega = \sqrt{2}$. The system width parameter (Γ_s) was chosen to be the optimal width for the quartic symmetric double well system, as described in ref 31.

B. Normalization. The hybrid method was invented in order to improve upon the normalization of the frozen Gaussian method without having to pay the price of computing the full monodromy matrices for system and bath. The time dependent normalization of the hybrid SCIVR propagator, using the factorized initial conditions as described above, is

$$|K_0|(t) = \sqrt{Tr_b\left[\frac{e^{-\beta\hat{H}_b}}{Q_b}\langle\Psi|\hat{K}_0^\dagger(t)\hat{K}_0(t)|\Psi\rangle\right]} \qquad (3.8)$$

with $\beta = 1/k_BT$ being the inverse temperature and $Q_b = Tr_b[e^{-\beta\hat{H}_b}]$ is the partition function of the bath. Due to the choice of factorized initial conditions, the bath initial conditions are Gaussian in the bath phase space variables, while the system initial conditions are Gaussian in the system phase space variables due to the choice of the Gaussian wave packet. Therefore, the Monte Carlo computation of the normalization is straightforward.

The time dependent normalization is shown in Figure 1. These results were obtained using a Monte Carlo sample size of $8.1\cdot10^7$ for the three cases, respectively. The respective standard deviations are ≤ 0.01. Panel a of the figure shows the normalization for the case of no system bath coupling, for which the hybrid propagator reduces to the HK system propagator, so that the unitarity presented in the figure is just a test of the unitarity of the Herman−Kluk propagator for the one-dimensional quartic oscillator. As is evident from the figure the Herman−Kluk propagator is close to unity for the whole time interval presented in the figure. Panel b shows the normalization of the hybrid propagator (solid line) for weak friction ($\eta = 0.5$) and panel c for strong friction ($\eta = 3.0$). In panels b and c we show also the normalization that would have been obtained (dashed line) had we made use of the prefactor free frozen Gaussian propagator of

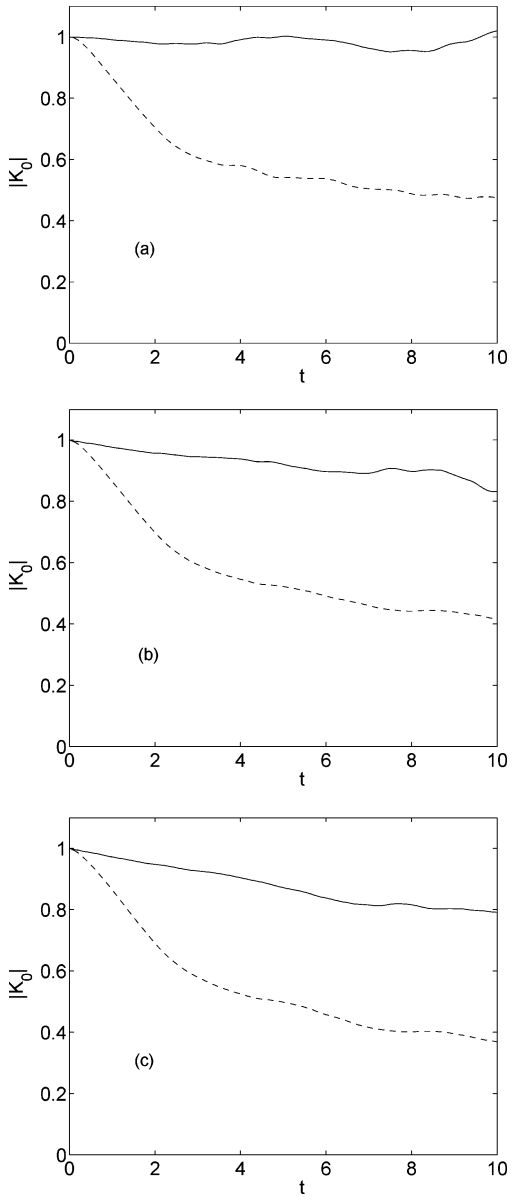


Figure 1. The time dependent normalization $|K_0|$ of the hybrid propagator for a quartic oscillator coupled bilinearly to a bath of five harmonic oscillators. Panels a−c correspond to the friction coefficient values $\eta = 0$, 0.5, 3, respectively. The solid and dashed lines correspond to the normalization obtained from the hybrid and prefactor free propagators, respectively.

Heller. As expected, use of the hybrid propagator leads to an impressive improvement in the unitarity property.

C. Thermal Overlap Function. The time dependent thermally averaged overlap function of the wave packet, which includes the first k terms in the SCIVR series, is defined as

$$A_k(t) = Tr_b\left[\frac{e^{-\beta\hat{H}_b}}{Q_b}\langle\Psi|\sum_{j=0}^{k}\hat{K}_j(t)|\Psi\rangle\right] \tag{3.9}$$

The absolute value of the thermal overlap function is plotted in Figure 2 for the same three values of the friction coefficient as before. Numerical details about the Monte Carlo averaging are given in the Appendix. Results are presented for both the Herman−Kluk propagator used for all degrees of freedom as well as the hybrid method, in which the HK form is used only for the system variables. The Monte Carlo sample size used for A_0 was 1.3×10^6 points, enough to ensure an accuracy of 0.01, while for A_1 1.74×10^7 points were used to keep the same accuracy. The CPU time needed for the HK propagator was about 10 times greater than the CPU time used for the hybrid propagator. As can be seen from the figure, the converged results are very close to each other, justifying the use of the hybrid method.

The results shown in Figure 2 for $\eta = 0$ have already been published in ref 31; they serve to remind us that for the system alone, we know that convergence is obtained by including only the first two terms in the series. At $\eta = 0$ we could compare the SCIVR series results with the numerically exact result as obtained from matrix diagonalization of the Hamiltonian. However, in the presence of friction we do not know the numerically exact results from an independent computation. The results in the figure show though that the differences between the zeroth-order and the first-order terms are sufficiently small even in the presence of friction so that we can safely assume that there is no need to go to higher order terms in the series. This is then the first example of the convergence of the SCIVR series method for real time dependent quantum dynamics on a system with six degrees of freedom.

D. The Thermal Autocorrelation Function. As a last example, we consider a more physically measurable quantity, the thermal autocorrelation function of the Gaussian wave packet, defined as

$$M(t) = Tr\left[\frac{e^{-\beta\hat{H}_b}}{Q_b}|\Psi\rangle\langle\Psi|\,\hat{K}^{\dagger}(t)|\Psi\rangle\langle\Psi|\,\hat{K}(t)\right] \tag{3.10}$$

Computation of the thermal autocorrelation function is more expensive than the overlap function since here the propagator appears twice. With respect to the SCIVR series, we will use the notation

$$M(t) = M_0(t) + \Delta_1 M(t) + \Delta_2 M(t) + ... \tag{3.11}$$

such that $\Delta_j M(t)$ denotes the term obtained by using all terms of order $\hat{C}^j$ in the SCIVR series. The results obtained for $M(t)$, based on the hybrid propagator, are shown in Figure 3 for the strong friction case of $\eta = 3.0$. They are compared

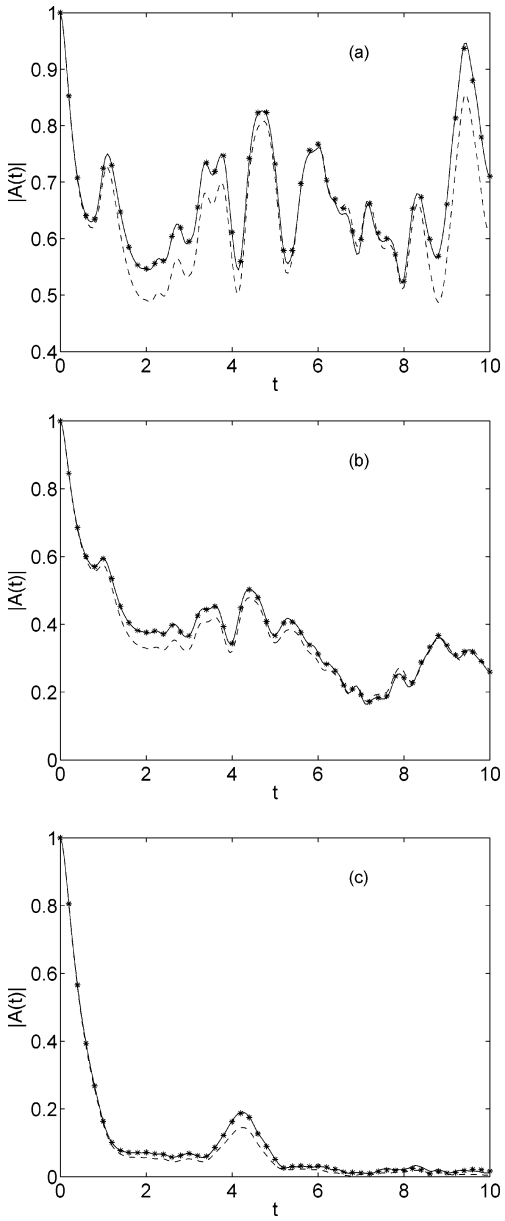


Figure 2. The time dependence of the absolute value of the thermal overlap function. Panels a−c correspond to the friction coefficient values $\eta = 0$, 0.5, 3, respectively. In all panels, the dashed and solid lines correspond to the results obtained with the zeroth-order and first-order hybrid propagator, respectively. The stars denote results obtained with the first-order HK propagator. The good agreement between the stars and the solid lines demonstrates that the hybrid propagator is as good as the HK propagator for this problem. The small differences between the zeroth-order and first-order lines indicate that the results have converged, using only the first-order term in the SCIVR series.

with the numerically exact results obtained using basis set techniques[38] and the "classical" autocorrelation function obtained by replacing the SCIVR dynamics with classical mechanics. From the figure we note, that here too, it suffices to go to first order in the SCIVR series. Higher order terms are negligible, in fact we also computed the second order correction up to the time $t = 7$ with no appreciable difference to that obtained with the first-order term only. Not surpris-

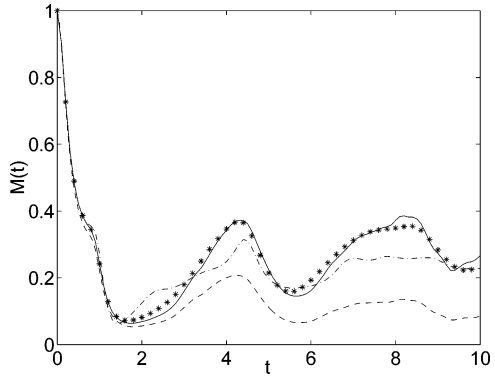


Figure 3. The time dependence of the thermal autocorrelation function for strong friction $\eta = 3.0$ using the hybrid propagator. The dashed and solid lines correspond to results obtained using the zeroth-, and zeroth-, and first-order terms in the SCIVR series, respectively. The stars are the numerically exact results of ref 38. The dashed dotted line shows the results obtained with classical mechanics propagation. Note the good agreement between the first-order SCIVR series results and the exact results.

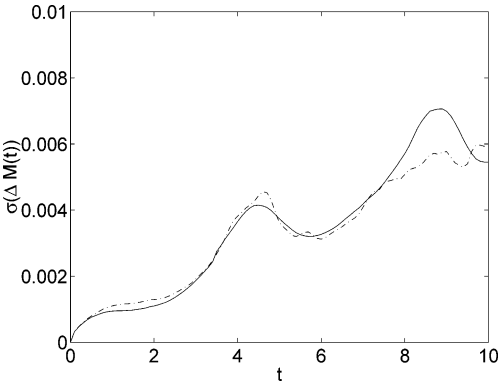


Figure 4. The time dependent variances for the leading order terms in the SCIVR series computation presented in Figure 3. The solid line is for the first-order result, the dashed−dotted line is for the zeroth-order result. The variance $\sigma(\Delta M) = \sqrt{(\langle \Delta M^2 \rangle - \langle \Delta M \rangle^2)/N}$ where N is the number of Monte Carlo points used.

ingly, we find that the deviation of the converged quantum autocorrelation function from the classical function is not very big, and at strong friction one expects dissipative quantum systems to become similar to the classical.

As for the error anlaysis, for M_0 we used a sample size of $6.5 \cdot 10^7$ points. For $\Delta_1 M$ the sample size was $6.8 \cdot 10^9$ points and for $\Delta_2 M$, $1.1 \cdot 10^{10}$ points. The step size in time was 0.01. The variances for the zeroth- and first-order terms are plotted in Figure 4. We note that the standard deviation for the first-order term is greater than the differences between $M_1(t)$ and the numerically exact result of ref 38.

For the thermal autocorrelation function, use of the Herman−Kluk propagator instead of the hybrid propagator would have been numerically prohibitive, the computational power of the 64 processor pentium IV PC farm at our disposal would not have been sufficient to converge the computation within a period of less than six months. The reason for this is clear. For the HK propagator, computation of the prefactor requires the solution of $2N + 1 + (2N)^2 = 157$ coupled equations of motion for our model with $5 + 1$ degrees of freedom. For the hybrid method one needs $2N + 1 + (2N_s)^2 = 17$ coupled equations. The added expense also grows as a power law for each additional term in the SCIVR series, since each additional term requires an additional propagator with its independent prefactor. It is thus not surprising that the hybrid method leads to a considerable saving in computational time.

IV. Discussion

In this paper we have presented a new hybrid SCIVR series representation of the quantum propagator which may be considered as a compromise between the series representation based on the Herman−Kluk and the Heller prefactor free frozen Gaussian propagators. The hybrid zeroth-order propagator conserves unitarity significantly better than the Heller propagator, but the numerical effort is much smaller than that needed for the Herman−Kluk propagator. The hybrid method would seem to be ideally suited for dissipative systems where there is a clear identification of a system and a bath. Its usefulness for more general systems remains to be tested.

This paper presents also for the first time the application of the SCIVR series method to a system with 6 degrees of freedom. We found that as in our experience with one- and two-dimensional systems, the series converges rapidly and that only the first two terms in the series suffice for obtaining a converged result.

The HK propagator is known to be problematic for strongly chaotic systems, since the HK prefactor then gives an exponentially growing in time term which makes it very difficult to converge a Monte Carlo computation. The hybrid method has the potential of overcoming this problem, by judicious choice of the "system" and the "bath" so that the system part would not include the strongly chaotic trajectories.

The computations presented in this paper were for a finite time, of the order of three oscillations of the particle in a well of the quartic potential. Integration for longer times would have necessitated going to higher order terms in the SCIVR series expansion, making the computation prohibitive for our modest computational resources. We also note that all our results are for $T = 0$, which may considered as a "tough" example as the system would tend to be more classical as the temperature is increased.

Finally, although we have used the factorization approximation for the bath, this is not essential. One may treat the full thermal Hamiltonian using imaginary time path integral methods.[39,40] This would make the programming and computation somewhat more expensive but should not make it impossible. In this paper, we were interested to demonstrate the utility of the hybrid method and the fact that the SCIVR series can be computed and converged for "large" systems. Doing away with the factorization approximation is a topic for future work.

Acknowledgment. This work has been supported by grants from the U.S. Israel Binational Science Foundation and the Israel Science Foundation.

V. Appendix: Monte Carlo Evaluation of Terms in the SCIVR Series for the Thermal Overlap Function

In this Appendix we will describe the Monte Carlo methodology we used to evaluate the thermal overlap function. First we note that the trace of the harmonic bath Hamiltonian and the initial Gaussian wave packet with the coherent state at time t gives a Gaussian in the phase space variables at time t:

$$Tr_b[e^{-\beta\hat{H}_b}\langle\Psi|g(\mathbf{Y}, t)\rangle\langle g(\mathbf{Y}, 0)|\Psi\rangle] = \langle g(\mathbf{Y}_b, 0)|e^{-\beta\hat{H}_b}|g(\mathbf{Y}_b, t)\rangle\langle\Psi|g(\mathbf{Y}_s, t)\rangle\langle g(\mathbf{Y}_s, 0)|\Psi\rangle \quad (A.1)$$

The rather lengthy analytic expression for the bath matrix element $\langle g(\mathbf{Y}_b, 0)|e^{-\beta\hat{H}_b}|g(\mathbf{Y}_b, t)\rangle$ is given in eq 3.20 of ref 13 and provides a Gaussian weighting for the initial conditons of the bath. The system term $\langle g(\mathbf{Y}_s, 0)|\Psi\rangle$ provides a Gaussian weighting for the system initial condition since

$$\langle g(\mathbf{Y}_s, t)|\Psi\rangle = \Gamma_1^{1/4} e^{-\Gamma_2(q_t - x_0)^2 - (p_t^2)/(4\Gamma_a\hbar^2) + (i)/(2\Gamma_a\hbar)p_t(q_t - x_0)} \quad (A.2)$$

where $\Gamma_a = (\Gamma_s + \gamma)/2, \Gamma_1 = \gamma\Gamma_s/\Gamma_a^2$, $\Gamma_2 = \gamma\Gamma_s/(4\Gamma_a)$.

The computation of the leading order term is now straightforward. Each Gaussian in either coordinate or momentum is converted in standard fashion to a variable defined on the [0, 1] interval and then generated randomly. One then integrates trajectories forward from time 0 to the final time and collects the data. For the leading order term one can use each time step in the integration to obtain the thermal overlap function at each intermediate time.

The first-order term is somewhat more complex, since now one has two phase space integrations, one coming from the SCIVR operator $\hat{K}_0(t)$, the other coming from the correction operator $\hat{C}(t)$. Both of them are rewritten as

$$\hat{K}_0(t - t') = \int_{-\infty}^{\infty} d\mathbf{Y}\, E_0(\mathbf{Y}, t - t')|g(\mathbf{Y}, t - t')\rangle\langle g(\mathbf{Y}, 0)| \quad (A.3)$$

$$\hat{C}(t') = \int_{-\infty}^{\infty} d\mathbf{Y}^c E_0(\mathbf{Y}^c, t')\Delta V(\hat{\mathbf{q}}, t')|g(\mathbf{Y}^c, t')\rangle\langle g(\mathbf{Y}^c, 0)| \quad (A.4)$$

and we have used the notation

$$E_0(\mathbf{Y}, t) = \frac{1}{(2\pi\hbar)^N} R(\mathbf{Y}_s, t) e^{(i/\hbar)[S_s(\mathbf{Y}_s, t) + S_{b,s}(\mathbf{Y}, t)]} \quad (A.5)$$

The integrand of the first-order term may thus be rewritten as

$$Tr_b[e^{-\beta\hat{H}_b}\langle\Psi|\hat{K}_0(t - t')\hat{C}(t')|\Psi\rangle] = \int_{-\infty}^{\infty} d\mathbf{Y}\int_{-\infty}^{\infty} d\mathbf{Y}^c E_0(\mathbf{Y}, t - t')E_0(\mathbf{Y}^c, t') \times \langle g(\mathbf{Y}, 0)|\Delta V(\hat{\mathbf{q}}, t')|g(\mathbf{Y}^c, t')\rangle Tr_b[e^{-\beta\hat{H}_b}\langle\Psi|g(\mathbf{Y}, t - t')\rangle\langle g(\mathbf{Y}^c, 0)|\Psi\rangle] \quad (A.6)$$

As in the case of the leading order term, the trace term in the expression is again a Gaussian in both the initial and final time phase space coordinates and momenta

$$Tr_b[e^{-\beta\hat{H}_b}\langle\Psi|g(\mathbf{Y}, t - t')\rangle\langle g(\mathbf{Y}^c, 0)|\Psi\rangle] = \langle g(\mathbf{Y}_b^c, 0)|e^{-\beta\hat{H}_b}|g(\mathbf{Y}_b, t - t')\rangle\langle\Psi|g(\mathbf{Y}_s, t - t')\rangle\langle g(\mathbf{Y}_s^c, 0)|\Psi\rangle \quad (A.7)$$

thus providing a Gaussian weighting to the phase space variables included in $\mathbf{Y}$ and $\mathbf{Y}^c$. Since the phase space volume element is invariant in time ($d\mathbf{Y} = d\mathbf{Y}_t$), one can replace the integration over initial points in phase space with an integration over the final points. This then leads to Gaussian final conditions for the phase space points associated with the propagator $\hat{K}_0$ as well as Gaussian initial conditions for the phase space associated with the correction operator $\hat{C}(t)$. This means that for the first-order correction we have two sets of trajectories, those run backward in time associated with the SCIVR propagator and those run forward in time associated with the correction operator.

Operationally, one propagates a randomly chosen phase space point backward in time from the final time t to the initial time 0, and one has a different initial condition from which a trajectory is propagated from time 0 to time t. These two trajectories then contribute a vector of points for all the intermediate times t'. This process is then repeated a large number of times, to obtain the converged integrand which is then time integrated by Simpson's rule.

The same procedure becomes more costly when considering the integrand of the second-order contribution to the SCIVR series which now involves a product of three operators:

$$Tr_b[e^{-\beta\hat{H}_b}\langle\Psi|\hat{K}_0(t - t' - t'')\hat{C}(t'')\hat{C}(t')|\Psi\rangle] = \int_{-\infty}^{\infty} d\mathbf{Y}\int_{-\infty}^{\infty} d\mathbf{Y}^{c''}\int_{-\infty}^{\infty} d\mathbf{Y}^{c'} E_0(\mathbf{Y}, t - t'' - t')E_0(\mathbf{Y}^{c''}, t'')E_0(\mathbf{Y}^{c'}, t') \langle g(\mathbf{Y}^{c''}, 0)|\Delta V(\hat{\mathbf{q}}, t')|g(\mathbf{Y}^{c'}, t')\rangle \times \langle g(\mathbf{Y}, 0)|\Delta V(\hat{\mathbf{q}}, t'')|g(\mathbf{Y}^{c''}, t'')\rangle Tr_b[e^{-\beta\hat{H}_b}\langle\Psi|g(\mathbf{Y}, t - t'' - t')\rangle\langle g(\mathbf{Y}^{c'}, 0)|\Psi\rangle] \quad (A.8)$$

Now we have three independent phase space volumes, one coming from the propagator $\hat{K}_0$, one coming from the correction operator $\hat{C}(t'')$, and the third one from $\hat{C}(t')$. As in the case of the zeroth- and first-order contributions, the term $Tr_b[e^{-\beta\hat{H}_b}\langle\Psi|g(\mathbf{Y}, t - t'' - t')\rangle\langle g(\mathbf{Y}^{c'}, 0)\Psi\rangle]$ is a Gaussian, providing a Gaussian weighting to two phase space volumes. However, the third intermediate Gaussian weighting results from the time dependent overlap of the two correction operators, which we write in the following form

$$\langle g(\mathbf{Y}_1, 0)|\Delta V(\hat{\mathbf{q}}, t')|g(\mathbf{Y}_2, t')\rangle = \langle g(\mathbf{Y}_1, 0)|g(\mathbf{Y}_2, t')\rangle \frac{\langle g(\mathbf{Y}_1, 0)|\Delta V(\hat{\mathbf{q}}, t')|g(\mathbf{Y}_2, t')\rangle}{\langle g(\mathbf{Y}_1, 0)|g(\mathbf{Y}_2, t')\rangle} \quad (A.9)$$

The overlap of the two coherent states is known analytically and again gives a Gaussian distribution

$$\langle g(\mathbf{Y}_1, 0)|g(\mathbf{Y}_2, t')\rangle = \exp\Big[-\frac{1}{4}[\mathbf{q}_1(0) - \mathbf{q}_2(t')]^{\mathbf{T}}\mathbf{\Gamma}[\mathbf{q}_1(0) - \mathbf{q}_2(t')]\Big] \times \exp\Big[-\frac{1}{4\hbar}[\mathbf{p}_1(0) - \mathbf{p}_2(t')]^{\mathbf{T}}\mathbf{\Gamma}^{-1}[\mathbf{p}_1(0) - \mathbf{p}_2(t')] + \frac{i}{2\hbar}[\mathbf{q}_1(0) - \mathbf{q}_2(t')][\mathbf{p}_1(0) + \mathbf{p}_2(t')]\Big] \quad (A.10)$$

The problem is that this Gaussian distribution is time dependent. Thus one can no longer run trajectories from the end and initial time points only as before and store the time dependent data between, but one has to propagate trajectories at each intermediate time point from a newly generated random distribution which depends on the end point of a

previous trajectory at that time. Furthermore, one now has to collect the generated data into a matrix of times t' and t'' so that finally the double time integration may be carried out using the trapezoid rule.

Although becoming complicated, the actual algorithm, up to order two is straightforward. The computation of the third-order term though is much more prohibitive, since now one would have two intermediate sets of phase space variables and would have to store a matrix for three times, so that the storage requirements may become critical. In principle this could be overcome by treating also the time integration using Monte Carlo methods, but to date we have not tested how well one could converge such a computation.

References

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

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

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

(4) Campolieti, G.; Brumer, P. *J. Chem. Phys.* **1992**, *96*, 5969.

(5) Kay, K. G. *J. Chem. Phys.* **1994**, *100*, 4377.

(6) Kay, K. G. *J. Chem. Phys.* **1994**, *100*, 4432.

(7) Kay, K. G. *J. Chem. Phys.* **1994**, *101*, 2250.

(8) Garaschuk, S.; Tannor, D. J. *Chem. Phys. Lett.* **1996**, *262*, 477.

(9) Makri, N.; Thompson, K. *Chem. Phys. Lett.* **1998**, *291*, 101.

(10) de Sand, G. v.; Rost, J. M. *Phys. Rev. A* **1999**, *59*, R1723.

(11) Thompson, K.; Makri, N. *J. Chem. Phys.* **1999**, *110*, 1343.

(12) Shao, J.; Makri, N. *J. Phys. Chem. A* **1999**, *103*, 7753.

(13) Gelabert, R.; Gimenéz, X.; Thoss, M.; Wang, H.; Miller, W. H. *J. Chem. Phys.* **2001**, *114*, 2572.

(14) Sun, X.; Miller, W. H. *J. Chem. Phys.* **1999**, *110*, 6635.

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

(16) Thoss, M.; Wang, H.; Miller, W. H. *J. Chem. Phys.* **2001**, *114*, 9220.

(17) Shalashilin, D. V.; Child, M. S. *J. Chem. Phys.* **2001**, *115*, 5367.

(18) Wright, N. J.; Makri, N. *J. Phys Chem. B* **2004**, *108*, 6816.

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

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

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

(22) Elran, Y.; Brumer, P. *J. Chem. Phys.* **2004**, *121*, 6763.

(23) Kunikeev, S. D.; Atilgan, E.; Taylor, H. S.; Kaledin, A. L.; Main, J. *J. Chem. Phys.* **2004**, *120*, 6478.

(24) Shao, J.; Makri, N. *J. Phys. Chem. A* **1999**, *103*, 9479.

(25) Maitra, N. T. *J. Chem. Phys.* **2000**, *112*, 531.

(26) Gelabert, R.; Gimenez, X.; Thoss, M.; Wang, H.; Miller, W. H. *J. Phys. Chem. A* **2000**, *104*, 10321.

(27) Harabati, C.; Rost, J. M.; Grossmann, F. *J. Chem. Phys.* **2004**, *120*, 26.

(28) Guallar, V.; Batista, V. S.; Miller, W. H. *J. Chem. Phys.* **1999**, *110*, 9922.

(29) Ankerhold, J.; Saltzer, M.; Pollak, E. *J. Chem. Phys.* **2002**, *116*, 5925.

(30) Pollak, E.; Shao, J. *J. Phys. Chem. A* **2003**, *107*, 7112.

(31) Zhang, S.; Pollak, E. *J. Chem. Phys.* **2003**, *119*, 11058.

(32) Zhang, S.; Pollak, E. *Phys. Rev. Lett.* **2003**, *91*, 190201.

(33) Zhang, S.; Pollak, E. *J. Chem. Phys.* **2004**, *121*, 3384.

(34) Ovchinnikov, M.; Apkarian, V. A. *J. Chem. Phys.* **1996**, *105*, 10312.

(35) Sun, X.; Miller, W. H. *J. Chem. Phys.* **1997**, *106*, 916.

(36) Zhang, D.-H.; Pollak, E. *Phys. Rev. Lett.* **2004**, *93*, 140401.

(37) Pollak, E.; Miret-Artés, S. *J. Phys. A* **2004**, *37*, 9669.

(38) Zhang, D.-H.; Pollak, E. To be published.

(39) Berne, B. J.; Thirumulai, D. *Annu. Rev. Phys. Chem.* **1986**, *7*, 401.

(40) Ceperley, D. M. *Rev. Mod. Phys.* **1995**, *7*, 279.

CT0499074

JCTC Journal of Chemical Theory and Computation

Dynamics of the Staudinger Reaction

Wei Quan Tian† and Yan Alexander Wang*

Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

Received October 14, 2004

Abstract: The Staudinger reaction of phosphane and azide has been investigated by Atom-centered Density Matrix Propagation (ADMP) approach to ab initio molecular dynamics (AIMD) in combination with molecular orbital analysis within density functional theory. At room temperature, the reaction pathway with the cis initial attack dominates the Staudinger reaction. Electrostatic interaction, charge transfer, and covalent overlap are responsible for the initial attack and for the system to overcome the initial reaction barrier. The rotation of PH_3 and PH vibrations facilitate the isomerization of the system from cis intermediate to the last transition state, which indicates that small substituent groups on phosphane can facilitate the last stage of the Staudinger reaction. During the course of the reaction, the change of the average polarizability correlates positively to the change of the potential energy of the system, which clearly suggests that polar solvents can facilitate the overall reaction by stabilizing all transition states and reducing all reaction barriers.

I. Introduction

Staudinger reactions are widely utilized in organic chemistry[1−4] and biology.[5] In a Staudinger reaction, phosphane (**A**) reacts with azide (**B**) to produce phosphazene (**C**) and nitrogen gas[1] (as shown in Scheme 1). With density functional theory (DFT), we investigated reaction mechanisms of Staudinger reactions and identified four initial reaction pathways.[6] The cis-reaction pathway is the most accessible for Staudinger reactions.[6−8] Both of the cis- and trans-intermediates were observed in experiments;[9−11] the existence of the trans-intermediate is the result of isomerization from the cis-intermediate.[6,8] The initial trans-reaction barrier is much higher than the initial cis-reaction barrier.[6,8] The remaining two initial reaction pathways, gamma and concerted attacks (as shown in Scheme 1), were also studied.[6] The concerted initial reaction can be realized with appropriate substituent groups on phosphane and azide. The one-step gamma initial reaction has a reaction barrier lower than the trans-initial reaction barrier but higher than the cis-initial reaction barrier. The trans-initial reaction is always unfavorable because there is only electrostatic attraction between P and N_α to stabilize the transition state, while P can have electrostatic attractions with both N_α and N_γ in the cis- and concerted transition states.[6] From a long distance (e.g. 4 Å), P in phosphane has electrostatic attractions with N_α and N_γ and electrostatic repulsion with N_β in both the initial cis- and concerted transition states. The electrostatic attraction is much stronger than the electrostatic repulsion as indicated by the natural charges of these atoms (as shown in Scheme 1). In the initial concerted transition state, the large distortion of N_3 backbone in azide from nearly linear structure is an unfavorable factor compared with the more open initial cis-transition state. However, the preference of the cis-initial attack has not been fully understood dynamically. Even with conventional quantum mechanical (QM) calculations on the stationary points of the potential energy surface (PES), the dynamical information of preference of the initial reaction pathway is not clear, especially about the interaction of phosphane with azide during the initial reaction phase.

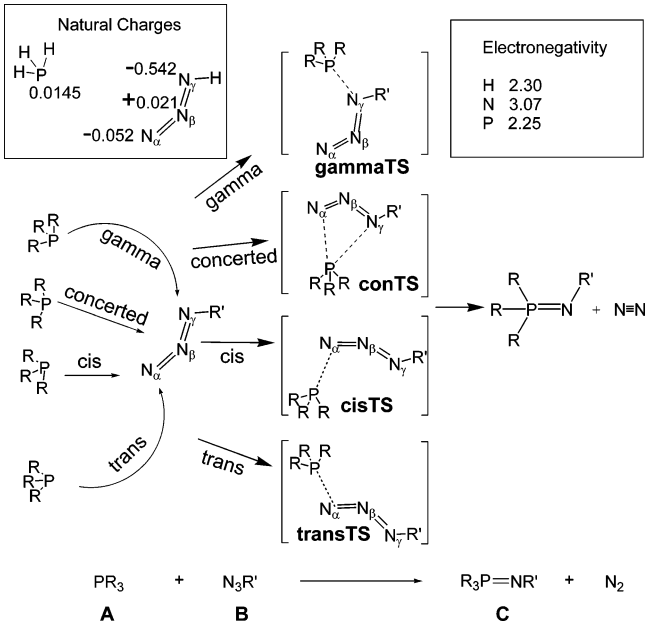
After the initial reaction, two possible cis-intermediates (**cis** or **cisclose** as illustrated in Scheme 2) form. Among these two cis-intermediates, only the **cis** can isomerize to a trans-intermediate (**trans**).[6] Both cis-intermediates can proceed forward to form another intermediate (**int**) and to reach the

* Corresponding author e-mail: yawang@chem.ubc.ca.

† Current address: Department of Molecular and Material Science, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, 6-1 Kasugakoen, Kasuga, Fukuoka 816-8580, Japan.

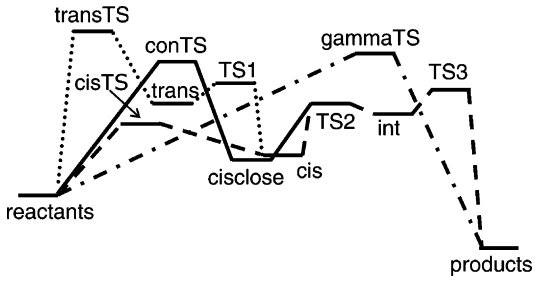
10.1021/ct049918x CCC: $30.25
Published on Web 04/19/2005

Scheme 1. Illustration of the Staudinger Reaction of Phosphane (PR_3) + Azide (N_3R')[a]



[a] The electronegativities are from ref 12; the charges are from nature charge analysis.

Scheme 2. Possible Reaction Pathways for the Staudinger Reaction[a]



[a] Adapted from ref 6.

final products (phosphazene and N_2).[6] Our studies[6] and other theoretical works[7,8] indicated that the PES is very flat at the N_2 dissociation region around **TS3** (as shown in Scheme 2). Intrinsic reaction coordinate[13] model can locate the minimum energy reaction pathway starting from a transition state; however, the dynamic detail of a reaction pathway starting from a minimum could not be explored with this model.

Molecular dynamics (MD) involving propagation of nuclei in molecule on the PES by solving Newton's equation of motion provides rich information about reactivity and dynamics of a system. The PES could be obtained by fitting to experimental or computational data. However, for polyatomic systems, the experimental data for dynamics are sparse and PES fitting is not a trivial task. Empirical force field has gained wide popularity especially in MD simulations of large systems (e.g. in biology,[14,15] solid-state physics and surface science[16]). Nonetheless, when the quantum effect is important, classical trajectory simulation with the empirical force field cannot produce qualitatively correct result. QM force field, also known as ab initio MD (AIMD) or quantum mechanical MD (QMMD), is a natural choice to overcome this difficulty. In AIMD, (classical) nuclei move on the electronic PES whose energy and derivatives are calculated directly from ab initio methods. Born−Oppenheimer MD (BOMD)[17] methods and extended Lagrangian MD (ELMD) methods are two major flavors of AIMD. In BOMD, the electronic structure calculation is fully converged at each nuclear configuration. While in ELMD, both the electronic wave function and the nuclei are treated as dynamical variables and are propagated simultaneously. The time-consuming feature of BOMD refrains its broad applications. ELMD, which produces comparable dynamics of nuclei to that from BOMD but with lower cost, has been embraced in both physics and chemistry communities, especially since the seminal work of Car and Parrinello.[18] Car-Parrinello MD (CPMD) is a prototype of ELMD. Aimed to treat condensed phases, CPMD employs pseudopotentials and a large number of plane-wave basis functions, which are natural choices for describing condensed phases.[19] It is possible to use atom-centered Gaussian basis function to carry out CPMD,[20−22] although a strict energy conservation problem remains to be solved.[20] In molecular systems of chemical reactions in solution or in gas phase, atom-centered basis functions are more chemically intuitive choices due to their localized nature. Recently, atom-centered density matrix propagation (ADMP) based MD emerged[23−26] and paved a new way to perform AIMD simulations. This new method provides a "novel and robust computational tool to perform AIMD",[24] especially for chemical reactions. Kohn−Sham molecular orbitals are propagated in conventional CPMD,[20−22] while it is the single-particle density matrix that is propagated with nuclei in ADMP.[23−26] Of course, other variables can also be propagated in AIMD.[27]

In present work, with ADMP, we will study the details of the initial Staudinger reaction of (PH_3 + N_3H) and subsequent isomerization of the cis-initial intermediate. The details of the reaction mechanisms of this reaction have been reported before.[6−8] In combination with molecular electronic theory, we will shed lights on the detailed molecular orbital interactions of the two reactants in the initial reaction stage.

II. Computational Methods

The initial applications[25] and analysis[26] on ADMP manifested that ADMP produces a similar PES to that of BOMD while possess some advantages over the plane-wave based CPMD, e.g. ADMP has no systematic bias due to the fictitious electronic mass in computing molecular properties.[25] In ADMP with orthonormal basis, an extended Lagrangian for a system is[23]

$$\mathcal{L} = \frac{1}{2}\mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{M}\mathbf{V}) + \frac{1}{2}\mu\mathrm{Tr}(\mathbf{W}\mathbf{W}) - E(\mathbf{R},\mathbf{P}) - \mathrm{Tr}[\Lambda(\mathbf{P}\mathbf{P}-\mathbf{P})] \quad (1)$$

where **M**, **R**, **V**, **P**, **W**, and μ are the nuclear masses, nuclear positions, nuclear velocities, density matrix, density matrix velocity, and fictitious mass for the electronic degrees of freedom, respectively. The first term in eq 1 is the kinetic energy of nuclei; the second term is the kinetic energy for the electronic degrees of freedom; the third term is the electronic energy of the system; and the fourth term is a constraints on the total number of electrons N_e and on the

idempotency of the density matrix with a Largrangian multiplier matrix Λ. In the orthonormal basis, the Euler−Lagrange equations of motion of the nuclei and of the density matrix are[23]

$$\mathbf{M}\frac{d^2R}{dt^2} = -\left.\frac{\partial E(R,P)}{\partial R}\right|_P \tag{2}$$

and

$$\mu\frac{d^2P}{dt^2} = -\left[\left.\frac{\partial E(R,P)}{\partial P}\right|_R + \Lambda P + P\Lambda - \Lambda\right] \tag{3}$$

The nuclei and density matrix are propagated with eqs 2 and 3, respectively. In CPDM, the molecular orbitals rather than the density matrix are propagated[19]

$$\mu\frac{d^2\psi}{dt^2} = -\left.\frac{\partial E(R,\psi)}{\partial \psi}\right|_R + \text{constraints} \tag{4}$$

Proper choosing of the fictitious mass of the electronic degrees of freedom and time step in ADMP simulations ensures a good energy conservation and the adiabaticity between the nuclear and electronic motions.[23,24] One important advantage of ADMP over the traditional CPMD is that ADMP can treat the isotope effect of hydrogen and deuterium,[24,25] which becomes an important dynamical factor in processes involving hydrogen atoms, e.g. proton transfer and PH bond vibrations and rotations. In ADMP simulations, the initial velocities of nuclei are randomly generated to simulate Boltzmann distribution.[23−25] The kinetic energy of a system also affects the adiabaticity between nuclei and electrons; this kinetic energy should be much smaller than the gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of the system to ensure that the dynamics is simulated close to the Born−Oppenheimer ground-state surface, well below the lowest excited electronic state.[24,25] Usually a thermostating method, such as Nosé-Hoover thermostats,[28] on the electronic subsystem is applied when the kinetic energy of the system is too high. MD simulation provides complementary information about the thermodynamic, dynamical properties, and microscopic motions of nuclei of a chemical reaction. However, due to the computing effort on the electronic energy of the system, ADMP is time-consuming for large systems, though the ONIOM model can be applied in some cases.[29] Furthermore, due to the nature of AIMD, the zero-point vibrational energy (ZPVE) correction is not incorporated for the PES.

Quantum chemical package Gaussian 03[30] has been employed for the calculations. Various stationary points (e.g. **cisTS**, **gammaTS**, **cis**, **TS1**, **TS2**, and **TS3**) on the PES from QM calculations[6] are used as starting points for MD simulations for the Staudinger reaction. To be consistent with previous QM calculations, B3LYP[31,32] with 6-31G(d) basis set is used for MD simulations. To get appropriate PES for the **TS3** region, the ZPVE correction is included in the static QM calculations.[6] However, no ZPVE corrections are considered in ADMP simulations in our present studies. The performance of B3LYP with the 6-31G(d) basis set[6,33] gives us confidence in applying this method in present MD simulations.

The coordinates of the system in the Staudinger reaction during ADMP simulations are extracted for single point QM calculations with natural bond orbital (NBO)[34] and molecular orbital analysis. In ADMP simulations, two different conditions are employed, adiabatic MD and thermostatic MD. In adiabatic MD, once the system is given an initial kinetic energy, the total energy is conserved during the entire simulation; most of the ADMP simulations are carried out under this condition. In thermostatic MD simulations, the kinetic energy of the system is kept constant. In all ADMP simulations, no ZPVE correction is made. For convenience, the initial kinetic energy for each ADMP simulation is chosen to be 0.01 eV (6.3 kcal/mol) for transition states or a multiple of 0.01 eV for minima. All of the kinetic energies are chosen to be much smaller than the HOMO−LUMO gap of all the stationary points investigated to keep the trajectories close to the ground-state PES and well below any excited-state PES. The smallest HOMO−LUMO gap is 3.08 eV (71.0 kcal/mol) for **transTS**, which is much larger than any of the initial kinetic energies added to the system at the stationary points. Last, due to the finite number of trajectories we have sampled, we do not claim that our results are statistically accurate; rather we are confident about the qualitative understanding derived from the ADMP simulations presented hereafter.

III. Results and Discussion

A. gammaTS of $PH_3 + N_3H$. Table 1 and Scheme 3 display the energies and reaction profiles of the Staudinger reaction, respectively. The ADMP simulations starting from **gammaTS** are performed with time step 0.2 fs for 400 fs; the system has 6.3 kcal/mol (0.01 eV, another value could also be chosen) initial kinetic energy with fictitious electronic mass of 0.1 amu Bohr^2. These ADMP simulations end with either the reactants ($PH_3 + N_3H$) or the products ($N_2 + H_3P{=}NH$) with **gammaTS** bridging in between.

B. cisTS of $PH_3 + N_3H$. With the same conditions, ADMP simulations are carried out for **cisTS** for 2 ps. The change of potential energy of the system during simulations and average polarizability of some selected points are shown in Figure 1. In the reverse reaction direction to reactants PH_3 and N_3H, **cisTS** completely dissociates to PH_3 and N_3H around 84 fs as shown in Figure 1; at this stage, the PN_α and PN_γ distances are 4.4 and 4.1 Å, respectively. After the dissociation, the potential energy of the system keeps constant with small fluctuations. The dissociation energy is 21 kcal/mol, i.e. the system needs 21 kcal/mol to initialize the Staudinger reaction; this is in good agreement with our recent DFT calculations if the ZPVE correction is not included.[15]

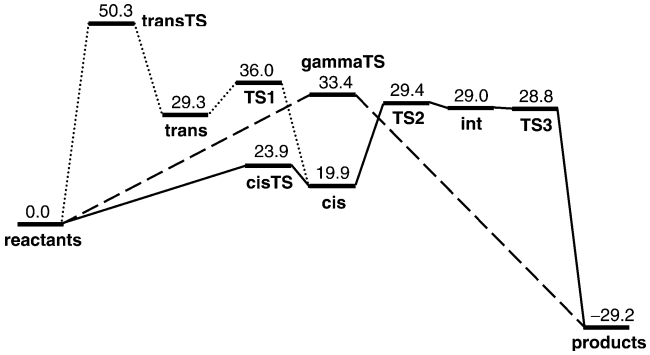
The average polarizability decreases along the reverse reaction trajectory; this indicates that solvent effects on the Staudinger reaction ($PH_3 + N_3H$) will get stronger as PH_3 approaches N_3H in the initial attack. The average polarizability of the system begins to increase at 40 fs, when the strong interaction between PH_3 and N_3H occurs. (This is further verified by charge transfer between these two reactants in Figure 11.)

Table 1. HOMO−LUMO Gaps (in eV) and the Relative Electronic Energies (with and without the ZPVE Correction) and the Relative Gibbs Free Energies (at 1 atm and 298 K) of the Stationary Points along the Staudinger Reaction[a]

	gammaTS	transTS	cisTS	trans	TS1	cis	TS2	int	TS3	products
HOMO−LUMO gap	4.24	3.08	5.07	5.35	4.90	5.80	5.56	5.54	5.67	
electronic (no ZPVE)	32.8	48.2	21.6	25.5	32.3	15.8	26.4	25.6	27.4	−30.1
electronic (with ZPVE)	33.4	50.3	23.9	29.3	36.0	19.9	29.4	29.0	28.8	−29.2
Gibbs	41.9	59.6	33.2	38.7	45.7	29.7	39.5	38.8	39.3	−28.8

[a] All relative energies are measured with respect to that of the reactants and in kcal/mol.

Scheme 3. Reaction Profiles of the Staudinger Reaction of $PH_3 + N_3H$[a]



[a] The numbers are relative electronic energy with the ZPVE correction (in kcal/mol) of stationary points (measured with respect to that of the reactants).

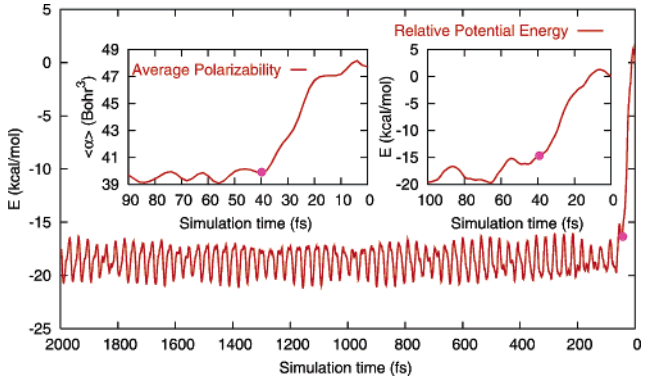


Figure 1. The relative potential energy (in kcal/mol) of the system to **cisTS** during ADMP simulation of 2 ps. The inserts are the average polarizability (in Bohr^3) of system during the first 90 fs simulation and the relative potential energy for the first 100 fs, respectively.

The potential energy, geometric changes, and molecular orbital (MO) interactions between N_3H and PH_3 of some selected points for the first 100 fs simulations are shown in Figure 2. From the geometries of the five points shown in Figure 2, one notices that bond distance difference between $R_{PN\alpha}$ and $R_{PN\gamma}$ increases as PH_3 approaches N_3H. This means that the initial approach of PH_3 to N_3H for the cis attack is not directly toward N_α: P approaches rather simultaneously to N_α and N_γ. As getting close to N_3H, PH_3 moves attractively toward N_α in N_3H due to the electrostatic attraction and covalent interaction. It is very evident from the structure and molecular orbtials of the five points in Figure 2 that the motion of PH_3 approaching N_3H is translation concurrently with rotation of PH_3. Due to the relative orientation of PH_3 to N_3H, the molecular orbital overlap begins to turn on at 40 fs with $R_{PN\alpha}$ and $R_{PN\gamma}$ distances of 2.9 and 3.2 Å, respectively, as PH_3 approaches N_3H; the electrostatic attraction should play a prominent role for the initial approach of PH_3 to N_3H (more discussed later).

At point **E** in Figure 2, there is no covalent interaction (MO overlap) between PH_3 and N_3H, and the three MOs shown are the lone pair electrons of N and P. As the two molecules get closer at **D**, the MOs of the lone pair electrons on P and N adjust themselves according to the relative orientation of PH_3 and N_3H and no MO overlap occurs. When the two molecules get closer at point **C**, orbital overlap and charge transfer between PH_3 and N_3H occur as indicated by the relevant MOs. P donates electrons to N_γ from its lone pair, and N_α back-donates the lone pair to P through orbital overlap. This explains the strong interaction of P with both N_α and N_γ in **cisTS**. As these two molecules get closer as manifested by points **B** and **A**, the covalent interactions get stronger (more MOs for **A**, **B**, **C**, and **D** are shown in Figure 2s of the Supporting Information), which help to overcome the initial reaction barrier to reach intermediate, **cis**. The empty d orbitals on P have very limited contribution to the interaction of PH_3 with N_3H according to the NBO analysis.

In another ADMP simulation of **cisTS** with the same conditions (time step 0.2 fs and 6.3 kcal/mol initial kinetic energy) for 2 ps, PH_3 attacks N_α and forms PN_α bond after 12 fs, thus rendering the formation of the **cis** intermediate. The system, trapped in the **cis** potential well, fluctuates around the structure of **cis** after reaching **cis** during the 2 ps ADMP simulations.

The same conditions are applied to ADMP simulations of **cis**, with 6.3 kcal/mol of initial kinetic energy. Under this condition, the system is trapped in the **cis** potential well for 4 ps; this indicates that the system does not have enough energy in the reaction coordinates to overcome the reaction barriers to return to reactants or to tautomerize to **trans** or isomerizes to **int**. Increasing the initial kinetic energy to 25.2 kcal/mol (0.04 eV/mol) in an ADMP simulation of 2 ps starting from **cis** drives the system out of the **cis** potential well. The changes of potential energy of the system along the trajectory are shown in Figure 3. The system tries to open the $PN_\alpha N_\beta N_\gamma$ four-membered ring to reach a local maximum **A** at 54 fs (as shown in Figure 3) with $A_{PN\alpha N\beta} = 145°$, $A_{N\alpha N\beta N\gamma} = 126°$, and $R_{PN\gamma} = 3.4$ Å and then returns to **cis** and moves out of the **cis** potential well to reach the structure at point **B** similar to **TS2** at 136 fs. The system then repeats the $PN_\alpha N_\beta N_\gamma$ ring-opening and ring-closing motions till 300 fs. There is a small potential energy plateau after 300 fs of ADMP simulation, where PH_3 rotates about the PN_α bond from 300 to 370 fs resembling **cis** → **TS2** motion (**D** → **E**). After 370 fs, the system returns to **cis** and internally rotates about the $N_\alpha N_\beta$ bond to reach 35° for the dihedral angle $D_{PN\alpha N\beta N\gamma}$ at point **F**. Along with this internal rotation, the system stretches PN_α, $N_\alpha N_\beta$, and $N_\beta N_\gamma$ bonds

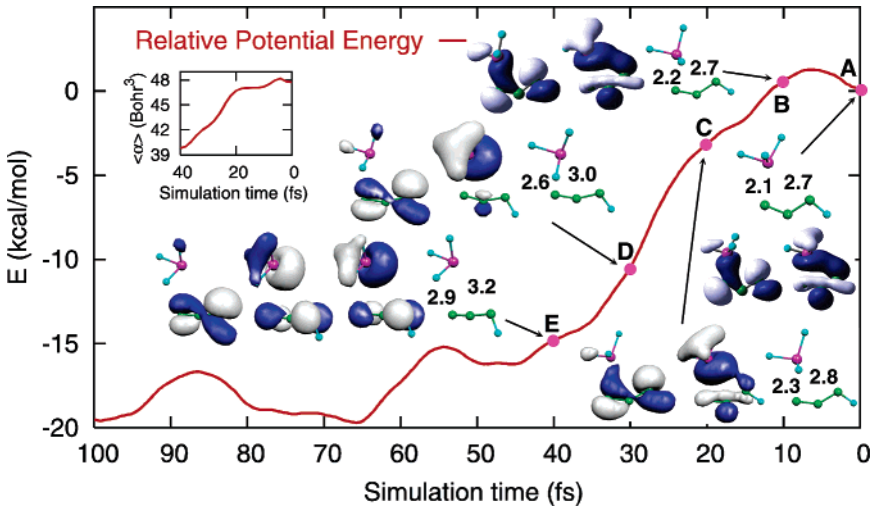


Figure 2. The relative potential energy (in kcal/mol) of the system to **cisTS** during the first 100 fs of a total 2 ps ADMP simulation with 0.2 fs time step and 6.3 kcal/mol initial kinetic energy. The pair of left and right numbers in the structures are the bond distances of P to N_α and N_γ, respectively. Simulation times for the selected points are **A** at 0 fs, **B** at 10 fs, **C** at 20 fs, **D** at 30 fs, and **E** at 40 fs. Only those molecular orbitals relevant to the interaction between PH_3 and N_3H are shown (from left to the right): HOMO-2, HOMO-1, and HOMO for point **E**, HOMO-2 and HOMO for points **D** and **C**, and HOMO-2 and HOMO-1 for points **B** and **A**. HOMO-n is the *n*th orbital below the HOMO. The molecular orbitals are drawn with isovalue 0.04 $Å^{-3}$. More molecular orbitals for points **A**, **B**, **C**, and **D** are shown in Figure 2s in the Supporting Information.

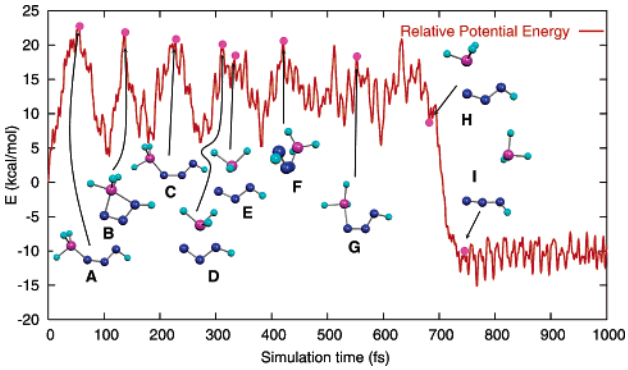


Figure 3. The relative potential energy (in kcal/mol) of the system to **cis** during the first 1 ps of a total 2 ps ADMP simulation, starting from **cis** with 0.2 fs time step and 25.2 kcal/mol initial kinetic energy. Simulation times for the selected points are **A** at 56 fs, **B** at 136 fs, **C** at 223 fs, **D** at 309 fs, **E** at 330 fs, **F** at 416 fs, **G** at 564 fs, **H** at 679 fs, and **I** at 743 fs.

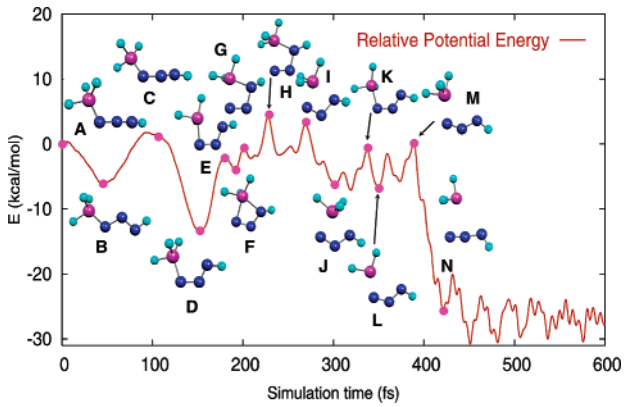


Figure 4. The relative potential energy (in kcal/mol) of the system to **TS1** during the first 600 fs of a total 2 ps ADMP simulation, starting from **TS1** with 0.2 fs time step and 6.3 kcal/mol initial kinetic energy. Simulation times for the selected points are **A** at 0 fs, **B** at 48 fs, **C** at 105 fs, **D** at 153 fs, **E** at 179 fs, **F** at 190 fs, **G** at 201 fs, **H** at 228 fs, **I** at 271 fs, **J** at 301 fs, **K** at 334 fs, **L** at 349 fs, **M** at 379 fs, and **N** at 421 fs.

from 410 to 600 fs to prepare for the right energy and momentum distributions for PH_3 and N_3H dissociation as represented by point **G**. After 640 fs, the system begins to dissociate back to PH_3 and N_3H, as indicated by the drastic drop of potential energy starting from point **H**. At 740 fs, the system completely dissociates to PH_3 and N_3H with $R_{PN\alpha} = 4.2$ Å and $R_{PN\gamma} = 3.3$ Å, as indicated by point **I**.

C. TS1 of $PH_3 + N_3H$. If the ADMP simulations start from **TS1** (**A**) (as shown in Figure 4), the system comes to **trans** (**B**) and then goes to **cis** (**D**) by overcoming **TS1** (**C**). The reaction barrier from **trans** to **cis** is 7.5 kcal/mol, and the reverse barrier is 11.9 kcal/mol from **cis** to **trans**, which are in good agreement with our previous QM calculations.[6] The deviation of these barriers from those of the QM studies is expected, since the MD simulation usually does not go through minimum energy path and does not include the ZPVE correction. By overcoming an 11.1 kcal/mol reaction barrier at **TS2** (**E**), the system reaches **int** (**F**) and then comes to a potential energy plateau involving the migration of PH_3 from N_α to N_γ and backward migration from N_γ to N_α as indicated by points **E**, **F**, **G**, **H**, **I**, **J**, and **K** till 380 fs. The most noticeable geometric change during the PH_3 migration is the rotation of PH_3, which is clearly manifested by the relative positions of the three H atoms on P of PH_3. This rotation is the driving force for the PH_3 migration, breaking the PN_α bond and forming the PN_γ bond. After 380 fs, the system uses 50 fs to dissociate back to PH_3 and N_3H. The tautomerization from **cis** to **trans** and isomerization from **cis** to **int** are competitive processes according to this ADMP simulation, and the isomerization from **cis** to **int** is favored energetically.

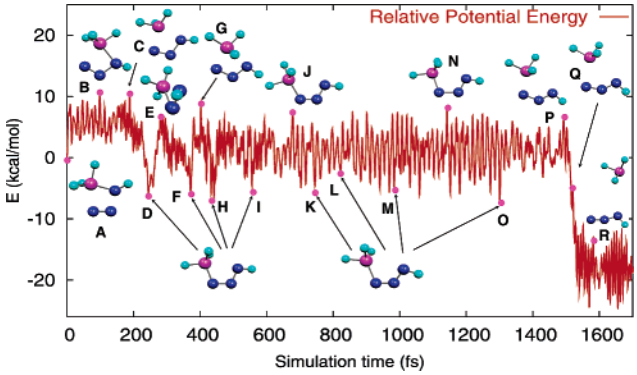


Figure 5. The relative potential energy (in kcal/mol) of the system to **TS3** during the first 1.6 ps of a total 2 ps ADMP simulation, starting from **TS3** with 0.2 fs time step and 12.6 kcal/mol initial kinetic energy. Simulation times for the selected points are **A** at 0 fs, **B** at 105 fs, **C** at 173 fs, **D** at 246 fs, **E** at 288 fs, **F** at 369 fs, **G** at 402 fs, **H** at 440 fs, **I** at 554 fs, **J** at 677 fs, **K** at 772 fs, **L** at 857 fs, **M** at 991 fs, **N** at 1140 fs, **O** at 1384 fs, **P** at 1510 fs, **Q** at 1532 fs, and **R** at 1590 fs.

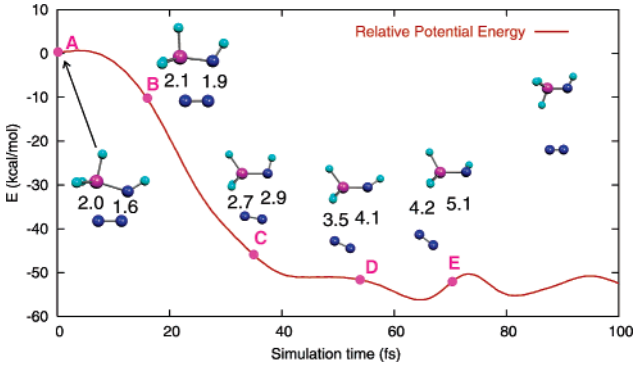


Figure 6. The relative potential energy (in kcal/mol) of the system to **TS3** during the first 100 fs of a total 400 fs ADMP simulation, starting from **TS3** with 0.2 fs time step and 6.3 kcal/mol initial kinetic energy. Simulation times for the selected points are **A** at 0 fs, **B** at 16 fs, **C** at 36 fs, **D** at 54 fs, and **E** at 70 fs. The upper-right structures are the final products, $H_3P{=}NH$ and N_2, drawn for comparison.

D. TS3 of $PH_3 + N_3H$. Previous QM calculations[6−8] have all predicted that the PES around **TS3** is very flat: the energy of **TS3** after the ZPVE correction is even lower than that of **int**[6,7] (as shown in Scheme 3). The only conclusion could be drawn here is that the energies of **TS3** and **int** are very close, and essentially there is only one reaction barrier from **cis** to the final products.[6] Starting from **TS3** with 6.3 kcal/mol initial kinetic energy and 0.2 fs time step in adiabatic ADMP simulations, the system is trapped in the **cis** potential well after it goes through **int** and gets over **TS2** within 4 ps. More initial kinetic energy might help the system overcome the reaction barrier to dissociate back to PH_3 and N_3H. A trajectory of ADMP simulation starting from **TS3** with time step 0.2 fs and 12.6 kcal/mol initial kinetic energy is shown in Figure 5. Around 250 fs, the system reaches **cis** (point **D** in Figure 5) after going through **int** (similar to point **B**) and **TS2** (similar to point **C**). The PES of this region of **TS2**, **int**, and **TS3** is very flat, consistent with previous QM predictions.[6−8] At 288 fs, the system tries to tautomerize to **TS1** by twisting dihedral angle $D_{PN\alpha N\beta N\gamma}$ to around 30° (point **E** in Figure 5) and returns to **cis** at 369 fs and switches to $PN_\alpha N_\beta N_\gamma$ ring-opening and ring-closing motions. The ring-opening and ring-closing motions companied with bond stretching take 700 fs before preparing the system to dissociate back to PH_3 and N_3H. At about 1.51 ps, the system begins to dissociate to PH_3 and N_3H by redistributing internal energy and momentum from **cisTS** (point **P**). Once the system goes over **cisTS**, the potential energy of the system drops quickly along the dissociation path. Starting from **TS3** with 0.2 fs time step and 6.3 kcal/mol initial kinetic energy, the ADMP simulation toward the N_2 dissociation is straightforward, and the system dissociates to N_2 and $H_3P{=}NH$ quickly and smoothly as shown in Figure 6.

During the N_2 dissociation, the most visible motions within the system are the internal rotation of NH about the PN_γ bond and the departure of N_α and N_β from the system. Within the first 40 fs, the major motions of the system are N_2 formation and detachment, which is in the forward direction from **TS3**.[6] N_α and N_β are not leaving at the same speed: N_β leaves faster than N_α does (as indicated by points **B** and **C** in Figure 6). After 40 fs, N_2 dissociates from $H_3P{=}NH$, and a portion of the residual energy redistributes into the internal rotation of NH about the PN bond in $H_3P{=}NH$, which is clearly indicated by the relative positions of the hydrogen atoms on P and the hydrogen atom on N (points **D** and **E** in Figure 6). In **TS3**, the hydrogen atom on N_γ is in the eclipse position to one of the hydrogen atoms on P and is in the staggered conformation to the remaining atoms on P in $H_3P{=}NH$. The dissociation energy of the system to $H_3P{=}NH$ and N_2 from **TS3** is about 55 kcal/mol, which is consistent with our previous QM studies.[6]

The ADMP simulations with 0.2 fs time step and 6.3 kcal/mol initial kinetic energies (12.6 kcal/mol for **TS3**) started from **gammaTS**, **cisTS**, **TS1**, and **TS3** reproduce the potential energy surface as predicted in our previous QM studies[6] and reveal details about the dynamics on the PES. ADMP simulations starting from **TS2** end with dissociation back to PH_3 and N_3H.

E. Thermostatic MD Simulations. In chemical reactions, the adiabatic condition might not be maintained, and most reactions are thermostatic. The thermostatic MD simulations should be more appropriate to uncover the reaction mechanism in an actual chemical environment. The final step of the Staudinger reaction only takes place at room temperature, and a complex forms at lower temperature before the whole reaction completes.[35] To reproduce the experimental condition, an ADMP simulation with 0.2 fs time step at 298 K is carried out starting from **TS3** for 2 ps. Shown in Figure 7, the trajectory of the potential energy is different from the adiabatic trajectory with 12.6 kcal/mol initial kinetic energy (as shown in Figure 5). However, the qualitative evolutions of the system in the two trajectories are similar, both isomerize to **cis** first, then undergo $PN_\alpha N_\beta N_\gamma$ ring opening followed by PH_3 shift, and finally dissociate to PH_3 and N_3H (shown by structures of the system during the simulations in Figures 5 and 7). The system dissociates back to PH_3 and N_3H after 600 fs in the thermostatic ADMP simulation, which is much faster than that in the adiabatic ADMP simulation. Under the same condition, an ADMP simulation starting from

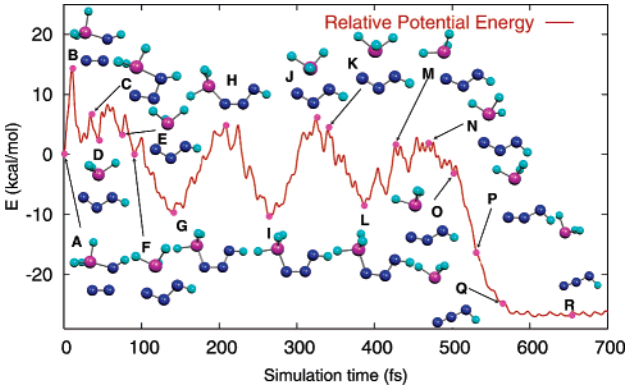


Figure 7. The reaction potential energy (in kcal/mol) of the system to **TS3** during the first 700 fs of a total 2 ps thermostatic ADMP simulation, starting from **TS3** at 298 K with 0.2 fs time step. Simulation times for the selected points are **A** at 0 fs, **B** at 11 fs, **C** at 32 fs, **D** at 43 fs, **E** at 72 fs, **F** at 88 fs, **G** at 138 fs, **H** at 207 fs, **I** at 267 fs, **J** at 323 fs, **K** at 341 fs, **L** at 385 fs, **M** at 427 fs, **N** at 467 fs, **O** at 506 fs, **P** at 530 fs, **Q** at 567 fs, and **R** at 643 fs.

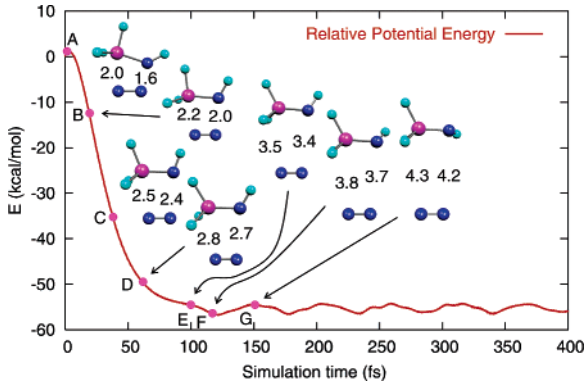


Figure 8. The relative potential energy (in kcal/mol) of the system to **TS3** during a total 400 fs thermostatic ADMP simulation, starting from **TS3** with 0.2 fs time step at 298 K. Simulation times for the selected points are **A** at 0 fs, **B** at 20 fs, **C** at 40 fs, **D** at 60 fs, **E** at 100 fs, **F** at 120 fs, and **G** at 150 fs. The pair of left and right numbers are PN_α and $N_\beta N_\gamma$ bond distances (in Å), respectively.

TS3 is carried out for 400 fs, and the system dissociates forward to $H_3P{=}NH$ and N_2 (as shown in Figure 8). Similar to the adiabatic dissociation (as shown in Figure 6), the system dissociates forward to $H_3P{=}NH$ and N_2, followed by the internal rotation of NH about the PN bond in $H_3P{=}NH$. In the thermostatic dissociation, N_α and N_β leave $H_3P{=}NH$ at the same speed. The thermostatic dissociation has a similar dissociation energy to that of the adiabatic dissociation (about 56 kcal/mol). From the above ADMP simulations, the complete reaction pathway of the Staudinger reaction is explored, and the ADMP simulations corroborate the previous QM predictions.[6−8]

To get more details on the evolution of the system during the Staudinger reaction, it is necessary to do structural analysis on the trajectory. The thermostatic ADMP simulation starting from **TS3** at 298 K is analyzed accordingly. $R_{PN\alpha}$ and $R_{PN\gamma}$ and the bond distance fluctuations of $R_{N\gamma H}$ and R_{PH} during the simulation are shown in Figure 9. The

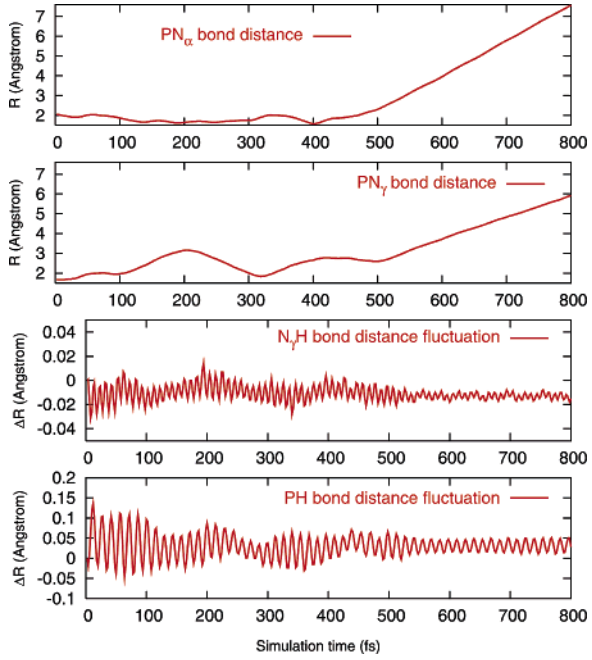


Figure 9. Bond distances $R_{PN\alpha}$ and $R_{PN\gamma}$ and bond distance changes of $R_{N\gamma H}$ and R_{PH} with respect to **TS3** during the first 800 fs of a total 2 ps thermostatic ADMP simulation, starting from **TS3** at 298 K. The hydrogen atom on P is at the same side of the hydrogen atom on N_γ in **TS3**.

hydrogen atom of R_{PH} on P is at the cis position to the hydrogen atom on N_γ, and all the structural data pertinent to hydrogen atom on P are based on this hydrogen atom. During the course of the reaction, bond distance $R_{PN\gamma}$ changes much more than $R_{PN\alpha}$ before dissociation back to PH_3 and N_3H. Initially, PH_3 approaches N_α and N_γ at the similar distance from far away—ca. 700 fs. From **TS3** to **cis**, bond distance $R_{N\gamma H}$ does not change much, while bond distance R_{PH} changes a lot (mainly stretching). It can be inferred that the PH bond stretching facilitates the PH_3 migration between N_α and N_γ. Figure 10 shows bond angle $A_{N\alpha N\beta N\gamma}$ and dihedral angles $D_{PN\alpha N\beta N\gamma}$, $D_{N\alpha N\beta N\gamma H}$, and $D_{HPN\alpha N\beta}$ of the trajectory. The value of $A_{N\alpha N\beta N\gamma}$ shows the linearity of the azide backbone during the simulation. When PH_3 and N_3H fall apart, $A_{N\alpha N\beta N\gamma}$ begins to increase: $A_{N\alpha N\beta N\gamma}$ reaches 175° at 470 fs from 120° at 475 fs. Dihedral angle $D_{PN\alpha N\beta N\gamma}$ indicates the planarity of the $PN_\alpha N_\beta N_\gamma$ four-membered ring and serves as a criterion for the system to tautomize from **cis** (close to 0°) to **trans** (close to 180°) through **TS1** (close to 90°). At 220 fs, the system tries to twist $D_{PN\alpha N\beta N\gamma}$ (point **H** in Figure 7) and returns to **cis** ca. 300 fs, which indicates that the system can tautomerize to **trans** from **cis** if provided with enough energy and proper energy and momentum distributions. The change of $D_{PN\alpha N\beta N\gamma}$ indicates that PH_3 does migrate from N_α to N_γ within the $N_\alpha N_\beta N_\gamma$ plane. Dihedral angle $D_{N\alpha N\beta N\gamma H}$ serves as an indicator for the involvement of H motion on N_γ in the Staudginer reaction: the change of this dihedral angle during the ADMP simulation indicates that the out-of-$N_\alpha N_\beta N_\gamma H$ plane motion of H takes place all along the Staudinger reaction, especially during the PH_3 migration from N_α to N_γ as indicated by point **F** (at 88 fs) and point **J** (ca. 320 fs) in Figure 7. The rotation of PH_3 group during the reaction is indicated by dihedral angle $D_{HPN\alpha N\beta}$ shown

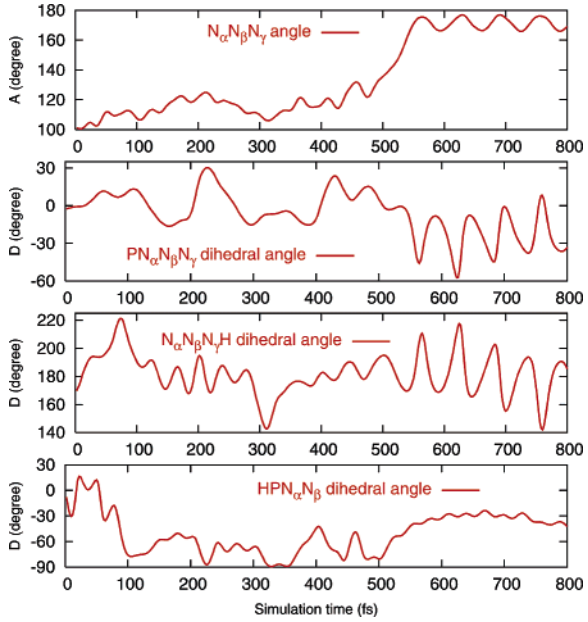


Figure 10. Bond angles $A_{N\alpha N\beta N\gamma}$ and dihedral angles $D_{PN\alpha N\beta N\gamma}$, $D_{N\alpha N\beta N\gamma H}$, and $D_{HPN\alpha N\beta}$ of the system with respect to **TS3** during the first 800 fs of a total 2 ps thermostatic ADMP simulation, starting from **TS3** at 298 K. The hydrogen atom on P is at the same side of the hydrogen atom on N_γ in **TS3**.

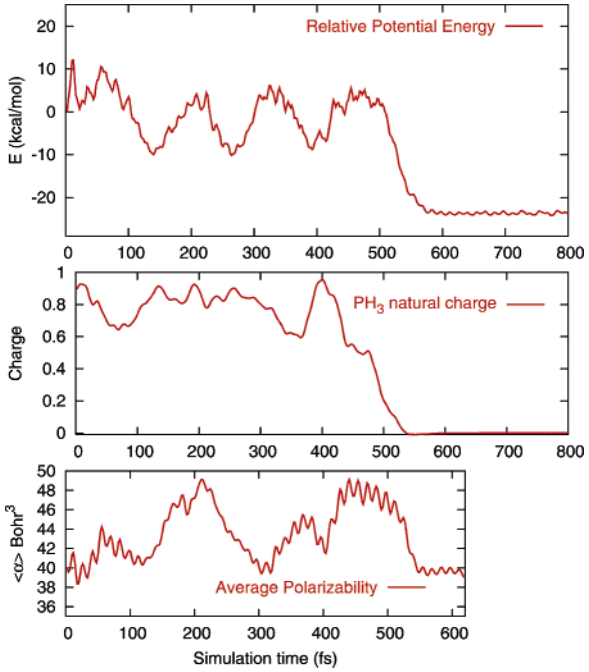


Figure 11. Relative potential energy and average polarizabily of the system and natural charge of PH_3 unit of the system, based on single-point and thermostatic ADMP calculations at B3LYP with 6-31G(d) basis set, starting from **TS3** with 0.2 fs time step at 298 K. The total simulation time is 2 ps. Only the first 800 fs is shown for relative energy of the system and the natural charge of PH_3, and only the first 620 fs is shown for the averaged polarizability. The structures of the stationary points on the potential energy surface are displayed in Figure 7.

in Figure 10. The overall change of $D_{HPN\alpha N\beta}$ is more than 90° during the isomerization from **TS3** to **cis,** which clearly indicates that PH_3 rotation serves as a driving force for the PH_3 migration between N_α and N_γ. $D_{HPN\alpha N\beta}$ changes through the entire reaction pathway. The structural analysis indicates that (1) the PH_3 leaves (or approaches) N_α and N_γ with similar speeds when the system dissociates (or forms), (2) P is not always within the $N_\alpha N_\beta N_\gamma$ plane during the Staudinger reaction, and (3) the out-of-plane motion of the hydrogen atom on N_γ and the rotation of PH_3 facilitate the migration of PH_3 between N_α and N_γ. Bulky substituent groups hinder the rotation and increase the rotation barrier from **cis** to **TS3**. This is indeed the case for large substituent groups, as predicted by our previous QM studies.[6]

ADMP simulation is an approximation to BOMD on the PES and should be parallel to that of BOMD.[26,36] We performed single-point calculations with the ADMP trajectory starting from **TS3** at the same level of theory as before.[6] The relative energies of the single-point calculations to **TS3** are plotted in Figure 11. The relative energies of the single-point calculations are very similar to those of ADMP simulations, which verify the validity of the ADMP simulations. Figure 11 also shows the natural charge of PH_3 group during the ADMP simulation. The plot of the natural charge of PH_3 indicates the charge transfer between PH_3 and N_3H subunits during the Staudinger reaction. Around the dissociation, the charge transfer between PH_3 and N_3H decreases as the two groups fall apart, and the overall charge on each group diminishes around 540 fs when PN_α and PN_γ bond distances are about 3.0 Å. The natural charge on PH_3 group around the dissociation (or attacking) region along the trajectory indicates the electrostatic attraction plays an important role for the initial Staudinger reaction. During the Staudinger reaction, the average polarizability of the system (as shown in Figure 11) changes along with the reaction course, and the region around **TS3** has relatively small polarizability because of its compact structure. The polarizability increases as the system tries to twist $D_{PN\alpha N\beta N\gamma}$ to reach **TS1**. **cis** has small polarizability. As **cis** begins to dissociate back to PH_3 and N_3H, the polarizability increases as the system overcomes the **cisTS** barrier and decreases after overcoming the barrier. The change of polarizability of the system during the reaction course indicates that the solvent effect on the system varies with the reaction course: the polar solvent will stabilize all transition states and thus facilitates the overall reaction by decreasing the reaction barriers.

Some geometric data (bond distances $R_{PN\alpha}$, $R_{N\beta N\gamma}$, R_{PH}, and $R_{N\gamma H}$ and dihedral angles $D_{HPN\alpha N\beta}$, $D_{HPN\gamma H}$, and $D_{N\alpha N\beta N\gamma H}$) are plotted in Figures 12 and 13 for the dissociation trajectory to $H_3P{=}NH$ and N_2 starting from **TS3** in the thermostatic ADMP simulation. As shown in Figure 12, the N_α and N_β leave $H_3P{=}NH$ at the same speed. Dihedral angle $D_{HPN\gamma H}$ changes 30° at the first 90 fs, indicating that the leaving of N_α and N_β is the major motion during the N_2 dissociation. Dihedral angle $D_{HPN\gamma H}$ changes from −15° to −100° after the dissociation of N_2 during the first 90 fs, which is responsible for the potential energy fluctuation of the system after N_2 dissociation. Figure 13 displays bond distances R_{PH}, $R_{N\gamma H}$, and $R_{PN\gamma}$ along the dissociation. $R_{PN\gamma}$ decreases to the equilibrium P=N bond distance (around 1.56 Å) in $H_3P{=}NH$ after 50 fs. PN_γ bond stretches with the internal rotation

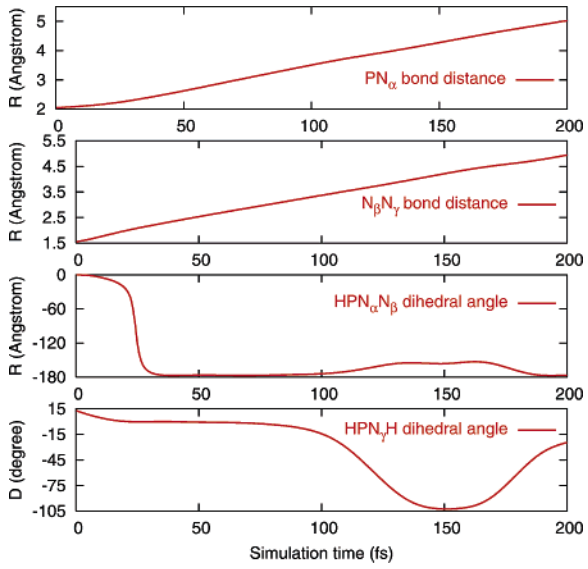


Figure 12. Bond distances $R_{PN\alpha}$ and $R_{PN\gamma}$ and dihedral angles $D_{HPN\alpha N\beta}$ and $D_{HPN\gamma H}$ with respect to **TS3** during the first 200 fs of a total 400 fs thermostatic ADMP simulation, starting from **TS3** at 298 K. The hydrogen atom on P is at the same side of the hydrogen atom on N_γ in **TS3**.

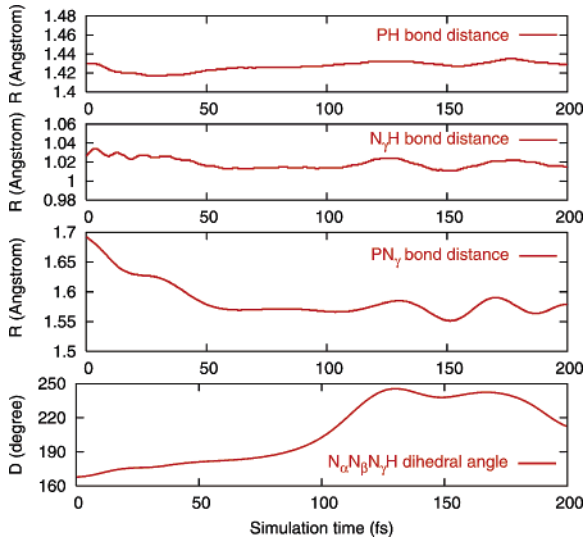


Figure 13. Bond distances R_{PH}, $R_{PN\gamma}$, and $R_{N\gamma H}$ and dihedral angle $D_{N\alpha N\beta N\gamma H}$ of the system with respect to **TS3** during the first 200 fs of a total 400 fs thermostatic ADMP simulation, starting from **TS3** at 298 K. The hydrogen atom on P is at the same side of the hydrogen atom on N_γ in **TS3**.

of $H_3P{=}NH$. From the changes of $D_{HPN\alpha N\beta}$, $D_{HPN\gamma H}$, and $D_{N\alpha N\beta N\gamma H}$, one can infer that it is $N_\gamma H$ that rotates about the PN bond in $H_3P{=}NH$ after the N_2 dissociation.

IV. Conclusions

In the present work, the Staudinger reaction of $PH_3 + N_3H$ has been simulated by ADMP molecular dynamics within DFT. The ADMP simulations starting from **cisTS**, **gamma-TS**, **cis**, **TS1**, **TS2**, and **TS3** reproduce the reaction pathways predicted by previous QM methods.[6−8] The details of the Staudinger reaction have been uncovered through the trajectories of the ADMP simulations:

(1) For the initial attack of PH_3 to N_3H, the P atom approaches N_α and N_γ with similar distances before significant interaction occurs between the two reactants. When PH_3 approaches N_3H ca. 3.0 Å, charge transfer from PH_3 to N_3H occurs, and the average polarizability of the system increases. This manifests that the strong electrostatic interaction occurs at this beginning stage and is the main driving force for the initial attack. The P atom interacts with both N_α and N_β when PH_3 approaches N_3H, as manifested by the molecular orbitals of the system during the initial reaction. The empty d orbitals of P play very limited role in this reaction.

(2) The Staudinger reaction goes through **cis**, **TS2**, **int**, and **TS3** to form phosphazene as predicted quantum mechanically before.[6−8] According to the ADMP simulations, **cis** can tautomerize to **trans** through **TS1** with proper conditions. The rotation of PH_3 and the stretching of PH bonds serve as dominant driving forces for the second phase of the Staudinger reaction from **cis** to **TS3**, according to the dihedral angle $D_{HPN\alpha N\beta}$ and the bond distance R_{PH} changes during the Staudinger reaction. Small substituent groups on P of phosphane, with faster PR_3 rotation and stronger PR stretching, will certainly facilitate the last stage (from **cis** to **TS3**) of the Staudinger reaction.

(3) The fact that the polarizability changes during the course of the reaction implies that different solvent effects are expected at different stages of the reaction. Appropriate solvent can alter the reaction course (e.g. tautormerization to **trans** from **cis**). Although the solvent effects are not taken into account in the present work explicitly, this omission should not qualitatively change the conclusion drawn here: polar solvent can facilitate the overall reaction by stabilizing all transition states and hence decreasing the reaction barriers. This understanding is based on the close correlation between the changes of the average polarizability and the potential energy of the system during the reaction. In our previous static quantum mechanical studies,[6] we compared the Staudinger reactions with different substituent groups on phosphane and azide in gas phase and in dimethyl sulfoxide and reached the same conclusion about the solvent effects for this reaction.

In summary, the Saudinger reaction of $PH_3 + N_3H$ has been studied with ab initio molecular dynamics. Our work demonstrates that the combination of quantum mechanical studies with ab initio molecular dynamics will enhance the strengths of both approaches and yield detailed mechanical and dynamical understanding of chemical reactions.

Acknowledgment. The financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada is gratefully acknowledged. We thank Professor Srinivasan S. Iyengar for a reprint of ref 26. W.Q.T. is grateful to Professor Yuriko Aoki at Kyushu University for her hospitality.

Supporting Information Available: Relevant molecular orbital diagrams of points **A**, **B**, **C**, and **D** in Figure 2 during an ADMP simulation starting from **cisTS** (Figure 2s). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Staudinger, H.; Meyer, J. *Helv. Chim. Acta* **1919**, *2*, 635.

(2) (a) Cordridge, D. E. C. *Phosphorus 2000: Chemistry, Biochemistry & Technology*; Elsevier: Amsterdam, 2000. (b) Irving, J. T. *Calcium and phosphorus metabolism*; Academic Press: New York, 1973.

(3) (a) Gololobov, Yu. G.; Zhmurova, I. N.; Kasukhin, L. F. *Tetrahedron* **1981**, *37*, 437. (b) Gololobov, Yu. G.; Zhmurova, I. N.; Kasukhin, L. F. *Tetrahedron* **1992**, *48*, 1353.

(4) Kato, H.; Ohmori K.; Suzuki, K. *Synlett* **2001**, *SI*, 1003.

(5) (a) Saxon, E.; Luchansky, S. J.; Hang, H. C.; Yu, C.; Lee, S. C.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2002**, 124, 14893. (b) Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007. (c) Kato, H.; Ohmori K.; Suzuki, K. *Synlett* **2001**, *SI*, 1003.

(6) Tian, W. Q.; Wang, Y. A. *J. Org. Chem.* **2004**, *69*, 4299.

(7) Widauer, C.; Grützmacher, H.; Shevchenko I.; Gramlich, V. *Eur. J. Inorg. Chem.* **1999**, 1659.

(8) Alajarin, M.; Conesa, C.; Rzepa, H. S. *J. Chem. Soc., Perkin Trans.* **1999**, *2*, 1811.

(9) Goerlich, R.; Farkens, M.; Fischer, A.; Jones, P. G.; Schmutzler, R. *Z. Anorg. Allg. Chem.* **1994**, *620*, 707.

(10) Molina, P.; López-Leonardo, C.; Llamas-Botía, J.; Foces-Foces, C.; Fernández-Castaño, C. *Tetrahedron* **1996**, *52*, 9629.

(11) Alajarín, M.; Molina, P.; López-Leonardo, C. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 67.

(12) Allen, L. C. *Int. J. Quantum Chem.* **1994**, *49*, 253.

(13) (a) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363. (b) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154. (c) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523.

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

(15) Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, T. F. K.; Mattos, C.; Michnick, S.; Nago, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(16) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A. III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10046.

(17) *Modern methods for multidimensional dynamics computation in chemistry*; Thompson, D. L., Ed.; World Scientific: Singapore, 1998.

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

(19) Marx, D.; Hutter, J. In *Modern methods and algorithms of quantum chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, 2000; Vol. 1, p 301.

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

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

(22) Lippert, G.; Hutter, J.; Parrinello, M. *Mol. Phys.* **1997**, *92*, 477.

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

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

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

(26) Iyengar, S. S.; Schlegel, H. B.; Voth, G. A.; Millam, J. M.; Scuseria, G. E.; Frisch, M. J. *Isr. J. Chem.* **2002**, *42*, 191.

(27) Gibson, D. A.; Ionova, I. V.; Carter, E. A. *Chem. Phys. Lett.* **1995**, *240*, 261.

(28) (a) Tuckerman, M. E. *J. Phys.: Condens. Matter* **2002**, *12*, R1297. (b) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511. (c) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695. (d) Martyna, G. J.; Klein, M. L.; Tuckerman, M. E. *J. Chem. Phys.* **1992**, *97*, 2635. (e) Tuckerman, M. E.; Parrinello, M. *J. Chem. Phys.* **1994**, *101*, 1302.

(29) Rega, N.; Iyengar, S. S.; Voth, G. A.; Schlegel, H. B.; Vreven, T.; Frisch, M. J. *J. Phys. Chem. B* **2004**, *108*, 4210.

(30) Gaussian 03, Revision B.05, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Cossi; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Pittsburgh, PA, 2003.

(31) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(32) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785. (b) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, 157, 200.

(33) (a) Sheng, Y.; Musaev, D. G.; Reddy, K. S.; McDonald, F. E.; Morokuma, K. *J. Am. Chem. Soc.* **2002**, *124*, 4149. (b) Vereecken, L.; Peeters, J.; Bettinger, H. F.; Kaiser, R. I.; Schleyer, P. v. R.; Schaefer, H. F., III *J. Am. Chem. Soc.* **2002**, 124, 2781. (c) Jiao, H.; Frapper, G.; Halet, J.-F.; Saillard, J.-Y. *J. Phys. Chem. A* **2001**, *105*, 5945. (d) Beno, B. R.; Wilsey, S.; Houk, K. N. *J. Am. Chem. Soc.* **1999**, *121*, 4816. (e) Batra, R.; Giese, B.; Spichty, M.; Gescheidt, G.; Houk, K. N. *J. Phys. Chem.* **1996**, *100*, 18371. (f) Jursic, B.; Zdravkovski, Z. *J. Chem. Soc., Perkin Trans. 2* **1995**, 1223.

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

(35) Leffler, J. E.; Temple, R. D. *J. Am. Chem. Soc.* **1967**, *89*, 5235.

(36) The single-point calculations based on the ADMP trajectory are certainly different from the BOMD simulations, since the BOMD calculates the force from the converged charge density and potential.

CT049918X

JCTC Journal of Chemical Theory and Computation

NMR Relaxation and Internal Dynamics of Ubiquitin from a 0.2 μs MD Simulation

Aart J. Nederveen and Alexandre M. J. J. Bonvin*

Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands

Received November 19, 2004

Abstract: A 0.2 μs molecular dynamics simulation of ubiquitin in water is presented, which allows us to assess both the global tumbling in solution and the internal dynamics. The latter reveals slow motions outside the classical NMR timewindow, in agreement with recent RDC and cross-correlation measurements. Analysis of back-calculated relaxation rates using the classical NMR model-free approach reproduces the amplitudes of internal motions expressed in the order parameter, while it severely underestimates the corresponding time scales present in the simulation.

I. Introduction

Motion in the picosecond and nanosecond time scale can be monitored in Nuclear Magnetic Resonance (NMR) relaxation studies.[1] Important characteristics of motions in these time scales can be studied with the help of molecular dynamics (MD) simulations. In the past decade significant progress has been made in the assessment of biomolecular NMR data with MD simulations (for a review see ref 2). In many cases, the amplitudes of motions found in MD simulation are in good agreement with the corresponding values derived from experimental data although, in general, the simulated values tend to be somewhat lower. For globular proteins, MD simulations are typically restricted to simulation time lengths of tens of nanoseconds. As a result the global tumbling cannot be sampled and needs to be artificially included when back-calculating NMR parameters. In addition, a proper description of internal motions in the nanosecond to microsecond time scale which are of biological relevance, especially in the context of folding, recognition, and enzymatic function, requires much longer simulation times.

Thanks to the continuously increasing computing power, MD simulations in the microsecond time scale are now within our reach. Duan and Kollman[3] presented a 1 μs MD simulation of a 36-residue peptide (HP-36) starting from an unfolded structure, showing that, within that time period, a marginally stable folded state could be reached. In the context of NMR, simulations in the microsecond time scale have been reported in the literature for small peptides. Feenstra et al.[4] performed several simulations of a nonapeptide in water up to 0.7 μs in an attempt to simulate NMR NOE/ROE intensities. Peter et al.[5] presented two 0.2 μs simulations of a heptapetide of β-amino acids in methanol from which they simulated ROESY spectra by calculating the exact spectral densities for the interproton vectors. In that study the separability of internal and global motions could be confirmed for most of the residues, indicating that the global tumbling of the molecule and the internal motions are uncorrelated. Although this decorrelation assumption underlies many of the standard methods for analyzing NMR relaxation data, its validity for larger proteins has recently been questioned.[2,6] In addition, it has been suggested that time scales in standard NMR relaxation studies are significantly underestimated.[7] Long MD simulations should allow for addressing this question.

In this paper we present the results of two long MD simulations (0.1 and 0.2 μs) of ubiquitin, a well-studied 76 amino acid globular protein. Human ubiquitin is one of the most conserved eukaryotic proteins. Its primary role is in intracellular, ATP-dependent protein degradation.[8] Several sets of NMR data, describing the dynamics of ubiquitin, are available in the literature, including $^{15}N^{1}H$ R_1, R_2, and NOE ratios,[9−11] order parameters derived from residual dipolar couplings (RDCs),[12,13] and cross-correlated relaxation data.[14−16]

* Corresponding author phone: +31.(0)30.2533859; fax: +31.(0)30.2537623; e-mail: a.m.j.j.bonvin@chem.uu.nl. Corresponding author address: Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

10.1021/ct0498829 CCC: $30.25
Published on Web 03/31/2005

The RDC and cross-correlation data provide evidence for slow motions in the nanosecond to microsecond time scale both in loops and secondary structure elements. Such slow motions are typically not monitored in standard NMR-relaxation analysis.

Our 0.2 μs MD simulation allows us to estimate the global tumbling of ubiquitin in solution and study its internal dynamics, revealing slow motions outside the classical NMR timewindow in agreement with the recent RDC and cross-correlation measurements. The R_1, R_2, and NOE relaxation rates are back-calculated from the spectral densities obtained after Fourier transformation of the internal correlation functions. These rates are then used as input for the classical NMR model-free analysis to check if this commonly used analysis can properly extract the amplitudes and time scales of the internal motions present in our simulation.

II. Materials and Methods

A. Simulations. Two independent trajectories are presented in this paper: a trajectory starting from the refined NMR structure of 1D3Z[17] taken from the DRESS database[18] (NMR trajectory) and a trajectory starting from the first monomer of the dimeric X-ray structure of 1AAR[19] (X-ray trajectory, see Figure 1). The latter trajectory shall be extensively investigated throughout this paper, whereas the former will be used for validation. Simulations and analysis were performed using the GROMACS package,[20] version 3.1.3, with the GROMOS96 43a1 force field.[21]

The solute was placed in a cubic box, with a minimum solute-box distance of 1.4 nm and solvated with SPC waters.[22] No counterions were used since the net charge of the system was zero (neutral pH). The entire system consists of 760 solute atoms and 10 257 and 11 393 water molecules for the X-ray and NMR systems, respectively. A short energy minimization was performed on the system with positional restraints on the protein. Then a MD equilibration stage was carried out consisting of five successive 20 ps runs with decreasing positional restraints force constants on the solute (K_{posre} = 1000, 1000, 100, 10, and 0 kJ mol^{-1} nm^{-2}). After that the production runs were started. For the X-ray trajectory 218 ns were simulated, whereas the total simulation time for the NMR trajectory amounted to 100 ns.

For the integration of the equations of motion a 2 fs time step was used. Positions and velocities were stored every 2 ps. In each simulation the temperature was maintained at 300 K by weakly coupling solute and solvent separately to an external bath using the Berendsen thermostat[23] with a relaxation time of 0.1 ps, whereas the pressure was maintained by weakly coupling the system to an external pressure bath[23] at 1 atm with a coupling constant of $\tau_p = 0.5$ ps. The neighbor list was updated every 10 integration steps. The twin-range method was applied for dealing with long-range interactions, the short-range cutoff being 1.0 nm and the long-range cutoff 1.4 nm, both for electrostatic and van der Waals interactions. Electrostatic interactions beyond the cutoff were treated with a generalized reaction field[24] using a dielectric constant of 54.0.

B. NMR Relaxation Theory. The NMR relaxation theory has been extensively described elsewhere.[25,26] In this paper we focus on the relaxation of the amide ^{15}N nuclear spin through dipolar interaction with the attached ^{1}H spin and ^{15}N Chemical Shift Anistropy (CSA). The relaxation in this case is governed by the correlation function of the ^{15}N-^{1}H spin-interaction vector $\vec{r}(t)$

$$C(t) = \frac{4\pi}{5}\sum_{m=-2}^{m=2}\left\langle \frac{Y_{2m}(\theta_\tau,\phi_\tau)Y_{2m}(\theta_{\tau+t},\phi_{\tau+t})}{r^3(\tau)r^3(\tau+t)}\right\rangle_\tau \quad (1)$$

where Y_{2m} are the second rank spherical harmonics, $r(t)$ is the interaction vector, and (θ_τ, ϕ_τ) is the polar angles of the interaction vector at time τ in the laboratory frame of reference. Using the addition theorem for spherical harmonics, this expression can be rewritten as

$$C(t) = \left\langle \frac{P_2(\cos\chi_{\tau,\tau+t})}{r^3(\tau)r^3(\tau+t)}\right\rangle_\tau \quad (2)$$

where $P_2(x) = {}^3/_2x^2 - {}^1/_2$ is the second-order Legendre polynomial and $\chi_{\tau,\tau+t}$ is the angle between the interaction vectors at times τ and $\tau + t$. Equation 2 is used in this paper for the calculation of the global correlation function for each backbone amide-proton pair from the MD trajectory.

The spectral density function $J(\omega)$ can be obtained by Fourier transformation of the correlation function. According to the relaxation theory of Bloch, Wangness, and Redfield[26] this function is probed at specific frequencies in NMR ^{15}N relaxation experiments for R_1, R_2, and NOE

$$R_1 = \frac{d^2}{4}[3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)] + \frac{c^2}{3}J(\omega_N) \quad (3)$$

$$R_2 = \frac{d^2}{8}[4J(0) + 3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{18}[4J(0) + 3J(\omega_N)] \quad (4)$$

$$\mathrm{NOE} = 1 + \frac{d^2}{4R_1}\frac{\gamma_H}{\gamma_N}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \quad (5)$$

where $d = \mu_0 h\gamma_H\gamma_N r_{NH}^{-3}/8\pi^2$, $c = \Delta\sigma\omega_N$, μ_0 is the permeability of free space, h is Planck's constant, γ_N and γ_H are the gyromagnetic ratios of ^{15}N and ^{1}H nuclei, $\Delta\sigma$ is the CSA of ^{15}N, r_{NH} is the N−H distance, and ω_N and ω_H are the Larmor frequencies of the ^{15}N and ^{1}H nuclei.

From NMR relaxation experiments the values of R_1, R_2, and NOE for nonproline residues can be determined. Two methods are in use for extracting the dynamic information from these rates. Spectral density mapping[29] can be used to directly determine the values of the spectral density functions that are probed by NMR. In this paper reduced spectral density mapping[30] is used for calculating the spectral density values from experimental relaxation data. This approach assumes that $J(\omega_H + \omega_N) = J(\omega_H) = J(\omega_H - \omega_N)$, which is valid if the spectral density function is flat at high frequencies.

The most popular method for analyzing NMR relaxation data of proteins is the Lipari-Szabo model-free approach.[31,32] This approach uses no specific model for describing the

internal motion of the interaction vector. It is based on the decorrelation assumption that splits up the global correlation function in an internal correlation function and the overall correlation function of the molecule:

$$C_{\text{global}}(t) = C_{\text{int}}(t)C_{\text{overall}}(t) \quad (6)$$

In the case of isotropic tumbling $C_{\text{overall}}(t)$ can be taken as e^{-t/τ_c}, where τ_c is the global correlation time, which gives an indication of the tumbling frequency of the protein in the solvent. $C_{\text{int}}(t)$ now describes the internal motion of the NH-interaction vector within the molecular frame of the molecule studied. The model-free analysis yields an order parameter S^2 describing the magnitude of the internal motions along with an estimation of the internal time scale τ_e. Nowadays five models are commonly used within the model-free approach, each of them fitting a specific form of the spectral density function $J(\omega)$ to the experimental data, leading to various combinations of parameters:[33]

- model 1: S^2,
- model 2: S^2,τ_e,
- model 3: S^2,R_{ex},
- model 4: S^2, τ_e, R_{ex}, and
- model 5: S^2_s, S^2_f, τ_s.

R_{ex} indicates conformational exchange on the millisecond and microsecond time scale. Model 5 delivers information on internal motions taking place at slow and fast time scales. Selection of the model that fits best the available experimental data is commonly based on χ^2-statistics and F-tests. In this paper the model-free analysis of back-calculated NMR relaxation data is carried out with the program TENSOR2.[33]

Throughout this paper the following physical constants were used: $\Delta\sigma = -170$ ppm and $r_{NH} = 0.102$ nm. We made use of three different experimental relaxation data sets of ubiquitin measured at 600 MHz, that were reported by Lienin et al.,[9] Fushman et al.,[11] and Lee et al.[10] From these data sets one averaged set was generated consisting of R_1, R_2, and NOE for 62 residues (see Supporting Information). The standard deviations calculated from the 3 different sets are in the same range as the reported experimental errors. This average set was used as input data for the reduced spectral density mapping and for the comparison with the relaxation data from the MD simulation.

C. Analysis of Internal Dynamics. The internal dynamics of ubiquitin is assessed by computing the internal correlation function of the NH-vectors from the MD simulation with eq 2, after superposition of all structures onto the backbone heavy atoms of the secondary structure elements of the starting structure. These comprise the following segments (as defined with the program DSSP[34]): 5 β-sheets (2−7, 12−16, 41−45, 48−49, 66−71) and 1 α-helix (23−34). When the internal correlation function converges, the value of the plateau, $C_{\text{int}}(\infty)$ can be estimated as[35]

$$C_{\text{int}}(\infty) = \frac{4\pi}{5}\sum_{m=-2}^{m=2}|\langle Y_{2m}(\theta,\phi)\rangle|^2 = \frac{1}{2}(3\sum_{i=1}^{3}\sum_{j=1}^{3}\langle\mu_i\mu_j\rangle^2 - 1) \quad (7)$$

where μ_i are the Cartesian coordinates of the normalized internuclear NH-vector in the molecular frame. This expression yields the well-known order parameter S^2 that is frequently used for estimating the magnitude of the internal dynamics in a protein.

For determining if the correlation function of a specific residue has converged we use the following criterion: $|C_{\text{int}}(\infty) - C_{\text{tail}}| < 0.005$ where C_{tail} is the mean value of the tail of the internal correlation function. For the 100 ns MD this tail was taken from 45 to 50 ns, whereas for the 218 ns from 90 to 100 ns[7].

If the internal correlation function converges, the effective internal correlation time τ_{int} for these motions can be determined by integrating the correlation function

$$\tau_{\text{int}} = \frac{1}{C_{\text{int}}(t_{\text{start}}) - C_{\text{int}}(\infty)}\int_{t_{\text{start}}}^{T}(C_{\text{int}}(t) - C_{\text{int}}(\infty))dt \quad (8)$$

where T is the timepoint where the function reaches the plateau value and $t_{\text{start}} = 2$ ps. Autocorrelation functions calculated from MD simulations tend to have a sharp initial drop, that is caused by librational motion of the bonds.[36,37] In calculating τ_{int} we corrected for this initial drop by starting the integration at the first timepoint, thus neglecting the ultrafast motions in the estimation of time scales of the internal motions.

D. Computer Cluster. All calculations were carried out on the TERAS supercomputer, a 1024 CPU system (SGI Origin 3800), at the national computer center SARA (http://www.sara.nl) in Amsterdam, The Netherlands. We used 16 processors in parallel. Per day, approximately 3.5 ns could be simulated if efficient use was made of the available CPU time. Effectively the total cost for both trajectories amounted to approximately 5 months of simulation.

III. Results and Discussion

A. Validation of the Trajectories. Both MD trajectories, the 100 ns one starting from the refined NMR structure (1D3Z[17]), denoted in the following as NMR, and the 218 ns one starting from the crystal structure (1AAR[19]), denoted in the following as X-ray, were validated in terms of secondary structure evolution, RMSD, energetics, and agreement with the experimental NOE data. Both trajectories equilibrated within 1−2 ns. The RMSD relative to the starting structure is presented in Figure 2: both trajectories are stable with the average RMSDs around 0.1 nm for the backbone heavy atoms belonging to secondary structure elements (see also Table 1). Note that the RMSD calculated for residues 2−71 is 0.1 nm lower on average than for the entire backbone (Table 2). This is due to the high flexibility of the C-terminal tail. An increase in the backbone RMSD can be observed in the X-ray simulation between 130 and 185 ns; this increase is due to a change in turn conformation at position 38−41 (also evident in Figure 3). After this event the structure recovers its initial conformation and remains stable until the end of the simulation. The two trajectories sample similar conformational space: the RMSD of one trajectory relative to the starting structure of the other trajectory does not indicate major differences (see Table 1). This is confirmed by an analysis of the pairwise RMSD matrix of the concatenated trajectories: no separate blocks that would indicate conformational differences between the trajectories can be distinguished (see Supporting Information). The RMS

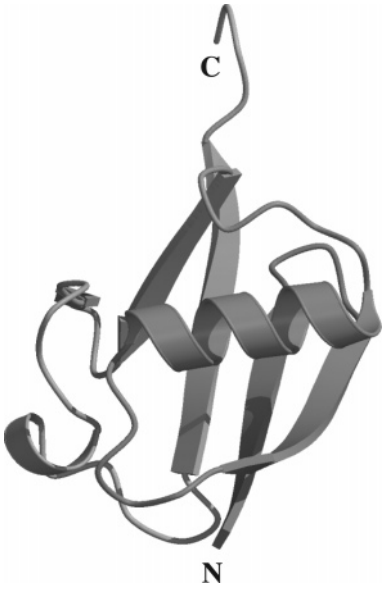


Figure 1. Cartoon representation of ubiquitin (PDB entry 1AAR[19]) showing its secondary structure elements. This figure was generated with Molscript[27] and Raster3D.[28]

fluctuations of the C^{α} atoms exhibit a similar profile for both trajectories (see Figure 2c), except for residues 62−64 that show larger fluctuations in the NMR trajectory and residues 37−39 that have larger fluctuations in the X-ray trajectory.

The time evolution of the secondary structure elements is depicted in Figure 3: in both trajectories the secondary structure is well preserved. In the X-ray trajectory the fourth β-sheet shows some instability in the 155 to 165 ns time interval. The same is observed for the onset of the third β-sheet from 100 ns on. In the time window from 130 to 185 ns there is a change in turn conformation at position 38−41.

Analysis of the energetics of both trajectories does not reveal any major difference (see Table 1).

We further validated our simulation against experimental NMR data by calculating NOE distances using an r^{-6} averaging. We used for this the set of 985 unambiguous distance restraints that are available from the Protein Data Bank[38] for entry 1D3Z.[17] Table 2 summarizes the results of this analysis for both the starting structures and the MD trajectories (see also Supporting Information for NOE violation matrices). In the NMR starting structure (refined 1D3Z from the DRESS database[18]) no violations larger than 0.05 nm are present, and in the NMR-trajectory only 4 restraints are violated by more than 0.05 nm (max. violation 0.16 nm). In the crystal structure (1AAR) 18 restraints are violated by more than 0.05 nm (max. violation 0.21 nm), after adding the protons to the original file, while 13 violations above 0.05 nm are found in the X-ray trajectory (max. violation 0.19 nm). None of the latter are present in the set of 18 violations for the starting structure. Two medium distance restraints, both connecting the helix to the subsequent loop are violated in both trajectories ($31H^{\alpha}$-35H and $30(H^{\gamma 2})_3$-$36H^{\beta}$) by 0.12 nm on average. Considering the very small number of violations, it can be concluded that both trajectories are in excellent agreement with the experimental NOE data.

B. Global Tumbling of the Molecule and Global Correlation Functions. The global tumbling time of the molecule in the solvent is determined by fitting all structures from the trajectory to the backbone heavy atoms of a reference structure, excluding the three highly mobile C-terminal residues. This reference structure was oriented such that its principal axes correspond to the *x*,*y*,*z* axes by calculating the inertia tensor using all heavy atoms excluding the three highly mobile C-terminal residues. Fitting then delivers a rotation matrix for each snapshot from the MD trajectory. The row vectors of the rotation matrix are the rotated unit vectors given in the coordinate system of the reference structure. The correlation functions of these row vectors are plotted in Figure 4a. The global tumbling appears to be much smaller than the experimental value of 4.03 ns.[9] Since the correlation functions are not exponential and not yet fully converged, it is difficult to exactly determine the correlation time for the global tumbling of the molecule from these plots. Integration of the curves yields an average value of 740 ps for the global correlation time. In addition, from this plot it can be concluded that there is a small amount of anisotropy in the global tumbling, since the correlation functions for the different direction are not exactly identical. When taking the estimates for the correlation times τ_1, τ_2, and τ_3 from Figure 4a and using Woessner equations for the

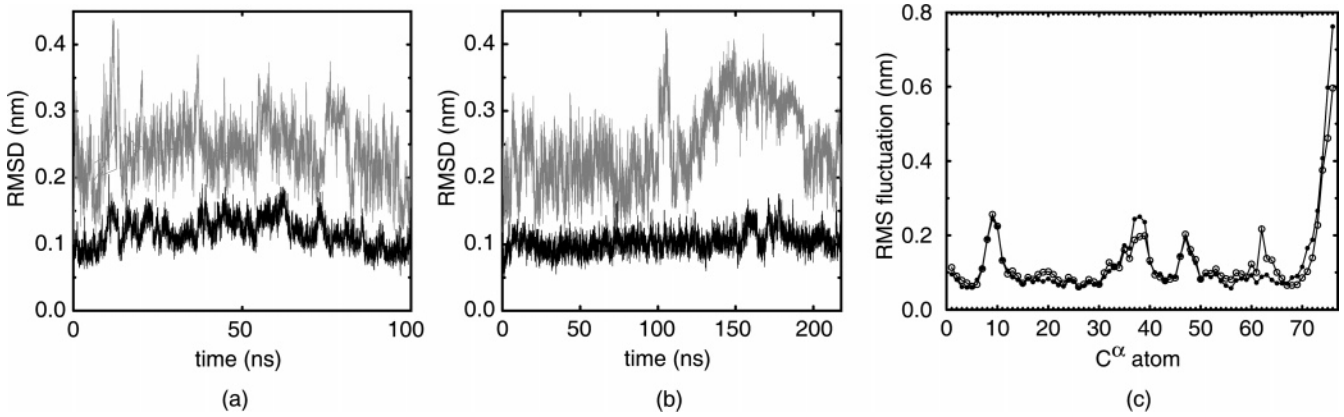


Figure 2. Time evolution of the RMSD relative to the starting structure for the heavy atoms of the entire backbone (grey) and backbone heavy atoms within secondary structure (black): (a) NMR trajectory, (b) X-ray trajectory, and (c) RMS fluctuation of C^{α} positions along the sequence for the NMR trajectory (○) and X-ray trajectory (●).

Table 1. Statistics of the NMR and X-ray MD Trajectories

	NMR	X-ray
heavy atom RMSDs (nm)[a]		
backbone	0.24 (0.05)	0.25 (0.06)
backbone residues 2−71	0.15 (0.02)	0.15 (0.02)
secondary structure	0.12 (0.02)	0.10 (0.02
secondary structure & relative to other[b]	0.12 (0.02)	0.11 (0.03)
Lennard-Jones (van der Waals) energy (kJ/mol)[c]		
ubiquitin internal	−2296 (44)	−2311 (47)
ubiquitin-solvent	−323 (74)	−324 (74)
Coulomb's electrostatic energy (kJ/mol)		
ubiquitin internal	−8966 (237)	−9238 (245)
ubiquitin-solvent	−10418 (486)	−9943 (509)

[a] Positional RMSD values are calculated relative to the starting structure of the trajectories, after superposition onto the secondary structure elements (see Materials and Methods section). Standard deviations are indicated in parentheses. [b] Average RMSDs of the NMR trajectory from the crystal structure and vice versa. [c] The nonbonded energies were calculated with the GROMOS96 43a1 force field using a twin range cutoff of 1.0 and 1.4 nm (see Materials and Methods section). The energies are the sum of short-range and long-range terms.

Table 2. NOE Violation Analysis for 1D3Z, 1AAR, NMR, and X-ray MD Trajectory

	1D3Z[a]	1AAR[b]	NMR trajectory[c]	X-ray trajectory
# violations[d] > 0.05 nm	04	18	4	13
# violations > 0.1 nm	0	7	1	6
maximum violation (nm)	0.02	0.21	0.16	0.19
RMS violations (nm)	0.002	0.017	0.008	0.014

[a] Refined NMR structure of 1D3Z taken from the DRESS database.[18] [b] The crystal structure corresponds to the first unit of the K48-linked ubiquitin dimer. [c] NOE distances in the trajectory are calculated using a r^{-6} averaging. [d] The violation analysis is based on a set of 985 unambigious NOE restraints obtained from the deposited data for entry 1D3Z in the PDB.

principal axis values of the rotational diffusion tensor,[39] we find an experimental value for the anisotropy of the diffusion tensor $D_{\parallel}/D_{\perp}$ of 1.2. The latter value is close to the anisotropy found experimentally: $D_{\parallel}/D_{\perp} = 1.17$.[40]

The discrepancy between the experimental and simulated tumbling times of the molecule in solution can be attributed in part to the SPC water model that is used in our simulation. It is well-known that, while the SPC water model displays good thermodynamic properties, its kinetic properties differ from the experimental values by a factor of 2 to 2.5: its diffusion constant is 5.28×10^{-9} m^2/s against 2.2×10^{-9} m^2/s experimentally and its viscosity is 0.58×10^{-3} Pa·s against 0.85×10^{-3} Pa·s experimentally at 300 K.[41] The fast motions of the water apparently influences the global tumbling of the molecule in the present simulation and possibly also the internal dynamics of exposed regions as will be discussed below. MD simulations using another water model, for example SPC/E, which has better kinetic properties, might lead to better estimates of the overall tumbling. Initials results from a short (10 ns) simulation in SPC/E indeed indicate slower overall tumbling by approximately a factor 2. One should however be aware that this might happen to the detriment of the protein structural stability as it is difficult to find a water model with both excellent thermodynamic and kinetic properties. Note that in another MD simulation using the CHARMM force field[42] similar correlation times in the subnanosecond time range were obtained for the overall tumbling of ubiquitin based on a completely different analysis approach using isotropic reorientational eigenmode dynamics (see Table 1 in ref 43). This indicates that reproducing experimental tumbling times might not be a specific problem related to our choice of force field and analysis methods but rather a more general issue in current biomolecular simulation methodology.

The global ^{15}N-^{1}H correlation functions could be assessed using eq 2. Typical plots for two residues, ASP21 and ALA46, are presented in Figure 4b. These correlation functions converge, although oscillations at longer time scales remain. Direct Fourier transformation of the global correlation functions is now possible to obtain the spectral density function (see inset in Figure 4b). To obtain a smooth spectral density function we use a exponential window function in the time domain ($\exp(-t/a)$, with a set to 8 ns) to remove the remaining long time scale oscillations.

C. Internal Dynamics. The internal dynamics of ubiquitin is assessed by calculating the time correlation functions of the NH-vectors with eq 2, after superposition of all MD snapshots onto the backbone heavy atoms of the secondary structure elements of the starting structure. Using the criterion for the convergence described in the Materials and Methods section, it appears that 50 residues in the 100 ns MD trajectory starting from the NMR structure are not converged, whereas only 20 residues are not converged in the 218 ns trajectory starting from the X-ray structure. Therefore we limit our discussion to the results from the X-ray trajectory. The unconverged residues are as follows: LEU8, THR12, ILE13, GLY35, ILE36, ASP39, GLN40, GLN41, ARG42, LEU43, PHE45, LYS48, GLN49, HIS68, LEU71, ARG72, LEU73, ARG74, GLY75, and GLY76. For these residues the internal dynamics is apparently not yet fully sampled, suggesting the presence of slow motions. Such lack of convergence has often been accounted for by poor sampling of dihedral transitions. Inspection of the time series for the scalar product of the interaction vector $\vec{\mu}(0)\cdot\vec{\mu}(t)$ for the unconverged residues indeed shows jumps that only occur once in the entire simulation window for some of these residues. In Figure 5a four examples of internal correlation functions are presented for short, intermediate, and long time scales together with an example of an unconverged correlation function. The time series of the scalar product for GLY53, ILE44, and GLN40 are shown in insets. Figure 5b

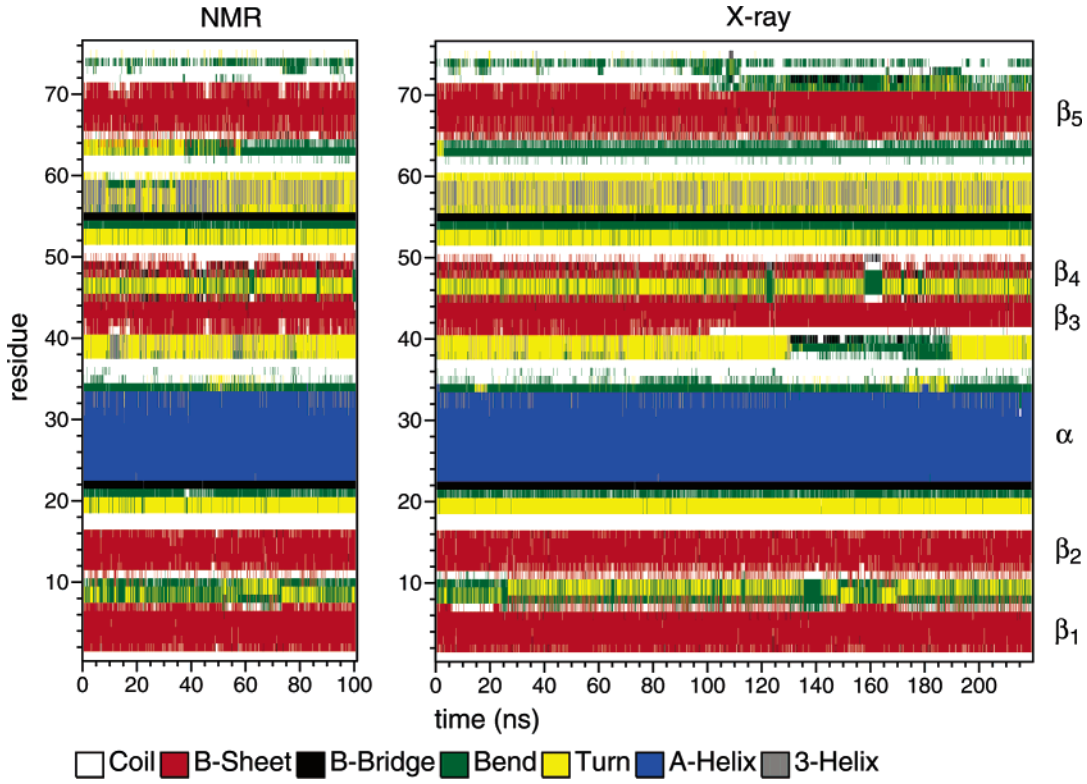


Figure 3. Secondary structure evolution of NMR and X-ray trajectories.

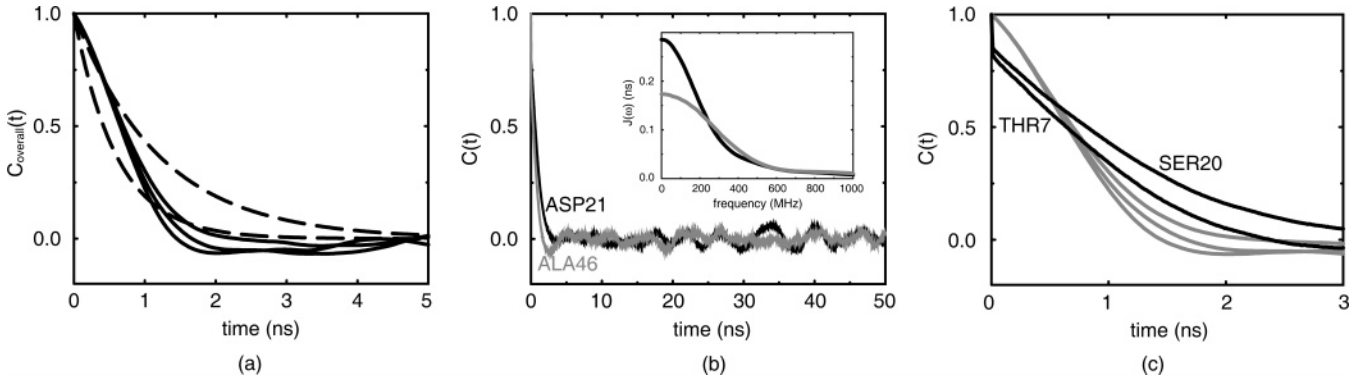


Figure 4. Assessment of global tumbling of the molecule. (a) Overall correlation functions in three directions. Exponential functions ($\exp(-t/\tau_c)$ dashed lines) with $\tau_c = 0.6$ ns and 1.2 ns are plotted for comparison. (b) Global correlation functions for residues ASP21 and ALA46 and their corresponding spectral density functions after Fourier transformation (inset). (c) Global correlation functions (black) for THR7 and SER20, together with overall correlation functions of the molecule (grey, as in panel (a)).

shows the time scales for the internal motions that were calculated using eq 8 for the 52 converging residues. Residues with time scales for the internal dynamics larger than 2 ns are found in the β-turn between the first two β-sheets, between the end of the helix and the third β-sheet and in the fifth β-sheet. Note that by our choice of t_{start} (eq 8) we are neglecting the effect of the ultrafast librational motions and are thus putting the focus on longer time scale motions.

There is no straightforward method for comparing these data with available experimental data. The available RDC order parameters[12,13] for ubiquitin indicate that motions are present beyond the classical NMR relaxation time window. It is however difficult to distinguish specific areas with slower motions. Peti et al.[12] suggested that the deviation between the RDC order parameter and the standard Lipari-Szabo order parameter is most profound in loops and β-sheets, which would partially fit our data. The cross-correlated data[14−16] are more specific than the RDC data but are not necessarily correlated with the mobility of the NH-interaction vectors. However, according to data for cross-correlated dipole−dipole interference between successive $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ vectors, the residues that featured slow local motion mainly reside in the first two β-sheets (2−7, 12, 13) and some residues between (8, 11), in the helix (26, 27, 29, 30) and in the third β-sheet (41−45).[14] Experimental data for cross-correlated chemical shift modulation showed slow motions for residues in the first (7), second (12, 14), third (41, 43), and fifth (66, 68, 69) β-sheet and in the helix (27, 28, 29, 31, 33) and for one residue between the first and second β-sheet (9).[16] Here again there is agreement with our data: the first two β-sheets together with the β-turn between

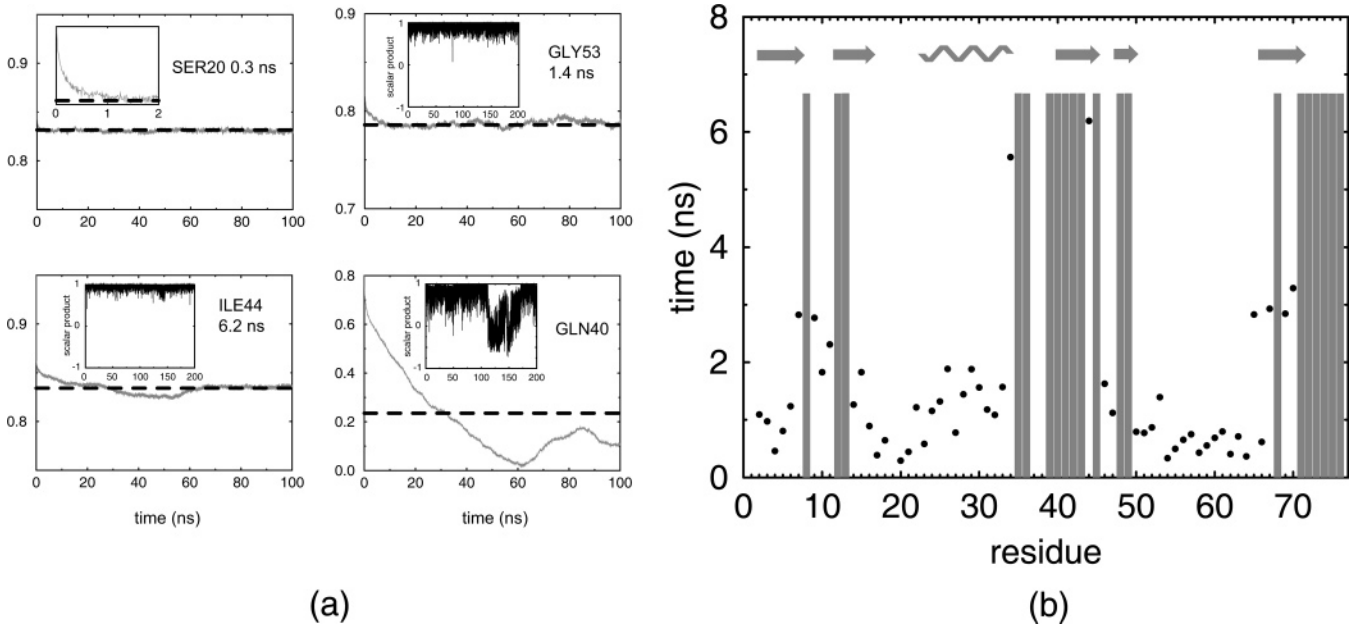


Figure 5. (a) Examples of internal correlation functions: short time scale (SER20), intermediate time scale (GLY53), long time scale (ILE44), and unconverged (GLN40) with $C_{int}(\infty)$ shown as dashed line and (insets) scalar product timeseries of interaction vector for GLY53, ILE44, and GLN40. (b) Correlation times of internal motions for 52 converged residues; unconverged residues are indicated with a gray bar. Secondary structure elements are depicted in top of the figure.

exhibit motions larger than 2 ns, whereas the same applies to the regions 34−45 (end of helix till end of third β-sheet) and 65−70 (fifth β-sheet).

Since we have calculated both the global correlation functions of each residue and the overall correlation function of the molecule, we can now determine if the decorrelation assumption (eq 6) holds for each residue in the present MD simulation. If this equation holds, the global correlation function of a specific residue should not be larger than the overall correlation function of the molecule at any timepoint. We checked this for the first 2 ns for each global correlation function. For 13 residues the global correlation function was larger than the overall correlation function, indicating that for these residues the decorrelation assumption does not hold. This set contains the following residues: THR7, LEU8, THR9, LYS11, THR12, ILE13, SER20, ASP21, PHE45, LYS48, ASP52, GLY53, and GLU64. Some of these residues belong to the residues with long time scale internal motions, indicating that the presence of long time scale motions can be correlated to a failure of the decorrelation assumption. In Figure 4c two examples are shown, THR7 and SER20, for which the decorrelation assumption does not hold.

D. Order Parameters. We can assess the order parameter S^2 calculated from the MD simulation using eq 7 for different timewindows. Figure 6a shows the order parameters calculated for 0.1, 1, 10, and 200 ns timewindows. For the first three cases the order parameter is averaged over all available time windows. For some residues the order parameter decreases with increasing timewindow; these correspond to residues with unconverged correlation functions or long correlation times for the internal motions (see Figure 5b).

In Figure 6b the order parameters derived from the 1 ns timewindow are compared to the experimentally derived order parameters from Lienin et al.[9] Only the 62 residues for which experimental data exists are now considered. The error bars for the MD values were taken as the standard deviation calculated from the averages over the different time windows. Overall the order parameter from the simulation follows the experimental values (correlation coefficient r=0.78), although the MD values are lower on average: 0.74 $\pm$ 0.14 compared to 0.81 $\pm$ 0.13 experimentally. Three residues considerably deviate: LYS11, LYS48, and ARG54, indicating that, in the simulation, those residues have more subnanosecond motions than experimentally derived. These three residues are solvent exposed, preceded by a glycine residue and located outside (LYS11, ARG54) or at the edge (LYS48) of secondary structure elements. Other glycine residues are present at the positions 35, 75, and 76 in the sequence. For GLY35 the deviation from the experimental order parameter is slightly less than in the other cases, possibly due to the fact that this residue is the least solvent exposed from all glycines in the sequence. Here again the overestimation of the amplitude of the fast internal motions may be affected by the too fast dynamics of the water molecules interacting with the protein within the SPC water model.

E. Back-Calculation of NMR Relaxation Data. The spectral density functions were obtained by Fourier transformation of the global correlation functions directly calculated from the MD simulations using eq 2. For comparing the simulated values of the spectral densities at the frequencies $\omega = 0$, ω_N, ω_H with the experimental values, the latter are calculated with the help of reduced spectral density mapping[30] from the averaged experimental set (see Supporting Information). Figure 7 shows the ratios J_{exp}/J_{md} for the frequencies $\omega = 0$, ω_N, ω_H for all residues. It appears that there is a large discrepancy between the simulated and the experimental values, especially at $\omega = 0$ and $\omega = \omega_H$, indicating that there are on average too many fast global motions and too little slow global motions in the simulation. The mean ratio between J_{exp}/J_{md} at $\omega = 0$ is 7.2, which

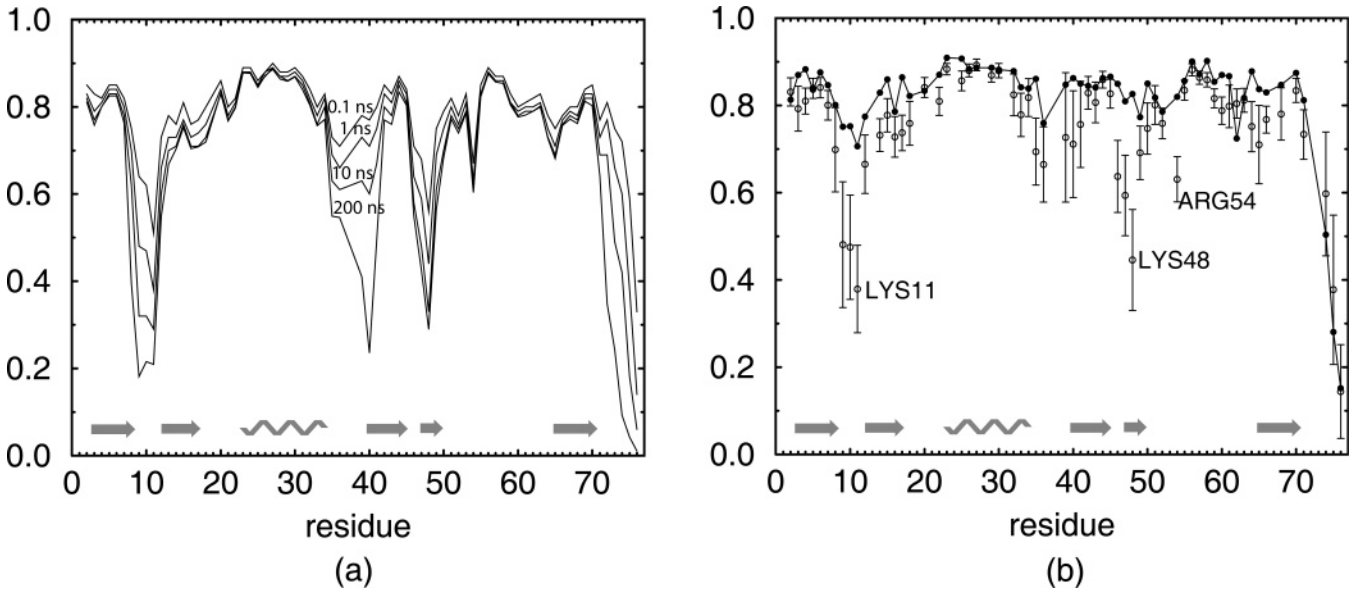


Figure 6. (a) Order parameters profiles for increasing timewindows and (b) comparison of experimental (●) and simulated (○) values averaged over the 1 ns timewindow.

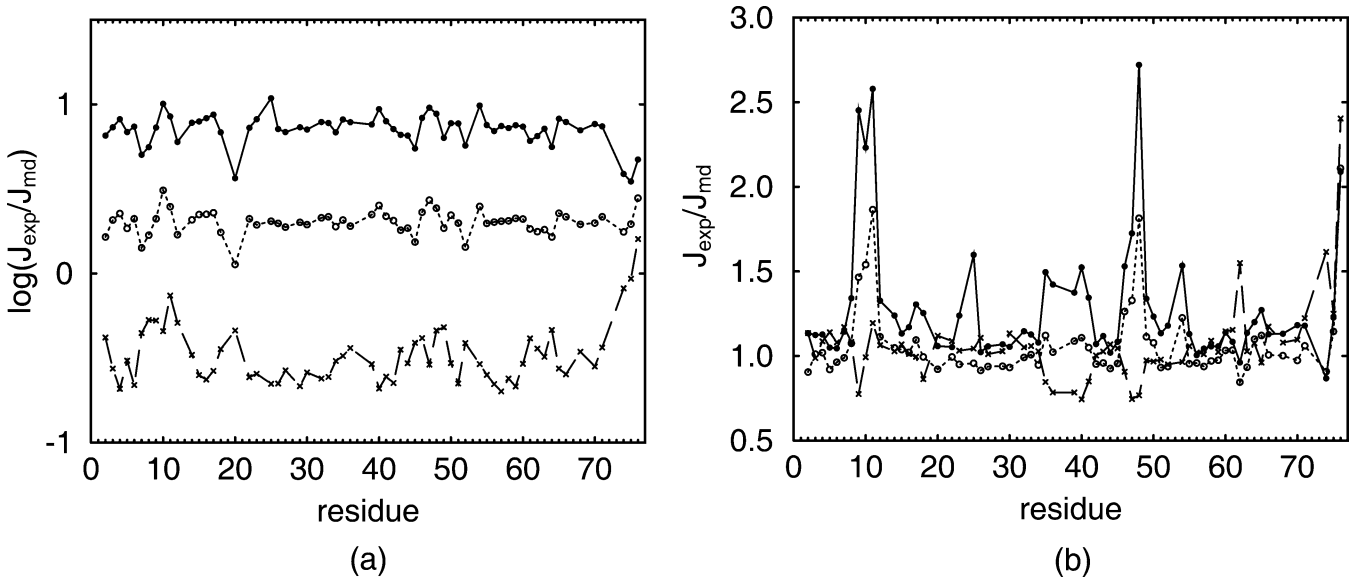


Figure 7. Ratio of spectral densities derived from the MD trajectories and the experimental data at frequencies 0 (●), ω_N (○), and ω_H (×) with (a) global tumbling as present in MD and (b) global tumbling $\tau_c = 4.03$ ns added. Note that the logarithm (base 10) of the ratio is indicated in the left panel (a).

reflects the factor by which the global tumbling is off from the experimental value.

Using eq 6 the global correlation function for each residue is reassessed assuming isotropic tumbling with $\tau_c = 4.03$ ns.[9] The second term in eq 6, $C_{overall}(t)$, thus becomes the isotropic function $\exp(-t/\tau_c)$. The first part is given by the internal correlation functions calculated using eq 2, after superposition of all structures onto the backbone heavy atoms of the secondary structure elements of the starting structure. After Fourier transformation of the resulting global correlation functions for each residue, spectral densities are obtained that are much closer to the experimental values for the spectral density as can be seen from Figure 7b. For both $J(0)$ and $J(\omega_N)$ large deviations are present for THR9, GLY10, LYS11, ALA46, GLY47, LYS48, and to some extent ARG54. The same applies for the segment 35−41. This can be attributed to the increased amplitude of the fast internal motions for those residues (also apparent in Figure 6b for the order parameters) which translates into an overestimation of $J(\omega_H)$ at the expense of $J(0)$ and $J(\omega_N)$ that become smaller since the integral over $J(\omega)$ must remain constant. The deviation of $J(0)$ for ASN25 originates from the large exchange contribution R_{ex} in the experimental data. The other small deviations of the J_{exp}/J_{sim} ratio from 1 may not be significant considering the amount of uncertainty that exists for the physical parameters used[2,37] and the experimental tumbling time.[10] Furthermore, small deviations from 1 may also originate from the anisotropy which is present in the experimental values[40] and which we omitted in our simulated values by assuming isotropic tumbling.

Direct calculation of NMR relaxation data is now feasible by substituting the spectral density values in eqs 3, 4, and 5.

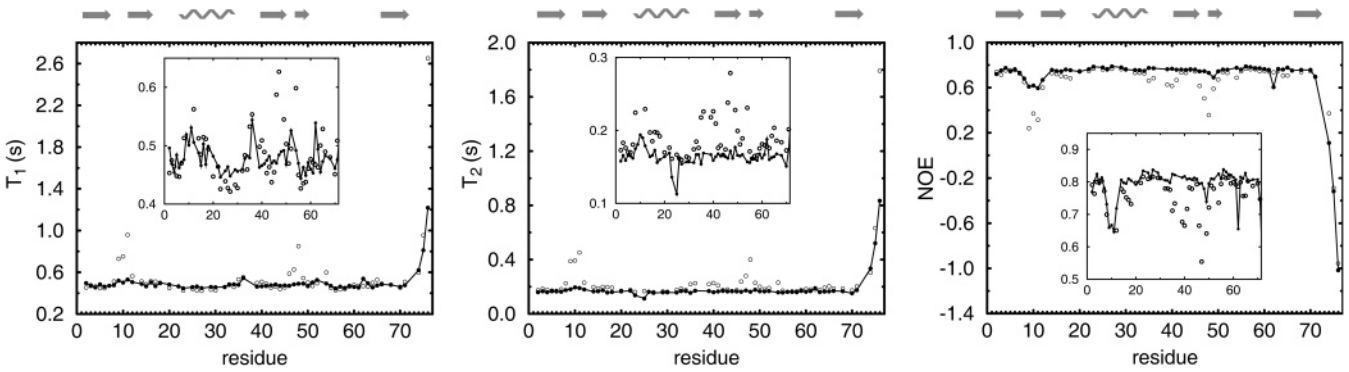


Figure 8. Experimental (●) and simulated (○) NMR relaxation data for T_1, T_2, and NOE. In the insets residues 1−71 are shown; relaxation parameters for residues 9−11 and 48 are off scale.

The quantitative agreement is poor if done for the values in Figure 7a (data not shown) since the global tumbling of the molecule is not reproduced in the simulation. Therefore we used the internal correlation functions from the simulation as before to derive the global correlation functions per residue using eq 6. The agreement now is quite good as can be noticed in Figure 8, although some residues still considerably deviate: 9−11, 35−41, 46−48, and 54. As discussed above, this deviation is mainly due to the presence of fast internal motions in the simulation. These fast motions are visible as the initial rapid decay from $C(0) = 1$ to a lower value. E.g. for LYS11 the value for both the internal and global correlation function is 0.67 at $t = 10$ ps. This indicates that substantial motion is present below this time scale that is responsible for the initial rapid decay of the correlation functions. In Figure 6 the aforementioned residues also deviate from the experimental S^2 values indicating again that too much fast internal motion is present on the subnanosecond time scale. The internal motions in the 2−6 ns range which we presented in Figure 5 do, however, not affect the relaxation data, since they are shielded by the global tumbling of the molecule: for example the T_1, T_2, and NOE values for residues GLU34 and ILE44 for which we found internal correlation times of 5.6 ns and 6.2 ns, respectively, are in good agreement with the experimental data.

F. Model-Free Analysis. Since we now have a complete and consistent picture of the dynamic properties of ubiquitin from our simulation including order parameters S^2, correlation times of internal motions τ_{int}, and predicted R_1, R_2, and NOE data, we can assess the robustness of model-free approaches to extract information on internal dynamics from relaxation data. For this we used the program TENSOR2[33] to extract order parameters and internal time scales from the relaxation data derived from the MD simulation (see Figure 8). We ran TENSOR2 with an experimental uncertainty of 2.5%. This value is a good approximation of the errors that are commonly present in NMR relaxation studies. A minimum absolute uncertainty of 0.01 is used for the NOE values.

We restricted our analysis to 43 residues that fulfilled the following criteria: i) The internal correlation function converges in the MD simulation. ii) The internal correlation time τ_{int} estimated in the MD is smaller than 2 ns. Larger motions are shielded by the global molecular tumbling (as shown above) and can consequently not be reproduced by TENSOR2.

Table 3. Result for Model-Free Analysis of Back-Calculated Relaxation Parameters from MD Trajectory

model	number	τ_{MF}[a]	τ_{int}[b] (ns)	$S_{MF}^2 - S_{MD}^2$
1	33[c]	n/a	0.9 ± 0.5	0.03
2	7[d]	19 ± 5 ps	0.6 ± 0.3	0.01
5	3[e]	1.4 ± 0.1 ns	1.5 ± 0.3	−0.15

[a] Time scales as calculated by TENSOR2. In case model 1 is selected, no time scale is assigned. [b] Internal time scales from MD simulation as calculated by eq 8 and presented in Figure 5. [c] GLN2, PHE4, VAL5, LYS6, THR14, LEU15, SER20, ASP21, THR22, ILE23, GLU24, ASN25, VAL26, LYS27, ALA28, LYS29, ILE30, GLN31, ASP32, LYS33, GLU51, ASP52, GLY53, THR55, LEU56, SER57, ASP58, TYR59, ASN60, ILE61, GLN62, LYS63, THR66. [d] ILE3, GLU16, VAL17, GLU18, LEU50, ARG54, GLU64. [e] GLY10, ALA46, GLY47.

For the selected residues TENSOR chose model 1 in 33 cases, whereas the models 2 and 5 were chosen less often: 7 and 3 times, respectively (see Table 3). This means that for 33 cases TENSOR2 is not capable of detecting the nanosecond time scale internal motion that is present for these residues. The same applies to the 7 residues for which model 2 is selected: the time scales of the internal motions are severely underestimated. Only in the 3 cases where model 5 is selected, the time scales are in agreement with the time scales presented in Figure 5. Note that this behavior is not correlated to the failure of the decorrelation assumption that was found for 13 residues (see above).

The failure of detecting nanosecond time scale motions is mainly due to the noise level in the experimental data: lowering the amount of uncertainty in the experimental data to the unrealistically low value of 0.5% favors the choice of model 5, and consequently motions in the nanosecond time scale are detected. However, such small uncertainties may not be realistic from an experimental point of view. In addition, longer time scale motions for all residues would as well be detected if the analysis was performed exclusively with model 5. This would however not be statistically justified because of the unrealistically high associated errors.

The findings for the time scales presented above are in agreement with results presented by Chen et al.[7] They also find that nanosecond time scale motions can go undetected in the model-free analysis and that the model-free time scales consequently underestimate the effective correlation times.

For the selected residues the order parameter is correctly reproduced, as illustrated in Figure 9. The differences be-

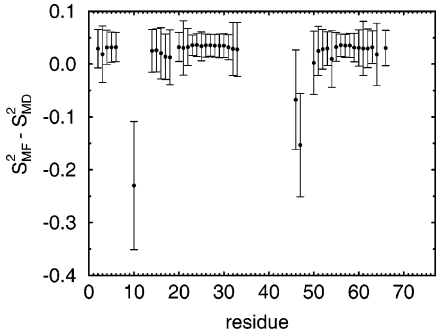


Figure 9. Difference between model-free order parameters (S^2_{MF}) and MD order parameters (S^2_{MD}). Errorbars are calculated as the square root of the quadratic sum of the individual errors.

tween the order parameters predicted by TENSOR2 S_{MF}^2 and the simulated order parameters S_{MD}^2 for the time window of 1 ns are close to zero, except for GLY10 and GLY47 (see Figure 6a). For these residues the predicted order parameters by TENSOR2 are close to the ones calculated for a 10 ns time window. For the residues fitted with model 1 the order parameter is slightly overestimated by an amount of 0.03 (see Table 3), although this difference is not significant for each of the residues (see errorbars in Figure 9).

IV. Conclusions

The present results from our molecular dynamics simulation of ubiquitin allows us to draw several conclusions. Most importantly we have shown that long MD simulations can yield insight in the internal dynamics beyond the time scale that is commonly studied in NMR relaxation studies. Motions on longer time scales than the overall tumbling do not affect the back-calculated NMR relaxation parameters T_1, T_2, and NOE and will consequently go undetected in the standard model-free analysis. We also have shown that time scales are in general underestimated in the standard model-free analysis, unless extremely precise experimental data can be recorded. Consequently nanosecond time scale motions often remain undetected. The observation of slow motions in the present simulation is corroborated by recent experimental studies on protein backbone motion based on residual dipolar coupling measurements and cross-correlation measurements.

One of the generally accepted assumptions that overall and internal motions are uncorrelated in globular proteins,[2] a requisitive for the model-free analysis, does not hold for quite a number of residues in our simulation. Here it should be noted that in the present simulation for many residues the internal correlation time is larger than the global correlation time. This situation is opposite from what is believed to be the case for globular proteins studied with NMR.[2] It remains thus undecided if global and internal motions can be split up following eq 6. Although this relation is widely used in the NMR field, evidence from simulation or experiment that it really holds is sparse.[2,6] To really test this assumption for globular proteins, longer MD simulations of at least 1 μs are needed.

An important point in extending the time scales of MD simulations for this purpose is that suitable parameters and solvent models should be used that reproduce the overall tumbling in solution. Most force fields, however, have been parametrized on much shorter time scales aiming at properly reproducing structural, short time scale dynamics, and thermodynamics properties. Obtaining parameters that, in addition, properly reproduce long time scale dynamics and overall tumbling remains a challenge. Such an effort will also strongly depend on the availability of high quality experimental data describing long time scale dynamics in biomolecules.

In conclusion, the present work underlines both the strengths and weaknesses of MD simulation methods and, to some extent, the experimental limitations of NMR relaxation studies, especially regarding the required accuracy, in properly describing long times scale dynamics.

Acknowledgment. We thank Dr. Chris Spronk (Radboud University) for providing us with the refined model of 1D3Z. Dr. Rafael Brüschweiler (Clark University) and Dr. David Fushman (University of Maryland) are acknowledged for providing their experimental relaxation data of ubiquitin. Rafael Brüschweiler, Martin Blackledge (IBS Grenoble), Jeanine Prompers (Eindhoven University), Klaartje Houben (Utrecht University), Hans Wienk (Utrecht University), and Christine Peter (NIH) are thanked for useful discussions. Aalt-Jan van Dijk (Utrecht University) is acknowledged for preparing Figure 1. David Fushman is thanked for careful reading of the manuscript. This work was supported by the European Community program NMRQUAL (contract number QLG2-CT-2000-0313).

Supporting Information Available: Cross-RMSD matrix, matrix for NOE violations, and averaged NMR relaxation data set (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Bruschweiler, R. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr. Opin. Struct. Biol.* **2003**, *13*, 175−183.

(2) Case, D. A. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* **2002**, *35*, 325−331.

(3) Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998**, *282*, 740−744.

(4) Feenstra, K. A.; Peter, C.; Scheek, R. M.; van Gunsteren, W. F.; Mark, A. E. A comparison of methods for calculating NMR cross-relaxation rates (NOESY and ROESY intensities) in small peptides. *J. Biomol. NMR* **2002**, *23*, 181−194.

(5) Peter, C.; Daura, X.; van Gunsteren, W. F. Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations. *J. Biomol. NMR.* **2001**, *20*, 297−310.

(6) Tugarinov, V.; Liang, Z.; Shapiro, Y. E.; Freed, J. H.; Meirovitch, E. A structural mode-coupling approach to 15N NMR relaxation in proteins. *J. Am. Chem. Soc.* **2001**, *123*, 3055−3063.

(7) Chen, J.; Brooks, C. L. R.; Wright, P. E. Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation. *J. Biomol. NMR* **2004**, *29*, 243−257.

(8) Joazeiro, C. A. P.; Hunter, T. BIOCHEMISTRY: Ubiquitination-More Than Two to Tango. *Science* **2000**, *289*, 2061−2062.

(9) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD computer simulation. *J. Am. Chem. Soc*. **1998**, *120*, 9870−9879.

(10) Lee, A. L.; Wand, A. J. Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation. *J. Biomol. NMR* **1999**, *13*, 101−112.

(11) Fushman, D.; Varadan, R.; Assfalg, M.; Walker, O. Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. *Prog. NMR Spectrosc*. **2004**, *44*, 189−214.

(12) Peti, W.; Meiler, J.; Bruschweiler, R.; Griesinger, C. Model-free analysis of protein backbone motion from residual dipolar couplings. *J. Am. Chem. Soc*. **2002**, *124*, 5822−5833.

(13) Tolman, J. R. A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *J. Am. Chem. Soc*. **2002**, *124*, 12020−12030.

(14) Fruh, D.; Tolman, J. R.; Bodenhausen, G.; Zwahlen, C. Cross-correlated chemical shift modulation: a signature of slow internal motions in proteins. *J. Am. Chem. Soc*. **2001**, *123*, 4810−4816.

(15) Deschamps, M.; Bodenhausen, G. Anisotropy of rotational diffusion dipole−dipole cross-correlated NMR relaxation and angles between bond vectors in proteins. *Chem. Phys. Chem*. **2001**, *8/9*, 539−543.

(16) Vugmeyster, L.; Perazzolo, C.; Wist, J.; Frueh, D.; Bodenhausen, G. Evidence of slow motions by cross-correlated chemical shift modulation in deuterated and protonated proteins. *J. Biomol. NMR* **2004**, *28*, 173−177.

(17) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc*. **1998**, *120*, 6836−6837.

(18) Nabuurs, S. B.; Nederveen, A. J.; Vranken, W.; Doreleijers, J. F.; Bonvin, A. M. J. J.; Vuister, G. W.; Vriend, G.; Spronk, C. A. E. M. DRESS: a database of REfined solution NMR structures. *Proteins* **2004**, *55*, 483−486.

(19) Cook, W. J.; Jeffrey, L. C.; Carson, M.; Chen, Z.; Pickart, C. M. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *J. Biol. Chem*. **1992**, *267*, 16467−16471.

(20) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm*. **1995**, *91*, 43−56.

(21) Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem*. **1998**, *19*, 535−547.

(22) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F. In *Intermolecular forces*; Reidel Publishing Company: Dordrecht, Interactions models for water in relation to protein hydration, 1981; pp 331−342.

(23) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys*. **1984**, *81*, 3684−3690.

(24) Tironi, I. G.; Sperb, R.; Smith, P. E.; Van Gunsteren, W. F. A generalized reaction field method for molecular-dynamics simulations. *J. Chem. Phys*. **1995**, *102*, 5451−5459.

(25) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.

(26) Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Skelton, N. J. *Protein NMR Spectroscopy, Principles & Practice*; Academic Press: San Diego, CA, 1996.

(27) Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr*. **1991**, *24*, 946−950.

(28) Merrit E. A.; Murphy M. E. P. Rasterd3D, Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr*. **1994**, *D50*, 869−873.

(29) Peng, J. W.; Wagner, G. Mapping of the spectral densities of N−H bond motions in Eglin-C using heteronuclear relaxation experiments. *Biochemistry* **1992**, *31*, 8571−8586.

(30) Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E. Spectral density function mapping using 15N relaxation data exclusively. *J. Biomol. NMR* **1995**, *6*, 153−162.

(31) Lipari, B.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc*. **1982**, *104*, 4546−4559.

(32) Lipari, B.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc*. **1982**, *104*, 4559−4570.

(33) Dosset, P.; Hus, J. C.; Blackledge, M.; Marion, D. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* **2000**, *16*, 23−28.

(34) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.

(35) Henry, E. R.; Szabo, A. Influence of vibrational motion on solid-state line shapes and NMR relaxation. *J. Chem. Phys*. **1985**, *82*, 4753−4761.

(36) Fushman, D.; Ohlenschlager, O.; Ruterjans, H. Determination of the backbone mobility of ribonuclease T1 and its 2′GMP complex using molecular dynamics simulations and NMR relaxation data. *J. Biomol. Struct. Dyn*. **1994**, *11*, 1377−1402.

(37) Pfeiffer, S.; Fushman, D.; Cowburn, D. Simulated and NMR-derived backbone dynamics of a protein with significant flexibility: a comparison of spectral densities for the beta ARK1 PH domain. *J. Am. Chem. Soc*. **2001**, *123*, 3021−3036.

(38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res*. **2000**, *28*, 235−242.

(39) Woessner, D. E. Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. *J. Chem. Phys*. **1962**, *37*, 647−654.

(40) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc*. **1995**, *117*, 12562−12566.

(41) Smith, P. E.; Van Gunsteren, W. F. The viscosity of SPC and SPC/E water at 277 and 300 K. *Chem. Phys. Lett*. **1993**, *215*, 315−318.

(42) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen,

D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wi?rkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(43) Prompers, J. J.; Bruschweiler, R. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc*. **2002**, *124*, 4522−4534.

CT0498829

JCTC Journal of Chemical Theory and Computation

Environmental Effects on Glycophorin A Folding and Structure Examined through Molecular Simulations

Hirsh Nanda,[†] Jonathan N. Sachs,[§] Horia I. Petrache,[‡] and Thomas B. Woolf*[,†]

Department of Physiology, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205, Laboratory of Physical and Structural Biology, NICHD, National Institutes of Health, Bethesda, Maryland 20892, and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520

Received September 29, 2004

Abstract: The human erythrocyte sialoglycoprotein glycophorin A (GpA) has been used extensively in experiment and simulations as a model of transmembrane helix−dimer formation, emphasizing the critical role of specific residue−residue interactions between helices in dimer stability. While the tertiary dimer structure is modulated by the hydrophobic lipid bilayer environment, we show that interactions of GpA with ordered interfacial water are commensurate to intrahelical forces. The role of lipid−water interface in stabilizing transmembrane proteins is not yet understood; however, dramatic water reordering in the presence of the transmembrane domains is observed from simulations and is possibly measurable by experiment. Interfacial interactions including anisotropic interactions with the polar headgroups might favor parallel association of transmembrane helices. To quantify forces capable of disrupting the GpA dimer, we generate folding/unfolding intermediates by replacement of the lipid bilayer with water, eliminating not only the native hydrophobic environment but also the native interfacial water region. Dramatic changes in the secondary, helical structures occur, with a transition from i,i+4 α-helix to i,i+5 π-helix and concomitant perturbation of the tertiary structure. Enforcing the native α-helix secondary structure by soft dihedral restraints restores the native tertiary structure, in essence substituting for the lack of the lipid−water interface. We suggest that differentiation between interactions within the lipid bilayer, including interactions with lipid headgroups and interfacial water can enrich our understanding of the thermodynamic stability of transmembrane domains.

Introduction

Transmembrane proteins provide a unique challenge to a full molecular understanding of their structure and function.[1,2] A fundamental problem is obtaining structural information, even with the availability of modern experimental techniques such as high resolution X-ray diffraction and solid-state NMR spectroscopy. This handicap has given the community just a limited number of membrane proteins for in-depth analysis of structure and folding, compared to globular proteins. At the time of this writing there are approximately 400 entries in the PDB for membrane proteins and >25 000 entries for globular proteins. On the other hand, membrane protein folding motifs may be simpler than their water soluble counterparts.[3−6] The membrane spanning region of proteins are usually closely packed bundles of α-helices or β-barrels. Furthermore, from primary sequence data alone, α-helical regions can often be identified through hydropathy scales,[7] and reliable methods for predicting membrane protein

* Corresponding author phone: (410)614-2643; fax: (410)614-4436; e-mail: woolf@groucho.med.jhmi.edu.
[†] Johns Hopkins University School of Medicine.
[‡] National Institute of Health.
[§] Yale University.

10.1021/ct049928y CCC: $30.25
Published on Web 04/09/2005

topology have been recently developed.[8,9] This makes some aspects of structural prediction easier. However, protein interactions with the anisotropic, heterogeneous environment of the membrane bilayer continue to challenge our understanding of folding and function.[10,11]

The association of the glycophorin A (GpA) transmembrane region has been studied intently as a model for elucidating how proteins fold in a lipid membrane environment.[3,12−14] The GpA transmembrane region consists of a homodimer formed by two parallel α-helices of more than 18 residues each. The α-helices are formed by the apolar residue sequences T_{74}LIIFGVMAGVIGTILLI$_{91}$, that self-associate within a hydrophobic lipid environment. The question becomes what relative contributions do protein−protein interactions and lipid−protein interactions have in driving protein assembly? Consequently, what types of interactions with the lipid bilayer modulate GpA structure and function?

Much of the literature on transmembrane protein folding has focused on specific protein−protein interactions. In particular, early mutagenesis work on the GpA dimer identified the key residue sequence, LIxxGVxxGVxxT, as being responsible for dimer formation.[15,16] Based on these findings, a model of the dimer structure was proposed[17] and later validated by NMR spectroscopy.[18,19] With the structure of the GpA dimer solved, molecular modeling has revealed that close residue contacts, required for an overall attractive van der Waals interaction, were allowed by the small, compliant Gly residues at the helical interface.[20−22] It is now believed that, in general, the highly conserved GxxxG motif [4] allows the close packing of transmembrane helices in other proteins to occur.

Insights gained from the above studies are thought to be generalizable to other transmembrane proteins. However, in addition to the GxxxG motif mentioned above, other statistically relevant sequence motifs were identified,[4] some suggesting alternative mechanisms for helix−helix associations in membranes. In particular, while the glycophorin model emphasized the importance of van der Waals interactions, other proteins use polar side chains buried in the hydrophobic region of the bilayer to electrostatically drive the association of transmembrane helix domains.[23−26] These examples highlight that membrane associated proteins may utilize aspects of the lipid environment to promote folding and function.

A number of experimental studies have shown that features of the lipid bilayer environment such as thickness, fluidity, and chemical composition affect the folding and function of membrane proteins. For example, disparate cellular processes such as the rhodopsin photocycle[27,28] and alamethicin channel formation[29] have been shown to depend on the monolayer intrinsic curvature, a distinct property of lipid assemblies. Lipid chain length has been shown to have a crucial affect on gramicidin channel formation.[30,31] Studies on GpA have shown that the type of detergent micelle can affect monomer affinity by several orders of magnitude although the dimer structure itself is not altered.[14,32]

Lipid bilayers are not simple slabs of hydrophobic matter. A large section of the lipid bilayer, the interfacial region, is significantly more polar than the hydrocarbon core. We therefore should not restrict our attention to interactions just within the hydrophobic core but instead take into account possible interactions with the polar interface. The role of the bilayer interface and its interplay with the apolar environment of the membrane in stabilizing the GpA dimer and influencing parallel helix packing is a central concern of this paper.

The use of molecular dynamics to gain insight into the behavior and function of globular proteins is by now well established.[33] Furthermore, the importance of the solvation environment in properly capturing protein dynamics has been shown in several studies.[34] This powerful tool is also capable of giving insight into the complicated lipid bilayer environment.[22,35] Both ensemble averages and dynamic properties of the system are obtainable through this method. Because all forces are calculated at the atomic level it is, in principle, possible to parse out the most relevant interactions responsible for the process of interest. Recently, MD computations have been successfully applied to membrane proteins, including bacteriorhodopsin,[36] rhodopsin,[37] gramicidin,[38−40] OmpA,[35,41] KcsA,[42] and GpA.[22,43] The role of protein−lipid interactions has been the main aspect of these studies.

The central thrust of the current molecular dynamics calculations is to address the effects of the lipid−water interface environment on GpA transmembrane structure. In particular, how are interactions with the interface and lipid hydrocarbon mediated to stabilize the dimer? To this end, we first analyze simulations of the GpA dimer in a dimyristoylphosphatidylcholine (DMPC) membrane (Figure 1a), where protein structure has been validated by solid-state NMR measurements,[19] to investigate the interactions of GpA with the various regions within the bilayer, including interfacial water. We find that interfacial water, which differs markedly from bulk due to interactions with the lipid headgroups, is further modified by strong interactions with the transmembrane GpA dimer. In addition, because of the parallel orientation of the GpA dimer along the bilayer normal, interfacial dynamics on the two sides are distinct.

An important framework for understanding membrane protein assembly is the two-stage model introduced by Popot and Engelman,[44] in which stable α-helices first form within the lipid bilayer and then tertiary association forms the final structure. The process of helical insertion into bilayers and spontaneous dimer formation using simulations is still beyond current computational abilities. However much can be learned about the folding process from simulations of dimer disruption. The GpA dimer appears to be a particularly stable construct. As we have reported earlier, dimer structure is modulated by bilayers of different lipid types, in a manner consistent with hydrophobic matching concepts.[22] However this modulation does not involve significantly different structural intermediates. It has also been shown that dimer association probability is affected by micelles of different detergent type but not dimer structure itself.[14,32] Since GpA in its native environment is in permanent interaction with interfacial water as well as lipid hydrocarbon, a simple way to disrupt the dimer is removal of lipids and replacement with water throughout the simulation box.

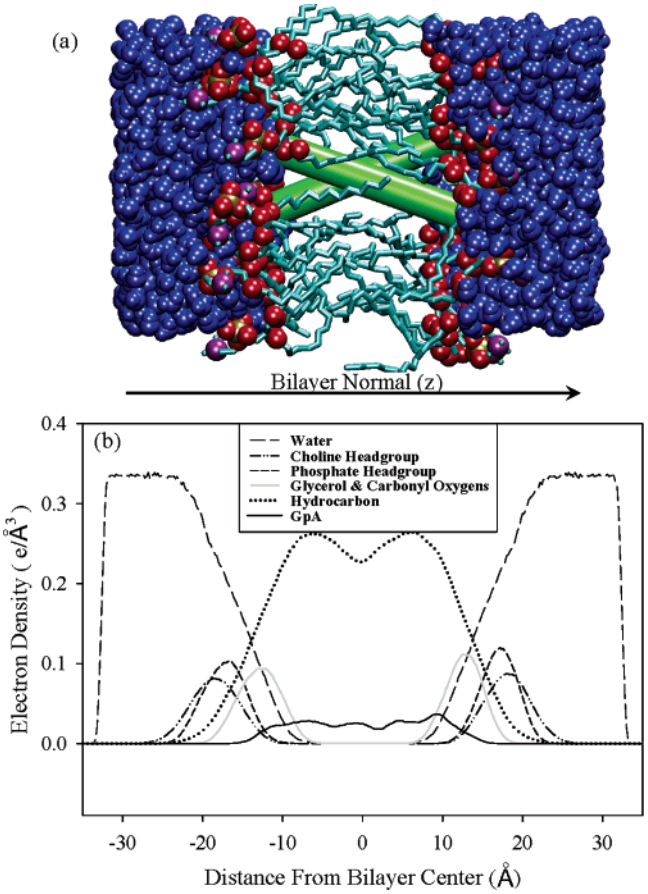


Figure 1. (a) Simulation snapshot of GpA in a DMPC bilayer. The protein is represented as cylinders (green), the lipid chains as lines (cyan), and the lipid headgroups and waters are shown as VdW spheres (water, blue; oxygen, red; phosphate, gold; nitrogen, violet). Hydrogens are omitted for clarity. (b) Electron density profile showing the distribution of molecular groups along the bilayer normal, averaged over the simulation.

By learning how the protein "falls apart" in the absence of the ordering potential of the membrane, we obtain insights into the role of the lipid/water interface in stabilizing the native protein structure. In addition, the water simulation allows us to quantify the asymmetry effect seen in the lipid simulation. Reassuringly, GpA structure is denaturated in water. Interestingly, however, the collapse of the tertiary and the secondary structure are very much interdependent. Indeed, by holding the backbone dihedral angles in the vicinity of the starting conformation (from NMR), we show that the tertiary structure is preserved, despite the absence of a native, bilayer environment to hold it in place. This suggests that, once helical monomeric structures are stabilized within both the hydrophobic core and the headgroup/water interface, dimer stability is assured.

Finally, allowing backbone dihedrals to evolve freely does not lead to direct unfolding into random coil conformations. Instead, exposure to water modulates backbone hydrogen bonding patterns, and alternate (metastable) dimer conformations are found throughout the length of the trajectory.

We suggest that interactions with the polar region of lipid bilayer, including direct and indirect perturbation of the interfacial water should be taken into account for molecular associations within transmembrane domains.

Materials and Methods

Bilayer Simulation Details. All bilayer simulations were performed using CHARMM versions 26 and 27[45] software and the CHARMM 27 parameter set. The DMPC pure bilayer system was constructed with 18 lipids per monolayer (36 lipids total), while the DMPC/GpA simulation was constructed with 45 lipids per monolayer (90 lipids total). Furthermore, 925 and 3756 waters were used in the pure DMPC and DMPC/GpA simulations, respectively, under the TIP3P water model. The construction of the membrane around the GpA molecule followed a procedure proposed by Woolf and Roux.[38,39] The NMR structure of GpA reported by MacKenzie et al.[18] was used as the initial structure for the GpA dimer.

Equilibration of both systems was performed initially through minimization and Langevin dynamics. The final set of equilibration and production was performed with a constant number of atoms (N), constant lateral surface area (A), constant normal pressure ($P = 1$ atm), and a constant temperature ($T = 298$ K) to give a NAP_NT ensemble. The total lateral surface area was 1075 Å^2 for the pure bilayer system and 3040 Å^2 for DMPC/GpA simulation. These values represent an area per lipid of 59.72 Å^2 that is consistent with experimental numbers[46] where for the DMPC/GpA system ~353 Å^2 was estimated for the dimer. All systems were run with periodic boundary conditions (PBC) using particle mesh Ewald for the treatment of long-range electrostatic interactions.[47] A cutoff of 12 Å was used for van der Waals interactions.[48] All bonds involving hydrogen atoms were fixed using the SHAKE algorithm[49] allowing a time step of 2 fs for dynamics. Lipid systems were simulated for a total of 1.5 ns and 22 ns for the DMPC and DMPC/GpA systems, respectively. The first 0.25 ns was taken as an equilibration period, and results are reported on the remaining simulation time. Further discussion involving ensemble choice and system construction can be found in ref 22.

GpA in Water Simulation Details. This set of simulations were performed using the CHARMM 27 software package with the CHARMM 27 parameter set. All simulations used periodic boundary conditions and particle mesh Ewald for long-range electrostatic interactions as well as a 12 Å cutoff for van der Waals interactions. Bonds to hydrogen atoms were fixed using the SHAKE algorithm allowing a 2 fs time step during dynamics.

The initial water box simulations used the same simulation conditions as the bilayer simulations. Namely, the NAP_NT ensemble was considered with a constant lateral area ($A = 1200$ Å^2), a constant normal pressure ($P_N = 1$ atm), and a constant temperature ($T = 298$ K). The simulation cell consisted of one GpA dimer molecule and 2060 waters. The system was constructed from an equilibrated water box with cell dimensions similar to those of the bilayer simulations. Waters that overlapped with the protein were deleted, and minimization was performed on the waters with all atoms of the protein fixed. This was then proceeded by 200 ps of

equilibration using constant normal pressure and constant temperature dynamics with the structure of the GpA dimer still fixed. The cell dimensions of the system were observed to have remained stable over the last 100 ps of equilibration. All restraints on the proteins were then removed, and two simulations were started each beginning with a different set of initial random velocities. The velocities were assigned using a Gaussian distribution such that the system was set to the proper temperature. These two GpA/water systems were then run out to the limit of available computational resources, 16 ns and 19 ns, respectively.

Construction of the large water boxes used an equilibrated, cubic, water box of larger dimensions with roughly double the volume of the bilayer simulations. Again the micelle structure of the GpA dimer was inserted into the system, and overlapping waters were deleted. This resulted in a system containing 5626 waters with cell dimensions of ~55.3 Å on each side. All three cell dimensions were allowed to vary under a constant pressure and temperature ensemble, *NPT*, with a pressure, $P = 1$ atm, and temperature, $T = 298$ K. The equilibration procedure was similar to that above; first all atoms of the GpA dimer were fixed, and then minimization on waters was performed followed by 200 ps of dynamics on the entire system. All restraints on the protein were then released, and three simulations were begun each starting with a different set of initial velocities. One simulation was run out for 5 ns of production. For the next two simulations a harmonic restraint was put on the dihedral angles of the backbone, and the systems were simulated for 2 ns each. The total 4 ns of simulation time was analyzed and reported in the results.

Hydrogen Bonding. There is no explicit hydrogen bonding term in the CHARMM 27 force field. For this reason backbone hydrogen bonding partners were identified by geometrical criteria. All hydrogens that were within a 90° angle cone of the hydrogen bond acceptor, carbonyl oxygen, and within 2.5 Å of the oxygen were considered potential hydrogen bonding partners. If more than one hydrogen satisfied the above criteria, they were further discriminated against using a smaller 135° angle cone and then considering the hydrogen with the shortest distance to the acceptor atom. The same algorithm was applied to finding water oxygen partners for the backbone amide hydrogen.

Results

The Native Lipid/Water Interface. There are distinct structural features along the bilayer normal of a membrane. Electron density profiles from a DMPC/GpA simulation in Figure 1b show the distribution of different molecular groups along this normal with the distributions weighted by the number of electrons for each molecular group. The membrane interface, defined as the region of the polar headgroups (phosphate and choline as well as the polar glycerol backbone and carbonyl oxygen), contributes to almost one-half of the bilayer thickness. Furthermore, there is a significant population of waters in this region. Some waters penetrate to within 10 Å of the bilayer center, well within the hydrocarbon region of the membrane and in contact with GpA.

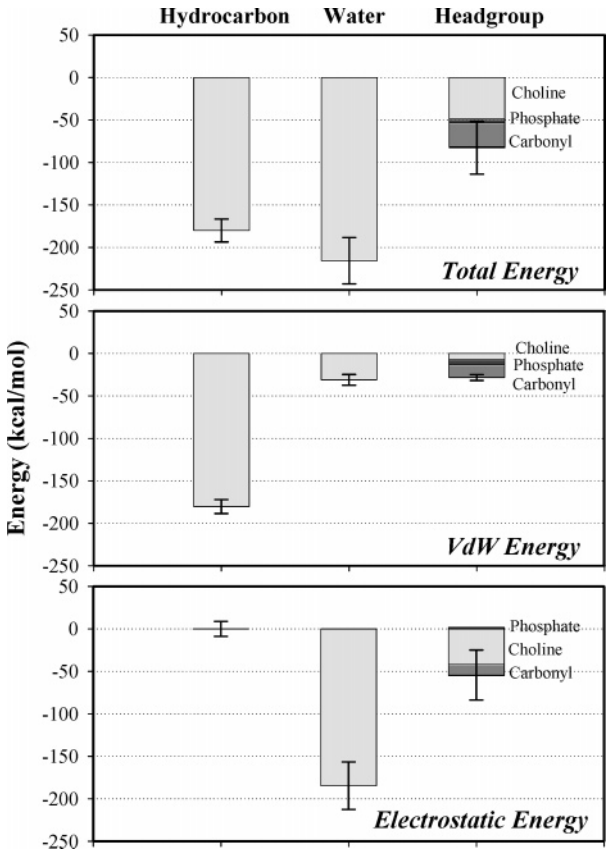


Figure 2. Average GpA interaction energy with bilayer environment in DMPC/GpA simulation. Interaction energy is decomposed into contributions from hydrocarbon, water, and headgroup. The headgroup term is further decomposed into phosphate, choline, and carbonyl\ester interactions. Mean square fluctuations are shown with error bars.

Figure 2a shows average interaction energies between the GpA dimer and the various polar and nonpolar groups of the lipid bilayer. It is striking that the protein−water interactions are the strongest. Interactions with the lipid hydrocarbon are also large, as expected, and are of roughly the same magnitude as water interactions. The combined contribution of the phosphate, choline, and glycerol/carbonyl groups (headgroup interactions) represent roughly 1/6 of the total interaction energy between the protein and its environment. However, if one considers the total membrane interface as the sum of the water and headgroup contributions, then interactions with the interface dominate over that of the lipid hydrocarbon. The interaction of individual helices with the lipid environment are statistically equivalent and half that of the total dimer interaction. Major differences in the breakdown into van der Waals and electrostatic components are also seen in Figure 2. As expected, the nonpolar lipid tails interact primarily through van der Waals interactions (Figure 2b), whereas the water interactions are electrostatically dominated (Figure 2c). The phosphate group has, on average, a small slightly repulsive electrostatic interaction with GpA, with large positive and negative fluctuations about the mean.

The strong interaction between the protein and water motivates a more detailed analysis of water behavior in the vicinity of the GpA dimer. Figure 3 compares water order

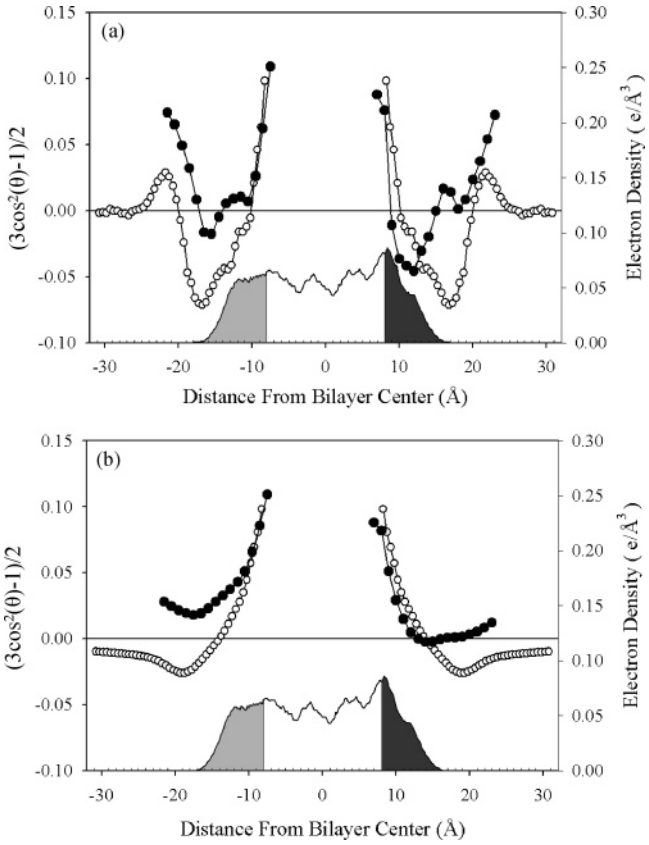


Figure 3. (a) Local second rank order parameters of water O−H bond orientation with respect to the membrane normal for simulations of pure DMPC (open circles) and DMPC/GpA (filled circles). For DMPC/GpA only water within a 6 Å radius of the protein was considered to emphasize the effect in the protein vicinity. GpA electron density profie is replotted from Figure 1b. The N-terminus amide hydrogen and the C-terminus carbonyl oxygen regions are shaded in light and dark gray, respectively. (b) Running averages (integrated profiles) of data in part (a).

parameters relative to the membrane normal with and without the GpA dimer. The average second-order Legendre polynomial, $(3\cos^2(\theta)-1)/2$, is plotted where θ is the angle of the O−H bond vector with the membrane normal. These second rank order parameters of the water O−H bonds are measurable by solid-state ^{2}H NMR in heavy water.[50] The NMR observable corresponds to running averages of these profiles as seen in Figure 3b. However, from simulations, local order parameters can be determined and are amenable to a more straightforward interpretation, Figure 3a. As shown by Figure 3, for a pure bilayer, interfacial water is ordered.[51−54] Away from the interface, water molecules have no preferred orientation (zero local order parameter), defining the bulk water region. As one moves into the membrane and the lipid headgroup environment, regions of the second rank order parameter with different signs are observed.[51−54] The parallel orientation of GpA causes significant alternation of water ordering, than that characteristic to lipid bilayers,[55] as shown by calculations within a 6 Å radius of the protein dimer (Figure 3). This includes the first two hydration shells of water around the protein as determined by radial distribution functions.[56] Furthermore, the parallel packing of the GpA homodimer introduces an asymmetry about the bilayer midplane, manifested through differences in water orientation at the two sides of the bilayer. There is a clear shift in water ordering at the N-terminus (negative z), when compared to ordering at the C-terminus (positive z). A possible interpretation is a tendency for water at the N-terminus to orient with the OH bond vector pointing away from the bilayer center. In contrast, at the C-terminus, water ordering is biased toward the bilayer center. The structural difference between the N- and C-termini of the GpA dimer is, obviously, the free amide hydrogens on one side and the carbonyl oxygens on the other. For reference to water order parameter profiles, these regions are highlighted in Figure 3.

Lipid Replacement. We now investigate the effect of removing the lipid molecules by filling the whole simulation box with water. This replacement not only eliminates the native hydrophobic environment of the GpA core but also destroys the water ordering characteristic of lipid/water interfaces (the native water environment of GpA). Because of these differences, major disruption of the GpA dimer is expected. However, despite radical structural changes, a helical secondary structure was still preserved in all water runs. Furthermore, the helices remained associated for the duration of the simulations. A comparison of the Cα root-mean-square deviation for the DMPC/GpA and water/GpA Run1 simulations show that initially there is a larger deviation in helix for the water run, Supplementary Figure 1 (see Supporting Information). However, as the protein stabilizes into the new helical structure the deviation plateaus. The nature of this hydrophobic association or aggregation in water will be described below. Without the shielding of the membrane, significant competition between intrahelical backbone hydrogen bonding and that with water occurs, as shown in Figure 4. GpA hydrogen bonding with water in the DMPC/GpA run is shown for comparison. In the bilayer setting, only the first four residues at the N-terminal side and the last four residues at the C-terminal side can hydrogen bond to water. The internal residues of the protein are shielded from water interactions by the lipid hydrocarbon region. The α-helices of the transmembrane dimer, stabilized by the i,i+4 intrahelical backbone hydrogen bonding motif, leaves the amide hydrogens of the first four residues at the N-terminus and the carbonyl oxygens of the last four residues at the C-terminus with unpaired donors and acceptors, respectively.

With the lipids replaced by water, increased hydrogen bonding between GpA and water is seen. Overall, there is a greater degree of water interactions with the backbone carbonyl oxygen. This is likely due to the ability of the water hydrogens to adopt favorable orientations while avoiding steric blocking from amino acid side chains. However, not all interior amino acids show the same degree of water hydrogen bonding. Reduced interactions with water are found for the residues at the helix−helix interface, in particular Gly79, Ala82, and Gly83. Other residues, Val80, Val84, and Ile85 exhibit minimal hydrogen bonding to water even though they are not involved in the helix−helix interface. These residues all contain β-branched side chains that can sterically block access to the protein backbone.

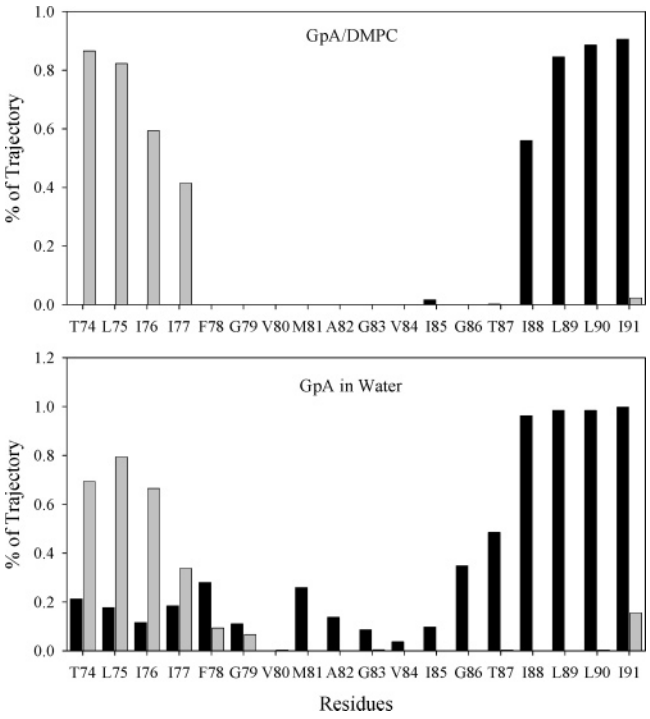


Figure 4. Fraction of simulation each residue spends forming backbone-water hydrogen bonds. Amide hydrogens to water bonds are shown in light gray; carbonyl oxygens to water bonds are shown in dark gray.

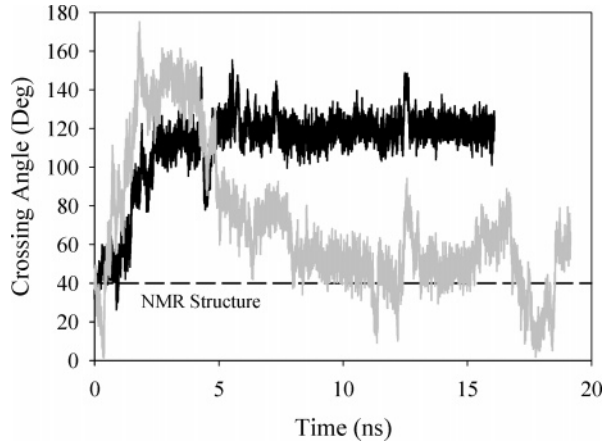


Figure 5. Time series of crossing angles for unconstrained water simulations, run1 (black) and run2 (grey). The experimental reference value[7] is indicated by the dashed line.

Despite preservation of a helical structure and dimer association, significant changes to secondary and tertiary structural elements were found in the water/GpA systems. The crossing angle between the two helices that make up the GpA dimer is a relevant parameter defining the tertiary structure of the protein. This angle is mediated by the interplay between interhelical interactions and interactions with the environment. Variations of this angle can be regarded as modifications of the net torque about the dimer crossing point. The time series of the helix crossing angle for the different simulations is plotted in Figure 5. As determined experimentally for GpA structures in micelles,[18] the GpA dimer has a crossing angle of about 40°. Analysis of the lipid simulation shows that the dimer structure differs only slightly within a DMPC bilayer, with a mean crossing angle of 44°. For the water simulations shown in Figure 5, significant deviations from these values are seen. Initially, within the first 2 ns of the simulation, both trajectories deviate rapidly and in a similar fashion from the NMR structure. At 2 ns, however, trajectories begin to diverge. One run stabilizes at a mean crossing angle of 120° for the duration of the simulation, ~12 ns. The helices in the second run, however, continue to rotate about each other, settling only slightly preferentially at approximately 50°. These water simulations were performed at the same system size as the lipid simulation for direct comparison of interaction energies. Simulation in an aqueous solution of a larger size showed similar trends, though changes occurred more slowly, implying that the smaller system dimensions do not inhibit further denaturation of GpA structure by water.

To further characterize the behavior of the GpA dimer in water, we recorded the intermonomer distance (between helical mass centers) and the length per residue for each helix. Compared to the DMPC/GpA simulation, the average center of mass distance in water increases by 1.5 Å. A decrease in the pitch per residue is also seen, to a value of 1.2 Å, compared to the α-helical 1.5 Å[57] found in DMPC. Interestingly, the pitch value that we find in water corresponds to a π-helix geometry. For a π-helix an increase in the helical radius of 0.5 Å is also expected.[57] This would result in at least a 1 Å increase in the center of mass distance between two packed helices, as we find in our simulations (also see above).

Constrained Secondary Structure. To study the interdependence of secondary and tertiary structures, we carried out simulations with the GpA backbone dihedral angles constrained by harmonic potentials in the vicinity of the NMR starting structures. The internally restricted GpA helices are otherwise free to move in space, in response to the denaturing, all-water environment. Under these conditions, we find an even more tightly packed dimer, with a center of mass between the two helices 0.7 Å less than in the DMPC bilayer. The most striking result is that the helix crossing angle remains roughly unchanged (stable at about 40°). Hence, preserving the secondary, α-helical structure resulted in conservation of the tertiary structure.

Is water ordering next to GpA also observed? We investigate this by comparison of the water dipole angle next to the dimer for the DMPC/GpA and constrained-backbone systems. Because in the pure water simulation there is no such reference frame as the bilayer normal, dipole angles are computed with respect to the helical axis for both the lipid and the water simulations. This calculation was performed within a cylinder with a 6 Å radius centered around the helix. The cylinder is extended beyond the ends of the helix into the bulk region of the system. For waters included in this cylinder, average water dipole orientations were calculated as a function of distance from the geometrical center of the helix, placed at zero (similar to the analysis in Figure 3).

Results are plotted in Figure 6. For the DMPC/GpA run, the water dipole orients away from the helix center initially. Further in, the orientation changes and the water dipole

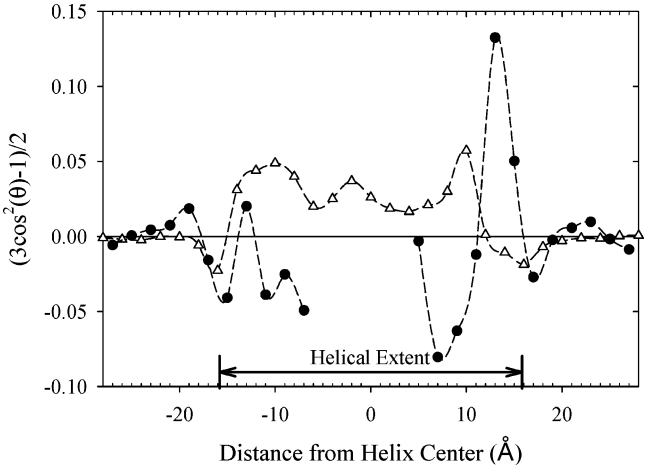


Figure 6. Water dipole order parameters with respect to the GpA monomeric axis for DMPC/GpA (filled circles) and constrained-backbone water/GpA (open triangles), averaged over both GpA helices. The helical extent is indicated at the bottom of the figure.

orients into the center. This behavior is similar to water orientation (measured by the OH bond vector) with respect to the membrane normal for a pure DMPC bilayer, Figure 3a. The bilayer strongly influences water orientation, and this is enhanced by the slight angle between the GpA helix axis and the membrane normal (<20°). From the constrained-backbone simulations, ordering of waters along the helix axis due solely to the protein (unbiased by the lipid environment) can also be observed. However, in the absence of the surrounding lipids, water ordering appears weaker and, interestingly, more symmetric than in the bilayer.

Comparison to NMR Measurements. Structural information on the membrane spanning region of GpA has come from both solution and solid-state NMR.[18,19] A set of 20 structures was determined (PDB code 1AFO) from NMR measurements in dodecylphosphocholine micelles. These structures were solved using a set of interhelical H−H NOE distance restraints along with J-coupling values sensitive to dihedral angles. Calculating the average interhelical H:H distances for the family of NMR structures and our simulations of GpA in a DMPC bilayer and GpA in water with a constrained-backbone gives us a basis for comparison to experiment. The hydrogen pairs considered correspond to those for which NOE distance restraints were assigned. For chemically equivalent hydrogens, the contribution of each hydrogen to the distance average was weighted by a factor of $1/r^6$. This is proportional to the drop off in NOE signal with respect to the distance, r. The results of these calculations can be seen in Figure 7a. There is generally good agreement between both the DMPC/GpA and constrained-backbone simulations with that of the NMR structure. However both simulations show a larger separation between hydrogens at the N-terminus and a greater mean square fluctuation at the C-terminus when compared to the NMR structures.

Further structural information has come from rotational resonance solid-state NMR of GpA in a DMPC bilayer.[19] This measurement gives a set of distance restraints between isotopically labeled carbons, shown in Figure 7b. While a

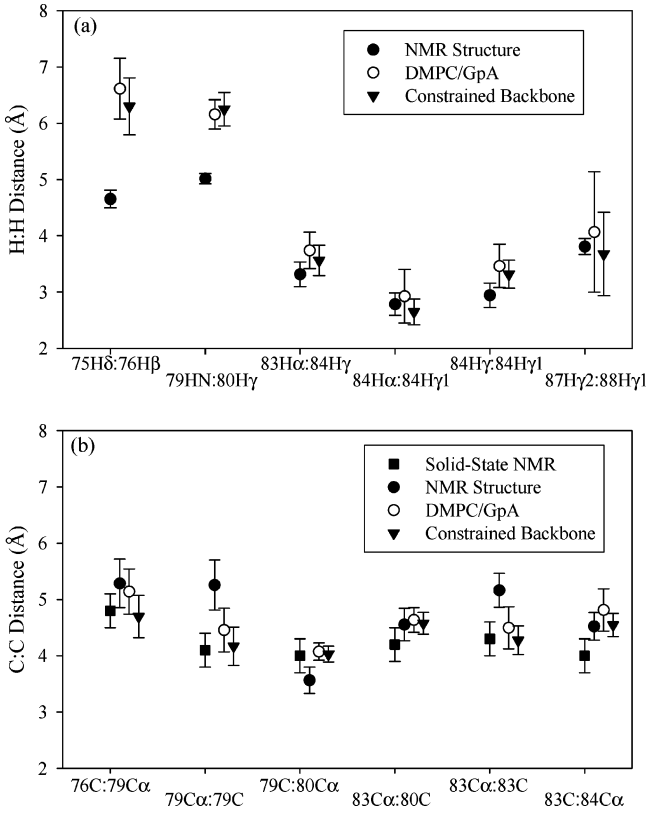


Figure 7. Comparison of simulations with NMR structural data. (a) Average interhelical H−H distances, corresponding to assigned NOE signals from solution NMR measurements.[7] Averages over 20 NMR structures (PDB code 1AFO) are shown with solid circles, DMPC/GpA simulation with open circles, and the constrained-backbone simulation with solid triangles. (b) Average interhelical C−C distances corresponding to solid-state NMR measurements (solid squares),[19] compared to the structures in part (a). Error bars represent mean square fluctuations.

formal protein structure was not solved, these average C−C distances can be compared between the various simulations and experiments. We find similar C−C distances between the bilayer simulation, the constrained-backbone water simulation, the solution NMR structure, and the solid-state NMR measurements. There appear to be subtle differences in C−C distances between the structure from the micelle and from the bilayer measurements. Both simulations analyzed here are in good agreement with the experimental data. The striking agreement between the constrained-backbone simulations and both the solution and solid-state NMR data emphasizes that the preservation of secondary structure results in preservation of tertiary packing between the two monomers.

Intermonomer Side-Chain Interactions. A detailed picture of how the environment affects interhelical interactions can be seen through residue−residue interaction energy matrices shown in Figure 8. In these plots, darker squares represent favorable or attractive residue interactions, and lighter squares represent unfavorable or repulsive residue interactions. Representative snapshots of the dimer structure for each system are also shown. The plot for the GpA simulation in DMPC and the constrained backbone simula-

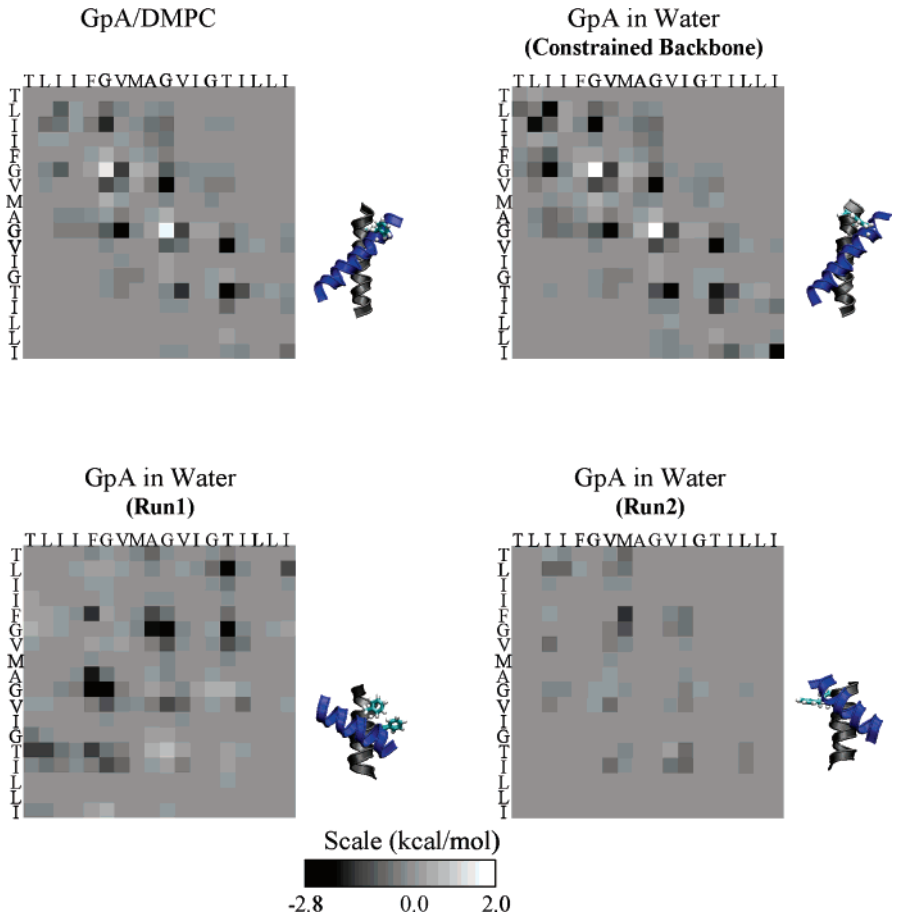


Figure 8. Average residue−residue interaction energy maps between GpA helices and representative snapshots from the simulations. Dark squares are favorable residue interactions and light squares are unfavorable residue interactions. Water/GpA Run1 is an average over the last 7 ns of the simulation. Water/GpA Run2 is an average taken from the 8−15 ns interval of the simulation.

tion in water (Figure 8, top panels) show the primary residues involved in the helix−helix interface to be consistent with the $L_{75}I_{76}xxG_{79}V_{80}xxG_{83}V_{84}xxT_{87}$ motif determined through numerous mutagenesis and structural studies.[15−17] The two glycines in the highly conserved GxxxG motif have a slightly unfavorable interresidue interaction with the corresponding glycines of the other helix. As we have shown previously for GpA in simulations of several different lipid types,[22] the repulsive Gly−Gly interaction is an electrostatic one, originating from the two closely packed backbone carbonyls and the two $C_\alpha H_2$ groups, from each helix glycine. The residue−residue interaction energy map of the constrained-backbone simulation is very similar to the bilayer simulation. Furthermore, interhelical backbone hydrogen bonding interactions between C_α-H···O═C groups of Gly79-Ile76, Val80-Gly79, and Gly83-Val80 appear to be conserved. This is consistent with the crossing angle result above, suggesting that native helix−helix packing interactions are conserved once native helical structures are preserved.

Analysis of free GpA in water (backbone restraints removed), for which π-helices are observed, presents markedly different interaction matrices (Figure 8, bottom panels). In one case, residue−residue interactions occur along the other diagonal of the matrix, indicative of an antiparallel helix packing motif. This is consistent with the 120° crossing angle reported above and shown by the representative snapshot in Figure 8. This alternate crossing angle of the GpA dimer appears to also be stabilized by a specific residue−residue interaction energy pattern. In the second case, corresponding to large fluctuations in the crossing angle (Figure 5), the interaction map resembles the upper panels, but the residue motif is less pronounced. The interactions are overall weaker, due to the fact that the center of mass distance between the helices is greater (see above). Here we show interaction energies averaged over the 8 to 15 ns time interval where the dimer crossing angle appeared most stable. Finally, the highlighted Phe groups in the snapshot serve as a dial showing that slightly different helical faces are in contact compared to the bilayer conformation

Backbone Structures. The distortion of the α-helical secondary structure observed in the water/GpA simulations implies a change in the intrahelical backbone hydrogen bonding motif. We looked at the time series of the backbone hydrogen bonding behavior with the total number of backbone hydrogen bonds divided into several categories. (1) The 3/10 helix i,i+3 motif, (2) α-helix, i,i+4 motif, (3) the π-helix, i,i+5 motif, and (4) water hydrogen bonding. The percent of all the GpA dimer backbone hydrogen bonds in each of these five categories was calculated as a function of time and plotted in Figure 9.

By design, the constrained-backbone simulation is not capable of significant secondary structural changes, so the

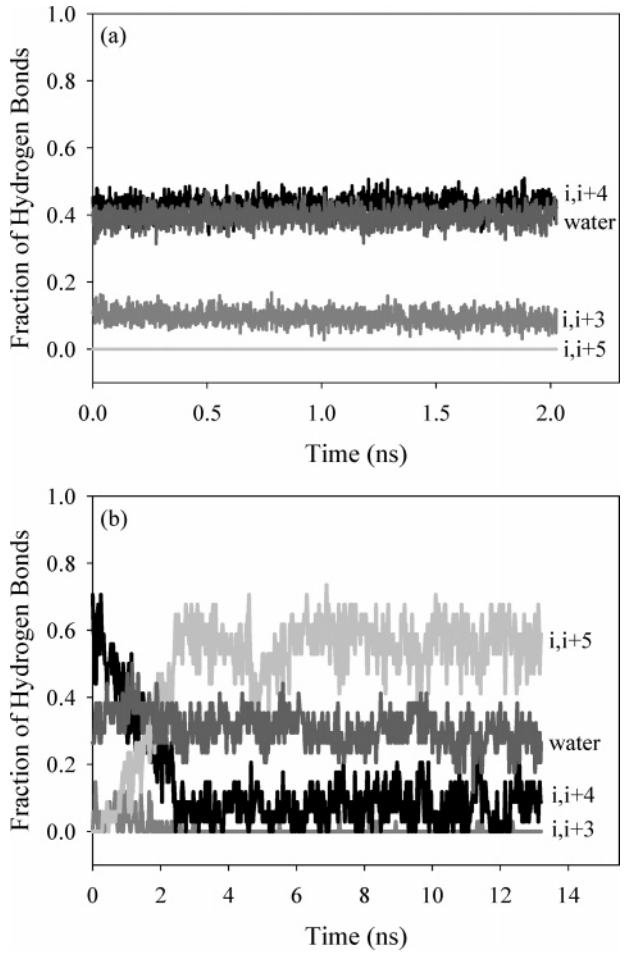


Figure 9. Time series of backbone hydrogen bonding patterns: (a) constrained backbone simulation and (b) water/GpA Run1 simulation.

time series of hydrogen bonding patterns in Figure 9a has a fairly constant behavior as given by the strength of the restraining potential. The α-helical, i,i+4, hydrogen bonding motif, and hydrogen bonding to water make up about 80% of all backbone hydrogen bonds. There is also a small degree of the 3/10 helix formation as well. Finally there is no indication of the π-helix associated hydrogen bonding pattern.

The behavior of the water/GpA Run1 is dramatically different (Figure 9b). Most noticeably, there is a rapid change within the first 3 ns, going from the i,i+4 hydrogen bonding motif to mostly i,i+5. The percentage of i,i+3 hydrogen bonding goes to zero. While fluctuations in hydrogen bonding patterns are greater for the water/GpA Run 1, the π-helix like conformation is stable throughout the majority of the trajectory. Furthermore from monitoring the time series of individual residue backbone hydrogen bonding, we find that Ala83-Gly84 and Ile86-Gly87 are the most common nucleation sites for π-helix formation (data not shown). Stabilization of a π-helix is found in both of the GpA in water simulations and represents a consistent affect of water on secondary structure.

The average backbone hydrogen bonding behavior of all simulations is given in Table 1. Averages for the free GpA dimer in water simulations were taken over the last 7 ns

Table 1. Fraction of GpA Backbone Hydrogen-Bonding Motifs Observed during Simulations[a]

system	i,i+3	i,i+4	i,i+5	water	other
constrained backbone	0.1	0.4	0	0.4	0.1
water/GpA (Run1)[b]	0	0.05	0.6	0.3	0.05
water/GpA (Run2)[b]	0	0.1	0.55	0.3	0.05
DMPC/GpA	0.05	0.7	0	0.2	0.05
DMPC/GpA (term res)	0	0.4	0	0.4	0.2

[a] Fluctuations about the mean were similar, in the range of ±0.02 to 0.05. [b] Average taken over the last 7 ns of the simulation.

Table 2. GpA Helix Interaction Energies in kcal/mol[a]

system	DMPC/ GpA	constrained backbone	water/ GpA (Run1)[b]	water/ GpA (Run2)[b]
electrostatic self-energy	−21.3	−13.7	−2.4	−3.4
environment interaction energy	−478.4 (−215.6)[c]	−490.6	−529.6	−554.3

[a] Errors about mean energy values were small and in the range of ±0.1 to 0.5 kcal/mol. [b] Average taken over last 7 ns of simulation. [c] Interaction with waters only.

where hydrogen bonding behavior appears to have stabilized. The constrained-backbone simulations show a much greater degree of hydrogen bonding with water than the simulation of DMPC/GpA. However if one just looks at the hydrogen bonding behavior of the 4 terminal residues at both the N- and C-terminus, the ratio of hydrogen bonding patterns are in better agreement. These terminal residues are significantly more exposed to water as was shown in Figures 1b and 4. Simulations of the free GpA molecule in water show similar behavior, with the i,i+5 helix conformation becoming strongly favored.

The shift in backbone structure corresponds to a change in the energetics of helix stability. This can be seen from the electrostatic self-energies of the protein helices in the different simulations (Table 2). GpA in the π-helix conformations have electrostatic self-energies nearly an order of magnitude less than in the α-helical structures in DMPC and the constrained-backbone simulations. Electrostatic self-energies are stronger in the bilayer environment. There is an inverse relationship between protein self-energy and GpA interaction with its environment, also shown in Table 2.

Discussion

Using GpA as a model system, the analysis described here begins to build a framework for the critical role of the bilayer interface in stabilizing membrane protein structure. We have shown that transmembrane domains interact strongly with the polar headgroup region of the lipid bilayer, possibly implicated in favoring a parallel helical orientation rather than antiparallel. This should not come as a surprise, based on accumulating evidence on bilayer structures, especially with regard to the spatial extent of the headgroup region (Figure 1b). The molecular distributions shown in Figure 1b are consistent with X-ray and neutron scattering as well as NMR spectroscopy.[22,58−62] While the folding of membrane proteins is primarily dictated by the hydrophobic environment at the bilayer center, the stability of its functional form can be influenced by interactions at the lipid/water interface.

We have also shown that the preservation of the secondary structure maintains the integrity of the tertiary structure even when the native lipid environment is (admittedly artificially) replaced by water. This suggests, although does not directly prove, that once helical structures are formed within the bilayer, dimer association is favored, consistent with the two-stage model of Popot and Engelman.[44] It is important to note that secondary structures were constrained internally through soft harmonic potentials on backbone dihedral angles only. These soft potentials did not prevent overall helical motions (translation or rotation), nor did they act on side-chain atoms. The strength of the constraining potential was such that distance fluctuations shown in Figure 7 were comparable in magnitude to experimental measurements in micelles and bilayers. The dependence of tertiary structure on the preservation of secondary structure is further seen when the constraints on the backbone are removed.

Full exposure of the free GpA dimer to an aqueous environment destabilizes both the native tertiary and the secondary structures. While this modification can be readily labeled "hydrophobic collapse", we see a surprising degree of order in the collapsed protein. On the time-scale of these simulations one cannot assume that the highly structured collapsed protein material is in a thermodynamically stable denatured state. However, it may be possible that the sampled structures represent meta-stable unfolding/folding intermediates. Because the dimer is driven to an alternative, structured configuration, simulations in water show how an aqueous environment disrupts the protein and indirectly suggest how the membrane environment holds it stable. The convenient structural parameters to evaluate the GpA dimer are the helix−helix crossing angle, the center of mass distance between the helices, and the length/residue of the helix. Following these indicators, we see that there are several paths to disruption of the tertiary dimer structure, but modification of secondary structure is consistent.

The phase space of the denatured state is large, and it is not surprising that the water/GpA Run1 and Run2 explore different crossing angles. The differences in the crossing angle are likely an indication that the helical association behaves more like nonspecific aggregation. This is expected because disruption of the helical secondary structure breaks the native packing surface between helices. Even when crossing angles in the water/GpA Run2 approach those of the GpA in DMPC there are no native contacts between the two helices as seen in residue−residue interactions of Figure 8.

A helix motif was clearly preserved in all simulations of the free protein in water, despite the disruption of the α-helix form. The pitch per residue of 1.2 Å that we obtain is characteristic of a π-helix conformation. The π-helix was first proposed as another theoretical construct of an energetically stable helix conformation in addition to the α-helix.[57] Rare occurrences of this helix in nature have been shown in globular proteins and proposed as transient structures in membrane proteins.[63] The π-helix structure is thought to be stabilized by an i,i+5 backbone hydrogen bonding pattern. This would allow for a more loosely wound helix than an α-helix which has an i,i+4 hydrogen bonding motif. Analysis shows an initial, rapid changing from i,i+4 α-helix motif to an i,i+5 motif on a time scale of 2−3 ns (Figure 9).

Nucleation of π-helix formation was predominantly found at the glycine residues. In globular proteins glycines are known to be helix breakers.[64] The absence of a side chain in glycine residues allows greater conformational flexibility to the protein backbone, although their affect on helicity is position dependent.[64] Despite the effect on water soluble α-helices, glycines are commonly found in membranes proteins.[4] Previous simulations of transmembrane proteins in an aqueous setting have also reported on the occurrence of π-helices. These include simulations of the alamethicin channel-forming peptide and a poly-alanine helix in water,[65,66] and transmembrane segments of bacteriorhodopsin, which were simulated using Langevin dynamics in a bulk medium of high dielectric.[67]

It is of interest to compare these results with previous experimental work on the stability of poly-Ala helices. Poly-Ala peptides do not form stable α-helices by themselves in aqueous solution. However, the introduction of polar and especially charged guest residues are able to stabilize water soluble monomer helices.[68,69] Stability is increased by the greater number of substitutions. This behavior has been attributed to the desolvation of the alanine backbone NH and CO groups due to the preferential hydration of the polar residues.[70] These hydration deprived residues form tighter intrahelical hydrogen bonds, driving α-helix formation. Parallels can be drawn to membrane proteins where the lipid hydrocarbon environment severely dehydrates the backbone of the transmembrane helix.

A framework for α-helix stabilization in membranes through compensation of electrostatic interactions and entropy is suggested. Simulations of poly-Ala in various dielectric mediums show that helix instability can be induced by raising the bulk dielectric constant of the media.[70] As mentioned previously, a direct α-helix to π-helix transition can be induced in bacteriorhodopsin helices by raising the dielectric of the bulk media.[67] A low dielectric medium would favor short, rigid hydrogen bonds of an α-helix, whereas a higher dielectric would favor longer bond lengths (such as a π-helix) in order to allow for greater conformational flexibility. The results presented here may provide a more detailed molecular picture of how an aqueous setting disrupts a transmembrane α-helix. In the polar aqueous solution, restrictions on hydrogen bonding are relaxed allowing for increased water hydrogen bonding interactions and increased backbone and side-chain flexibility. In contrast, the lipid hydrocarbon region appears to be important for suppressing fluctuations out of the native secondary structure of the protein. Namely, by increasing the energetic importance of backbone hydrogen bonding interactions, the α-helix conformation is stabilized. This then drives the association of the transmembrane helices, because the α-helix conformation presents an optimized, unique face that stabilizes interhelical association, mainly through van der Waals, specific residue−residue interactions.[18]

While the two-stage model gives an important framework for understanding transmembrane protein assembly, the role of the interface environment on protein structure, not just

helix association, remains to be addressed. As shown by the molecular distributions in Figure 1, about 1/5 of the peptide length lies within the headgroup region where it is exposed to water. Including the peptide open ends in the membrane plane, a total of 2/5 of GpA surface is exposed to water. A peptide with primarily apolar side chains still has an amphipathic quality due to its polar backbone. An α-helix, in particular, is capable of significant hydrogen bonding interactions with water because of the i,i+4 motif leaves four residues each at the N- and the C-termini with unsatisfied hydrogen bonds. Thus, water molecules on either side of the membrane can hydrogen bond to backbone amide hydrogens and to the carbonyl oxygen groups. This may explain why glycophorin interactions with water are shown to be predominantly electrostatic. Satisfying this hydrogen bond pairing is of particular importance to membrane proteins. The cost of burying an unpaired hydrogen bond into a hydrophobic region is thought to be extremely unfavorable, 1.5 kcal per hydrogen bond,[2] though this can vary with respect to the position within the bilayer. We see significant water hydrogen bonding for the first four and last four residues of each helix, even when penetration into the hydrocarbon region is required to do so. In the case of the glycophorin dimer, we find that interactions with interfacial water can easily dominate over other relevant terms.

These results also point out the need to recognize some assumptions in using the GpA transmembrane domain as a model system for understanding membrane protein assemblies in both the experimental and simulation literature. By studying only the transmembrane domain the protein is truncated at the bilayer interface, whereas the full membrane protein would extend into the aqueous environment on both sides of the bilayer. The artificial truncation of the protein results in free terminal groups with unpaired backbone hydrogen bonds and may increase interactions with water and the membrane interface. However the interfacial region is indeed very broad, and the extended (untruncated) protein would actually have an increased surface area exposed to the polar region of the membrane. In fact, charged residues are often found near the N- and C-termini of membrane spanning helix domains where they are thought to mediate hairpin turns.[71] The presence of charged residues such as Lys or Arg as well as aromatics such as Trp or Tyr would significantly influence a proteins interactions with the membrane environment in general and the polar headgroups more specifically. The GpA protein sequence studied lacks these types of residues and represents interactions with purely apolar side chain groups. However, the value of the GpA transmembrane domain as a model system has been shown through many insights gained from numerous experimental and simulation studies. Finally helices that naturally end in the membrane region may take advantage of water in the capping of terminal hydrogen bonds.

In discussing interfacial interactions, one recognizes that water in the membrane interface has very different behavior from bulk water. As shown in Figure 3, interaction with lipid headgroups induces water ordering, measurable by solid-state NMR[52,53] and Raman spectroscopy.[72] Relative to the *local* membrane surface, water order profiles alternate between oxygens mainly pointing out toward the aqueous phase and a reversal with water oxygens pointing in toward the bilayer center. On a larger scale, that allows for membrane shape fluctuation, such profiles are smoothed out, as detected by ^{2}H NMR.[50,73,74] Experimentally, water is found to be strongly perturbed at distances of about 5 Å from the bilayer surface, corresponding to 10−15 waters per lipid headgroup.[73,74] A similar number is obtained in our simulation from the order profiles. This water fraction is also seen in Figure 1; it corresponds to that part of the water density profile that drops below the plateau value. When lipid headgroups are displaced by GpA, this water hydrates peptide termini.[75]

Analogies between water ordering next to lipid membranes and at interfaces with nonpolar liquids, air, and even solid surfaces can be drawn from literature to lend some understanding to interfacial water properties.[76,77] Ordering is generated, or at least enhanced, by the fact that interfacial waters cannot satisfy all hydrogen bonding sites, as it does in isotropic bulk. On average, waters at interfaces have one unsatisfied hydrogen bond per molecule.[54,76] Compensation through stronger hydrogen bonding networks at the interfaces can occur, including bonding with other solute molecules present at the interface. With an α-helix presenting unpaired partners, water-backbone hydrogen bonding is energetically favored by both the solute (peptide) and the solvent (water).

In summary, an important role of the membrane interface in the folding of membrane proteins[2] is suggested by our simulations. Consistent with the two stage model, the polypeptide is thought to first fold into an α-helix while inserting into the membrane. Second, intermonomer association occurs once α-helical domains are stabilized by the hydrocarbon environment. Dimer formation appears to be driven by strong, specific residue−residue interactions (Figure 8). Furthermore, favorable packing interactions occur between the apolar side chains and the lipid chain.[2] The role of the lipid−water interface, however, in stabilizing or modulating membrane protein structure is not as well understood in this scheme. With regard to the dimer crossing-angle, it is striking that the native GpA dimer is parallel rather than antiparallel, as this arrangement breaks the symmetry about the center of the lipid bilayer. In particular, as we show in Figure 3, interfacial water is differentially perturbed at the C- and N-termini, hence the bilayer sides are qualitatively different. Why is the parallel configuration preferred in the bilayer setting? When the constraints imposed by the lipid bilayer are removed, the dimer crossing-angle may increase if not flip (Figure 5). The constrained backbone simulations in water suggest that strong and specific packing interactions dictate the crossing angle between helices. Furthermore, perhaps this asymmetry aspect should be viewed in its full three-dimensional setting. Viewed in the membrane plane, the GpA dimer has a antiparallel orientation, suggesting that the alteration of a lipid headgroup arrangement in the membrane plane may be worth addressing.

Acknowledgment. We thank Vikas Nanda and Alfonso Leyva for helpful discussions and critical reading of the manuscript. This research was supported in part by a grant

from the NIH (R01GM064746). J.N.S. was funded by the Whitaker Foundation for Biomedical Engineering.

Supporting Information Available: Cα root-mean-square deviation (RMSD) for DMPC/GpA simulation and water/GpA Run1 simulation (Supplementary Figure 1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Popot, J. L.; Engelman, D. M. Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* **2000**, *69*, 881−922.

(2) White, S. H.; Wimley, W. C. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319−365.

(3) Russ, W. P.; Engelman, D. M. The GxxxG motif: A framework for transmembrane helix-helix association. *J. Mol. Biol.* **2000**, *296* (3), 911−919.

(4) Senes, A.; Gerstein, M.; Engelman, D. M. Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J. Mol. Biol.* **2000**, *296* (3), 921−936.

(5) Ubarretxena-Belandia, I.; Engelman, D. M. Helical membrane proteins: diversity of functions in the context of simple architecture. *Curr. Opin. Struct. Biol.* **2001**, *11* (3), 370−376.

(6) Curran, A. R.; Engelman, D. M. Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.* **2003**, *13* (4), 412−417.

(7) Jayasinghe, S.; Hristova, K.; White, S. H. Energetics, stability, and prediction of transmembrane helices. *J. Mol. Biol.* **2001**, *312* (5), 927−934.

(8) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305* (3), 567−580.

(9) Melen, K.; Krogh, A.; von Heijne, G. Reliability measures for membrane protein topology prediction algorithms. *J. Mol. Biol.* **2003**, *327* (3), 735−744.

(10) Gouaux, E.; White, S. H. Membranes − Lipids lost, lipids regained − Editorial overview. *Curr. Opin. Struct. Biol.* **2001**, *11* (4), 393−396.

(11) Gouaux, E.; White, S. H. Membranes − Proteins and membranes − a fusion of new ideas − Editorial overview. *Curr. Opin. Struct. Biol.* **2003**, *13* (4), 401−403.

(12) Fleming, K. G.; Ackerman, A. L.; Engelman, D. M. The effect of point mutations on the free energy of transmembrane alpha-helix dimerization. *J. Mol. Biol.* **1997**, *272* (2), 266−275.

(13) Liu, W.; Crocker, E.; Siminovitch, D. J.; Smith, S. O. Role of side-chain conformational entropy in transmembrane helix dimerization of glycophorin A. *Biophys. J.* **2003**, *84* (2), 1263−1271.

(14) Fisher, L. E.; Engelman, D. M.; Sturgis, J. N. Effect of detergents on the association of the glycophorin A transmembrane helix. *Biophys. J.* **2003**, *85* (5), 3097−3105.

(15) Lemmon, M. A.; Flanagan, J. M.; Hunt, J. F.; Adair, B. D.; Bormann, B. J.; Dempsey, C. E.; Engelman, D. M. Glycophorin-a Dimerization Is Driven by Specific Interactions between Transmembrane Alpha-Helices. *J. Biol. Chem.* **1992**, *267* (11), 7683−7689.

(16) Lemmon, M. A.; Flanagan, J. M.; Treutlein, H. R.; Zhang, J.; Engelman, D. M. Sequence Specificity in the Dimerization of Transmembrane Alpha-Helices. *Biochemistry-Us* **1992**, *31* (51), 12719−12725.

(17) Treutlein, H. R.; Lemmon, M. A.; Engelman, D. M.; Brunger, A. T. The Glycophorin-a Transmembrane Domain Dimer − Sequence-Specific Propensity for a Right-Hanqded Supercoil of Helices. *Biochemistry-Us* **1992**, *31* (51), 12726−12733.

(18) MacKenzie, K. R.; Prestegard, J. H.; Engelman, D. M. A transmembrane helix dimer: Structure and implications. *Science* **1997**, *276* (5309), 131−133.

(19) Smith, S. O.; Song, D.; Shekar, S.; Groesbeek, M.; Ziliox, M.; Aimoto, S. Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. *Biochemistry-Us* **2001**, *40* (22), 6553−6558.

(20) Lemmon, M. A.; Engelman, D. M. Specificity and Promiscuity in Membrane Helix Interactions. *Febs. Lett.* **1994**, *346* (1), 17−20.

(21) Adams, P. D.; Engelman, D. M.; Brunger, A. T. Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins* **1996**, *26* (3), 257−261.

(22) Petrache, H. I.; Grossfield, A.; MacKenzie, K. R.; Engelman, D. M.; Woolf, T. B. Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: Molecular dynamics calculations. *J. Mol. Biol.* **2000**, *302* (3), 727−746.

(23) Bowie, J. U. Understanding membrane protein structure by design. *Nat. Struct. Biol.* **2000**, *7* (2), 91−94.

(24) Choma, C.; Gratkowski, H.; Lear, J. D.; DeGrado, W. F. Asparagine-mediated self-association of a model transmembrane helix. *Nat. Struct. Biol.* **2000**, *7* (2), 161−166.

(25) Zhou, F. X.; Cocco, M. J.; Russ, W. P.; Brunger, A. T.; Engelman, D. M. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat. Struct. Biol.* **2000**, *7* (2), 154−160.

(26) Zhou, F. X.; Merianos, H. J.; Brunger, A. T.; Engelman, D. M. Polar residues drive association of polyleucine transmembrane helices. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (5), 2250−2255.

(27) Brown, M. F. Modulation of Rhodopsin Function by Properties of the Membrane Bilayer. *Chem. Phys. Lipids* **1994**, *73* (1−2), 159−180.

(28) Litman, B. J.; Mitchell, D. C. A role for phospholipid polyunsaturation in modulating membrane protein function. *Lipids* **1996**, *31*, S193-S197.

(29) Bezrukov, S. M.; Rand, R. P.; Vodyanoy, I.; Parsegian, V. A. Lipid packing stress and polypeptide aggregation: alamethicin channel probed by proton titration of lipid charge. *Faraday Discuss.* **1998**, *111*, 173−183.

(30) Greathouse, D. V.; Hinton, J. F.; Kim, K. S.; Koeppe, R. E. Gramicidin-a Short-Chain Phospholipid Dispersions − Chain-Length Dependence of Gramicidin Conformation and Lipid Organization. *Biochemistry-Us* **1994**, *33* (14), 4291−4299.

(31) Koeppe, R. E.; Andersen, O. S. Engineering the gramicidin channel. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 231−258.

(32) Fisher, L. E.; Engelman, D. M.; Sturgis, J. N. Detergents modulate dimerization but not helicity, of the glycophorin A transmembrane domain. *J. Mol. Biol.* **1999**, *293* (3), 639−651.

(33) 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.

(34) Novotny, J.; Bruccoleri, R.; Karplus, M. An Analysis of Incorrectly Folded Protein Models − Implications for Structure Predictions. *J. Mol. Biol.* **1984**, *177* (4), 787−818.

(35) Bond, P. J.; Sansom, M. S. P. Membrane protein dynamics versus environment: Simulations of OmpA in a micelle and in a bilayer. *J. Mol. Biol.* **2003**, *329* (5), 1035−1053.

(36) Edholm, O.; Berger, O.; Jahnig, F. Structure and Fluctuations of Bacteriorhodopsin in the Purple Membrane − a Molecular-Dynamics Study. *J. Mol. Biol.* **1995**, *250* (1), 94−111.

(37) Crozier, P. S.; Stevens, M. J.; Forrest, L. R.; Woolf, T. B. Molecular dynamics simulation of dark-adapted rhodopsin in an explicit membrane bilayer: Coupling between local retinal and larger scale conformational change. *J. Mol. Biol.* **2003**, *333* (3), 493−514.

(38) Woolf, T. B.; Roux, B. Molecular-Dynamics Simulation of the Gramicidin Channel in a Phospholipid-Bilayer. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (24), 11631−11635.

(39) Woolf, T. B.; Roux, B. Structure, energetics, and dynamics of lipid−protein interactions: A molecular dynamics study of the gramicidin A channel in a DMPC bilayer. *Proteins* **1996**, *24* (1), 92−114.

(40) Woolf, T. B.; Roux, B. The binding site of sodium in the gramicidin A channel: Comparison of molecular dynamics with solid-state NMR data. *Biophys. J.* **1997**, *72* (5), 1930−1945.

(41) Bond, P. J.; Faraldo-Gomez, J. D.; Sansom, M. S. P. OmpA: A pore or not a pore? Simulation and modeling studies. *Biophys. J.* **2002**, *83* (2), 763−775.

(42) Domene, C.; Sansom, M. S. P. Potassium channel, ions, and water: Simulation studies based on the high-resolution X-ray structure of KcsA. *Biophys. J.* **2003**, *85* (5), 2787−2800.

(43) Braun, R.; Engelman, D. M.; Schulten K. Molecular dynamics simulations of micelle formation around dimeric glycophorin A transmembrane helices. *Biophys. J.* **2004**, *87* (2), 754−763.

(44) Popot, J. L.; Engelman, D. M. Membrane-Protein Folding and Oligomerization − the 2-Stage Model. *Biochemistry-Us* **1990**, *29* (17), 4031−4037.

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

(46) Petrache, H. I.; Tristram-Nagle, S.; Nagle, J. F. Fluid phase structure of EPC and DMPC bilayers. *Chem. Phys. Lipids* **1998**, *95* (1), 83−94.

(47) Sagui, C.; Darden, T. A. Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155−179.

(48) Feller, S. E.; Pastor, R. W. Constant surface tension simulations of lipid bilayers: The sensitivity of surface areas and compressibilities. *J. Chem. Phys.* **1999**, *111* (3), 1281−1287.

(49) Vangunsteren, W. F.; Berendsen, H. J. C. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol Phys.* **1977**, *34* (5), 1311−1327.

(50) Volke, F.; Eisenblatter, S.; Galle, J.; Klose G. Dynamic Properties of Water at Phosphatidylcholine Lipid-Bilayer Surfaces as Seen by Deuterium and Pulsed-Field Gradient Proton Nmr. *Chem. Phys. Lipids* **1994**, *70* (2), 121−131.

(51) Aman, K.; Lindahl, E.; Edholm, O.; Hakansson, P.; Westlund, P. O. Structure and dynamics of interfacial water in an L-alpha phase lipid bilayer from molecular dynamics simulations. *Biophys. J.* **2003**, *84* (1), 102−115.

(52) Halle, B.; Wennerstrom, H. Interpretation of Magnetic-Resonance Data from Water Nuclei in Heterogeneous Systems. *J. Chem. Phys.* **1981**, *75* (4), 1928−1943.

(53) Arnold, K.; Pratsch, L.; Gawrisch K. Effect of Poly(Ethylene Glycol) on Phospholipid Hydration and Polarity of the External Phase. *Biochim. Biophys. Acta* **1983**, *728* (1), 121−128.

(54) Pratt, L. R.; Pohorille, A. Hydrophobic effects and modeling of biophysical aqueous solution interfaces. *Chem. Rev.* **2002**, *102* (8), 2671−2691.

(55) Gawrisch, K.; Ruston, D.; Zimmerberg, J.; Parsegian, V. A.; Rand, R. P.; Fuller, N. Membrane Dipole Potentials, Hydration Forces, and the Ordering of Water at Membrane Surfaces. *Biophys. J.* **1992**, *61* (5), 1213−1223.

(56) Sidhu, K. S.; Goodfellow, J. M.; Turner, J. Z. Effect of molecular shape and electrostatic interactions on the water layer around polar and apolar groups in solution. *J. Chem. Phys.* **1999**, *110* (16), 7943−7950.

(57) Creighton, T. E. *Proteins: Structures and Molecular Properties*; W. H. Freeman and Company: Heidelberg, 1996; 507 p.

(58) Wiener, M. C.; King, G. I.; White, S. H. Structure of a Fluid Dioleoylphosphatidylcholine Bilayer Determined by Joint Refinement of X-ray and Neutron-Diffraction Data. 1. Scaling of Neutron Data and the Distributions of Double-Bonds and Water. *Biophys. J.* **1991**, *60* (3), 568−576.

(59) Wiener, M. C.; White, S. H. Structure of a Fluid Dioleoylphosphatidylcholine Bilayer Determined by Joint Refinement of X-ray and Neutron-Diffraction Data. 2. Distribution and Packing of Terminal Methyl-Groups. *Biophys. J.* **1992**, *61* (2), 428−433.

(60) Wiener, M. C.; White, S. H. Structure of a Fluid Dioleoylphosphatidylcholine Bilayer Determined by Joint Refinement of X-ray and Neutron-Diffraction Data. 3. Complete Structure. *Biophys. J.* **1992**, *61* (2), 434−447.

(61) Nagle, J. F.; Tristram-Nagle, S. Structure of lipid bilayers. *BBA-Rev Biomembranes* **2000**, *1469* (3), 159−195.

(62) Nagle, J. F.; Tristram-Nagle, S. Lipid bilayer structure. *Curr. Opin. Struct. Biol.* **2000**, *10* (4), 474−480.

(63) Fodje, M. N.; Al-Karadaghi, S. Occurrence, conformational features and amino acid propensities for the pi-helix. *Protein Eng.* **2002**, *15* (5), 353−358.

(64) Chakrabartty, A.; Schellman, J. A.; Baldwin, R. L. Large Differences in the Helix Propensities of Alanine and Glycine. *Nature* **1991**, *351* (6327), 586−588.

(65) Tieleman, D. P.; Sansom, M. S. P.; Berendsen, H. J. C. Alamethicin helices in a bilayer and in solution: Molecular dynamics simulations. *Biophys. J.* **1999**, *76* (1), 40−49.

(66) Tieleman, D. P.; Woolley, G. A.; Sansom, M. S. P. Alamethicin as model system for ion selectivity: Computational studies. *Biophys. J.* **2000**, *78* (1), 174a-174a.

(67) Korzhnev, D. M.; Orekhov, V. Y.; Arseniev, A. S.; Gratias, R.; Kessler, H. Mechanism of the unfolding of transmembrane alpha-helical segment (1−36)-bacteriorhodopsin studied by molecular dynamics simulations. *J. Phys. Chem. B* **1999**, *103* (33), 7036−7043.

(68) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. Unusually Stable Helix Formation in Short Alanine-Based Peptides. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (14), 5286−5290.

(69) Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. A Neutral, Water-Soluble, Alpha-Helical Peptide − the Effect of Ionic-Strength on the Helix Coil Equilibrium. *J. Am. Chem. Soc.* **1991**, *113* (13), 5102−5104.

(70) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. Physical reasons for the unusual alpha-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (24), 13075−13079.

(71) Nilsson, I.; Johnson, A. E.; von Heijne, G. How hydrophobic is alanine? *J. Biol. Chem.* **2003**, *278* (32), 29389−29393.

(72) Cheng, J. X.; Pautot, S.; Weitz, D. A.; Xie, X. S. Ordering of water molecules between phospholipid bilayers visualized by coherent anti-Stokes Raman scattering microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (17), 9826−9830.

(73) Gawrisch, K.; Richter, W.; Mops, A.; Balgavy, P.; Arnold, K.; Klose, G. The Influence of Water Concentration on the Structure of Egg-Yolk Phospholipid Water Dispersions. *Stud. Biophys.* **1985**, *108* (1), 5-16.

(74) Petrache, H. I.; Tristram-Nagle, S.; Gawrisch, K.; Harries, D.; Parsegian, V. A.; Nagle, J. F. Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophys. J.* **2004**, *86* (3), 1574−1586.

(75) Ho, C.; Stubbs, C. D. Hydration at the Membrane Protein−Lipid Interface. *Biophys. J.* **1992**, *63* (4), 897−902.

(76) Wang, H. B.; Carlson, E.; Henderson, D.; Rowley, R. L. Molecular dynamics simulation of the liquid−liquid interface for immiscible and partially miscible mixtures. *Mol. Simulat.* **2003**, *29* (12), 777−785.

(77) Tikhonov, A. M.; Schlossman, M. L. Surfactant and water ordering in triacontanol monolayers at the water-hexane interface. *J. Phys. Chem. B* **2003**, *107* (15), 3344−3347.

CT049928Y

JCTC Journal of Chemical Theory and Computation

Monte Carlo Sampling of a Markov Web

Georgios C. Boulougouris and Daan Frenkel*

FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands

Received November 3, 2004

Abstract: The efficiency of Markov-Chain Monte Carlo simulations can be enhanced by exploiting information about trial moves that would normally be rejected. The original presentation of this approach was limited to a specific MC sampling scheme. Here we present a general derivation of a method to improve the sampling efficiency of Monte Carlo simulations by collecting information about the microstates that can be linked directly to the sampled point via an independent Markov transition matrix. As an illustration, we show that our approach greatly enhances the efficiency of a scheme to compute the density of states of a square-well fluid.

Markov-Chain Monte Carlo (MCMC) methods have been used extensively for numerical evaluation of thermodynamic properties of molecular systems.[1−10] The core of any MCMC program is an algorithm that generates a Markov chain of configurations. By a judicious choice of the transition probability from one point in the chain to the next, the overall probability of visiting each microstate can be made proportional to its statistical weight ρ (e.g. the Boltzmann weight, in the case of thermal systems).

MCMC algorithms traditionally construct a Markov-Chain using two steps:[2] a. Starting from the current (old) state (o) a trial move is attempted to a new state (n) according to a trial probability α (for example, displacement, rotation or regrowth of a particle[4]). b. The trial state is then accepted or rejected according to an acceptance rule that ensures detailed balance (or, at least, the less strict balance[3]) between sampled microstates.

$$\pi_{ij} = \alpha_{ij} P_{\mathrm{acc}}(ij) \; \forall \; i \neq j, \; \pi_{ii} = 1 - \sum_j \pi_{ij} \tag{1}$$

$$\sum_j \pi_{ij} = 1, \; \forall i \tag{2}$$

Here, and in what follows, we use the term "balance" to describe algorithms that leave the equilibrium distribution invariant. That is when we apply one Monte Carlo step to an equilibrium distribution of initial states, the average flux into any given state i is exactly balanced by the total flux out of that state. As a consequence, the probability density is not changed by such the Monte Carlo algorithm. The stronger "detailed balance" condition states that, for any pair of states i and j in an equilibrium ensemble, the average flux from i to j is equal and opposite to the flux from j to i. Clearly, the latter condition can only be satisfied if the algorithm satisfies microscopic reversibility: i.e., if j can be reached from i, then there is a finite probability to carry out the reverse move from j to i.

In eq 1 π_{ij} is the ijth element of the Markovian transition matrix. For $i \neq j$, π_{ij} is given by the product of two terms: the probability of attempting a trial move from state i to state j and the probability of accepting j as the new state. The transition matrix is a stochastic matrix therefore its elements must obey eq 2.

In a Monte Carlo simulation, thermodynamic properties are evaluated as the expectation values of the corresponding instantaneous properties. When a trial move is not accepted, the instantaneous property of the old state has to be recounted in the calculated average, and no information about the rejected state is included in the computation of averages. Let us now consider MC algorithms that satisfy detailed balance. The condition of detailed balance is satisfied whenever eq 3 holds for any pair of old states (o) and new states (n).

$$\rho_o \pi_{on} = \rho_o \alpha_{on} P_{\mathrm{acc}}(on) = \rho_n \alpha_{no} P_{acc}(no) = \rho_n \pi_{no} \tag{3}$$

Equations 1−3 do not uniquely define $P_{acc}(on)$: there is therefore a relative freedom in the choice of the functional

* Corresponding author phone: +31-20-6081234; fax: +31-20-6684106; e-mail: frenkel@amolf.nl.

10.1021/ct049900m CCC: $30.25
Published on Web 03/03/2005

form. The most popular choice is the Metropolis rule.[5] An alternative (and usually less efficient) choice is the "symmetric" rule proposed by Barker.[6]

In special cases it is possible to choose the trial probability α_{ij} in such a way that all trial moves are accepted.[7−10] Usually, however, $P_{acc}(on) \leq 1$, and there is a probability $1 - P_{acc}(on)$ that the trial move will be rejected. In that case the new state is rejected, and all information about it is discarded. Recently, one of us[11] showed that information about the rejected states can be included in the calculation of equilibrium properties. However, the method discussed in ref 11 was restricted to one specific Monte Carlo scheme.[6] Below we present a more general derivation that allows the evaluation of equilibrium properties by combining importance sampling and local sampling of microstates using any combination of valid MCMC schemes. To make this explicit, we consider two sets of transition probabilities (that may or may not be the same): the first describes a conventional sampling of microstates while the second describes the local sampling of a group of microstates that are connected to the individual microstates of the first Markov chain. We use the term "*Markovian web*" to denote the set of microstates thus connected.

We assume that the Markov chain that connects the states of the Markovian web satisfies detailed balance. Then eq 3 applies for any given pair of connected microstates $\{o,n\}$. From eqs 2 and 3 the balance condition may be derived:[2]

$$\sum_m \rho_m \pi_{mn}^{web} = \sum_m \rho_n \pi_{nm}^{web} = \rho_n \sum_m \pi_{nm}^{web} = \rho_n \tag{4}$$

Let us now consider the sampling of an equilibrium property A of the system:

$$\langle A \rangle_\rho = \sum_n A_n \rho_n / \sum_n \rho_n \tag{5}$$

We now substitute ρ_n from eq 4. By changing the order of summation (integration in the continuum case) over all microstates we get

$$\frac{\sum_n A_n \rho_n}{\sum_n \rho_n} = \frac{\sum_n \sum_m A_n \rho_m \pi_{mn}^{web}}{\sum_n \sum_m \rho_m \pi_{nm}^{web}} = \frac{\sum_m \rho_m \sum_n A_n \pi_{mn}^{web}}{\sum_m \rho_m \sum_n \pi_{mn}^{web}}$$

$$= \frac{\sum_m \rho_m \sum_n A_n \pi_{mn}^{web}}{\sum_m \rho_m} \leftrightarrow \langle A \rangle_\rho = \langle \sum_n \pi_{mn}^{web} A_n \rangle_{\rho_m} \tag{6}$$

where the last step required the use of eq 2. Equation 6 is exact and is in principle sufficient to describe how the expectation value of a property can be evaluated by combining importance sampling and integration over the local states of a Markovian web; i.e. for every state $\{m\}$ sampled via importance sampling an integration is performed over all $\{n\}$ states for which $\pi_{mn}^{web} \neq 0$. Equation 6 is quite general, and it may be implemented in many ways. In the present paper, we focus on an application to Monte Carlo sampling of the equilibrium properties of a simple off-lattice system. The transition matrix of the Markov chain that is used to generate the Markovian web can be separated in two terms as in the case of MCMC discussed above (eq 1). The sum in eq 6 may be broken into two terms using eq (1):

$$\sum_n A_n \pi_{mn}^{web} = \sum_{n \neq m} A_n \pi_{mn}^{web} + A_m (1 - \sum_{n \neq m} \pi_{mn}^{web}) \tag{7}$$

Substituting (7) into (6) and using the fact that $\sum_{n \neq m} \alpha_{mn} = 1$, a new expression for the average properties is derived:

$$\langle A \rangle_\rho = \langle \sum_{n \neq m} \alpha_{mn} [A_n P_{acc}(mn) + A_m (1 - P_{acc}(mn))] \rangle_{\rho_m} \tag{8}$$

The summation in eq 8 may be performed in many ways. One possible approach is to express the sum as a weighted average over the trial probability α_{ij}.

$$\langle A \rangle_\rho = \langle \langle A_n P_{acc}(mn) + A_m (1 - P_{acc}(mn)) \rangle_{\alpha_{mn}} \rangle_{\rho_m} \tag{9}$$

Equation 9 implies that the sum required for eq 6 can be computed as a weighted average, where the weight is the trial probability α_{ij}. In practice, this means that we perform not only an importance sampling of the states n with corresponding weights ρ_m but also we generate a finite subset of the trial states n, with weights α_{mn}.

We stress that eq 6 is general, whereas eq 9 represents one of the many possible implementations of the general case.

An important aspect of the proposed method is that the transition matrix of the Markovian web is completely independent from the Markovian matrix that is used to generate the importance sampling (the importance sampling could even have been generated by a completely different method, e.g. constant-temperature Molecular Dynamics). In other words: the two transition matrices may differ in their trial probability or in their acceptance probability (e.g. use of Metropolis or the symmetric rule) or in both.

It is advantageous to consider situations where the number of trial states is very large. Such a situation arises for instance when the system under consideration can be decomposed into noninteracting subsystems. An example is a system consisting of particles with intermolecular interactions that can be truncated beyond a finite cutoff distance r_c. If we divide the system into cells with diameter $d \geq r_c$, then nonadjacent cells will not interact. This means that the probability of acceptance of a trial move in one such cell does not depend on the possible outcome of a trial move inside all other nonadjacent cells. This makes it possible to construct "parallel" trial moves that attempt to move particles in a set of N noninteracting cells. In normal MCMC, the result of such a parallel move is that the system ends up in one specific final state (out of the 2^N possible states). With the present algorithm we can do much better. By generating M trial states in each of the subsystems we can evaluate the term $\sum A_n \pi_{mn}^{web}$ by summing all $(M+1)^N$ possible combinations of trial states for the subsystems. The number of trial microstates can be very large. Hence, special care should be taken to compute averages over trial states efficiently. Since the subsystems are independent, one can calculate the

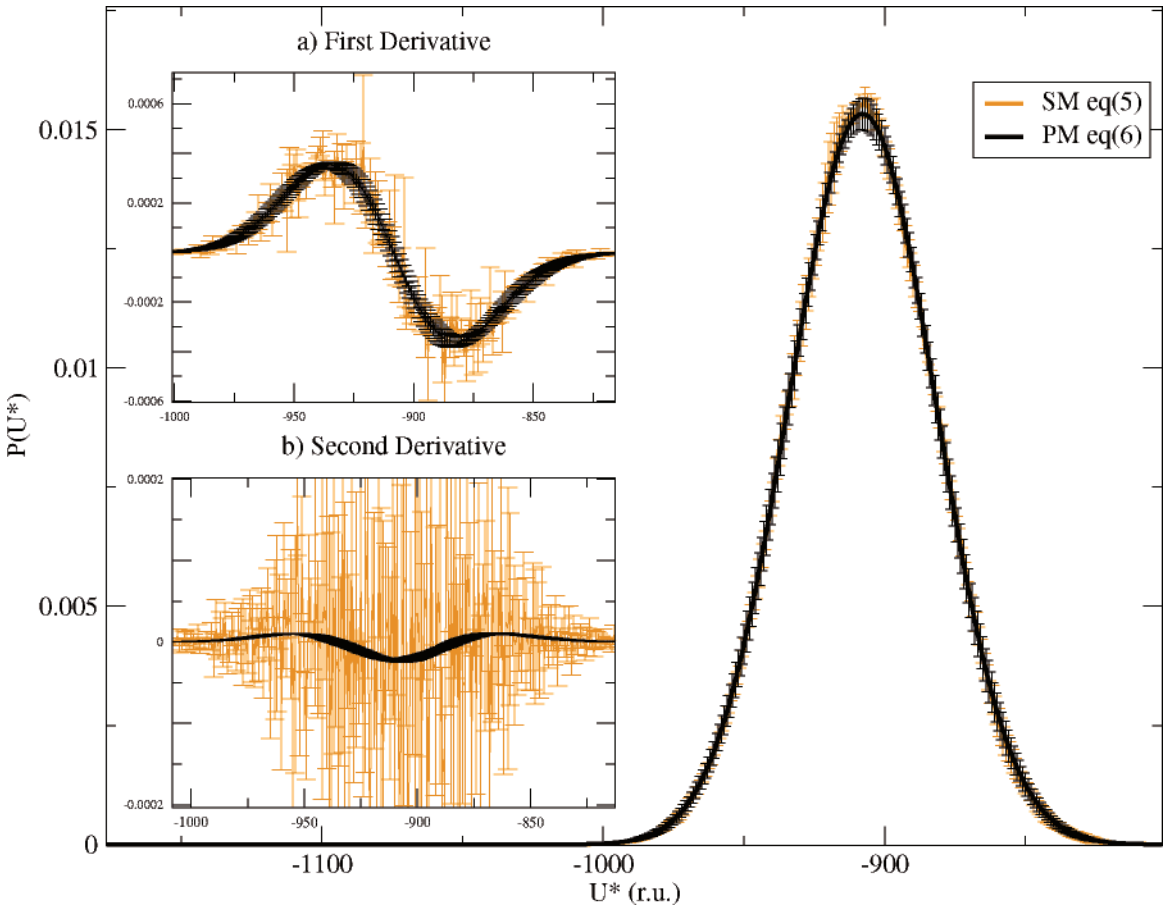


Figure 1. Main figure: Comparison of $P(U)$, the probability density of the potential energy, for a system of 2048 SW particles at $\rho^* = 0.542$ as obtained from conventional Monte Carlo sampling of serial moves (SM) using eq 5 and from parallel displacement moves of 256 particles, using eq 6 (PM). Insets: a) The first derivative of the probability density with respect to the potential energy. b) The second derivative of the probability density with respect to the potential energy. Same color coding.

probability of finding the total system in a specific macrostate (e.g. the collection of microstates with the same potential energy) as the convolution of the probabilities of the subsystems. It should be noted that the use of the convolution or even the independence between the subsystems is not a necessary requirement, but it is expected to be crucial in the case of a large number of possible trial states.

In some cases, the efficiency of the scheme may be enhanced by iterating the steps that led to the derivation of eq 6. In that case, we construct a Markov web inside a Markov web:

$$\langle A \rangle_\rho = \frac{\sum_k \rho_k A_k}{\sum_k \rho_k} = \frac{\sum_i \rho_i \sum_j \pi_{ij}^{(1)} \sum_k \pi_{jk}^{(2)} A_k}{\sum_i \rho_i} \qquad (10)$$

In our simulations, we use such a nested Markov web because we combine a set of parallel single-particle trial displacements with a set of parallel tempering moves where we attempt to swap (again in a parallel move) systems with different temperatures.[12]

To illustrate the power of the proposed approach, we have applied the above algorithm to the calculation of the energy histogram of a square well (SW) system. Although the SW potential is one of the simplest intermolecular potentials, the model exhibits glassy behavior for systems with short ranged interactions at low temperatures. This makes Monte Carlo sampling extremely difficult. Recently[13] we proposed an algorithm that was designed to overcome these sampling problems, allowing us to sample configuration space much more efficiently than using traditional methods. The method is an extension of the approach used in ref 14 and is based on the inclusion−exclusion theorem. It explicitly computes the accessible volumes in which a test particle would experience the square-well attraction of zero, one, two, etc. other particles. The technical details of this method are described in ref 13. What matters for the present discussion is that this algorithm, rather than considering a single trial displacement of a particle, computes the Boltzmann weights of a large number of trial positions. In a conventional algorithm, all trial states but one would be discarded. Here, we consider this algorithm precisely because it generates many trial states for a single particle move. In addition, as the square-well interactions that we consider are short ranged, we can perform N independent trial moves in parallel.

In Figure 1 we present a comparison between the traditional averaging, and the proposed methodology for the calculation of the energy histogram, and its numerical derivatives with respect to the energy, in a SW fluid with a

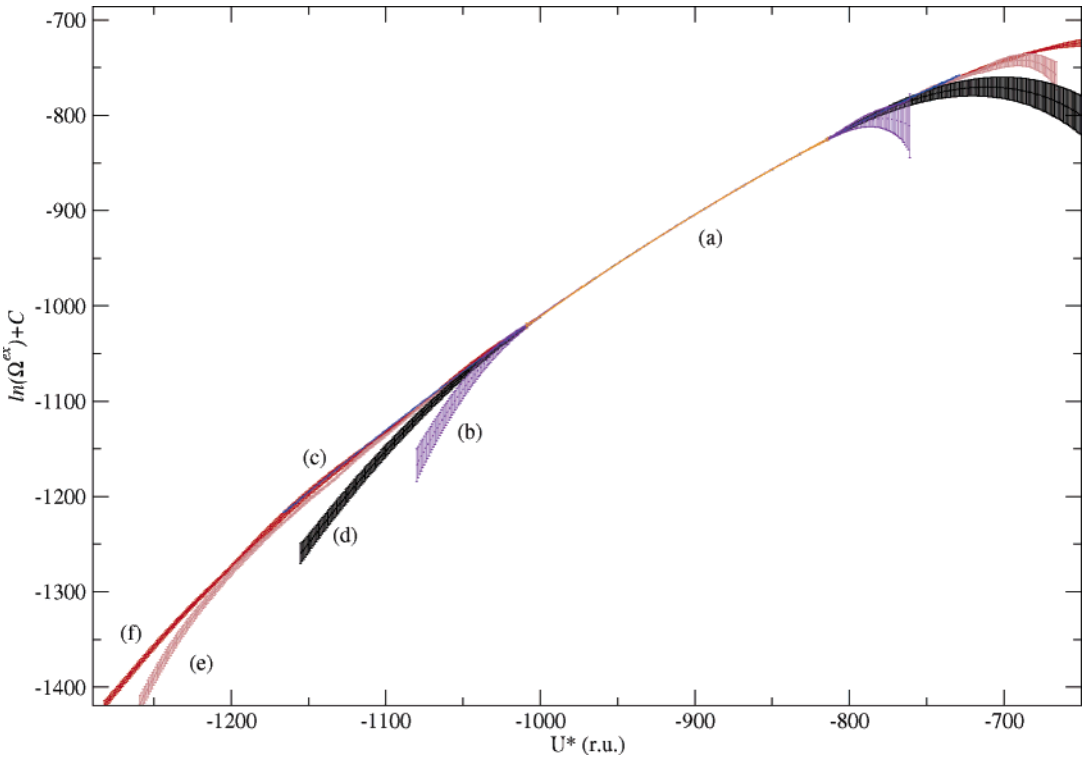


Figure 2. The logarithm of the excess density of states $\Omega^{ex}(U^*)$ for 2048 SW particles at $\rho^* = 0.542$ as obtained from a single run at $T^* = 1$: (a) (orange) averaging using eq 5, (b) (indigo) from parallel integration over 64 particle displacements using eq 6; (c) (blue) from multiple histogram reweighting of 5 systems with neighboring temperatures (see text); (d) from parallel integration over 256 particle displacements using eq 6. Curves (e) (brown) and (f) (red) were obtained by combining the parallel particle trial displacements with a set of parallel tempering moves between all 5 systems using eq 10: (e) and (f) correspond to 64 and 256 parallel particle displacements, respectively. The potential energy is expressed in units of the depth of the square well.

short-ranged attraction with a well width of $\lambda\sigma = 1.025\sigma$. Single-particle trial displacements were performed in a volume of $(0.40\sigma)^3$. The system contains 2048 SW particles at reduced density $\rho^* = 0.542$. The algorithm of ref 13 is used for the single-particle displacements. In our parallel calculations reported in Figure 1 we used 256 parallel displacements. Figure 1 shows that the use of the parallel algorithm enables us to achieve better statistics compared to the conventional algorithm, since it enables us to collect information for v^m points for every important sampled point, where v is the test volume in which we attempt the displacement and m is the number of parallel moves attempted. This is evident from the insets of Figure 1 where the numerical derivatives of the energy histogram are evaluated. These derivatives are related to thermodynamic derivatives, and the error in their evaluation is dramatically reduced by the parallel averaging scheme.

Figure 2 shows the excess density of states $\Omega^{ex}(U)$ for a system of 2048 SW particles with $\lambda = 1.025$, at reduced density $\rho^* = 0.542$. The excess density of states was computed from $P(U)$, the histogram of the potential energy of the system, using

$$\ln[\Omega^{ex}(U)] = \ln[P(U)e^{+\beta U}] + C \qquad (11)$$

For the sake of comparison, the orange curve *a* shows the results of a single run at $T^* = 1.0$, using conventional averaging (eq 5). In addition, we used the multiple-histogram method to obtain $\Omega^{ex}(U)$ over a wider energy range (curve *c*). This curve was obtained by combining the histograms obtained in 5 simulations at reduced temperatures $T^* = \{0.8,0.9,1,1.1,1.2\}$. The remaining curves were obtained at $T^* = 1.0$, using different versions of the Markovian-web algorithm: Curve *b* (yellow) was obtained using eq 6. Curves *b* (indigo) and *d* (black) are also based on eq 6, but in this case the trial moves consisted of 64 (*b*) and 256 (*d*) parallel trial displacements. The curves *e* and *f* were obtained using a "nested" web (eq 10). To obtain these curves, we combined the parallel displacement moves with parallel tempering trial moves in which we attempted all possible swaps of the temperatures of five simulations at reduced temperatures $T^* = \{0.8,0.9,1,1.1,1.2\}$.[12] Curve *e* corresponds to 64 parallel particle displacement and curve *f* to 256. In the figure, we have only included the sampling due to trial moves that originate from the system at $T^* = 1.0$. The figure shows that the present algorithm greatly improves the sampling of the low- and high-energy wings of the density of states. But, of course, at some point local sampling − no matter how good − is inadequate, and the estimated density of states starts to deviate from the more accurate curve obtained (at a higher cost) from a multiple-histogram calculation.

In this paper we argued that by using the basic properties of Markov chains it is possible to construct a sampling scheme where for each point of the phase space sampled via a given Markov-Chain or other equivalent means of importance sampling (e.g. Molecular Dynamics for the case of molecular systems), a subsampling is performed of the

points connected to the sampled point via an independent Markov web. By Markov web we mean the set of all points in the configuration space that are linked to the instantaneous configuration of the system via an independent Markovian transition matrix. By making use of the condition of detailed balance (or balance) of the transition matrix, we can sample these points according to their Boltzmann weight, even though the generation of the trial states is correlated to the old state. Using this approach, the sampling efficiency of the Monte Carlo simulation can be greatly enhanced.

If the intermolecular interactions are short ranged, it is straightforward to generate 2^n trial states by constructing parallel trial moves of n noninteracting particles. However, the potential advantage of the method is not limited to systems with short-ranged interactions. What is essential is the existence of a large number of potential final states. For instance, Liu and Luijten[10] have developed a rejection-free cluster algorithm that works for systems with long-ranged interactions. In the conventional version of this algorithm, only one particular cluster state would be sampled. We expect that the present algorithm should make it possible to sample all possible cluster states. However, we have not yet attempted to do so.

We stress that the present algorithm improves the local sampling of configuration space, but not the rate at which the systems moves through configuration space. In case that diffusion in configuration space is slow (e.g. in glassy systems), the present algorithm must be combined with existing schemes, such as parallel tempering, that help the system escape from local minima in configuration space. Interestingly, the parallel-tempering algorithm itself can also be viewed as a scheme that generates many trial states yet accepts only one.

The present algorithm differs fundamentally from existing schemes that use information about virtual moves to construct information about the density of macrostates. Starting with the Widom particle-insertion scheme to compute chemical potentials (see e.g. ref 2), many schemes have been developed that employ virtual moves to construct a density of *macrostates* (see e.g. refs 15−17). However, in the present scheme, we use virtual moves to collect information about the properties of individual *microstates*. In the example that we showed, we used the present scheme to compute a density of macrostates, and, for this specific application, existing schemes might also be used. However, the difference is that the present method allows us to compute many distributions of macrostates at the same time, because the sampling considers microstates. Furthermore the proposed scheme can be used to evaluate directly the ensemble average of any desired property, whereas most of the older methods are restricted to thermodynamic properties related to the density of macrostates.

Acknowledgment. G.B. would like to thank I. Coluzza and S. Tindemans for helpful discussions. This work is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie (FOM)", which is financially supported by the "Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)". The support by the European Community through a Marie Curie individual Fellowship with contract No. HPMF-CT-2002-02089, is gratefully acknowledged.

References

(1) Spyriouni, T.; Economou, I. G.; Theodorou, D. N. *J. Am. Chem. Soc.* **1999**, *121*, 3407−3413. Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. *Phys. Rev. E* **2002**, *66*, 056703. Harmandaris et al. *Macromolecules* **2003**, *36*, 1376−1387.

(2) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: New York, 2002. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.

(3) Manousiouthakis, V. I.; Deem, M. W. *J. Chem. Phys.* **1999**, *110*, 2753−2756.

(4) Karayiannis, N. C.; Mavrantzas, V. G.; Theodorou, D. N. *Phys. Rev. Lett.* **2002**, *88*, 105503. Zervopoulou, E.; Mavrantzas, V. G.; Theodorou, D. N. *J. Chem. Phys.* **2001**, *115*, 2860−2875.

(5) Metropolis et al. *J. Chem. Phys.* **1953**, *21*, 1087−1092.

(6) Barker, A. A. *Aust. J. Phys.* **1965**, *18*, 119−133.

(7) Bortz, A. B.; Kalos, M. H.; Lebowitz, J. L. *J. Comput. Phys.* **1975**, *17*, 10−18.

(8) Schulz, B. J.; Binder, K.; Müller, M. *Int. J. Mod. Phys. C* **2002**, *13*, 477−494.

(9) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1987**, *58*, 86−88.

(10) Liu, J.; Luijten, E. *Phys. Rev. Lett.* **2004**, *92*, 035504.

(11) Frenkel, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17571−17575.

(12) Coluzza, I.; Frenkel, D. Submitted for publication.

(13) Boulougouris, G. C.; Frenkel, D. Submitted for publication.

(14) Miller, M.; Frenkel, D. *Phys. Rev. Lett.* **2003**, *90*, 135702.

(15) Wang, F. G.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, 056101. Wang, F. G.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050. Dang, L. X.; Chang, T. M.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **2002**, *117*, 3522−3523.

(16) Fitzgerald, M.; Picard, R. R.; Silver, R. N. *Europhys. Lett.* **1999**, *46*, 282, 6269. Fitzgerald, M.; Picard, R. R.; Silver, R. N. *J. Stat. Phys.* **2000**, *98*, 321. Errington, J. R. *J. Chem. Phys.* **2003**, *118*, 22, 9915−9925. Errington, J. R. *Phys. Rev. E* **2003**, *67*, 012102.

(17) Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **2003**, *119*, 9406−9411.

CT049900M

JCTC Journal of Chemical Theory and Computation

Origins of Rotational Barriers in Hydrogen Peroxide and Hydrazine

Lingchun Song,[†] Minghong Liu,[‡] Wei Wu,[†] Qianer Zhang,[†] and Yirong Mo*[,†,‡]

Department of Chemistry, State Key Laboratory for Physical Chemistry of Solid Surfaces, Center for Theoretical Chemistry, Xiamen University, Xiamen, Fujian 361005, P. R. China, and Department of Chemistry, Western Michigan University, Kalamazoo, Michigan 49008

Received December 21, 2004

Abstract: Compared with their isoelectronic system ethane, both hydrogen peroxide and hydrazine exhibit a double well torsional energy curve where skew conformers are favored over trans conformers and cis conformers are energy-maximum states. Clearly, the involvement of the lone oxygen and nitrogen pairs, or more specifically, the enhanced stabilizing n→σ* negative hyperconjugation effect and destabilizing repulsion among lone pairs, complicates the conformational analysis. In this work, the modern ab initio valence bond (VB) method is employed to quantitatively investigate the torsional energy curves of hydrogen peroxide and hydrazine in terms of hyperconjugative stabilization, steric repulsion, and structural and electronic relaxations. It is found that if the hyperconjugation effect is completely quenched, the trans conformers will be favored, while the cis conformers are the only transition state pertaining to the torsional motion in the potential energy surfaces of H_2O_2 and N_2H_4. Although usually the steric effect includes the contributions from the electronic and geometric changes, our energy decomposition analysis shows that even the steric effect favors the skew conformers, while the electronic and geometric changes stabilize the trans conformers. Thus, we conclude that both the hyperconjugative and steric interactions lower the energy of skew conformers and eventually form low barriers from skew to trans conformers and high barriers from skew to cis conformers in both H_2O_2 and N_2H_4. Comparison between the VB and the natural bond orbital (NBO) results show similarities and discrepancies between the two methods.

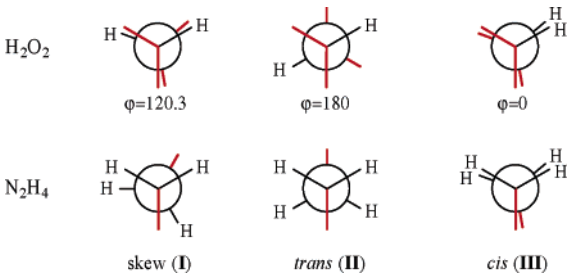
Introduction

Internal rotations about single bonds, which play an important role in conformational analyses, have been well explored experimentally and theoretically, and the barriers are often elucidated in terms of electronic and steric effects.[1] Although a stringent separation of these two often competing effects is generally difficult, physical intuitions led to the repulsion model which states that, in most if not all cases, the repulsive Pauli exchange interactions dominate the rotation barriers.[2,3] Theoretically, however, the role of hyperconjugative interaction in internal rotations has been well recognized.[4,5] For instance, Mulliken even presented theoretical details to analyze the hyperconjugation in ethane soon after the barrier was experimentally determined (a historical review on ethane rotation barrier can be found in ref 6), but he also cautioned that the hyperconjugation in ethane is only of second order and "should have little or no direct effect in restricting free rotation".[4] The magnitude of hyperconjugation between two saturated groups was unknown until a few recent works particularly in the ethane case based on the natural bond orbital (NBO) method.[7−10] These works showed that it is the attractive σ_{CH}→σ_{CH}* hyperconjugative interactions be-

* Corresponding author phone: (269)387-2916; fax: (269)387-2909; e-mail: ymo@wmich.edu.
[†] Xiamen University.
[‡] Western Michigan University.

10.1021/ct049843x CCC: $30.25
Published on Web 04/07/2005

Chart 1



tween two methyl groups that stabilizes the staggered conformer. This hyperconjugation model even shows that the Pauli exchange energy actually stabilizes the eclipsed conformer of ethane relative to the staggered conformer.[10] In other words, the eclipsed structure would be preferred if hyperconjugative interactions were quenched. In contrast to the hyperconjugation model, most recently, two groups using very different methods consistently demonstrated that the steric hindrance dominates the ethane rotation barrier, although the hyperconjugative interaction does favor the staggered conformer.[11,12] However, we note that this controversy is not ended yet.[13] Arguably, this repulsion explanation for the ethane rotation barrier was supported by a recent finely designed experiment.[14]

Considering that the lone pair and antibond orbital (n→σ*) type of hyperconjugation should be more pronounced than the σ→σ* type due to the generally higher energy level for lone pairs than σ bond electrons, we plan to use the modern ab initio valence bond (VB) method to probe the rotational barriers in hydrogen peroxide and hydrazine and examine the impact of the n→σ* negative hyperconjugation effect (or anomeric effect in the skew structures[15]) on the molecular energetics. The existence of n→σ* negative hyperconjugation as well as its influence on molecular structure and chemical reactivity have been well recognized both experimentally[16] and computationally.[17] Similarly, numerous computational studies have been conducted on hydrogen peroxide[18−21] and hydrazine.[20−22] Both computations and experiments[23,24] show that skew structures **I** (see Chart 1 where bold red lines represent lone pairs) in both systems are energy-minimum states with low barriers to trans (staggered) structures **II** but considerable barriers to cis (eclipsed) structures **III**. Notably, Smits and Altona used nonorthogonal and strictly localized molecular orbitals (MOs) and a single Slater determinant to study internal rotations for a few simple systems including H_2O_2 and N_2H_4.[25] Our goal in this paper is to employ the modern ab initio VB method to investigate the electronic and steric effects in these conformers and elucidate the driving forces for their torsional energy curves.

Computational Methods

To delineate the electronic and steric effects, it is essential to remove the delocalization tails from molecular orbitals and derive the localized bond orbitals since the electronic effect, or more specifically the hyperconjugation effect involved in the rotation about a single bond, results from the interaction between occupied bond orbitals (e.g., σ_{CH} in ethane) in one end and antibond orbitals (e.g., σ_{CH}*) in the other end of the single bond, which subsequently stabilizes the systems. The hyperconjugation energy is generally defined as the energy change owing to the mixture among the bond and antibond orbitals which leads to the delocalized orbitals.[26] However, in the MO theoretical calculations only delocalized orbitals can be obtained self-consistently. Assumptions have to be adopted to derive localized bond orbitals. If the latter are not optimal, the hyperconjugation energy might be significantly overestimated, and consequently the electronic effect will be biased against the steric effect in the analysis of the origin of conformational differences.[11,12] By using VB theory to construct both the localized (Lewis structure) and delocalized wave functions independently and self-consistently and compute the hyperconjugative stabilization explicitly in both the staggered and eclipsed structures of ethane and its congeners, we found that the steric strain dominates both rigid and fully relaxed rotations in all the group IVB ethane congeners, although the hyperconjugation effect does favor the staggered structures.[12,27]

Compared with the MO theory which interprets the electronic effect in terms of molecular orbital interactions, the VB theory is established on resonance structures, and thus a molecule is described by a set of resonance structures.[2,28] The electronic effect within the VB theory is measured by the magnitude of the contributions of certain ionic resonance structures to the most stable resonance structure (or Lewis structure as in the NBO method). Each resonance structure can be represented by a Heilter-London-Slater-Pauling (HLSP) function, which can be expanded into a linear combination of 2^m Slater determinants (m is the number of covalent bonds in the target system) or by its equivalent spin-free form such as the bonded tableau are used in our approach,[29] or in a preferable form with bond orbitals as (hereby we assume the total spin quantum number $S = 0$)

$$\Phi_k = N_k \hat{A}(\phi_{1,2}\phi_{3,4}\cdots\phi_{2m-1,2m}) \tag{1}$$

where $\hat{A}$ is the antisymmetrizer, N_k is the normalization constant, and $\phi_{i,j}$ is simply a bond function corresponding to the bond between orbitals φ_i and φ_j

$$\phi_{i,j} = \hat{A}\{\varphi_i\varphi_j[\alpha(i)\beta(j) - \beta(i)\alpha(j)]\} \tag{2}$$

where the spins of electrons (α and β) are explicitly considered and both φ_i and φ_j are expanded in the basis functions on the two bonding atoms and called bond-distorted orbitals (BDOs).[30] Eq 2 shows that a bond orbital is not only a singlet spin eigenfunction but also comprised of two Slater determinants. The wave function for the molecule is expressed as a linear combination of all VB functions

$$\Psi = \sum_K C_K \Phi_K \tag{3}$$

Despite the fact that VB concepts are much more familiar to chemists and the VB theory was proposed earlier than the MO theory, currently the ab initio VB method lags far behind the MO methods due to the extremely high computational demand in the evaluation of overlap and the

Hamiltonian matrix elements among the VB functions. In comparison, the orthogonality restraint among MOs makes the MO-based methods extremely efficient. However, it should be emphasized that the MO and VB theories are supplementary and consistent rather than conflicting and competing.[31] During the past decade, modern ab initio VB methods have been developed with a few practical codes available.[32−35] Applications have distinctively demonstrated the importance of the VB approaches in gaining new insights into molecular structures, properties, and reactivity, which provide an understanding of the results obtained from the MO computations from a different perspective. In the quest for efficient algorithms in the ab initio VB computations, we introduced a new function called paired-permanent-determinant (PPD), which is an algebrant.[34] The introduction of PPD simplifies the CPU and memory requirements and an algorithm of 2 × (*N*-2) expansion is used in the present version of our code, XMVB.[35,36]

For hydrogen peroxide and hydrazine, initially we define the VB wave functions for their respective covalent resonance (or Lewis) structures as

$$\Phi_L(H_2O_2) = N_1\hat{A}(K_{O_1}K_{O_2}\sigma_{O_1H_1}\sigma_{O_2H_2}n_{1O_1}n_{2O_1}n_{1O_2}n_{2O_2}) \quad (4a)$$

$$\Phi_L(N_2H_4) = N_1\hat{A}(K_{N_1}K_{N_2}\sigma_{N_1H_1}\sigma_{N_1H_2}\sigma_{N_2H_3}\sigma_{N_2H_4}n_{N_1}n_{N_2}) \quad (4b)$$

where K and n refer to the core orbitals and lone pairs, respectively, and each bond orbital σ_{ij} is a BDO localized on the two bonding atoms i and j only.[30] Obviously, a lone pair orbital is solely localized on one O or N atom. In principle, wave functions for ionic resonance structures can be written out similarly, and the simultaneous optimization of the orbitals and configuration coefficients results in the overall wave function Ψ as shown in eq 3. However, the most significant feature of modern ab initio VB methods is the flexibility of the one-electron orbitals that are allowed to expand and optimize either in the full space of basis functions of the system or in its subspaces.[30,33] The use of delocalized overlap-enhanced orbitals (OEOs),[37] which are expanded in the whole molecular space much like regular MOs, provides the key to the construction of VB functions of considerable accuracy and compactness. Since normally neutral resonance structures are much more stable than ionic resonance structures, using OEOs and only neutral covalent structures can recover the electron correlations overwhelmingly. E.g., for benzene Cooper et al. have demonstrated that over 93% of electron correlation can be recovered by adopting two Kékule and three Dewar structures.[38] For the current cases of hydrogen peroxide and hydrazine, ionic structures are clearly unfavorable, and high energy gaps between the stable covalent structure and the unstable ionic structures are anticipated since σ bonds usually cost much more energy to break than π bonds. Thus, we skip the ionic structures by using delocalized OEOs and define the wave function for H_2O_2 and N_2H_4 as

$$\Psi(H_2O_2) = N'\hat{A}(K_{O_1}K_{O_2}\sigma_{O_1H_1}'\sigma_{O_2H_2}'n_{1O_1}'n_{2O_1}'n_{1O_2}'n_{2O_2}') \quad (5a)$$

$$\Psi(N_2H_4) = N'\hat{A}(K_{N_1}K_{N_2}\sigma_{N_1H_1}'\sigma_{N_1H_2}'\sigma_{N_2H_3}'\sigma_{N_2H_4}'n_{N_1}'n_{N_2}') \quad (5b)$$

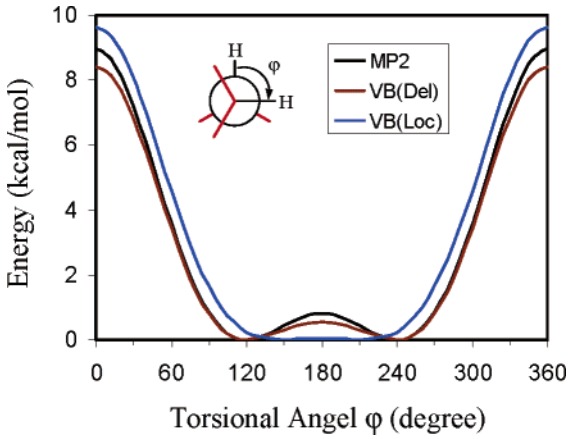


Figure 1. Comparison of energy profiles for the hydrogen peroxide rotation where the hyperconjugation effect is considered (black and brown lines) or screened out (blue line).

where the orbital σ_{ij}' contains φ_i' and φ_j' like eq 2, but the latter are expanded in the whole basis space of the molecule, rather than the subspace of two bonding atoms as φ_i and φ_j in Φ_L. Similarly, $\{n'\}$ are delocalized over the whole molecule too. Once the localized and delocalized wave functions are defined as in eqs 4 and 5, the hyperconjugative stabilization energy E_{hc} can be estimated as the energy difference between them

$$E_{hc} = E(\Psi) - E(\Phi_L) = \langle\Psi|\hat{H}|\Psi\rangle - \langle\Phi_L|\hat{H}|\Phi_L\rangle \quad (6)$$

where $\hat{H}$ is the Hamiltonian operator. It should be pointed out that both Ψ and Φ_L can be expanded into $2^2 = 4$ Slater determinants for hydrogen peroxide and $2^4 = 16$ Slater determinants for hydrazine, and in our code all orbitals in eqs 4 and 5 are optimized self-consistently.

To fully examine the hyperconjugative interactions in hydrogen peroxide and hydrazine along the rotational trajectories, geometries are fully optimized at each torsional angle at the MP2/6-311G(d,p) (6D) level, followed by the VB calculations with the same basis set. MP2 geometry optimizations as well as the primitive integrals that are required for VB calculations were carried out using Gaussian98,[39] while VB calculations were performed with our XMVB program.[35]

Results and Discussion

Torsional Potential Energy Surfaces (PESs). Figures 1 and 2 plot the energy profiles (black and brown curves) arising from internal rotations about the central O−O bond in H_2O_2 and the N−N bond in N_2H_4, respectively. Relative energies of the energy minimum and maximum conformers from the MP2 and VB calculations are compiled and compared with the experimental data in Table 1. Both systems have similar PESs and two energetically favored rotational positions, thus the internal rotations are hindered. In both molecules, skew conformers are the energy-minimum states and even more stable than the trans conformers which may have the least steric repulsion between the two OH or NH_2 moieties from the perspective of the simple repulsion model. However, the barriers from skew to trans conformers are slightly under-

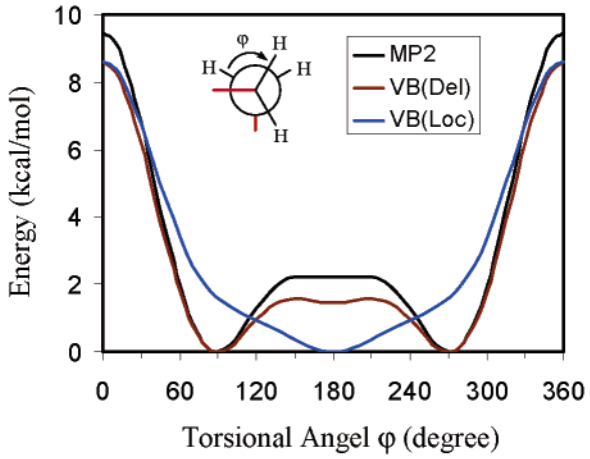


Figure 2. Comparison of energy profiles for the hydrazine rotation where the hyperconjugation effect is considered (black and brown lines) or screened out (blue line).

Table 1. Comparison of Relative Energies (Rotational Barriers) with the MP2 and VB Methods and 6-311G(d,p) Basis Set (kcal/mol)

molecule	φ (°)	conformer	ΔE_{rot}(MP2)	ΔE_{rot}(VB)	ΔE_{rot}(expt)
H_2O_2	120.1	skew (C_2)	0.0	0.0	0.0
	180.0	trans (C_{2h})	0.82	0.54	1.11[a]
	0.0	cis (C_{2v})	8.95	8.42	7.33[a]
N_2H_4	89.1	skew (C_2)	0.0	0.0	0.0
	180.0	trans (C_2)	2.21	1.46	3.14[b]
	0.0	cis (C_{2v})	9.48	8.61	

[a] Reference 40. [b] Reference 24.

estimated at both the MP2/6-311G(d,p) and VB/6-311G(d,p) levels compared with high-level computations as well as experimental data. E.g., Chung-Phillips and Jebber obtained 1.20 and 3.27 kcal/mol at the MP2/6-311+G(3df,2p) level for H_2O_2 and N_2H_4,[21] while Halpern and Glendening derived a value of 1.04 kcal/mol for H_2O_2 at the CCSD(T)/CBS level.[19] The far-IR spectrum yields a trans barrier of 1.10 kcal/mol for H_2O_2[40] and 3.14 kcal/mol for N_2H_4[24]. These results are in contrast to their isoelectronic system ethane, which favors the trans conformer as the energy-minimum state, and no stable skew conformer is found. Obviously, the discrepancy comes from the involvement of lone pairs in H_2O_2 and N_2H_4, whose exact role will be investigated in the next sections.

Compared with ethane, both H_2O_2 and N_2H_4 have a much higher rotational barrier from the energy-minimum state to the cis conformer, and our data are 8.41 and 8.60 kcal/mol which are comparable to 7.79 and 8.96 kcal/mol at the MP2/6-311+G(3df,2p) level for H_2O_2 and N_2H_4.[21] The CCSD(T)/CBS level leads to a barrier of 7.13 kcal/mol for H_2O_2[19] in agreement with the experimental value of 7.33 kcal/mol.[40]

Depicted also in Figures 1 and 2 are the torsional energy profiles (blue curves) of the Lewis structures (eq 4) where the hyperconjugative interactions are quenched. Thus, the energy profiles of Φ_L mostly correspond to effects due to the Pauli exchange and electrostatic repulsion interactions or steric effects. The structural effect which mainly concerns the change of the central bond lengths may also be involved

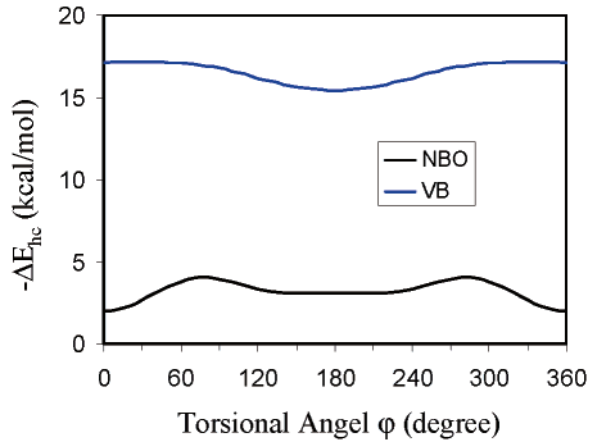


Figure 3. Comparison of the hyperconjugation effect during the hydrogen peroxide (H_2O_2) rotation estimated by the NBO method (black line) and VB method (blue line).

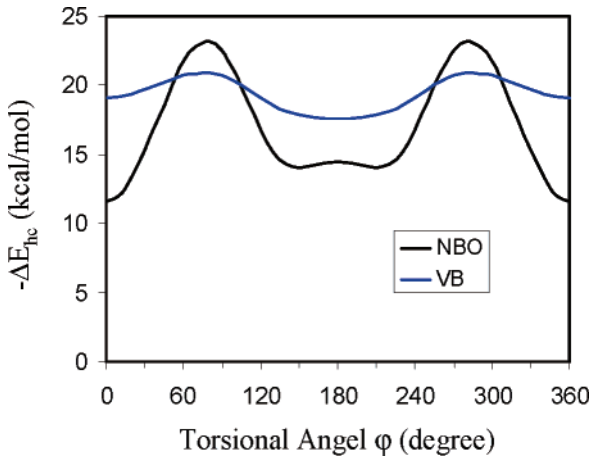


Figure 4. Comparison of the hyperconjugation effect during the hydrazine (N_2H_4) rotation estimated by the NBO method (black line) and VB method (blue line).

in the energy curves of Lewis structures and will be elaborately explored in the following. Much similar to ethane, our results demonstrate that when hyperconjugative interactions in H_2O_2 and N_2H_4 are excluded, the trans conformers are favored over the cis conformers by 9.62 and 8.62 kcal/mol, while the minimums at the skew conformers disappear in the potential energy surfaces pertaining to torsional motions. These findings once again confirmed that it is the lone pairs that complicate the overall torsional potential energy surfaces of H_2O_2 and N_2H_4.

Hyperconjugative Interaction. Since our focus in this work is the relative contributions of the hyperconjugation and steric effects to the torsional energy profiles, we performed ab initio VB calculations with the 6-311G(d,p) basis set where the one-electron orbitals are either delocalized OEOs (eq 5) or bond-localized BDOs (eq 4), and the hyperconjugation energy is derived with eq 6. For comparison hyperconjugation energies are also computed with the NBO method at the HF level. Figures 3 and 4 showed the variation of the VB and NBO hyperconjugation energies with respect to the torsional angle. The NBO delocalization energies are computed based on the deletion of off-diagonal elements between X−H bond orbitals and lone pair(s) on X

Table 2. Hyperconjugation Energies (kcal/mol) in H_2O_2 and N_2H_4 with the ab Initio VB Method and NBO Method at the MP2/6-311G(d,p) Optimal Geometries

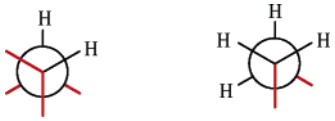
molecule	conformer	$E(\Psi)$ (au)	$E(\Phi_L)$ (au)	E_{hc} (VB)	E_{hc} (NBO)
H_2O_2	skew (C_2)	−150.90545	−150.87967	−16.18	−3.32
	trans (C_{2h})	−150.90459	−150.88000	−15.43	−3.13
	cis (C_{2v})	−150.89204	−150.86472	−17.14	−2.02
N_2H_4	skew (C_2)	−111.30849	−111.27550	−20.70	−22.57
	trans (C_2)	−111.30617	−111.27807	−17.63	−14.47
	cis (C_{2v})	−111.29478	−111.26434	−19.10	−11.55

(X = O, N) in one moiety and X−H antibond orbitals in the other moiety. A total of 6 elements for H_2O_2 and 12 elements for N_2H_4 are deleted. Table 2 listed the VB total energies and the subsequent VB hyperconjugation energies as well as the NBO hyperconjugation energies for the skew, trans, and cis conformers.

The comparison of the VB and NBO hyperconjugation energies reveals some interesting similarities and differences between the two methods. Both methods result in similar fluctuation patterns of the hyperconjugation energies with respect to the torsional angle, e.g., at the torsional angle around 80°, the hyperconjugative interactions reach the maximum, although the VB calculations result in a very flat plateau in the range of $\varphi = 0°-80°$ (and 280°−360°) for H_2O_2, while the NBO method leads to a decreasing of the hyperconjugative interaction with the reduction of the torsional angel from 80°, and the cis conformer of H_2O_2 has the weakest hyperconjugation effect. Nevertheless, both VB and NBO agree that the trans conformers of H_2O_2 and N_2H_4 have weaker hyperconjugative interaction than the skew conformers, strongly indicating that the preference of the skew conformer over the trans conformers is at least partially contributed by the favorable hyperconjugative stabilization in the formers. Although the VB and NBO hyperconjugation energies have a comparable fluctuation magnitude (around 2 kcal/mol), a huge discrepancy is observed in their absolute values (Figure 3). In the VB calculations, part of the hyperconjugation energies actually should be ascribed to the geminal bond−antibond interactions between OH or NH_2 groups and the O−O or N−N bond and antibond, which has been discussed in the previous works[12,27] and first pointed out by Reed and Weinhold.[9] These geminal interactions are essentially invariable with respect to internal rotation due to symmetry, and thus only the difference among various conformers is the focal point in our current investigation of the nature of rotational barriers. However, a similar computational strategy showed that the VB and NBO hyperconjugation energies are of the same magnitude in N_2H_4, although the NBO data now fluctuate significantly compared with the VB data. Comparably, in the isoelectronic system C_2H_6, the NBO hyperconjugation energies in the staggered and eclipsed structures are −24.88 and −18.36 kcal/mol, whereas the VB data are −13.1 and −12.1 kcal/mol. Based on the VB calculations, we found that the overall hyperconjugation effect between two groups connected by the central single bond increases in the order of $C_2H_6 < H_2O_2 < N_2H_4$. This order may suggest that the n→σ* negative hyperconjugation effect is stronger than the σ→σ* hyperconjugation effect. For the same type n→σ* or σ→σ* interactions, however, the effect weakens with the increasing of the electronegativity of the X atom in the σ_{XH} or σ_{XH}*.

Typically, for the sake of simplicity we assume that the rotational barriers are composed of steric and hyperconjugative energy terms, and the steric energy is thus simply taken as the difference between the overall rotational barriers and hyperconjugative energies which are shown in Tables 1 and 2 and plotted in Figures 1−4. On the basis of this approach, both the VB and NBO methods predict negligible energy changes around the trans conformer of H_2O_2 (φ = 120°−240°), and the cis conformer would be the energy-maximum state, if there were no hyperconjugation effect. Disagreement between the VB and the NBO methods shows up in the N_2H_4 case, where the VB method consistently predicts an energy peak at the cis conformer as already demonstrated by the energy curve of the Lewis structure in Figure 2, but the NBO method favors the skew structure as the energy-maximum state, while the cis conformer is a local minimum in the steric energy profile.

Interestingly, it is clear that the hyperconjugation effect is not solely dominated by single n→σ* negative hyperconjugative effect as otherwise the following conformers would be preferred as one lone pair is in a perfect antiparallel position (φ = 60°) with a X−H bond, as there is a better overlap between the lone pair orbital and σ_{XH}* (X = O or N).



Steric Hindrance and Structural Effect. Although in the above we assumed that the rotational barrier comes from the attractive electronic effect (hyperconjugation) and the repulsion steric effect and the steric effect can be estimated by subtracting the hyperconjugation energy from the overall rotational barrier, it is desirable to evaluate the steric energy independently. Moreover, some recent studies supporting the attractive explanation for rotation heights around single bonds also showed that the interpretation of the barrier is influenced by the slight change of the central bond. In other words, rigid rotations and relaxed rotations albeit that they have very close barriers could have very different rotation mechanisms.[8,10,41] Thus, it is illuminating to discern the electronic and geometric relaxations, particularly in the central O−O or N−N bond, in the process of rotations. In the simple treatment, these effects are usually merged into the steric effect. In the analysis on ethane and its congeners, we proposed a solution to probe the energetic variation by freezing the bond orbitals $\{\phi_{i,j}\}$ during the rotation when the hyperconjugation effect is deactivated.[12,27] As a consequence, the steric repulsion and electronic and geometric relaxations can be individually enumerated as the molecule undergoes changes from one conformer to others.

Here we adopted the above strategy to systematically explore the hyperconjugation, steric, structural relaxation, and electronic relaxation effects in the rotation. Taking the skew

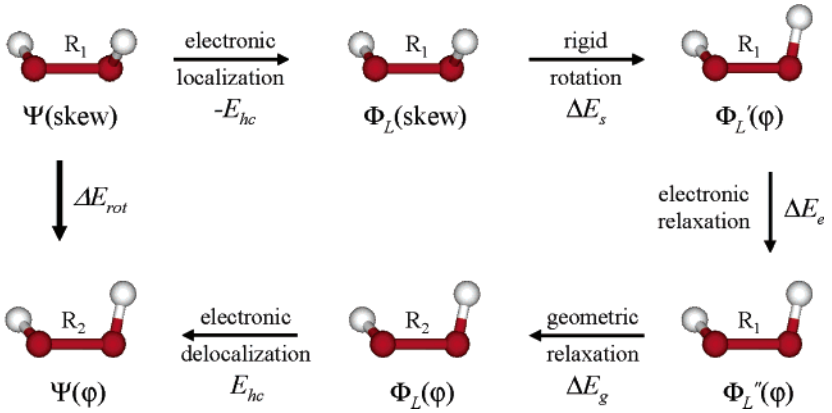


Figure 5. A stepwise decomposition scheme to explore the geometric impact on the rotation barrier in hydrogen peroxide.

conformers as references, we decompose the overall rotational process into the following five successive steps as shown in Figure 5: (1) At the optimal geometry of the skew conformer of H_2O_2 or N_2H_4, deactivate the hyperconjugation effect. The energy variation in this step is the reverse of the hyperconjugation energy defined by eq 6 in the skew conformer ($-E_{hc} = 16.18$ for H_2O_2 or 20.70 kcal/mol for N_2H_4 as listed in Table 2). (2) Freeze all orbitals and structural parameters except the dihedral angle φ, rotate one OH or NH_2 group by a certain amount, and recompute the total energy by fixing all bond orbitals without any SCF calculations. In this step, a Jacobi 2×2 matrix transformation is applied to the *p* and *d* orbitals of the rotated OH or NH_2 group. Energy changes (ΔE_s) in this step come from the steric repulsion in the rigid rotation as no other factors are involved except from the Pauli repulsion and electrostatic interactions between the two adjacent groups. (3) Relax the frozen orbitals in the above step. The reoptimization (relaxation) of the bond orbitals inherited from the skew conformer will stabilize the rotated structure by ΔE_e, and we define this step as the electronic relaxation step. (4) Relax the geometry to the optimal rotated structure at this torsional angle φ. The geometric change is dominated by the variation of the central bond and the accompanied energy reduction (ΔE_g) results from the structural effect. (5) Delocalize the electrons between the two OH or NH_2 groups which is the hyperconjugation effect in the rotated structure and the energy change is $E_{hc}(\varphi)$. After summing up all the above five energy terms, the overall energy change by transforming the structures of H_2O_2 and N_2H_4 from the skew conformer to a rotated conformer with torsional angle φ is

$$\Delta E_{rot} = -E_{hc}(\text{skew}) + \Delta E_s + \Delta E_e + \Delta E_g + E_{hc}(\varphi)$$
$$= \Delta E_{hc} + \Delta E_s + \Delta E_e + \Delta E_g \qquad (7)$$

Figures 6 and 7 show the changes of various energy terms in eq 7 with respect to the torsional angle φ and values for skew, trans, and cis conformers are collected in Table 3. On the basis of the above energy decomposition, we found that unlike the cases of ethane and its congeners,[12,27] both electronic and geometric relaxations make noticeable energy changes in the rotation, although the overall energy curve is solely dominated by the steric effect. However, since the changes from skew to trans conformers require much less energy than the skew→cis rotations, we found that actually the steric effect does not favor the trans conformers. Rather, it is the electronic and structural relaxations that stabilize the trans conformers and make the trans conformers as the energy-minimum states when the hyperconjugation effect is

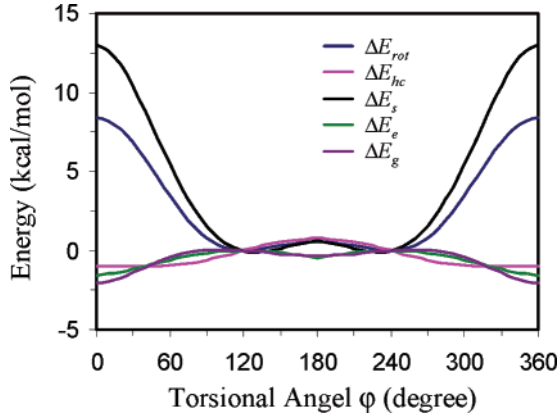


Figure 6. Decomposition of the rotational energy (ΔE_{rot}) in terms of hyperconjugation effect (ΔE_{hc}), steric repulsion (ΔE_s), electronic relaxation (ΔE_e), and geometric relaxation (ΔE_g) for hydrogen peroxide.

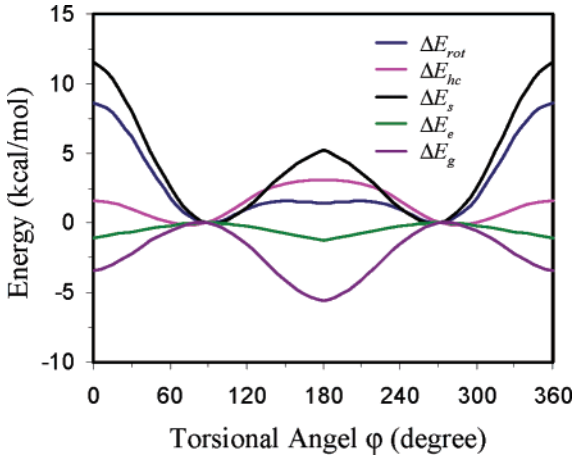


Figure 7. Decomposition of the rotational energy (ΔE_{rot}) in terms of hyperconjugation effect (ΔE_{hc}), steric repulsion (ΔE_s), electronic relaxation (ΔE_e), and geometric relaxation (ΔE_g) for hydrazine.

Table 3. Energy Decomposition of the Rotational Barriers in Terms of Hyperconjugation, Steric, Electronic Relaxation, and Geometric Relaxation Effects (kcal/mol)

	skew → cis					skew → trans				
molecule	ΔE_{rot}	ΔE_{hc}	ΔE_s	ΔE_e	ΔE_g	ΔE_{rot}	ΔE_{hc}	ΔE_s	ΔE_e	ΔE_g
H_2O_2	8.42	−0.96	13.01	−1.59	−2.04	0.54	0.75	0.61	−0.45	−0.37
N_2H_4	8.61	1.60	11.52	−1.08	−3.43	1.46	3.07	5.20	−1.25	−5.56

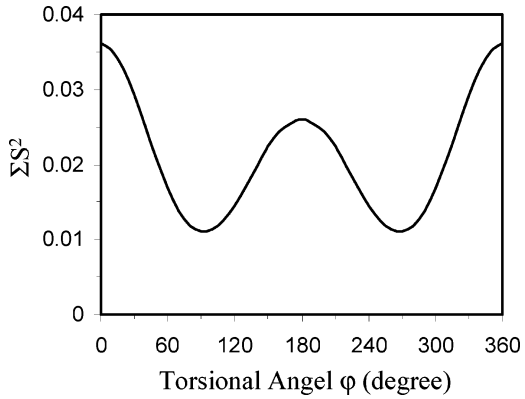


Figure 8. Fluctuation of the sum of squares of overlap integrals among the oxygen lone pairs in hydrogen peroxide with respect to the torsional angle.

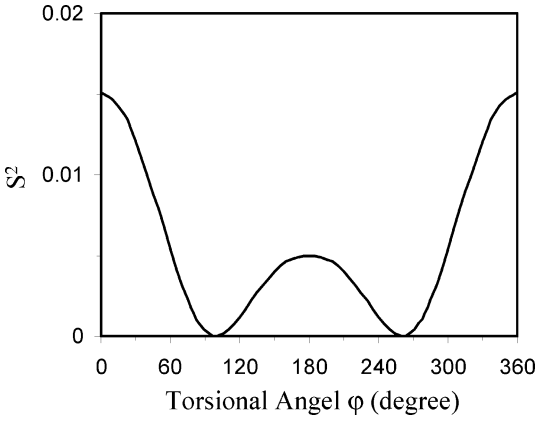


Figure 9. Fluctuation of the square of the overlap integral between the nitrogen lone pairs in hydrazine with respect to the torsional angle.

deactivated. This finding is echoed by the variation of the sum of the squares of the overlap integrals among the lone pairs in the VB wave function of Lewis structures which can roughly measure the Pauli repulsion among the lone pairs in H_2O_2 and N_2H_4 (see Figures 8 and 9).[42] Interesting, Figures 8 and 9 have similar shapes to the torsional potential energy curves, indicating the repulsion among the lone pairs dominate the torsional energies. However, we must remember that when the rotational barriers are small, the electronic and geometrical relaxation may be involved and make the simple binary theory of hyperconjugation and steric repulsion not work well.

Figures 6 and 7 also reveal that the structural changes during the rotation have a much more significant impact on the energetics of N_2H_4 than that of H_2O_2. This is inconsistent with the much larger central bond lengthening in N_2H_4 than in H_2O_2. For instance, the MP2/6-311G(d,p) optimizations show that the N−N bond lengthens by 0.043 Å and 0.037 Å from the skew to the trans and cis conformers in N_2H_4, compared with only 0.009 Å and 0.006 Å for the O−O bond in H_2O_2.

Conclusion

Ab initio VB calculations and subsequent analyses indicate that the preference of the skew conformers of hydrogen peroxide and hydrazine over both the trans and cis conformers comes from enhanced attractive hyperconjugative interactions and reduced repulsive steric interactions between the two bonding OH or NH_2 groups. Although the quenching of the hyperconjugation effect leads to the stabilization of the trans conformers as energy minimum states and the disappearance of the skew conformers in the torsional potential energy surfaces, it is the remarkable electronic and geometric relaxations rather than the steric hindrances that contribute to this interesting feature. In contrast to H_2O_2 and N_2H_4, ethane and its congeners show little electronic relaxations and small geometric relaxations.[23] Although conventionally it is approximated that rotational barriers are comprised of electronic and steric contributions, when the barriers are small or there are noticeable structural changes, it is pivotal to separate the structural relaxation contribution from the steric repulsion and estimate all energy terms individually. Our complete ab initio VB analyses demonstrate that the steric energies in H_2O_2 and N_2H_4 follow the same trends as the overall rotational barriers and show double well curves. It is also evident that the steric strain is dominated by the repulsion among the lone electron pairs on oxygen atoms in H_2O_2 and nitrogen atoms in N_2H_4, as the overlap integrals among the lone pair orbitals reveals the exact same fluctuation patterns.

Acknowledgment. The research at XMU is supported by the Natural Science Foundation of China (Grant Nos. 20225311, 20373052, and 20021002). Support from the Western Michigan University is also gratefully acknowledged (Y.M.).

References

(1) Payne, P. W.; Allen, L. C. In *Modern Theoretical Chemistry*; Schaefer H. F., III, Ed.; Plenum Press: New York and London, 1977; Vol. 4, pp 29. Wiberg, K. B. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; John Wiley & Sons: Berlin, 1998; p 2518.

(2) Pauling, L. C. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960.

(3) Sovers, O. J.; Kern, C. W.; Pitzer, R. M.; Karplus, M. *J. Chem. Phys.* **1968**, *49*, 2592. Hoyland, J. R. *J. Am. Chem. Soc.* **1968**, *90*, 2227.

(4) Mulliken, R. S. *J. Chem. Phys.* **1939**, *7*, 339.

(5) Lowe, J. *J. Am. Chem. Soc.* **1970**, *92*, 3799. England, W.; Gordon, M. S. *J. Am. Chem. Soc.* **1971**, *93*, 4649. Epiotis, N. D.; Cherry, W. R.; Shaik, S.; Yates, R. L.; Bernardi, F. *Topics in Current Chemistry: Structural Theory of Organic Chemistry*; 1977; Vol. 70.

(6) Schreiner, P. R. *Angew. Chem., Int. Ed.* **2002**, *41*, 3579.

(7) Brunck, T. K.; Weinhold, F. *J. Am. Chem. Soc.* **1979**, *101*, 1700. Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899. Goodman, L.; Gu, H. *J. Chem. Phys.* **1998**, *109*, 72. Goodman, L.; Gu, H.; Pophristic, V. *J. Chem. Phys.* **1999**, *110*, 4268. Goodman, L.; Pophristic, V.; Weinhold, F. *Acc. Chem. Res.* **1999**, *32*, 983.

(8) Corcoran, C. T.; Weinhold, F. *J. Chem. Phys.* **1980**, *72*, 2866.

(9) Reed, A. E.; Weinhold, F. *Isr. J. Chem.* **1991**, *31*, 277.

(10) Pophristic, V.; Goodman, L. *Nature* **2001**, *411*, 565.

(11) Bickelhaupt, F. M.; Baerends, E. J. *Angew. Chem., Int. Ed.* **2003**, *42*, 4183.

(12) Mo, Y.; Wu, W.; Song, L.; Lin, M.; Zhang, Q.; Gao, J. *Angew. Chem., Int. Ed.* **2004**, *43*, 1986.

(13) Weinhold, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 4188.

(14) Bohn, R. K. *J. Phys. Chem. A* **2004**, *108*, 6814.

(15) Kirby, A. J. *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*; Springer: Berlin, 1983. Deslongchamps, P. *Stereoelectronic Effects in Organic Chemistry*; Pergamon: Oxford, 1983. *The Anomeric Effect and Associated Stereoelectronic Effects*; Thatcher, G. R., Ed.; American Chemical Society: Washington, DC, 1993.

(16) Raban, M.; Kost, D. *J. Am. Chem. Soc.* **1972**, *94*, 3234. Forsyth, D. A.; Yang, J. R. *J. Am. Chem. Soc.* **1986**, *108*, 2157. Kluger, R.; Brandl, M. *J. Org. Chem.* **1986**, *51*, 3964. Rahman, M. M.; Lemal, D. M.; Dailey, W. P. *J. Am. Chem. Soc.* **1988**, *110*, 1964. King, J. F.; Rathore, R.; Guo, Z.; Li, M.; Payne, N. C. *J. Am. Chem. Soc* **2000**, *122*, 10308. King, J. F.; Li, M.; Cheng, A. Z.; Dave, V.; Payne, N. C. *Can. J. Chem.* **2003**, *81*, 638. Hetenyi, A.; Martinek, T. A.; Lazar, L.; Zalan, Z.; Fueloep, F. *J. Org. Chem.* **2003**, *68*, 5705. Hoge, B.; Thosen, C.; Herrmann, T.; Pantenburg, I. *Inorg. Chem.* **2003**, *42*, 3633. Bansal, R. K.; Gupta, N.; Singh, S.; Karaghiosoff, K.; Mayer, P.; Vogt, M. *Tetrahedron Lett.* **2004**, *45*, 7771.

(17) Reed, A. E.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1990**, *112*, 1434. Wiberg, K. B.; Rablen, P. R. *J. Am. Chem. Soc.* **1993**, *115*, 614. Schneider, W. F.; Nance, B. I.; Wallington, T. J. *J. Am. Chem. Soc.* **1995**, *117*, 478. Raabe, G.; Gais, H.-J.; Fleischhauer, J. *J. Am. Chem. Soc.* **1996**, *118*, 4622. Mo, Y.; Zhang, Y.; Gao, J. *J. Am. Chem. Soc.* **1999**, *121*, 5737. Scherer, W.; Sirsch, P.; Shorokhov, D.; McGrady, G. S.; Mason, S. A.; Gardiner, M. G. *Euro. J. Chem.* **2002**, *8*, 2324. Kormos, B. L.; Cramer, C. J. *Inorg. Chem.* **2003**, *42*, 6691. Lill, S. O. N.; Rauhut, G.; Anders, E. *Chem.: Euro. J.* **2003**, *9*, 3143.

(18) Fink, W. H.; Allen, L. C. *J. Chem. Phys.* **1967**, *46*, 2261. Davidson, R. B.; Allen, L. C. *J. Chem. Phys.* **1971**, *55*, 519. Dunning, T. H., Jr.; Winter, N. W. *J. Chem. Phys.* **1975**, *63*. Christiansen, P. A.; Palke, W. E. *J. Chem. Phys.* **1977**, *67*, 57. Howard, R. E.; Levy, M.; Shull, H.; Hagstrom, S. *J. Chem. Phys.* **1977**, *66*, 5189. Rodwell, W. R.; Carlsen, N. R.; Radom, L. *Chem. Phys.* **1978**, *31*, 177. Cremer, D. *J. Chem. Phys.* **1978**, *69*, 4440. Block, R.; Jansen, L. *J. Chem. Phys.* **1985**, *82*, 3322.

(19) Halpern, A. M.; Glendening, E. D. *J. Chem. Phys.* **2004**, *121*, 273.

(20) Pedersen, L.; Morokuma, K. *J. Chem. Phys.* **1967**, *46*, 3941. Lombardi, E.; Tarantini, G.; Pirola, L.; Torsellini, P. *J. Chem. Phys.* **1976**, *64*, 5229.

(21) Chung-Phillips, A.; Jebber, K. A. *J. Chem. Phys.* **2001**, *102*, 7080.

(22) Fink, W. H.; Pan, D. C.; Allen, L. C. *J. Chem. Phys.* **1967**, *47*, 895. Wagner, E. L. *Theor. Chim. Acta* **1971**, *23*, 115. Grana, A. M.; Mosquera, R. A. *J. Mol. Struct.* **2000**, *556*, 69.

(23) Hunt, R. H.; Leacock, R. A.; Peters, C. W.; Hecht, K. T. *J. Chem. Phys.* **1965**, *42*, 1931.

(24) Kasuya, T.; Kojima, T. *J. Phys. Soc. Jpn.* **1963**, *18*, 364.

(25) Gordon, M. S. *J. Am. Chem. Soc.* **1969**, *91*, 3122. Smits, G. F.; Altona, C. *Theor. Chim. Acta* **1986**, *67*, 461. Smits, G. F.; Krol, M. C.; Altona, C. *Mol. Phys.* **1988**, *63*, 921.

(26) Cramer, C. J. In *Encyclopedia of Computational Chemistry*; Schelyer, P. v. R., Ed.; John Wiley & Sons: Berlin, 1998; p 1294.

(27) Song, L.; Lin, Y.; Wu, W.; Zhang, Q.; Mo, Y. *J. Phys. Chem. A* **2005**, *ASAP, DOI: 10.1021/jp044700s*.

(28) Wheland, G. W. *Resonance in Organic Chemistry*; Wiley: New York, 1955.

(29) Li, X.; Zhang, Q. *Int. J. Quantum Chem.* **1989**, *36*, 599. Zhang, Q.; Li, X. *J. Mol. Struct.* **1989**, *189*, 413. Wu, W.; Mo, Y.; Cao, Z.; Zhang, Q. In *Valence Bond Theory*; Cooper, D. L., Ed.; Elsevier: Amsterdam, 2002; Vol. 10, p 143.

(30) Mo, Y.; Lin, Z.; Wu, W.; Zhang, Q. *J. Phys. Chem.* **1996**, *100*, 11569.

(31) Hoffman, R.; Shaik, S.; Hiberty, P. C. *Acc. Chem. Res.* **2003**, *36*, 750.

(32) *Valence Bond Theory*; Cooper, D. L., Ed.; Elsevier: Amsterdam, 2002. Mcweeny, R. *Int. J. Quantum Chem.* **1999**, *74*, 87. Dijkstra, F.; van Lenthe, J. H. *J. Chem. Phys.* **2000**, *113*, 2100. Thorsteinsson, T.; Cooper, D. L. *J. Math. Chem.* **1998**, *23*, 105.

(33) Cooper, D. L.; Gerratt, J.; Raimondi, M. *Chem. Rev.* **1991**, *91*, 929. Hiberty, P. C. *THEOCHEM* **1997**, *398*−*399*, 35.

(34) Wu, W.; Wu, A.; Mo, Y.; Lin, M.; Zhang, Q. *Int. J. Quantum Chem.* **1998**, *67*, 287.

(35) Wu, W.; Song, L.; Mo, Y.; Zhang, Q. XMVB - An ab initio spin-free valence bond (VB) program; Xiamen University, Xiamen, China, 2000.

(36) Song, L.; Mo, Y.; Zhang, Q.; Wu, W. *J. Comput. Chem.* **2005**, 26, 514.

(37) Coulson, C. A.; Fischer, I. *Philos. Mag.* **1949**, *40*, 386.

(38) Cooper, D. L.; Gerratt, J.; Raimondi, M. *Nature* **1986**, *323*, 699.

(39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A. J.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari,

K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; 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. A.9 ed.; Gaussian, Inc.: Pittsburgh, PA, 1998.

(40) Flaud, J. M.; Camy-Peyret, C.; Johns, J. W. C.; Carli, B. *J. Chem. Phys.* **1989**, *91*, 1504.

(41) Bader, R. F. W.; Cheeseman, J. R.; Laidig, K. E.; Wiberg, K. B.; Breneman, C. *J. Am. Chem. Soc.* **1990**, *112*, 6530. Pophristic, V.; Goodman, L.; Wu, C. T. *J. Phys. Chem. A* **2001**, *105*, 7454.

(42) The exchange integral between two doubly occupied orbitals can be approximated as $(ij|ij) = (1/4)S_{ij}^2[(ii|ii) + 2(ii|jj) + (jj|jj)]$. See: Murrell, J. N.; Harget, A. J. *Semiempirical Self-Consistent-Field Molecular Orbital Theory of Molecules*; Wiley-Interscience: London, 1972; p 147.

CT049843X

JCTC Journal of Chemical Theory and Computation

Properties of Atoms in Molecules: Caged Atoms and the Ehrenfest Force

Richard F. W. Bader* and De-Cai Fang†

Department of Chemistry, McMaster University, Hamilton, Ontario L8S 4M1, Canada

Received December 24, 2004

Abstract: This paper uses the properties of atom X enclosed within an adamantane cage, denoted by $X@C_{10}H_{16}$, as a vehicle to introduce the Ehrenfest force into the discussion of bonding, the properties being determined by the physics of an open system. This is the force acting on an atom in a molecule and determining the potential energy appearing in Slater's molecular virial theorem. The Ehrenfest force acting across the interatomic surface of a bonded pair atoms—atoms linked by a bond path—is attractive, each atom being drawn toward the other, and the associated surface virial that measures the contribution to the energy arising from the formation of the surface is stabilizing. It is the Ehrenfest force that determines the adhesive properties of surfaces. The endothermicity of formation for X = He or Ne is not a result of instabilities incurred in the interaction of X with the four methine carbons to which it is bonded, interactions that are stabilizing both in terms of the changes in the atomic energies and in the surface virials. The exothermicity for X = Be^{2+}, B^{3+}, and Al^{3+} is a consequence of the transfer of electron density from the hydrogen atoms to the carbon and X atoms, the exothermicity increasing with charge transfer despite an increase in the contained volume of X.

Introduction

Hydrocarbon complexes containing cage-centered atoms, termed endohedral atoms, have been synthesized and, as recently reviewed by Moran et al., are of both experimental and theoretical interest.[1] The smallest hydrocarbon cage complex so far synthesized was obtained by subjecting dodecahedrane $C_{20}H_{20}$ to a molecular beam of high energy He atoms.[2] The resulting cage complex, referred to as $He@C_{20}H_{20}$, is predicted to possess an energy 38 kcal/mol in excess of the isolated reactants.[3] Moran et al.,[1] in their theoretical study of endohedral hydrocarbon cage complexes, found that adamantane, $C_{10}H_{16}$, was the smallest cage capable of encapsulating a wide variety of endohedral atoms and ions, including H, He, Li^+, Be^{2+}, Mg^{2+}, and Ne. Of the $X@C_{10}H_{16}$ complexes, they found that only Be^{2+} gave an exothermic inclusion energy equal to −235.5 kcal/mol, the endothermicity of the other complexes ranging from 60 to 365 kcal/mol for X ranging from Li^+ to Ne.

The bonding in the cage complexes raises a number of interesting questions: a) What factors contribute to the stabilization of the complex? b) Do repulsive forces act within the complex, particularly between X and the carbons of the cage? c) In the endothermic complexes, is the excess energy stored in the interactions between X and the cage or among the atoms of the cage? These questions are answered unequivocally by the quantum theory of atoms in molecules, QTAIM, which describes an atom in a molecule using the physics of an open system.[4,5] This paper applies QTAIM to the adamantane complexes containing the endohedral atoms Be^{2+}, Li^+, He, Ne, and B^{3+}.

Computational Details

The virial theorem plays an important role in the physics of an open system, requiring that one employ wave functions wherein the electronic coordinates are scaled in a self-consistent manner. This is accomplished using a self-consistent virial scaling (SCVS) procedure written by Keith,[6] in conjunction with a Hartree−Fock (H−F) or extended H−F procedure. The calculations by Moran et al.[1] used B3LYP/6-31G(d) hybrid HF/DFT level of theory to account for the correlation necessary for the correct geometry predictions.

* Corresponding author phone: (905)529-7070 ext 23499; e-mail: bader@mcmaster.ca.

† Permanent address: Department of Chemistry, Beijing Normal University, Beijing 100875, China.

10.1021/ct049839l CCC: $30.25
Published on Web 04/02/2005

Table 1. Total Energies (au), Binding Energies (kcal/mol), and $-V/T$ at MP2(full)/6-311++G(2d,2p)+SCVS, Together with the Integration Errors[a]

	MP2+SCVS	$-V/T$	$E - \sum_n E(n)$	$n - \sum_n N(n)$
adam	−389.91206	2.000030	0.37	0.0004
Be^{2+}@adam	−403.89214 (−223.17)	2.000001	0.33	−0.0006
Li^{+}@adam	−397.06746 (70.20)	2.000020	0.71	0.0017
He@adam	−392.55518 (155.86)	2.000022	0.80	0.0028
Ne@adam	−518.07666 (395.31)	2.000008	0.84	−0.0032
B^{3+}@adam	−413.22802 (−825.85)	2.000008	0.91	−0.0006
Al^{3+}@adam	−630.53887 (−318.12)	2.000007	0.23	0.0023

[a] Integration error in E in kcal/mol. The total energies for atoms in au: $Be^{2+}(^1S)$, −13.62444; $Li^{+}(^1S)$, −7.26729; $He(^1S)$, −2.89150; $Ne(^1S)$, −128.79456; $B^{3+}(^1S)$, −21.99988; $Al^{3+}(^1S)$, −240.11985; $C(^3P)$, −37.77103; $H(^2S)$, −0.49993.

The present approach uses MP2(FULL) calculations, using GAUSSIAN98[7] which yield an equivalent description of the correlation, in conjunction with the SCVS procedure and employing the 6-311++G(2d,2p) basis set. The energy changes for the complex formation in the two approaches are comparable, the MP2/SCVS calculations yielding an exothermicity of formation for $Be^{2+}@C_{10}H_{16}$ 15 kcal/mol greater than the DFT result and an endothermicity of formation for $He@C_{10}H_{16}$ greater by 4 kcal/mol. The MP2 C−C and C−H bond lengths in adamantine are on the average 0.02 au shorter than the DFT values, using the smaller basis set. A B3LYP calculation using the present large basis set yields an endothermicity smaller by 2 kcal/mol and a C−C bond length greater by 0.023 au. The adamantane molecule and its complexes possess T_d symmetry, and none of the complexes exhibit imaginary frequencies.

The effect of the scaling of the electronic coordinates is to lower the total energies in amounts ranging from 0.0028 au for adamantane to 0.0036 au for $Ne@C_{10}H_{16}$. Table 1 lists the MP(full)/SCVS total energies of the separated atoms, of adamantane and of all of the complexes, together with ΔE, the energy of formation of the complexes with respect to the separated adamantane and X. Also listed are the $-V/T$ ratios together with the differences between the sum of the integrated energies and populations with the molecular values. The energy difference caused by the deviation of the calculated $-V/T$ ratios from the exact value of 2 equals 0.7 kcal/mol for the energy of Ne in $Ne@C_{10}H_{16}$, and the errors in the sums of the integrated energies and populations are acceptably small.

Definition of Molecular Structure

The quantum boundary condition for the definition of a *proper* open system is stated in terms of a surface $S(\mathbf{r})$ through which there is no flux in the gradient vector field of the density ρ, as expressed in eq 1[4]

$$\nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = 0 \quad \forall(\mathbf{r}) \in S(\mathbf{r}) \tag{1}$$

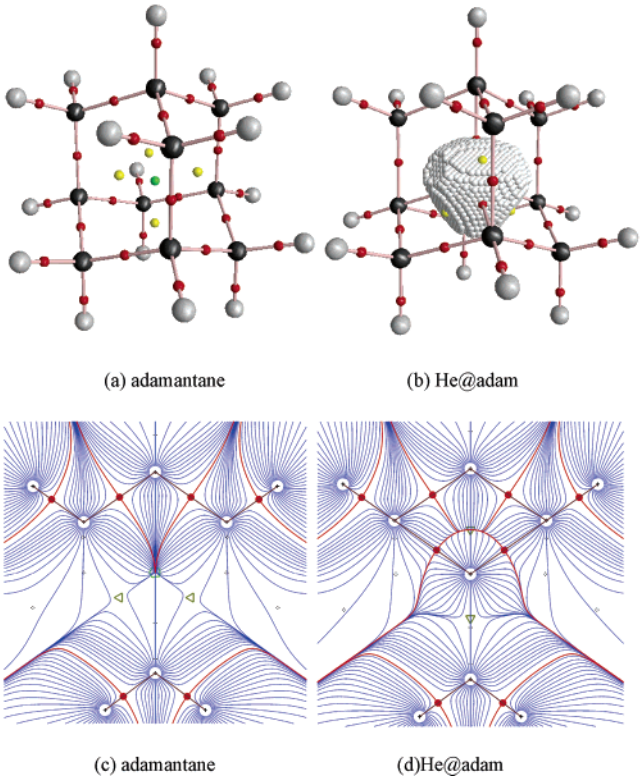


Figure 1. Molecular graphs for adamantane (LHS) and $He@C_6H_{10}$. The He atom is shown bounded by its four interatomic surfaces. CPs are denoted by dots; red for bond, yellow for ring, and green for cage. The characteristic set [n,b,r,c] for each molecule satisfies the Poincarè-Hopf rule, n − b + r − c = 1. The lower diagrams are of the gradient vector fields of the electron density in σ_d symmetry planes with the same color scheme for the CPs. Bond paths are drawn in brown, and the intersections of the interatomic surfaces with the σ_d plane are shown in red. Two X|C1 and the two C1|C2 interatomic surfaces intersect at the upper ring CP in the diagram for $He@C_6H_{10}$, the atoms involved in one of the six four-membered rings.

The same boundary condition applies to the isolated molecule, a particular case of an open system, and the same equations of motion define the properties associated with an observable. In general the atomic surface S(A) for atom A is composed of a number of interatomic surfaces S(A|B), one for each bonded atom B. An interatomic surface is defined by the gradient vectors of ρ that *terminate* at a bond critical point (CP), ρ being a maximum in the surface at the CP. Thus ρ possesses two negative curvatures at the CP. The third curvature is positive, ρ being a minimum at the CP in a direction perpendicular to the surface yielding a (3,−1) CP. Two unique trajectories thus *originate* at the CP and terminate at each of the neighboring nuclei. They define a line along which the density is a maximum with respect to any neighboring line. In an equilibrium geometry of a bound state or within the attractive region of a potential well, the line is termed a *bond path*, and atoms so linked are bonded to one another and share a common interatomic surface.[8] The gradient vector fields for a σ_d symmetry plane in adamantane and in $He@C_{10}H_{16}$ are displayed in Figure 1 to show how the atomic boundaries and molecular structure

are determined by the flow of trajectories in the associated gradient vector fields. One notes that the identification of an atom in a molecule as an open quantum system and the criterion for bonding are inseparable, both being physical consequences of the topology of the density displayed at a (3,−1) or bond CP, a ubiquitous feature of charge distributions.[9]

A *molecular graph*, the linked network of bond paths, defines a system's molecular structure, as illustrated in Figure 1 for adamantane, a cage structure composed of 26 nuclei, 28 bond CPs, 4 ring CPs, and 1 cage CP, to give the characteristic set [26,28,4,1]. Each six-membered ring consists of alternating CH and CH_2 groups in a chair conformation. Also shown is the molecular graph for $He@C_{10}H_{16}$ wherein the cage CP is replaced by the He atom giving a structure with the characteristic set [27,32,6,0]. All of the adamantane complexes possess a structure homeomorphic with that for $He@C_{10}H_{16}$, one in which X is bonded to the four methine carbon atoms resulting in the formation of six bonded rings of atoms each consisting of X linked to two CH groups that are in turn bonded to a CH_2 group. The He atom is shown enclosed by its four interatomic surfaces that impart to it a tetrahedral shape. Its atomic boundary is totally enclosed within the cage, as is true for each of the X atoms.

The density exhibits a structural homeomorphism with the virial field, the field that determines the distribution of the potential energy density in real space. Thus the physical presence of bonding signifies an accompanying energetic stabilization, as every bond path in an energetically stable structure is mirrored by a *virial path*, a line linking the same nuclei along which the electronic potential energy density is maximally stabilizing.[10] Thus coexisting with every molecular graph, is a shadow graph—the virial graph—indicating the presence of a corresponding set of lines, again defined in real space, delineating the lowering in energy associated with the formation of the structure defined by the molecular graph. For a system close to a singularity, it is possible for the geometries at which the two fields cross a structural boundary to differ. The systems under study in the present paper are removed from any such singularities.

The properties of the density and energy density at the bond CPs are given in Table 2. The values of ρ_b, the density at the bond CP, for X|C1 are on the boundary of values that are characteristic of shared and closed-shell interactions. The Laplacian, $\nabla^2\rho_b$, is positive, characteristic of a closed-shell interaction, but the energy density, H_b is negative, characteristic of a shared interaction. The energy density equals the sum of the positive definite form of kinetic energy density $G(\mathbf{r}) > 0$ and the virial field $\mathcal{V}(\mathbf{r}) < 0$ at the CP, and, unlike a closed-shell interaction, it is not dominated by the kinetic energy as happens in the approach of two closed-shell systems. Thus the indices for the X|C1 interactions are intermediate between closed-shell and shared interactions and are, in this respect, similar to those found for bonding between metal atoms or between a metal atom and a ligand, as in the transition metal carbonyls, for example.[11,12] The indices for the adamantane cage atoms are as anticipated for shared interactions. The value of ρ_b decreases for C1|C2, as does the magnitude of $\nabla^2\rho_b$, when the cage is complexed, with just the reverse occurring for the C1|H1 and C2|H2 interactions.

Table 2. Bond Critical Point Data in Atomic Units and Delocalization Indices[a]

C1|X	ρ_b	$\nabla^2\rho_b$	H_b	δ(A,B)
Be^{2+}	0.081	0.453	−0.024	0.081
Li^+	0.070	0.471	−0.004	0.077
He	0.081	0.382	−0.017	0.115
Ne	0.084	0.500	−0.009	0.168
C1|C2	ρ_b	$\nabla^2\rho_b$	H_b	δ(A,B)
adam	0.240	−0.494	−0.188	0.824
Be^{2+}	0.201	−0.286	−0.131	0.806
Li^+	0.207	−0.323	−0.137	0.796
He	0.214	−0.370	−0.147	0.780
Ne	0.169	−0.206	−0.091	0.725
C1|H1	ρ_b	$\nabla^2\rho_b$	H_b	δ(A,B)
adam	0.283	−1.030	−0.308	0.819
Be^{2+}	0.292	−1.140	−0.324	0.791
Li^+	0.294	−1.127	−0.325	0.820
He	0.284	−1.043	−0.308	0.824
Ne	0.284	−1.052	−0.307	0.833
C2|H2	ρ_b	$\nabla^2\rho_b$	H_b	δ(A,B)
adam	0.280	−1.011	−0.303	0.824
Be^{2+}	0.290	−1.098	−0.316	0.810
Li^+	0.287	−1.064	−0.312	0.825
He	0.280	−1.014	−0.303	0.831
Ne	0.281	−1.026	−0.305	0.842

[a] C1 methine C and H1 its bonded H; C2 methylene C and H2 one of its bonded H.

The delocalization index δ(A|B) measures the extent of the exchange of electrons between the basins of atoms A and B,[13] electrons that exchange being indistinguishable.[14] A value of unity implies the sharing of one Lewis pair, a value approached for the C|C and C|H interactions. Because of the delocalization of the electrons over other atoms in the molecule, the values fall short of the idealized values of unity obtained in a diatomic molecule. The indices were calculated from the natural orbitals obtained at the MP2 level using the procedure developed by Wang and Werstiuk.[15,16] The value of δ(X|C1) is greater for He and Ne than for Be and Li whose densities are contracted because of their positive charge, +1.7e on Be and +0.81e on Li. Both He and Ne possess density in excess of their closed-shell configurations, bearing charges of −0.10e and −0.11e, respectively. The same double integration of the exchange density over the basins of two atoms, when weighted by the operator $-1/r_{12}$, determines the contribution of the exchange energy to the reduction in the Coulomb repulsion between the two atomic basins.[11,17] While the exchange between the basins of the X and C1 atoms is small, the reduction in the Coulomb repulsion between C1 and X is greatest for the He and Ne complexes.

The Ehrenfest, Feynman, and Virial Theorems

The electrostatic force, the only force operative in a field-free molecule, determines the potential energy operators in

the Hamiltonian. The resulting wave function enables one to determine the average electrostatic forces that act on the electrons and on the nuclei. The force acting on the electrons, the Ehrenfest force,[18] is obtained from the equation of motion for the electronic momentum operator $\hat{\mathbf{p}}$, while the force acting on a nucleus, the Feynman force,[19] is obtained from a corresponding equation for the nuclear gradient operator. Both forces contribute to the virial of the forces acting on the electrons as described by the virial theorem, the equation of motion for the virial operator $\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}$. The molecular virial theorem relates the virial of the Ehrenfest force acting on the electrons to their kinetic energy.[20] Thus through the Ehrenfest and Feynman theorems, one has the tools that are needed to describe the forces acting in a molecule and through the virial theorem, to relate these forces to the molecule's energy and its kinetic and potential contributions.

Ehrenfest Force Acting on an Atom

The Ehrenfest force determines the force acting on an entire atom in a molecule,[4] as opposed to the Feynman force, which acts only on its nucleus. The force acting over the body of an atom in a molecule—its basin—is, as for a classical open system, equal and opposite to the pressure acting on each element of its bounding surface. Thus it may be represented in terms of the surface integral of the quantum stress tensor, a quantum analogue of pressure. The quantum stress tensor was first introduced by Schrödinger in its relativistic form in 1927[21] and its properties discussed by Pauli in 1933, as described in his book.[22] While this force is balanced by equal and opposite surface forces exerted by the neighboring atoms in a stationary state, it is shown to be an important quantity in determining the attractive or repulsive nature of the interactions between atoms in a molecule, being particularly applicable to atoms enclosed within a molecular cage.

The Ehrenfest force is given by the equation of motion—the time rate-of-change—for $\hat{\mathbf{p}} = -i\hbar\nabla$, the momentum operator for an electron, yielding the force acting on the electron density in a proper open system A, eq 2[5,23,24]

$$m\int_A d\mathbf{r}(\partial\mathbf{J}/\partial t) = \int_A d\mathbf{r}\int d\tau'\Psi^*(-\nabla_{\mathbf{r}}\hat{\mathrm{V}})\Psi + \oint dS(A,\mathbf{r}_s)\vec{\sigma}(\mathbf{r})\mathbf{n}(\mathbf{r}) \quad (2)$$

An isolated molecule presents a special limiting case of the physics of an open system, one for which the surface integral in eq 2 vanishes. $\mathbf{J}$ is the vector current density, the momentum density divided by the electronic mass m and thus it is the electronic velocity. Its time derivative gives the electronic acceleration which, when multiplied by the mass, gives the force acting on the electrons in the atom. The first term on the RHS is the expectation value of the commutator $(i/\hbar)[\hat{\mathrm{H}}, \hat{\mathbf{p}}]$ which equals the negative gradient of the full potential energy operator with respect to the coordinates of the electron at $\mathbf{r}$. The operator $-\nabla_{\mathbf{r}}\hat{\mathrm{V}}$ determines the force exerted on the electron at $\mathbf{r}$ by all of the remaining electrons and the nuclei in fixed positions. By taking the expectation value of this force in the manner denoted by $N\int d\tau'$, that is, by summing over spins and integrating over the coordinates of all electrons but the one at $\mathbf{r}$, one obtains an expression for $\vec{\mathscr{F}}(\mathbf{r})$, the force exerted on an electron at position $\mathbf{r}$ by the *average distribution* of the remaining electrons and by the rigid nuclear framework—the force exerted on the electron density,[4]

$$\vec{\mathscr{F}}(\mathbf{r}) = N\int d\tau'\Psi^*(-\nabla_r\hat{\mathrm{V}})\Psi \quad (3)$$

the same integration procedure applied to $\Psi^*\Psi$ determining the electron density $\rho(\mathbf{r})$. The force density $\vec{\mathscr{F}}(\mathbf{r})$, even though it involves two-electron interactions, is expressible as a density in real space, as found for all properties of a proper open system. Integration of $\vec{\mathscr{F}}(\mathbf{r})$ over atom A, as denoted by $\int_A d\mathbf{r}$ in eq 4, yields $\vec{\mathscr{F}}(\mathrm{A})$, the Ehrenfest force acting on the electron density over the basin of atom A. Eq 4 also indicates an alternative notation for the definition of an open system property.

$$\vec{\mathscr{F}}(A) = \int_A d\mathbf{r}\vec{\mathscr{F}}(\mathbf{r}) = 1/2\{(i/\hbar)\langle\psi,[\hat{\mathrm{H}}, \hat{\mathbf{p}}]\psi\rangle_{\mathrm{A}} + cc\} \quad (4)$$

The physics of an open system is distinguished by the presence of a surface contribution to the time rate-of-change of an observable $\hat{\mathcal{O}}$, one that corresponds to the flux in the current density of $\hat{\mathcal{O}}$ through the system's surface. Thus the final term on the RHS of eq 2 is the force exerted on the surface of the atom by the 'momentum flux density' that is expressed in terms of the quantum stress tensor, eq 5

$$\vec{\sigma}(\mathbf{r}) = (\hbar^2/4m)\{(\nabla\nabla + \nabla'\nabla') - (\nabla\nabla' + \nabla'\nabla)\}\Gamma^{(1)}(\mathbf{r},\mathbf{r}')|_{\mathbf{r}=\mathbf{r}'} \quad (5)$$

where $\Gamma^{(1)}(\mathbf{r},\mathbf{r}')$ is the one-electron density matrix. The stress tensor has the dimensions of pressure, force/unit area (F/L^2), or energy density (E/L^3). The equation of motion for an open system given in eq 2 applies to any proper open system, including as a special limiting case, the total closed system for which the surface integral vanishes. In this case one obtains Ehrenfest's second relation, that $d\langle\hat{\mathbf{p}}\rangle/dt = \langle-\nabla\hat{\mathrm{V}}\rangle = \langle\vec{\mathscr{F}}(\mathbf{r})\rangle$, and it is for this reason that $\vec{\mathscr{F}}(\mathrm{A})$ is identified as the Ehrenfest force acting on atom A. This force vanishes for a stationary state, and consequently, it has not played a role in discussions of the quantum mechanics of a total system. However, it is the dominant force in determining the mechanics of an open system, the physics of an atom in a molecule.

For a molecule in a stationary state, eq 2 reduces to

$$\vec{\mathscr{F}}(\mathrm{A}) = -\oint dS(\mathrm{A};\mathbf{r}_s)\vec{\sigma}(\mathbf{r})\mathbf{n}_{\mathrm{A}}(\mathbf{r}) = -\sum_{\mathrm{B}\neq\mathrm{A}}\oint dS(\mathrm{A}|\mathrm{B};\mathbf{r}_s)\vec{\sigma}(\mathbf{r})\cdot\mathbf{n}_{\mathrm{A}}(\mathbf{r}) = -\sum_{\mathrm{B}\neq\mathrm{A}}\vec{\mathscr{F}}(\mathrm{A}|\mathrm{B}) \quad (6)$$

and the force acting over the basin of the atom is equal and opposite to the force exerted on its surface $S(\mathrm{A})$. In general the atomic surface $S(\mathrm{A})$ in eq 6 is composed of a number of interatomic surfaces and the force on A may be correspondingly expressed as a sum of contributions $\vec{\mathscr{F}}(\mathrm{A}|\mathrm{B})$, one from each interatomic surface $S(\mathrm{A}|\mathrm{B})$ and where necessarily $\vec{\mathscr{F}}(\mathrm{A}|\mathrm{B}) = -\vec{\mathscr{F}}(\mathrm{B}|\mathrm{A})$.

In a diatomic AB, the principal contributions to $\vec{\mathscr{F}}(\mathrm{A})$, the force on the density of atom A, are from the nucleus and the electrons of B, the nucleus of B exerting an attractive force and the electrons on B a repulsive one. The contribution to the Ehrenfest force exerted on the density of A by its

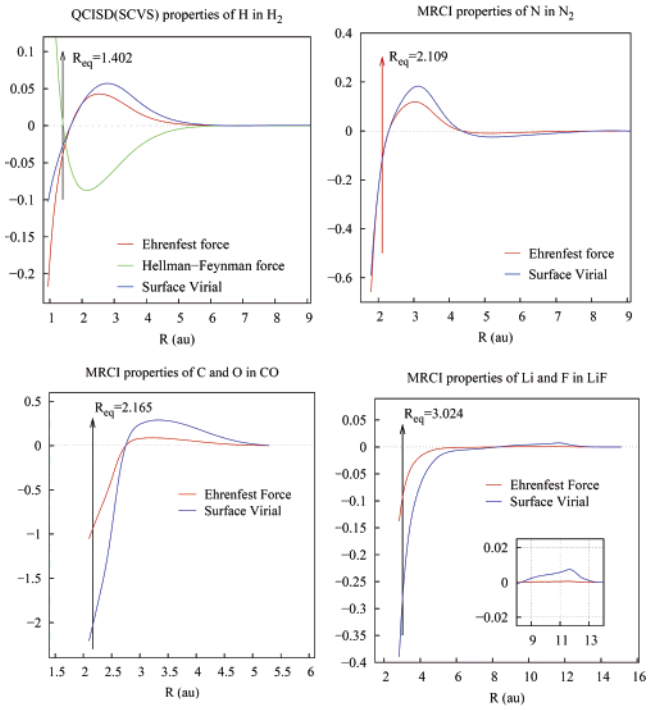


Figure 2. Variation of the Ehrenfest force $\vec{\mathcal{F}}$(A), eq 4, and the surface virial $\mathcal{V}_s$(A||B), eq 11, with internuclear separation R for H_2, N_2, CO, and LiF. QCISD denotes a wave function obtained from quadratic CI, including all singles and doubles, while MRCI denotes a multireference CI. All quantities are in atomic units. The equilibrium separation R_{eq} is indicated in each diagram. The Feynman force on a proton is also shown in the diagram for H_2. Note that $\vec{\mathcal{F}}(H)$ is maximally repulsive where the Feynman force is maximally attractive. The common zeroes in $\vec{\mathcal{F}}$(A) and $\mathcal{V}_s$(A||B) determine the ranges in R over which both quantities are repulsive and destabilizing or attractive and stabilizing, the latter occurring for separations greater than R_{eq}. Stabilization in the ionic system LiF occurs at a larger value of $R \sim 8$ au than for the shared or polar interactions. The transfer of charge from Li to F is essentially complete for $R < 10$ au.

own nucleus may be attractive or repulsive, that is, drawing it toward or away from B, depending upon the dipolar polarization of the density on A, and similarly for the intra-atomic electronic repulsion. The Ehrenfest force thus includes the electrostatic force exerted by the nuclei on the electron density, the attractive contribution to the Feynman force on the nuclei but not the contribution from the nuclear repulsive force, which in the case of $\vec{\mathcal{F}}$(A) is replaced by the repulsions between the electrons. The behavior of $\vec{\mathcal{F}}$(A) as a function of internuclear separation is illustrated in Figure 2 for H_2, N_2, CO, and LiF determined from highly correlated wave functions.[25] Table 3 gives the magnitudes of the Ehrenfest force acting on one of the atoms at R_e, which is attractive in each case, that is, the force on atom A is directed at atom B. The large value operative in CO is unique to polar, as opposed to shared or ionic bonding.

In general, for large separations in the approach of neutral atoms, $\vec{\mathcal{F}}$(A) > 0 and is repulsive, although it is slightly attractive for $R > 4.5$ au in N_2. The curves exhibit a maximum value—a maximum in the repulsion between the

Table 3. Ehrenfest Forces, Atomic Volumes, and Radii in Au[a]

AB	H_2	N_2	CO	LiF
$\lvert\vec{\mathcal{F}}(A)\rvert$	0.0395	0.0701	0.9788	0.0992
$X@C_{10}H_{16}$	Be^{2+}	Li^+	He	Ne
$\lvert\vec{\mathcal{F}}(X)\rvert$	0.1160	0.1099	0.1488	0.2171
$v(X)$	7.23	10.22	15.14	32.09
ΔR(C|C)	0.159	0.145	0.118	0.370
$r_b(X)$	1.090	1.225	1.358	1.730
r_b(C1)	2.002	1.862	1.685	1.618

[a] R(C|C) = 2.891 au in adamantane.

two atomic basins—in the region of the inflection point in the potential energy curve, the point where the Feynman force is maximally attractive. For separations slightly greater than R_e, $\vec{\mathcal{F}}$(A) undergoes a change in sign and a rapid decrease in value, and each atom experiences a force drawing it toward its neighbor. The increasingly attractive nature of $\vec{\mathcal{F}}$(A) is the result of the accumulation of density in the binding region between the nuclei. The attraction of the nucleus external to A for the density of A is the principal stabilizing force in the formation of a molecule and its virial is the principal source of energy lowering. For example, in H_2, N_2, and CO at R_e, the largest single contribution to $\vec{\mathcal{F}}$(A) is the attraction of the B nucleus for the density on A,[25] and the largest single contribution to the lowering of the energy in these molecules is the potential energy of interaction of the nucleus of one atom interacting with the density of the other.[11,25] The creation of attractive Feynman forces on the nuclei on the approach of two neutral atoms causes the potential energy to initially increase and the kinetic energy to decrease. Thus the curves for $V(R)$ versus R for H_2, N_2, and CO bear a similarity to the curves shown for corresponding behavior of $\vec{\mathcal{F}}$(A) in terms of the values of R where they attain their maximum value and for which they change sign and become attractive.[25] In general, $\vec{\mathcal{F}}$(A) parallels the potential energy of interactions between the atoms. For separations substantially less than R_e, the density on A will be increasingly polarized away from B causing both the electron−nuclear and electron−electron forces to make repulsive contributions to $\vec{\mathcal{F}}$(A), and they, in conjunction with the electron−electron repulsion term for B, may result in the Ehrenfest force becoming repulsive. Because of the absence of the nuclear force of repulsion, this will occur for shorter distances than for the Feynman force, including the interactions between rare gas atoms.

The Ehrenfest force acting on the surface in LiF exhibits only a small repulsive increase for distances in excess of 12 au, becoming at first weakly and then strongly attractive as a result of the precipitous transfer of approximately one electron from Li to F that is essentially complete by 10 au, a picture in accord with the 'harpoon mechanism'. Thus in an ionic system, the attractive electrostatic interactions outweigh the repulsive ones following charge transfer.

Geometric Consequences of X Insertion and the Ehrenfest Force

The above introduction to $\vec{\mathcal{F}}$(A) provides the necessary background for the discussion of the forces exerted on X

and its bonded methine carbon atoms caused by the insertion of X into the adamantane cage. A methine C atom and its bonded H are labeled C1 and H1, while a methylene carbon and a bonded H are labeled C2 and H2. The magnitude of the Ehrenfest force on atom X is given in Table 3 together with the volume of X, v(X), the latter being a well-defined quantity as the X atoms are totally encased within the adamanatane cage. Also listed are the increases in the value of R(C1|C2) from the value of 2.891 au in adamantane and the bonded radii r_b(X) and r_b(C1), the position of the X|C1 bond CP relative to the nuclei.

First and foremost, the Ehrenfest force on X is attractive in every case, drawing atoms X and C1 toward one another. As discussed below, the sign of $\vec{\mathcal{F}}$(X|C1), the force exerted on the X|C1 surface, determines the sign of the surface virial, the contribution to the potential energy resulting from the formation of the X|C1 interatomic surface, and an attractive Ehrenfest force results in a stabilizing energy for the formation of the X|C1 surface. Surprisingly, the increase in R(C1|C2) varies inversely with the volume of X for the first three members, becoming larger only for the insertion of the Ne atom whose volume is more than four times v(Be). The increase in size of Li and He over Be is instead accommodated by a decrease in the bonded radius of C1, r_b(C1). With the exception of its small reversal between the two ions, the magnitude of the attractive Ehrenfest force increases with the volume of X, increasing as the X|C1 bond CP moves closer to C1. Not only is there no Ehrenfest force of repulsion between the X and C1 atoms, its attractive nature increases with the size of the inserted atom, from Be^{2+} to He to Ne and the accompanying change from exothermicty to an increasing endothermicity with v(X) does not have its origin in an increasing 'antibonding'[26] in the C1|X interaction. This behavior of the Ehrenfest force is reflected in the exceptional decrease in the energy of the X atoms, particularly He and Be, upon insertion and the conclusion is substantiated by the study of the energy changes of the individual atoms presented in a later section.

The Ehrenfest Force and the Interactions between Macroscopic Systems

The Ehrenfest force on the electron density plays a role that must be distinguished from the more familiar Feynman force acting on the nuclei. For example, for large separations between neutral atoms the Ehrenfest force is repulsive, whereas the Feynman forces are attractive, the former correctly reflecting the initial increase in the potential energy. The behavior of the two forces may be compared in Figure 2 for the hydrogen molecule. The Feynman force eventually vanishes at R_e and then becomes repulsive, while the Ehrenfest force becomes increasingly attractive, reflecting the decrease in the potential energy required by the virial theorem on bonding. The Feynman force is invoked in situations where the Ehrenfest force is in fact the operative force, these situations arising in cases that require the physics of an open system for their statement and solution. An example is the operation of the atomic force microscope (AFM).[27] The AFM and the surface under study are two components of a total system separated by a zero-flux surface. It is the force transmitted from the tip of the probe to the attached cantilever arm of the AFM whose deflections are measured as the tip scans the surface of a sample. The force $\vec{\mathcal{F}}$ measured by the AFM is determined by the pressure that the sample exerts on each element of the surface of separation. The Feynman forced does not equal $\vec{\mathcal{F}}$. Instead, it measures the force $\mathbf{F}_\alpha$ required to displace the nucleus α of the atom in the tip of the AMF. One is interested in the force not just on the nucleus of one atom in the tip of the probe, but rather in the force $\vec{\mathcal{F}}$ that is exerted on all of the atoms that make up the open system "probe plus cantilever". This force is determined by the pressure exerted on every element of the surface separating the tip of the probe from the sample, as given in eq 6 in terms of the surface integral of the stress tensor. It is the Ehrenfest force $\vec{\mathcal{F}}$ that the cantilever arm exerts on the attached spring, displacing it from its equilibrium position, the displacement measured in the AFM. The force $\vec{\mathcal{F}}$ is exerted on a surface determined by the zero-flux interatomic surfaces separating the atoms in the tip from those in the sample, and thus its response is a consequence of the atomic form of matter.[27]

The definition of pressure requires the existence of a surface upon which the pressure is exerted placing it within the realm of the physics of an open system. A scaling procedure demonstrates that the expectation value of the pressure−volume product of a proper open system is proportional to its surface virial, the virial of the Ehrenfest force exerted on its surface.[28] Thus the thermodynamic pressure is determined by the virial of the force resulting from the electronic momentum flux through its zero-flux surface, the same quantity that determines the force $\vec{\mathcal{F}}$ acting on the open system. The pressure determined in this manner is a consequence of the mechanics of the interaction between the open system and its confining walls as opposed to previous treatments based on the analogy with the classical virial theorem for a contained gas wherein the pressure−volume product was incorrectly related to the virial of the "wall forces".

These examples demonstrate that the Ehrenfest force is the one acting on and between macroscopic systems, controlling in addition their relative motion. It is thus the one that determines the force required to separate two surfaces and the study of the adhesive properties of surfaces should be directed toward an understanding and prediction of the Ehrenfest force that is established when two surfaces are brought into contact.

Energy Changes

The virial theorem plays a central role in defining the energy of an open system.[4,5] Its statement for an open system is of the same form as that given in eq 2 for the Ehrenfest force: in this case the time rate of change of $\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}$ is equated to the expectation value of $(i/\hbar)[\hat{H}, \hat{\mathbf{r}}\cdot\hat{\mathbf{p}}]$ plus a surface term $\mathcal{V}_s$(A), denoting the flux in the associated current density that is expressible as $\mathbf{r}\cdot\sigma(\mathbf{r})$. The expectation value of the commutator yields $2T$(A) + $\mathcal{V}_b$(A), twice the electronic kinetic energy T plus the virial of the Ehrenfest force exerted over the basin of the atom. For a stationary state the atomic statement of the virial theorem yields $2T$(A) + $\mathcal{V}_b$(A) +

$\mathcal{V}'_s(A) = 0$. Expressing by $\mathcal{V}'(A)$ the total virial for an open system A, $\mathcal{V}'(A) = \mathcal{V}'_b(A) + \mathcal{V}'_s(A)$, the virial theorem for a stationary state may be stated as

$$-2T(A) = \mathcal{V}(A) = \mathcal{V}'_b(A) + \mathcal{V}'_s(A) \quad (7)$$

The virials of the Ehrenfest force exerted over the basin and the surface of the atom with the origin for the coordinate **r** placed at the nucleus of atom A are given in eq 8

$$\mathcal{V}'_b(A) = -\int_A \mathbf{r}_A \cdot \vec{\mathcal{F}}(\mathbf{r}) d\mathbf{r} = -\int_A \mathbf{r}_A \cdot \nabla \cdot \vec{\sigma}(\mathbf{r}) d\mathbf{r}_s(A)$$

$$\mathcal{V}'_s(A) = \oint dS(A, \mathbf{r}_s) \mathbf{r}_A \cdot \vec{\sigma}(\mathbf{r}) \cdot \mathbf{n}_A(\mathbf{r}) \quad (8)$$

An alternative expression for $\mathcal{V}'_b(A)$ in a stationary state in terms of the stress tensor is also given in eq 8. One notes that $\mathcal{V}'_b(A)$ is the virial of the Ehrenfest force acting over the basin of the atom and $\mathcal{V}'_s(A)$ is the virial of the Ehrenfest force acting on its surface. The current in the surface virial has the dimensions of F/L or surface tension, and $\mathcal{V}'_s(A)$ is thus a measure of the surface energy associated with the interaction of A with its bonded neighbors.

We shall be concerned with molecules in equilibrium geometries wherein no forces act on the nuclei and $\sum_A \mathcal{V}'(A) = V$, the total potential energy of the molecule. Under these conditions, one may employ the virial theorem to relate the change in the total energy to the changes in the kinetic energy T and the potential energy V with no contributions from the virials of external forces. Thus $\Delta E = -\Delta T = {}^1/_2 \Delta V$ where Δ denotes the difference between any two states of the system for which the forces on the nuclei vanish. The change in energy for some total system, ΔE, may be expressed in terms of the changes in the potential energy contributions of the individual atoms in one of two ways. First, one may equate the change ΔV to the sum of the changes in the electron−nuclear attractive energies ΔV_{en} plus a term ΔV_r grouping the two repulsive contributions from the electron−electron and nuclear−nuclear terms, as done in eq 9

$$\Delta E = {}^1/_2 \sum_A \{\Delta V_{ne}(A) + \Delta V_{ee}(A) + \Delta V_{nn}(A)\} = {}^1/_2 \sum_A \{\Delta V_{ne}(A) + \Delta V_r(A)\} \quad (9)$$

Alternatively, one may equate ΔE to the changes in the basin and surface virials

$$\Delta E = {}^1/_2 \sum_A \Delta \mathcal{V}'(A) = {}^1/_2 \sum_A \{\Delta \mathcal{V}'_b(A) + \Delta \mathcal{V}'_s(A)\} \quad (10)$$

Equation 10 divides $\Delta E(A)$ into two contributions: one from the change in the virial of the forces acting over the basin of the atom (with $\mathcal{V}'_b(A) = \mathcal{V}'(A) = V(A)$ for an isolated atom) and another from the formation of its interatomic surface (with $\Delta \mathcal{V}'_s(A) = \mathcal{V}'_s(A)$ for a change referenced to an isolated atom). The energy decomposition given in eq 9 is discussed first for the cage complexes.

Changes in Atomic Energies for Formation of Endohedral Cages

The electron−nuclear potential energy for atom A, the quantity $V_{en}(A)$, is given by the sum of two contributions: the term $V_{en}{}^{\circ}(A)$, the interaction of the electron density in the basin of atom A (the density internal to A) with its own nucleus together with the term $V_{en}{}^{e}(A)$, the interaction of the same density with all the nuclei external to A. $V_{en}{}^{e}(A)$ *yields the stabilizing interaction of the density of atom A with the nuclei of all the atoms that, together with A, are brought together in the formation of the molecule.* Thus $V_{en}{}^{e}(A)$, which is zero for an isolated atom, is necessarily negative, and it is the contributions of all the atoms to the decrease in $V_{en}{}^{e}$ that dominates the energy of molecular formation.

$Be^{2+}@C_{10}H_{16}$. The changes in the atomic populations $N(A)$, atomic volumes $v(A)$, and total energies $E(A)$ are listed in Table 4 for the formation of each complex from separated adamantane and X. Also listed are the changes in the potential energy in terms of their attractive internal $\Delta V_{en}{}^{\circ}(A)$ and external $\Delta V_{en}{}^{e}(A)$, and repulsive $\Delta V_r(A)$ contributions. The volume change $\Delta v(X)$ is with respect to the volume of a free X atom or ion determined by its 0.001 au density envelope, the van der Waals envelope.[29] The changes for the atoms of adamantane are for volumes determined by the intersection of their interatomic surfaces with the 0.001 au density envelope that bounds their nonbonded regions.

The principal contribution to the exothermic bonding energy of 223 kcal/mol in $Be^{2+}@C_{10}H_{16}$ is from the Be atom, with additional stabilizing changes from the C1 and C2 atoms, all of which are the recipients of electronic charge. Both the internal and external contributions to the potential energy of the carbons decrease and are sufficient to overcome the increase in the repulsive energy and lead to overall stabilization of the atoms. The hydrogens lose electronic charge, their volumes decrease, and their energies increase, contributing +652 kcal/mol to the total energy change, and overall, the insertion of Be^{2+} causes the energy of the adamantane cage to increase by 179 kcal/mol.

The largest single stabilizing contribution to the change in the potential energy of interaction is from the decrease in the external potential energy of the Be atom, the term $\Delta V_{en}{}^{e}(Be)$, a result of the interaction of its density with the nuclei of the atoms of the enclosing cage. This is a general result for all of the inserted X atoms. Electron density is transferred to X in each complex, $\Delta N(X)$ ranging from 0.28e for Be to 0.11e for Ne. Hence the internal contribution $\Delta V_{ne}{}^{\circ}(X)$ is stabilizing but of much smaller magnitude than $|\Delta V_{en}{}^{e}(X)|$, and the bonding of the X atom is primarily the result of the stabilization of its density with the nuclei of the surrounding cage. The attractive interactions are sufficient to overcome the increase in the repulsive contributions to the energy of X in all four complexes. The bonding of X in the endohedral cage complexes is similar to that found in the transition metal carbonyls and metallocenes such as ferrocene, where the principal source of the bonding is from dominant and exceptional decrease in the energy of interaction of the atomic density on the transition metal atom with the nuclei of the surrounding ligand atoms, the contribution from the term $\Delta V_{ne}{}^{e}(M)$.[11]

$Li^{+}@C_{10}H_{16}$. As with the Be complex, the dominant contribution to the energy of formation of the complex is from the decrease in energy of Li and this primarily from the interaction of its density with the nuclei of the cage, the

Table 4. Changes in the Atomic Properties of X@adam from Adamantane and X[a]

A	ΔN(A)	Δ_ν(A)	ΔE(A)	ΔV_{ne}°(A)	ΔV_{ne}^{e}(A)	ΔV_r(A)
X = Be^{2+}	0.2824	−6.56	−401.95	−1.4823	−48.5204	48.7215
C1	0.3311	8.54	−89.08	−0.9646	−9.0994	9.7804
C2	0.1451	5.95	−19.35	−0.3684	−5.2714	5.5785
H1	−0.2057	−11.14	57.48	0.1852	2.6228	−2.6248
H2	−0.1378	−8.28	35.11	0.1163	1.4954	−1.4997
X = Li^{+}	0.1953	−14.36	−260.26	−0.8064	−46.8285	46.8052
C1	0.1424	5.98	−10.27	−0.3212	−3.7599	4.0485
C2	0.0839	4.42	8.71	−0.1526	−2.9247	3.1051
H1	−0.1076	−6.29	28.47	0.0907	1.3433	−1.3433
H2	−0.0699	−4.25	17.09	0.0568	0.6927	−0.6951
X = He	0.1011	−48.60	−327.77	−0.6362	−45.1671	44.7584
C1	−0.0188	2.63	57.26	0.2816	0.2700	−0.3690
C2	0.0105	2.94	33.61	0.1125	−0.5264	0.5211
H1	−0.0035	0.65	4.06	0.0070	0.0311	−0.0250
H2	−0.0064	0.19	3.02	0.0077	0.0011	0.0009
X = Ne	0.1095	−76.22	−903.45	−3.4059	−201.2768	201.8030
C1	−0.0055	11.83	149.55	0.6652	−6.6310	6.4428
C2	0.0311	9.85	94.09	0.3349	−7.5827	7.5480
H1	−0.0118	2.09	9.79	0.0206	−0.5856	0.5963
H2	−0.0185	1.10	8.04	0.0215	−0.7195	0.7237
X = B^{3+}	0.7881	2.69	−1014.74	−4.6331	−59.3031	60.7019
C1	0.5537	10.32	−186.31	−1.6891	−16.1303	17.2260
C2	0.1149	6.83	−15.46	−0.2784	−5.8874	6.1169
H1	−0.3003	−15.32	88.95	0.2876	3.8547	−3.8588
H2	−0.2075	−11.63	56.04	0.1835	2.3680	−2.3728
X = Al^{3+}	0.4681	−12.57	−1151.25	−6.4149	−207.7750	210.5206
C1	0.4126	18.20	−44.74	−0.9032	−20.6561	21.4171
C2	0.2508	13.96	−1.74	−0.4965	−16.2341	16.7254
H1	−0.2893	−14.53	86.34	0.2725	3.1363	−3.1335
H2	−0.2056	−10.87	56.48	0.1828	1.5751	−1.5778

[a] ΔE(A) in kcal/mol, Δv(A) and ΔV(A) in au.

term ΔV_{ne}^{e}(Li). Each carbon bonded to Li is stabilized by 10 kcal/mol in forming the complex, while each methylene carbon is destabilized by 9 kcal/mol. The loss of electronic charge from the hydrogens is one-half that transferred in the Be complex as is the associated increase in their energy equal to +319 kcal/mol. The adamantane cage is destabilized by 330 kcal/mol, an increase of 151 kcal/mol over $Be^{2+}@C_{10}H_{16}$ and overall the energy of formation of $Li^{+}@C_{10}H_{16}$ is endothermic by 70 kcal/mol.

$He@C_{10}H_{16}$. The energy of formation of this complex is endothermic by 156 kcal/mol. The dominant stabilizing contribution to the energy of formation is from the interaction of the density of the He atom with the nuclei of the cage, the term V_{ne}^{e}(He), whose magnitude outstrips the increase in the repulsive contribution, ΔV_r(He), to the atomic energy change. One notes that the value of ΔV_r(He) is less than ΔV_r(Be), He undergoing a smaller increase in its repulsive interactions with the cage than Be. The transfer of electronic charge to X is least for He, and correspondingly it undergoes the smallest of the stabilizations caused by decreases in internal energy, the quantity ΔV_{ne}°(He). The energies of both the methine and methylene carbons increase, the former losing a small amount of electronic charge the latter gaining a still lesser amount.

The potential energy contributions to the energy changes for a methine carbon C1 bonded to He are of most interest. The interaction results in a *decrease* in the repulsive energy of C1, but its energy undergoes a net increase as a result of the decrease in magnitude of its attractive contributions. The value of ΔV_{ne}^{e}(C1) is the only one such contribution greater than zero for all of the complexes, and this is a result of the smaller nuclear charge on He compared to the other X atoms, in particular that for Ne with $Z = 10$, where $\Delta V_{ne}^{e}(C1) = -6.6$ au. The increase in the repulsive energies of the C2 atoms, ΔV_r(C2), is also less than for the insertion of Be^{2+} and the sum of the increases in repulsive energies for C1 and C2 in $He@C_{10}H_{16}$ is 70.9 au smaller than the corresponding sum for $Be^{2+}@C_{10}H_{16}$. This observation coupled with the decrease in the value of ΔV_r(He) compared to ΔV_r(Be) demonstrates that the claim that the endothermicity of $He@C_{10}H_{16}$ is a result of 'antibonding' of He with the methine carbons of the adamantane cage has no basis in physics.[26]

The energy of the carbon framework does increase, but not because of increased repulsions with the caged atom, but rather because of destabilizations in their mutual interactions. The small loss of electronic charge from the C1 atoms and the reorganization of their densities causes the internal contributions to their potential energies, together with those for the C2 atoms, to increase and overall the energies of the carbons increase by 431 kcal/mol. The small loss of density from the H atoms causes their energy to increase by 53 kcal/mol, and the adamantane cage is destabilized by 484 kcal/mol, an increase of 305 kcal/mol over $Be^{2+}@C_{10}H_{16}$.

Thus the endothermicity of $He@C_{10}H_{16}$ compared to the exothermicity of $Be^{2+}@C_{10}H_{16}$ is not a result of a repulsion between the X atom and the surrounding cage, a quantity that is actually less for He than for Be^{2+}, but because of the *decrease* in the attractive contributions to the energies of the carbons of the cage.

$Ne@C_{10}H_{16}$. The Ne complex has an endothermicity of 395 kcal/mol. The single largest change in the atomic energies for the Ne atom, as for the X atom in the other complexes, is a result of the stabilizing decrease in the energy of interaction of the density of the Ne atom with the nuclei of the cage. The endothermic energy of formation is again a result of the destabilization of the cage, and primarily from the increase in energy of the carbon atoms, the methine carbons in particular, whose energies increase by 1164 kcal/mol. Unlike the He complex where the repulsive energies of the methine carbons actually decrease on bonding, both sets of carbons in the Ne complex have positive $\Delta V_r(C)$ values that are only slightly less than the magnitude of the decreases in the values of $\Delta V_{en}^e(C)$. The internal contributions $\Delta V_{en}^\circ(C)$ are also positive, and the carbons are destabilized in their interactions with the Ne atom. The discussion of the data for $B^{3+}C_{10}H_{16}$ is reserved for the demonstration of the dominant role in the stabilization of the complexes played by the charge transfer from the hydrogens to the cage interior.

Bonding as Viewed through the Changes in the Basin and Surface Virials. While the virial $\mathcal{V}(A)$ is independent of the choice of origin, its expression in terms of basin and surface contributions is not. However by placing the origin at the position of the nucleus, as done in eq 8, one can obtain a unique expression for the potential energy change as a sum of basin contributions and interatomic surface terms, the latter depending only upon the internuclear separation between the bonded nuclei. The surface terms determine the contribution to the potential energy of interaction stemming from the formation of each surface A|B. Just as the Ehrenfest force on the surface of A can be equated to a sum of contributions, one from each of the interatomic surfaces bounding A, eq 6, so $\mathcal{V}'_s(A)$ can be similarly expressed, with the contribution from a bonded neighbor B denoted by $\mathcal{V}'_s(A|B)$. Summing the contributions to the energy of the surface $S(A|B)$ between A and B and defining $\mathbf{R}_{ab} = \mathbf{r}_a - \mathbf{r}_b$, the vector from the nucleus of B to that of A, and noting that $\mathbf{n}_A = -\mathbf{n}_B$, one obtains[4]

$$V_s(A||B) = \mathcal{V}'_s(A|B) + \mathcal{V}'_s(B|A) = \mathbf{R}_{ab}\cdot\oint dS(A|B;\mathbf{r}_s)\vec{\sigma}(\mathbf{r})\cdot\mathbf{n}_A(\mathbf{r}) = \mathbf{R}_{ab}\cdot\vec{\mathcal{F}}_b(A|B) \quad (11)$$

This expression equates the sum of the surface virials from two atoms sharing a common surface $S(A|B)$ to the scalar product of the Ehrenfest force acting on the surface with the vector $\mathbf{R}_{ab}$, and the dependence of the two surface integrals on individual origins is replaced by the physical separation between the nuclei of the bonded atoms. Thus the energy of formation of a molecule from separated atoms can be expressed in terms of contributions from the changes in the potential energy within each atomic basin plus a contribution from the virial of the Ehrenfest force $\vec{\mathcal{F}}(A|B)$ acting on the interatomic surface between each pair of bonded atoms denoted by $\mathcal{V}'_s(A||B)$, eq 12

$$\Delta E = {}^1/_2\sum_A \mathcal{V}'_b(A) + {}^1/_2\sum_{A|B} \mathcal{V}'_s(A||B) \quad (12)$$

The Ehrenfest force $\vec{\mathcal{F}}(A|B)$ is attractive when the force exerted on the density of atom A is directed at B, the situation found for all of the bonded interactions in the cage complexes. The combined surface virial $\mathcal{V}'_s(A||B)$ is negative in such a case, and the formation of the interatomic surface $S(A|B)$ contributes to the stability of the system.

The data for the implementation of eq 12 are given in Table 5 which also lists the separate values of $\Delta\mathcal{V}'_s(A)$. The changes in both the basin and surface virials are stabilizing for the X atom in all four complexes, with the magnitude of $\Delta\mathcal{V}'_s(X)$ increasing in the order Be → Ne. The basin virials for the carbons are destabilized by relatively small amounts in the Be and Li complexes, and their net stabilization is a result of the changes in their surface virials. The carbons in the He and Ne complexes undergo destabilizing increases in both their basin and surface virials, as do the hydrogen atoms in all four complexes.

The energies of the surface virials are given in kcal/mol to better gauge their relative contribution to energies of formation of the complexes. The energy of the X|C1 surface, as determined by the ${}^1/_2\mathcal{V}'_s(X||C1)$, is stabilizing in every complex, ranging from −112 kcal/mol in $Be^{2+}@C_{10}H_{16}$ to −247 kcal/mol in $Ne@C_{10}H_{16}$. The energy of the C1|C2 surface decreases slightly in the formation of $Be^{2+}@C_{10}H_{16}$ but increases in the remaining complexes, as do the energies of the C1|H1 and C2|H2 surfaces. In summary, the Ehrenfest force, which is attractive for all of the bonded interactions in these molecules, decreases in magnitude for all of the interactions within the adamantane cage with the exception of the C1|C2 surface in $Be^{2+}@C_{10}H_{16}$. Thus the adamantane cage is destabilized in the formation of the complexes by the destabilization of the basin virials of both the C and H atoms and by increases in the energies of the surface integrals internal to the cage, the latter resulting in a decrease in the stability of their bonded interactions. Contrawise, the energy of formation of the atomic surface bounding the X atom is stabilizing in every case, as is the energy of formation of the X|C1 surface, and these interactions contribute to the stabilization of the complexes.

Discussion and Conclusions

The role of the Ehrenfest force in the physics of an open system brings to the fore the unifying understanding provided by the virial theorem. Slater refers to Feynman's electrostatic theorem and the virial theorem as 'two of the most powerful theorems applicable to molecules and solids'.[30] To this one should add the Ehrenfest force, since it is the virial of this force that determines the electronic virial in Slater's molecular theorem.[20] The Ehrenfest force includes the electrostatic force of attraction of the density by the nuclei but not the nuclear repulsion. Consequently, the nuclear contribution to $\mathcal{F}$ equals the difference between the net Feynman force acting on the nuclei and the nuclear repulsive forces, thereby

Table 5. Changes in Atomic Basin and Surface Virials and with Surface Virials for Individual Bonded Interactions[a]

A	$\Delta\mathcal{V}_b$(A) au	$\Delta\mathcal{V}'_s$(A) au	$\mathcal{V}'_s$ (X||C1) kcal/mol	$\Delta\mathcal{V}'_s$ (C1||C2) kcal/mol	$\Delta\mathcal{V}'_s$ (C1||H1) kcal/mol	$\Delta\mathcal{V}'_s$ (C2||H2) kcal/mol
X = Be^{2+}	−0.6824	−0.5983	−223.52	−38.72	32.13	17.95
C1	0.0337	−0.3196				
C2	0.0236	−0.0872				
H1	0.1177	0.0655				
H2	0.0695	0.0425				
X = Li^+	−0.1609	−0.6682	−212.35	22.72	17.32	9.41
C1	0.0461	−0.0791				
C2	0.0423	−0.0155				
H1	0.0515	0.0392				
H2	0.0312	0.0232				
X = He	−0.0268	−1.0177	−282.63	55.53	3.20	2.82
C1	0.1656	0.0167				
C2	0.0842	0.0222				
H1	0.0018	0.0112				
H2	0.0032	0.0065				
X = Ne	−0.8497	−2.0274	−494.42	124.06	8.85	7.22
C1	0.3418	0.1339				
C2	0.1961	0.1007				
H1	0.0069	0.0242				
H2	0.0092	0.0164				

[a] For adamantine, $\mathcal{V}'_s$(C1||C2) = −0.3174; $\mathcal{V}'_s$(C1||H1) = −0.1318; $\mathcal{V}'_s$(C2||H2) = −0.1314.

accounting for the appearance of V_{nn} and the virial of the Feynman forces in the virial of the forces exerted on the electrons.[4]

The use of these theorems provides answers to the questions raised in the Introduction concerning the possible presence of repulsive interactions between X and the carbons of the cage[26] and of the storage of the excess energy in the endothermic complexes. The Ehrenfest force in a stationary state is balanced locally by the divergence of the stress tensor, $\vec{\mathcal{F}}(\mathbf{r}) + \nabla\cdot\vec{\vec{\sigma}}(\mathbf{r}) = 0$, as its integrated value $\vec{\mathcal{F}}$(A) is by the surface integral of $\vec{\vec{\sigma}}(\mathbf{r})$, eq 6. However, as discussed in the Introduction, $\vec{\mathcal{F}}$(A) enables one to determine the force acting over the basin of an atom and to determine how this force is balanced on the surface it shares with each of its neighbors. This is new information regarding the mechanics of bonding between atoms, information that underlies the accompanying energy changes because it is the virial of the Ehrenfest force that determines the potential energy changes. It is the nature of this force, that within the regime of equilibrium geometries, it is always found to be attractive, drawing each atom toward its neighbor such that the resultant force on the atom is zero. Thus one may extend the definition of bonding between atoms denoted by a bond path: *the presence of a line of maximum density linking a pair of nuclei in an equilibrium geometry of a bound state or one lying within the attractive region of a potential well—atoms linked by a bond path—implies not only the absence of repulsive Feynman forces on the nuclei but also the presence of attractive Ehrenfest forces acting across the interatomic surface shared by the bonded atoms.* This observation underlies the mirroring of a bond path by a virial path, a line of maximally stabilizing potential energy density.[10] This definition certainly applies in the case of the endohedral complexes, where the Ehrenfest force $\vec{\mathcal{F}}(X|C1)$ acting across each X|C1 surface is attractive and the interactions denoted by the corresponding bond paths are a consequence of attractive forces acting between the X and C1 atoms.

Since the force $\vec{\mathcal{F}}(X|C1)$ is attractive, the derived surface virial $\mathcal{V}'_s(X||C1)$, eq 11, a quantity that determines the contribution to the energy change resulting from the formation of the surface X|C1, is stabilizing. While $\mathcal{V}'_s$(A||B) is not a bond energy for A|B—a quantity not defined within the physics of an open system—its change during the formation of the complex provides a quantitative gauge of the distribution of the energy increase over the bonded interactions caused by an endothermic insertion of X. The values of $\mathcal{V}'_s$(A||B), Table 5, demonstrate that the energy increases incurred by the insertion of Li^+, He, and Ne are mirrored solely in the surface virials of the atoms comprising the adamantane cage, the formation of the X|C1 surfaces yielding stabilizing contributions in all cases. These changes are reflected in the energy changes for the individual atoms, Table 4, with only the energies of X and of C1 in $Li^+@C_6H_{10}$ contributing to the lowering of the energy of formation in the endothermic reactions.

The principal factor contributing to the stabilization of an endohedral complex may be determined by a comparison of the atomic properties for the formation of exothermic $Be^{2+}@C_{10}H_{16}$ with the endothermicity of the remaining complexes. The exothermicity has its origin in the transfer of charge from the hydrogens to the interior of the cage, to the carbons, and to the positively charged Be atom, rather than being a consequence of the smaller volume of the Be atom. The similar but much smaller flow of density from the hydrogens to the interior of the cage caused by the smaller positive charge on Li is insufficient to overcome the attendant destabilization resulting form the larger volume of Li. If the charge transfer to the cage interior is indeed paramount in leading to stabilization of the insertion complex, one would predict an even larger exothermicity for the insertion of a

B^{3+} ion and this is indeed the case. The formation of $B^{3+}@C_{10}H_{16}$ is predicted to be exothermic by 826 kcal/mol compared to an exothermicity of 223 kcal/mol for $Be^{2+}@C_{10}H_{16}$. There is a transfer of 3.7e from the hydrogens to the cage interior in $B^{3+}@C_{10}H_{16}$ compared to 2.5e in $Be^{2+}@C_{10}H_{16}$, with 0.79e being transferred to B, Table 4. While the volume of a free B^{3+} ion is less than that of free Be^{2+}, the volume of the boron atom in the complex is actually greater than that for Be, 9.6 compared to 7.2 au, a consequence of a greater transfer of electronic charge to the boron atom. The data in Table 4 show that the pattern of changes in the atomic energies for the formation of the boron complex mimic those for the Be complex, but all are of greater magnitude. One notes in particular the increase in the magnitude of the internal potential energy change for B compared to Be, all increases being a consequence of the flow of density from the hydrogens to the cage interior. The same increase in exothermicity of formation with increase of positive charge on X is found for $Al^{3+}@C_{10}H_{16}$ for which ΔE is predicted to equal -318 kcal/mol, a consequence of a transfer of 3.6e of electronic charges from the hydrogens to the interior of the cage. This increase in the stability of $Al^{3+}@C_{10}H_{16}$ compared to $Be^{2+}@C_{10}H_{16}$ is obtained despite the volume of the Al atom in the complex being greater than that for the Be atom, $v(\mathrm{Al}) = 17.3$ au compared to $v(\mathrm{Be}) = 7.2$ au and in spite it exhibiting the largest increase in the C1−C2 separation, equal to 0.39 au.

The theory of atoms in molecules provides a comprehensive understanding of the relative stabilities of endohedral complexes and of the mechanism of energy storage within the complex, all in terms of the atomic expectation values for populations, energies and forces obtained from the physics of an open system.

Haaland et al.[26] in a critique of the statement that the presence of a bond path linking two atoms implies that the atoms are bonded to one another state "Most chemists would probably agree that the defining property of a chemical bond is the existence of a positive bond rupture energy, that is, the energy of the molecule is lower than the energy of the fragments,...". This is followed by the division of the endothermicity for the formation of $He@C_{10}H_{16}$ by four to account for the presence of the four bond paths linking He to the methine carbons, to give "a negative He−C mean bond energy of − 39 kcal/mol". In their words "the interaction is *antibonding*". There is no explanation of the force responsible for this 'antibonding', occurring as it does in the absence of Feynman forces on the nuclei and wherein the Ehrenfest forces between the bonded atoms are attractive.

The response to this criticism requires that one correct for a number of misunderstandings of the theory by the authors. First, concerning the question posed in the title to their paper "Is the presence of an atomic interaction line in an equilibrium geometry sufficient condition for the existence of a bond?" and their reference to 'a chemical bond' in rationalizing the definition of a bond energy; there is no attempt to define a 'bond' in QTAIM since a bond has no definable nor identifiable physical form. Rather theory accounts for 'bonding between atoms', the presence of a bond path denoting that the atoms so linked are 'bonded to one another'. All of the properties that are traditionally associated with the presence of a 'bond' such as 'bond energy' or 'bond length' are, in fact, consequences of the occurrence of 'bonding between atoms'. As illustrated in this paper, one may bring all of quantum mechanics to bear on the forces and energetics of bonding using the physics of an open system. There is no unique, unambiguous definition of a 'bond energy' in a polyatomic molecule in traditional chemistry. The authors raise anew previous claims of bond paths being present in cases where the interactions are 'clearly repulsive',[31,32] claims that have been previously dealt with.[8,33,34]

Their conclusion that the X|C1 interaction is 'antibonding' is at variance with the observation that the Ehrenfest force $\vec{\mathscr{T}}(\mathrm{X|C1})$ is attractive and that the formation of the X|C1 surface makes a stabilizing contribution to the energy of formation, with $\mathscr{V}_s(\mathrm{X||C1}) < 0$. The destabilizing contributions to the total energies of insertion of X are confined to the interactions between the atoms of the adamantane cage, as evidenced by the changes in their atomic energies and the basin and surface virials. Indeed, as noted above, the repulsive contribution ΔV_r to the atomic energy changes for X and the carbon atoms are less for endothermic $He@C_{10}H_{16}$ than for exothermic $Be^{2+}@C_{10}H_{16}$. The argument given by Haaland et al. to rationalize the stabilizing decrease in the energy of the He atom upon insertion with its 'antibonding' character avoids all appeal to quantum mechanics and the forces operative in a molecular system and instead ascribes it to the 'density of the He dropping abruptly to zero at its four interatomic surfaces', a statement displaying a complete lack of understanding of the physics of an open system and its theorems. The stabilization of the X atoms in the complexes is a result of simple physics; of the attractive interaction of the density of X with the nuclei of the atoms in the adamantane cage, the values of $V_{en}^{e}(\mathrm{X})$ given in Table 4. While arbitrary models cannot be used to criticize conclusions that are obtained from physics, one may employ QTAIM in a complementary manner to determine viability of orbital models—the recovery of the Dewar-Chatt-Duncanson dπ-pπ* model[35,36] of back-bonding in terms of the quadrupole polarizations of the ligand atoms being a recent example.[11,37]

If one generalizes the argument of Haaland et al., then one must postulate the presence of undefined 'antibonding interactions' in all stable configurations other than the one of lowest energy. A molecule at thermal equilibrium in a stationary state has no memory of how it attained that state, from one of higher or lower energy. To postulate the presence of unknown repulsive forces in a molecule in an equilibrium geometry based on an anthropomorphic bias contributes nothing to the physical understanding of bonding.

Acknowledgment. We wish to express our thanks to Dr. J. Hernández-Trujillo for constructing the plots of the Ehrenfest force and surface virials by the retrieval of data from an earlier investigation.[25]

References

(1) Moran, D.; Woodcock, H. L.; Chen, Z.; Schaefer, H. F., III; Schleyer P. v. R. *J. Am. Chem. Soc.* **2003**, *125*, 11442−11451.

(2) Cross, R. J.; Saunders, M.; Prinzbach, H. *Org. Lett.* **1999**, *1*, 1479.

(3) Moran, D.; Stahl, F.; Jemmis, E. D.; Schaefer, H. F. III; Schleyer, P. v. R. *J. Phys. Chem. A* **2002**, *106*, 5144.

(4) Bader, R. F. W. *Atoms in Molecules: a Quantum Theory*; Oxford University Press: Oxford UK, 1990.

(5) Bader, R. F. W. *Phys. Rev.* **1994**, *B 49*, 13348−13356.

(6) Keith, T. A. SCVS. 1992.

(7) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., 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. *Gaussian98, Revision A.9*; Gaussian Inc.: Pittsburgh, PA, 1998.

(8) Bader, R. F. W. *J. Phys. Chem. A* **1998**, *102*, 7314−7323.

(9) Bader, R. F. W.; Nguyen-Dang, T. T.; Tal, Y. *Rep. Prog. Phys.* **1981**, *44*, 893−948.

(10) Keith, T. A.; Bader, R. F. W.; Aray, Y. *Int. J. Quantum Chem.* **1996**, *57*, 183−198.

(11) Cortés-Guzmán, F.; Bader, R. F. W. *Coord. Chem. Rev.* **2005**, *249*, 633−662.

(12) Macchi, P.; Sironi, A. *Coord. Chem. Rev.* **2003**, *238−239*, 383−412.

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

(14) Bader, R. F. W.; Stephens, M. E. *J. Am. Chem. Soc.* **1975**, *97*, 7391−7399.

(15) Wang, Y.-G.; Werstiuk, N. H. *J. Comput. Chem.* **2003,** *24*, 379−385.

(16) Wang, Y.-G.; Matta. C.; Werstiuk, N. H. *J. Comput. Chem.* **2003**; *24*, 1720−1729.

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

(18) Ehrenfest, P. *Z. Phys.* **1927**, *45*, 455.

(19) Feynman, R. P. *Phys. Rev.* **1939**, *56*, 340−343.

(20) Slater, J. C. *J. Chem. Phys.* **1933**, *1*, 687.

(21) Schrödinger, E. *Ann. D. Phys.* **1927**, *82*, 265.

(22) Pauli, W. *Encyclopedia of Physics;* Springer: Berlin, 1958; Vol. 5.

(23) Bader, R. F. W. *J. Chem. Phys.* **1980**, *73*, 2871−2883.

(24) Bader, R. F. W.; Nguyen-Dang, T. T. *Ad. Quantum Chem.* **1981**, *14*, 63−124.

(25) Hernández-Trujillo, J.; Bader, R. F. W. *J. Phys. Chem. A* **2000**, *104*, 1779−1794.

(26) Haaland, A.; Shorokhov, D. J.; Tverdova, N. V. *Chem. Eur. J.* **2004**, *10*, 4416−4421. See Corrigendum *Chem. Eur. J.* **2004**, *10*, 6210.

(27) Bader, R. F. W. *Phys. Rev. B* **2000**, *61*, 7795−7802.

(28) Bader, R. F. W.; Austen, M. A. *J. Chem. Phys.* **1997**, *107*, 4271−4285.

(29) Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. *J. Am. Chem. Soc.* **1987**, *109*, 7968−7979.

(30) Slater, J. C. *J. Chem. Phys.* **1972**, *57*, 2389.

(31) Cioslowski, J.; Edgington, L.; Stefanov, B. B. *J. Am. Chem. Soc.* **1995**, *117*, 10381.

(32) Abramov, Y. A. *J. Phys. Chem. A* **1997**, *101*, 5725.

(33) Matta, C. F.; Hernández-Trujillo, J.; Tang, T.-H.; Bader, R. F. W. *Chem. Eur. J.* **2003**, *9*, 1940−1951.

(34) The bond paths found between Cl atoms in substituted cyclohexanes and judged by Cioslowski et al.[31] to be repulsive are similar in all respects to those linking neighboring atoms in solid molecular chlorine, interactions that account for its layered structure. If these interactions were repulsive, solid chlorine would not exist. Those who doubt the universal identification of a bond path with bonding by citing the presence of bond paths between anions in an ionic crystal[32] should differentiate between classical models of bonding and its proper quantum mechanical description.[8] The description of an ionic crystal as one whose cohesive energy is determined by electrostatic forces between the ions together with short-range core repulsions is a classical model without foundation in quantum mechanics. Matter is a distribution of charge in real space—of pointlike nuclei embedded in the diffuse density of electronic charge, $\rho(\mathbf{r})$, defined as the expectation value of the density operator. The properties of $\rho(\mathbf{r})$ are determined by the local statements of the theorems of quantum mechanics, the most important being the Ehrenfest force theorem, determining its equation of motion, and the virial theorem, governing its energy. There is nothing in the quantum description of the energy corresponding to the classical model of interacting ions. In particular there are no resultant Ehrenfest forces acting on the atoms and no net forces, neither attractive nor repulsive, acting on the atomic nuclei, since the Feynman forces are zero and any displacement of a nucleus results in the creation of a force restoring it and its associated density to its equilibrium position. The model is useful for pedagogical purposes, but it should not be used to criticize the quantum mechanical description of bonding.

(35) Dewar, M. *Bull. Soc. Chim. Fr.* **1951**, C79.

(36) Chatt, J.; Duncanson, L. A. *J. Chem. Soc.* **1953**, 2329.

(37) Bader, R. F. W.; Matta, C. F.; Cortés-Guzmàn, F. *Organometallics* **2004**, *23*, 6253−6263.

CT049839L

JCTC Journal of Chemical Theory and Computation

Benchmark Databases for Nonbonded Interactions and Their Use To Test Density Functional Theory

Yan Zhao and Donald G. Truhlar*

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

Received December 17, 2004

Abstract: We present four benchmark databases of binding energies for nonbonded complexes. Four types of nonbonded interactions are considered: hydrogen bonding, charge transfer, dipole interactions, and weak interactions. We tested 44 DFT methods and 1 WFT method against the new databases; one of the DFT methods (PBE1KCIS) is new, and all of the other methods are from the literature. Among the tested methods, the PBE, PBE1PBE, B3P86, MPW1K, B97-1, and BHandHLYP functionals give the best performance for hydrogen bonding. MPWB1K, MP2, MPW1B95, MPW1K, and BHandHLYP give the best performances for charge-transfer interactions, and MPW3LYP, B97-1, PBE1KCIS, B98, and PBE1PBE give the best performance for dipole interactions. Finally, MP2, B97-1, MPWB1K, PBE1KCIS, and MPW1B95 give the best performance for weak interactions. Overall, MPWB1K is the best of all the tested DFT methods, with a relative error (highly averaged) of only 11%, and MPW1K, PBE1PBE, and B98 are the best of the tested DFT methods that do not contain kinetic energy density. Moving up the rungs of Jacob's ladder for nonempirical DFT, PBE improves significantly over the LSDA, and TPSS improve slightly over PBE (on average) for nonbonded interactions.

1. Introduction

One can classify interatomic interactions as bonded or nonbonded. One can further subdivide bonded interactions into ionic, metallic, covalent, coordinate covalent, and partial bonds (as at transition states), and one can subdivide nonbonded interactions into charge-transfer interactions, hydrogen bonds, dipolar interactions, dispersion (London forces), and so forth. Mixed cases are also possible, such as polar covalent (e.g., an HF bond is about 50% ionic and 50% covalent[1]) or a much more complicated range of possibilities[2] for nonbonded interactions. Nevertheless the distinctions and the broadly defined categories of interactions are useful for understanding chemical phenomena and for testing the abilities of approximate theories and models to interpret chemical phenomena.

Density functional theory (DFT[3−87] and wave function theory (WFT)[31,34,44,55,87−118] have been widely compared for their abilities to treat bonds and transition states, but comparisons are less complete for nonbonded interactions. There are two reasons for this. First, it has been realized for a long time that DFT, at least with the early functionals, is less accurate for nonbonded interactions than for bonded ones,[16−18] and this can be understood in part by the fact that current functionals are not designed to treat dispersion interactions, which are sometimes dominant in nonbonded interactions. Second, no standard databases (analogous to the G3 database[105,106,109] or Database/3[113] for bond energies, ionization potentials, and electron affinities; the latter also includes partial bond strengths as measured by barrier heights) are available for nonbonded interactions. The purpose of the present article is to remedy the latter problem and to use newly created databases for nonbonded interactions for a systematic comparison of DFT and WFT methods.

Considerable insight into how DFT works can be obtained by detailed analysis of the functionals and the Kohn−Sham electron density. In particular it should be recognized that, for molecules, the separation of exchange-correlation effects into exchange and correlation are different in WFT and

* Corresponding author e-mail: truhlar@umn.edu.

10.1021/ct049851d CCC: $30.25
Published on Web 04/23/2005

DFT.[10,21,28,46,49] In particular, DFT exchange includes a certain amount of what is called nondynamical (also called static or internal or, in certain contexts, left-right) correlation in WFT.[10,28,43,49] Handy and Cohen[49] have shown that DFT with an exchange functional (Becke88 or B88X[8]) but no correlation functional gave lower energies than Hartree−Fock for the multicenter system (for example, molecules), and they concluded that local exchange functionals must introduce nondynamical correlation. Furthermore, He et al.[46] found that even for closed-shell systems that are well described without nondynamical correlation (so-called single-reference systems), densities obtained by DFT with an exchange functional but no correlation functional look more like those obtained with fourth-order perturbation theory (MP4) than those obtained by uncorrelated Hartree−Fock.[46] They concluded that "even though the DFT exchange functional does not include any Coulomb correlation effects by construction, it simulates orbital relaxation, pair correlation, ...".[46] However the resulting electron density is too high in the van der Waals region, and correlation functionals contract the density toward high-density regions (where there is more favorable correlation energy), thus improving the description of van der Waals interactions.[46] Since correlation functionals make up for deficiencies in exchange functionals, and since the exchange functional gives a much larger contribution to molecular interactions than the correlation one, it is important that the correlation functional be well matched to the exchange functional with which it is used.

Although the usual DFT functionals do not contain dipolar dispersion interactions, there is some debate as to whether DFT methods, with either the usual functionals or new ones, might nevertheless produce useful results for the attractive interaction between rare gas atoms.[119−122] Furthermore DFT, even with the usual functionals, does contain the polarizabilities.[123] Our goal in the present paper is not, however, to pursue lines of research based on explicit inclusion of dipole polarization but rather to check which of the density functionals in current widespread use disqualify themselves by predicting unrealistic interaction potentials in regimes where the real interaction potentials are dominated by dispersion forces or other nonbonded interactions, and which density functionals yield reasonable results in such situations, for whatever reason.

In addition to lacking explicit R^{-6} terms, DFT (without Hartree−Fock exchange) predicts no interaction energy for molecules so far apart that they do not overlap (because the density is the same as for infinitely separated molecules). At the equilibrium distance of nonbonded complexes, the lack of explicit R^{-6} terms need not be a serious issue because the higher terms (R^{-8}, etc.) in the asymptotic expansion are not negligible.[124,125] Furthermore, the overlap and exchange forces are also not negligible at the equilibrium internuclear distance of nonbonded complexes.[124,126] Thus DFT is not excluded as a potentially useful theory for nonbonded interactions, as is sometimes claimed.

In summary, our goal is to understand the performance of existing density functionals for nonbonded interactions and to compare this performance to that of WFT with the same basis sets. We therefore develop four new databases for such testing:

- A hydrogen bond database
- A charge-transfer complex database
- A dipole complex database
- A weak interaction database

Whereas hydrogen bonds are dominated by electrostatic and polarization (also called induction) interactions (with a smaller contribution from charge transfer), charge-transfer complexes derive a considerable portion of their stabilization from electron transfer between the two centers. Dipole complexes involve much smaller amounts of intermolecular charge transfer and have no hydrogen bonds. Weak complexes are defined here as those that are dominated by dispersion interactions.

In the literature, there are many theoretical studies of hydrogen bonds,[29,31,39,44,47,54,69,81,84,99,101,110−112,115,127,128] charge-transfer complexes,[18,19,27,29,34,70,116] and weak interactions.[29,44,54,62,83,92,93,96,114] However there are very few studies[31,39,47,68] of dipolar interaction complexes. Several studies[39,47,68] treated $(HCl)_2$ dimer as a hydrogen bond complex, but in the present study we will treat $(HCl)_2$ dimer as a dipole interaction complex since there is no classical hydrogen bond in $(HCl)_2$ dimer.

The databases are used to test several types of DFT: (i) the local spin density approximation (LSDA, in which the density functional depends only on density), (ii) the generalized gradient approximation (GGA, in which the density functional depends on density and its reduced gradient), (iii) meta GGA (in which the functional also depends kinetic energy density), (iv) hybrid GGA (a combination of GGA with Hartree−Fock exchange), and (v) hybrid meta GGA (a combination of meta GGA with Hartree−Fock exchange). In addition we study one level of WFT: Møller−Plesset second-order perturbation theory[87] (MP2).

Section 2 explains the theories, databases, and functionals used in the present work. Section 3 presents results and discussion, and section 4 has concluding remarks.

2. Theory and Databases

2.1. Weizmann 1 (W1) Theory. It is difficult to extract the zero-point-exclusive binding energies D_e from experiment for nonbonded complexes due to the uncertainties in the experimental ground-state dissociation energy D_0 and due to the uncertain effect of anharmonicity on the zero point vibrational energy of these loose complexes. To obtain the best estimates for the binding energies in the new database, we employed the W1 method for most of the nonbonded complexes, and we also took some theoretical and experimental results from the literature.

W1 theory was developed by Martin and Oliveira, and it is a method designed to extrapolate to the complete basis limit of a CCSD(T)[89] calculation. Thus W1 theory should be good enough for obtaining best estimates of binding energies of these nonbonded complexes. Boese et al.[68,80] have already used W1 and W2 theory to calculate best estimates for some hydrogen bonding dimers, and we will employ W1 theory for several more nonbonded complexes in the present

Table 1. Components of W1 Calculations for Binding Energies D_e (kcal/mol)

complex	SCF[a] limit	CCSD[b] limit	(T)[c] limit	core corr & relativistic	final D_e
			Hydrogen Bonding		
$(HCONH_2)_2$	9.98	3.93	0.97	0.06	14.94
$(HCOOH)_2$	12.05	3.15	0.94	0.00	16.15
			Charge Transfer		
$C_2H_4\cdots F_2$	−1.24	1.90	0.40		1.06
$NH_3\cdots F_2$	−1.21	2.45	0.57		1.81
$C_2H_2\cdots ClF$	−0.64	3.47	0.98		3.81
$HCN\cdots ClF$	1.73	2.50	0.64		4.86
$NH_3\cdots Cl_2$	0.62	3.27	0.98		4.88
$H_2O\cdots ClF$	2.33	2.34	0.68		5.36
$NH_3\cdots ClF$	2.37	6.34	1.90		10.62
			Dipole Interaction		
$(H_2S)_2$	−0.18	1.51	0.32	0.00	1.65
$HCl\cdots H_2S$	0.70	2.14	0.49	0.02	3.35
$CH_3Cl\cdots HCl$	0.39	2.57	0.59		3.55
$HCN\cdots CH_3SH$	1.53	1.64	0.42		3.59
$CH_3SH\cdots HCl$	0.10	3.62	0.43		4.16
			Weak Interaction		
$CH_4\cdots Ne$	−0.09	0.27	0.03	0.01	0.22
$(CH_4)_2$	−0.53	0.91	0.13	0.00	0.51
$(C_2H_2)_2$	0.27	0.90	0.19	−0.01	1.34
$(C_2H_4)_2$	−0.92	1.99	0.35	−0.01	1.42

[a] Hartree−Fock. [b] Coupled clusters theory with single and double excitations. [c] Quasiperturbative triple excitations.

work. The strengths and limitations of W1 theory have been described elsewhere.[97,104,107,108,118]

2.2. HB6/04 Database. The hydrogen bond database consists of binding energies of six hydrogen bonding dimers, namely $(NH_3)_2$, $(HF)_2$, $(H_2O)_2$, $NH_3\cdots H_2O$, $(HCONH_2)_2$, and $(HCOOH)_2$. The binding energies of $(NH_3)_2$, $(HF)_2$, $(H_2O)_2$, and $NH_3\cdots H_2O$ are taken from Boese and Martin's[80] W2 calculations. The best estimates of D_e for $(HCONH_2)_2$ and $(HCOOH)_2$ are calculated here by the W1 theory. This database is called the HB6/04 database.

2.3. CT7/04 Database. The charge transfer (CT) database consists of binding energies of seven charge-transfer complexes, in particular $C_2H_4\cdots F_2$, $NH_3\cdots F_2$, $C_2H_2\cdots ClF$, $HCN\cdots ClF$, $NH_3\cdots Cl_2$, $H_2O\cdots ClF$, and $NH_3\cdots ClF$. The best estimates of D_e for all complexes in the charge-transfer database are calculated here by the W1 model. This database is called the CT7/04 database.

2.4. DI6/04 Database. The dipole interaction (DI) database consists of binding energies of six dipole inteaction complexes: $(H_2S)_2$, $(HCl)_2$, $HCl\cdots H_2S$, $CH_3Cl\cdots HCl$, $CH_3SH\cdots HCN$, and $CH_3SH\cdots HCl$. The binding energy of $(HCl)_2$ is taken from Boese and Martin's[80] W2 calculation. The best estimates of D_e for the other complexes in the dipole interaction database are calculated here by the W1 theory. This database is called the DI6/04 database.

2.5. WI9/04 Database. The weak interaction database consists of binding energies of weak interaction complexes, namely HeNe, HeAr, Ne_2, NeAr, $CH_4\cdots Ne$, $C_6H_6\cdots Ne$, $(CH_4)_2$, $(C_2H_2)_2$, and $(C_2H_4)_2$. The binding energies of HeNe, HeAr, Ne_2, and NeAr are taken from Ogilvie and Wang's[129,130] analysis. The binding energy of $C_6H_6\cdots Ne$ is taken from Cappelleti et al.'s[131] experimental study. The best estimates of D_e for $CH_4\cdots Ne$, $(CH_4)_2$, $(C_2H_2)_2$, and $(C_2H_4)_2$ are calculated by W1 theory. This database is called the WI9/04 database.

2.6. AE6 Benchmark Database. We parametrized one new hybrid meta GGA method, namely PBE1KCIS (see Table 3). It has one parameter, the fraction X of Hartree−Fock exchange, and this was optimized against the AE6[132] benchmark database of atomization energies for six covalently bonded nonmetallic molecules. We have previously used this database as a training set to optimize the MPW1B95,[78] TPSS1KCIS,[85] and MPW1KCIS[86] methods.[75] The AE6 database is listed for reference in the Supporting Information.

To parametrize the PBE1KCIS model, we optimize the fraction of Hartree−Fock exchange, X, to minimize the root-mean-square error (RMSE) for the six data in the AE6 database. The optimized X parameter for the PBE1KCIS method is given in Table 3.

2.7. Theoretical Methods Tested. We tested a number of DFT-type methods against the new four-part (HB6, CT7, DI6, WI9) nonbonded-interaction database. In particular, we assessed three LSDAs: SVWN3,[5,133] SVWN5,[5,133] and SPWL.[12,133] We tested twelve GGAs: BP86,[7,8] BLYP,[8,9] BPW91,[8,11] BPBE,[8,22] mPWPBE,[22,134] G96LYP,[9,20] HCTH,[33] mPWLYP,[9,29] mPWPW91,[29] OLYP,[9,49] PBE,[22] and XLYP.[73] We tested seven meta GGA methods: BB95,[21] mPWB95,[21,29]

Table 2. Benchmark Databases of Binding Energies D_e (kcal/mol) for Hydrogen Bonding (HB), Charge Transfer (CT), Dipole Interaction (DI), and Weak Interaction (WI)

HB6/04			CT7/04			DI6/04			WI9/04		
complex	D_e	ref	complex	D_e	ref	complex	D_e	ref	complex	D_e	ref
$(NH_3)_2$	3.15	80	$C_2H_4\cdots F_2$	1.06	this work	$(H_2S)_2$	1.66	this work	HeNe	0.04	130
$(HF)_2$	4.57	80	$NH_3\cdots F_2$	1.81	this work	$(HCl)_2$	2.01	80	HeAr	0.06	130
$(H_2O)_2$	4.97	80	$C_2H_2\cdots ClF$	3.81	this work	$HCl\cdots H_2S$	3.35	this work	Ne_2	0.08	129
$NH_3\cdots H_2O$	6.41	80	$HCN\cdots ClF$	4.86	this work	$CH_3Cl\cdots HCl$	3.55	this work	NeAr	0.13	130
$(HCONH_2)_2$	14.94	this work	$NH_3\cdots Cl_2$	4.88	this work	$HCN\cdots CH_3SH$	3.59	this work	$CH_4\cdots Ne$	0.22	this work
$(HCOOH)_2$	16.15	this work	$H_2O\cdots ClF$	5.36	this work	$CH_3SH\cdots HCl$	4.16	this work	$C_6H_6\cdots Ne$	0.47	131
			$NH_3\cdots ClF$	10.62	this work				$(CH_4)_2$	0.51	this work
									$(C_2H_2)_2$	1.34	this work
									$(C_2H_4)_2$	1.42	this work
average	8.37			4.63			3.05			0.47	

Table 3. Summary of the DFT Methods Tested (in Chronological Order)

method	X^a	year	type	ex functional[b] corr functional[c]	refs
SVWN3[d]	0	1981	LSDA	Slater's local ex VWN no.3	5, 133
SVWN5[e]	0	1981	LSDA	Slater's local ex VWN no.5	5, 133, 26
BP86	0	1988	GGA	Becke88 Perdew's 1986 GGA	7, 8
BLYP	0	1988	GGA	Becke88 Lee−Yang−Parr	8, 9
SPWL	0	1992	LSDA	Slater's local ex Perdew-Wang local	12, 133
BPW91	0	1992	GGA	Becke88 Perdew-Wang91	8, 11
B3P86	20	1993	hybrid GGA	Becke88 Perdew's 1986 GGA	7, 8
B3PW91	20	1993	hybrid GGA	Becke88 Perdew-Wang91	8, 11
BHandHLYP	50	1993	hybrid GGA	Becke88 Lee−Yang−Parr	8, 9, 135
B3LYP	20	1994	hybrid GGA	Becke88 Lee−Yang−Parr	8, 9, 15
G96LYP	0	1996	GGA	Gill96 Lee−Yang−Parr	9, 20
BB95	0	1996	meta GGA	Becke88 Becke95	8, 21
B1B95	28	1996	hybrid meta GGA	Becke88 Becke95	8, 21
PBE	0	1996	hybrid GGA	PBE ex PBE corr	22
PBE1PBE[f]	25	1996	hybrid GGA	PBE ex PBE corr	22
BPBE	0	1996	GGA	Becke88 PBE corr	8, 22
mPWPW91[g]	0	1998	GGA	modified PW91 Perdew-Wang91	11, 29
mPW1PW91[h]	25	1998	hybrid GGA	modified PW91 Perdew-Wang91	11, 29
mPWPBE	0	1998	GGA	modified PW91 PBE corr	22, 29
mPWLYP	0	1998	GGA	modified PW91 Lee−Yang−Parr	9, 29
mPWB95	0	1998	meta GGA	modified PW91 Becke95	21, 29
VSXC	0	1998	meta GGA	VSXC ex VSXC corr	32
HCTH	0	1998	GGA	HCTH ex HCTH corr	33
B97-1	21	1998	hybrid GGA	B97-1 ex B97-1 corr	33
B98	21.98	1998	hybrid GGA	B98 ex B98 corr	30
MPW1K	42.8	2000	hybrid GGA	modified PW91 Perdew-Wang91	29, 24
B97-2	21	2001	hybrid GGA	B97-2 ex B97-2 corr	33
OLYP	0	2001	GGA	OPTX Lee−Yang−Parr	9, 49, 50
O3LYP	11.61	2001	hybrid GGA	OPTX Lee−Yang−Parr	9, 49, 50
TPSS	0	2003	meta GGA	TPSS ex TPSS corr	69, 71
TPSSh	10	2003	hybrid meta GGA	TPSS ex TPSS corr	69, 71
MPWKCIS	0	2004	hybrid meta GGA	modified PW91 KCIS corr	21, 37, 38, 60
PBEKCIS	0	2004	hybrid meta GGA	TPSS ex KCIS corr	22, 37, 38, 60
TPSSKCIS	0	2004	hybrid meta GGA	TPSS ex KCIS corr	37, 38, 60, 69, 71

Table 3 (Continued)

method	X[a]	year	type	ex functional[b] corr functional[c]	refs
XLYP	0	2004	GGA	Becke88+PW91 Lee-Yang-Parr	8, 9, 11, 73
X3LYP	21.8	2004	hybrid GGA	Becke88+PW91 Lee−Yang−Parr	8, 9, 11, 73
BB1K	42	2004	hybrid meta GGA	Becke88 Becke95	8, 21, 75
MPW3LYP	21	2004	hybrid GGA	modified PW91 Lee−Yang−Parr	9, 29, 78
MPW1B95	31	2004	hybrid meta GGA	modified PW91 Becke95	21, 29, 78
MPWB1K	44	2004	hybrid meta GGA	modified PW91 Becke95	21, 29, 78
TPSS1KCIS	13	2004	hybrid meta GGA	TPSS ex KCIS corr	37, 38, 60, 69, 71, 85
MPW1KCIS	15	2004	hybrid meta GGA	modified PW91 KCIS corr	21, 37, 38, 60, 86
MPWKCIS1K	41	2004	hybrid meta GGA	modified PW91 KCIS corr	21, 37, 38, 60, 86
PBE1KCIS	22	2005	hybrid meta GGA	PBE ex KCIS corr	22, 37, 38, 60, this work

[a] X denotes the percentage of HF exchange in the functional. [b] Upper entry. [c] Lower entry. [d] Also called SVWN and SVWNIII. [e] Also called SVWN (expression V) where the final V is Roman numeral 5. [f] Also called PBE0. [g] Also called mPWPW. [h] Also called mPW1PW, mPW0, and MPW25.

mPWKCIS,[29,37,38,60,69,71] PBEKCIS,[22,37,38,60] TPSS,[69,71] TPSSKCIS,[37,38,60,69,71] and VSXC.[32] We assessed thirteen hybrid GGA methods: B3LYP,[9,14,15] B3P86,[7,14] B3PW91,[11,14] BHandHLYP,[135] B97-1,[33] B97-2,[53] B98,[30] mPW1PW91,[29] MPW1K,[45] MPW3LYP,[9,29,78] O3LYP,[49,50] PBE1PBE,[22]and X3LYP,[73] and we also assessed nine hybrid meta GGA methods: B1B95,[21] BB1K,[75] MPW1B95,[78] MPWB1K,[78] MPW1KCIS,[86] MPWKCIS1K,[86] PBE1KCIS, TPSS1KCIS,[37,38,60,69,71,85] and TPSSh.[69,71]

Since the theory behind the various DFT functionals is explained in the original paper, we refer the readers to the original references for those details; however, a few comments are needed about these LSDAs. All these use the same exchange functional, due to Dirac and Slater,[133] but they differ in the correlation functional. Although SVWN3 is incorporated (apparently by mistake) in the popular B3LYP functional, it is based upon the inaccurate random phase approximation for the electron gas of uniform density, where VWN5 and PWL are more properly (from a theoretical point of view) based on the homogeneous electron gas results of Ceperly and Alder.[136] The difference between VWN3 and VWN5 has been discussed, for example, by Hertwig and Koch.[26] A complete summary of all DFT methods considered in this article is given in Table 3.

We also tested one ab initio WFT method: MP2.[87]

2.8. Computational Methods. All MP2, W1, and DFT calculations were carried out using the *Gaussian03*[137] and MOLPRO[138] programs, but a few aspects of the calculations require some comments.

PBE1KCIS is not available with standard keywords in *Gaussian03*; the keywords required in *Gaussian03* to carry out the PBE1KCIS calculation are:

#PBEKCIS

IOp(3/76= 0780002200)

Some workers have had difficulty with SCF convergence when using the B95 correlation functional, and the usefulness of the functional has been questioned for this reason. However, we have discovered the source of this problem, and it is easily remedied without changing the functional. In the B95 functional,[21] the parallel-spin component of the correlation energy is given in terms of the uniform electron gas (UEG) expression by

$$E_C^{\sigma\sigma} = \frac{D_\sigma}{\frac{3}{5}(6\pi^2)^{2/3}\rho_\sigma^{5/3}} E_{C\sigma\sigma}^{UEG}$$

where D_σ is a function of kinetic energy density and the reduced gradient density, and ρ_σ is the electron density for spin σ. This energy expression vanishes when the density goes to zero, and *Gaussian03* sets it equal to zero when the ρ_σ is less than 10^{-15}; however, for values of the ρ_σ in the range $10^{-15}-10^{-6}$, $E_C^{\sigma\sigma}$ is still negligible, but eq 1 has numerical instabilities because it is $\sim$0/0. Changing the criterion for neglecting $E_C^{\sigma\sigma}$ from a ρ_σ of 10^{-15} to a ρ_σ of 10^{-6} solves the problem. This can be accomplished by changing the density tolerance variable DTol in subroutine bc95ss.f from 10^{-15} to 10^{-6}. We have checked that this eliminates problems for counterpoise calculations, ionization potential calculations, metal−metal bonds, and metal−ligand bonds. The effect on the computed results is negligible, for example, an average error of 10^{-8} hartrees for the data in Database/3[64,113] and $< 10^{-4}$ cm^{-1} for the vibrational frequencies of methanethiol.

All DFT calculations use ultrafine grids in *Gaussian03*.

Except in the final table, which involves DFT geometry optimization with 20 different functionals, geometries for all molecules in this paper are optimized at the MC-QCISD/3 level, where MC-QCISD is the multicoefficient QCISD method,[102,113] which is one of the most cost efficient of the multicoefficient correlation methods (MCCMs). Figures 1−4 show the geometries of the complexes studied in the present paper. All W1 and MP2 calculations use the MC-QCISD/3 geometries, and we also present results for all 44 DFT methods with these geometries. The MC-QCISD/3 geometries for all monomers and complexes in this paper can be obtained from the Truhlar group database Web site.[139]

We tested all DFT methods in Table 3 with three basis sets: DIDZ (which denotes 6-31+G(d,p)),[88] aug-cc-pVTZ,[90]

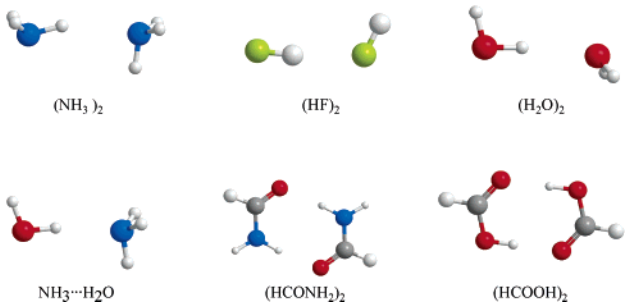


Figure 1. Geometries of the dimers in the HB6/04 database.

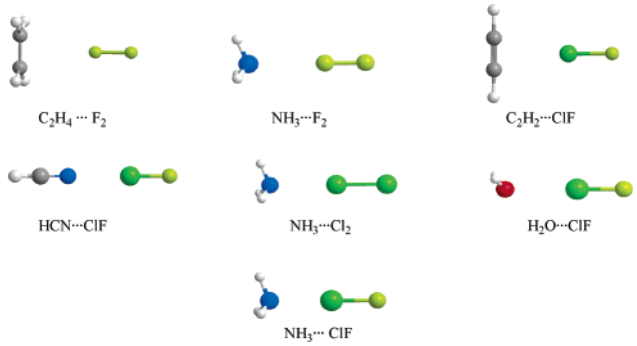


Figure 2. Geometries of the complexes in the CT7/04 database.

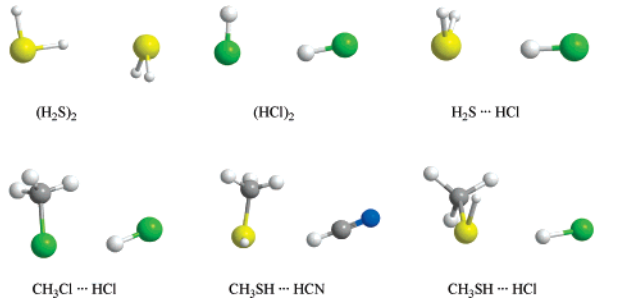


Figure 3. Geometries of the dimers in the DI6/04 database.

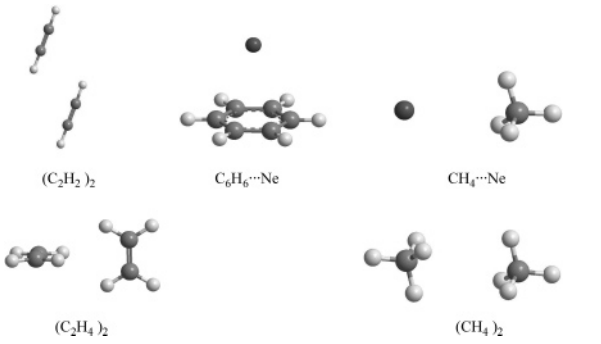


Figure 4. Geometries of the dimers in the WI9/04 database.

and MG3S. The MG3S basis[64] is the same as MG3 except that it omits diffuse functions on hydrogens. MG3 is the modified[99,101] G3Large[96] basis set. It is also called the G3LargeMP2[99] basis set, and it is the same as 6-311++G-(3d2f, 2df, 2p)[140] for H−Si but improved[95] for P−Ar.

2.9. Counterpoise Corrections. For all nonbonded complexes, we perform calculations with and without the counterpoise corrections[141,142] for basis set superposition error (BSSE).

3. Results and Discussion

3.1.W1 Results. Table 1 summarizes the 17 new W1 calculations carried out for the new databases. From this table, it can be seen that the Hartree−Fock (HF) component of the calculations underestimates all binding energies; the magnitudes of CCSD correlation contributions to the binding energies for the two hydrogen bonding dimers in Table 1 are in a range from 3 to 4 kcal/mol, and those for the seven charge-transfer complexes are in a range from 1.9 to 6.3 kcal/mol. The (T) correlation contributes about 1 kcal/mol to the binding energies for the two hydrogen bonding cases and from 0.4 to 1.9 kcal/mol for the charge-transfer cases. Note that HF theory gives negative binding energies for the $C_2H_4\cdots F_2$, $NH_3\cdots F_2$, and $C_2H_2\cdots ClF$ complexes, that is, the complex is not bound at the MC-QCISD/3 geometry.

The magnitudes of CCSD correlation contributions to the binding energies for the five dipole interaction complexes are in a range from 1.5 to 3.6 kcal/mol, whereas the magnitudes of the (T) correlation contributions are in a range from 0.3 to 0.6 kcal/mol, which makes the (T) contribution relatively less important for the dipole interaction cases than for the previous two types. The validity of various DFT methods might be correlated with the relative importance of (T) contributions because the relative importance of (T) contributions might be a rough measure of multireference character, and some DFT methods are better than others for multireference cases.

The magnitudes of CCSD correlation contributions to the binding energies for the four weak interaction complexes are in a range from 0.3 to 2 kcal/mol, and the magnitudes of (T) correlation contributions are in a range from 0.0 to 0.4 kcal/mol, again relatively small. Note that HF theory give negative binding energies for $CH_4\cdots Ne$, $(CH_4)_2$, and $(C_2H_4)_2$. This agrees with Tsuzuki and Luthi's theoretical study.[54] They have shown that HF predicts purely repulsive intermolecular potentials for the $(CH_4)_2$ and $(C_2H_4)_2$ dimers.

3.2. Benchmark Databases for Nonbonded Interactions. The new databases are presented in Table 2. The magnitudes of the binding energies are in the range 3.2 to 16.2 kcal/mol for the HB6/04 database, 1.1 to 10.6 kcal/mol for the CT7/04 database, 1.7 to 4.2 kcal/mol for the DI6/04 database, and 0.04 to 1.42 kcal/mol for the WI9/04 database. The last row gives the average binding energy for each type, but we note that each database in Table 2 contains strong, medium, and weak complexes for each particular kind of nonbonded interaction. Furthermore, the CT7/04, DI6/04, and WI9/04 databases contain complexes for the first and second row atoms. In the WI9/04 database, we have rare gas−rare gas complexes (HeAr, NeAr, etc.), a rare gas−π interaction complex (C_6H_6−Ne), an sp^2−sp^2 interaction complex ($(C_2H_4)_2$), an sp−sp interaction complex ($(C_2H_2)_2$), and a rare gas−sp^3 interaction (CH_4−Ne). These diverse data for nonbonded interactions were especially chosen to be suitable for testing theoretical methods.

3.3. Tests of Theoretical Methods. The mean errors of the tested methods are listed in Tables 4−8. In these tables we tabulate the mean unsigned error (MUE, also called mean absolute deviation) and mean signed error (MSE). We use "no-cp" to denote calculations without the counterpoise

correction for the BSSE, and we use "cp" to denote calculations that do include the counterpoise correction for the BSSE. The calculated binding energies with the MG3S basis set are listed in the Supporting Information.

To make it a little easier to put the large number of results in this paper in perspective, we define the following quantities:

$$\mathrm{MMUE} = [\mathrm{MUE(no\text{-}cp)} + \mathrm{MUE(cp)}]/2$$

$$\mathrm{MMMUE} = [\mathrm{MMUE(DIDZ)} + \mathrm{MMUE(aug\text{-}cc\text{-}pVTZ)} + \mathrm{MMUE(MG3S)}]/3$$

$$\mathrm{MMMMUE} = [\mathrm{MMMUE(HB)} + \mathrm{MMMUE(CT)} + \mathrm{MMMUE(DI)} + \mathrm{MMMUE(WI)}]/4$$

Our discussion will focus *mainly* on the highly averaged MMMUEs and MMMMUEs because they provide measures of broad usefulness for various kinds of calculations. Specialists interested in one or another subsets of the results are invited to make their own comparisons and draw their own conclusions.

3.3.1. Results for Hydrogen Bonding. Table 4 summarizes the results for the hydrogen bonding calculations. Among the LSDA and GGA methods, PBE gives the lowest MMMUE for binding energies in the HB6/04 database. VSXC is the best meta GGA, B3P86, and PBE1PBE are the best hybrid GGAs, and PBE1KCIS is the best hybrid meta GGA for hydrogen bonding calculation. Overall, PBE1PBE, B3PB86, and PBE give the best performance for calculating the binding energies for the hydrogen bonding dimers in the HB6/04 database. This result is consistent with the results of Ireta and co-workers.[82] They used ab initio pseudopotentials, a plane wave basis set, and periodic boundary conditions to integrate the Kohn−Sham equations for the calculation of the energetics of several hydrogen bonded dimers, and they found that PBE gives very good accuracy.

From Table 4, we can also see that B3P86 and MPW1K are the best performers for the DIDZ basis set. Good performance for the small basis is important because one of the attractive features of DFT is its applicability to large systems, for which larger basis sets can be cost prohibitive. Note that MP2 only works well with the aug-cc-pVTZ basis, and this will limit its application to large systems.

Table 4 shows that the two DFT methods that contain the OPTX exchange functional, OLYP and O3LYP, give very bad results for hydrogen bonding dimers. OLYP is even worse than LSDA. Note that in one of our previous papers, we found that OLYP performs well for atomization energy calculations on chemically bonded systems. From this point of view, we conclude that the OPTX exchange functional may have been fitted to a too restricted set of data; anyway it is not suitable for hydrogen bonding calculation.

3.3.2. Results for Charge Transfer. Table 5 summarizes the results for the charge-transfer calculations. Among the LSDA and GGA methods, G96LYP give the lowest MMMUE for calculating the binding energies in the CT7/04 database. BB95 is the best meta GGA, BHandHLYP and MPW1K are the best hybrid GGAs, and MPWB1K is the best hybrid meta GGA for charge-transfer calculations. Overall, MPWB1K and MPW1B95 give the best performance for calculating the binding energies of the charge-transfer complexes, and both of them outperform MP2. This is encouraging because MPWB1K and MPW1B95 are among the newest functionals, and so far they have proved to be good general-purpose functionals.

Table 5 shows that all LSDA, GGA, and meta GGA functionals systematically overestimate the binding energies of the complexes in the CT7/04 database. Ruiz et al.[19] have pointed out that "the wrong asymptotic behavior and the derivative discontinuity of the exchange and correlation energy in DFT lead to a reduction of the HOMO−LUMO gap ...". The small HOMO−LUMO gap leads to too much charge transfer and is the ultimate cause of the overestimation of the strength of the charge-transfer interaction. Inclusion of HF exchange in the DFT calculation improves the performance,[18] as shown here by the low MMMUE obtained by some hybrid and hybrid meta GGA methods.

From Table 5, we can see that MPWB1K and BHandHLYP are the best performers for the DIDZ basis set.

3.3.3. Results for Dipole Interaction. Table 6 summarizes the results for the dipole interaction complexes. Among the LSDA and GGA methods, mPWLYP gives the lowest MMMUE for calculating the binding energies in the DI6/04 database. PBEKCIS is the best meta GGA, B97-1, and MPW3LYP are the best hybrid GGAs, and PBE1KCIS is the best hybrid meta GGA for dipole interaction calculations. Overall, B97-1 and MPW3LYP give the best performance for calculating the binding energies for the dipole interaction complexes in the DI6/04 database.

Since hydrogen bonding interactions and dipole interactions are physically very similar, the good performers for hydrogen bonding, PBE and PBE1PBE, also work well for the dipole interaction calculation

From Table 6, we can see that MPW3LYP is the best performer for the DIDZ basis set.

3.3.4. Results for Weak Interaction. Table 7 summarizes the results for the weak interaction complexes. PBEKCIS is the best meta GGA, B97-1 is the best hybrid GGA, and MPWB1K is the best hybrid meta GGA for weak interactions. Overall, B97-1 and MP2 give the best performance for calculating the binding energies for the weak interaction complexes in the WI9/04 database.

In principle, a DFT method can yield the exact ground-state energy, including long range van der Waals (vdW) energies. However, in 1998, Kohn et al. pointed out, "the commonly used LDA and GGA, designed for non-uniform electron gases, fail to capture the essence of vdW energies".[143] It seems that this statement is true because the best method, B97-1, has a mean unsigned error (0.19 kcal/mol) that is 40% of the mean value of the binding energy (0.47 kcal/mol, see Table 2). However, even MP2 has a 34% error by this kind of assessment, so from the practical point of view, DFT is almost competitive in accuracy at a much lower cost. Other methods with errors of less than 50% are MPWB1K and MPW1B95. We have recently pointed out[78] that the X and mPW exchange functionals have similar

Table 4. Mean Errors (kcal/mol) for the HB6/04 Database[a,b]

	DIDZ					aug-cc-pVTZ					MG3S					
	MSE		MUE			MSE		MUE			MSE		MUE			
method	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	MMMUE
LSDA																
SPWL	4.41	4.29	4.41	4.29	4.35	5.44	4.72	5.44	4.72	5.08	4.64	4.20	4.64	4.20	4.42	4.62
SVWN5	5.45	4.73	5.45	4.73	5.09	4.42	4.30	4.42	4.30	4.36	4.65	4.21	4.65	4.21	4.43	4.63
SVWN3	5.69	4.98	5.69	4.98	5.34	4.67	4.55	4.67	4.55	4.61	4.90	4.46	4.90	4.46	4.68	4.87
GGA																
PBE	1.02	0.32	1.09	0.65	0.87	0.01	−0.09	0.26	0.25	0.26	0.22	−0.19	0.45	0.32	0.39	0.50
mPWLYP	0.70	−0.02	1.11	0.88	0.99	−0.29	−0.40	0.50	0.56	0.53	−0.08	−0.49	0.65	0.62	0.64	0.72
mPWPW91	0.23	−0.48	0.85	0.71	0.78	−0.52	−0.96	0.57	0.96	0.77	−0.52	−0.96	0.57	0.96	0.77	0.77
mPWPBE	0.20	−0.52	0.84	0.72	0.78	−0.80	−0.93	0.80	0.93	0.86	−0.56	−1.00	0.60	1.00	0.80	0.82
BP86	0.01	−0.65	0.74	0.78	0.76	−0.90	−1.01	0.90	1.01	0.95	−0.72	−1.10	0.72	1.10	0.91	0.88
XLYP	0.12	−0.56	1.07	0.85	0.96	−0.84	−0.96	0.84	0.96	0.90	−0.65	−1.04	0.74	1.04	0.89	0.92
BLYP	−0.42	−1.10	1.04	1.12	1.08	−1.37	−1.48	1.37	1.48	1.42	−1.18	−1.56	1.18	1.56	1.37	1.29
BPW91	−0.90	−1.58	1.13	1.58	1.36	−1.84	−1.97	1.84	1.97	1.91	−1.64	−2.05	1.64	2.05	1.84	1.70
BPBE	−0.93	−1.62	1.15	1.62	1.39	−1.89	−2.01	1.89	2.01	1.95	−1.68	−2.09	1.68	2.09	1.88	1.74
HCTH	−0.78	−1.51	1.47	1.51	1.49	−1.83	−1.94	1.83	1.94	1.88	−1.68	−2.12	1.68	2.12	1.90	1.76
G96LYP	−2.24	−2.87	2.24	2.87	2.55	−3.11	−3.22	3.11	3.22	3.17	−2.95	−3.30	2.95	3.30	3.13	2.95
OLYP	−2.67	−3.50	2.67	3.50	3.08	−3.83	−3.95	3.83	3.95	3.89	−3.60	−4.09	3.60	4.09	3.84	3.60
Meta GGA																
VSXC	0.26	−0.46	0.70	0.64	0.67	−0.46	−0.59	0.47	0.60	0.54	−0.34	−0.77	0.45	0.79	0.62	0.61
TPSS	0.42	−0.28	0.77	0.60	0.69	−0.59	−0.73	0.59	0.73	0.66	−0.37	−0.82	0.45	0.82	0.63	0.66
PBEKCIS	0.59	−0.09	1.09	0.86	0.98	−0.37	−0.47	0.48	0.54	0.51	−0.19	−0.58	0.62	0.62	0.62	0.70
TPSSKCIS	0.35	−0.36	0.90	0.70	0.80	−0.67	−0.80	0.67	0.80	0.74	−0.46	−0.89	0.55	0.89	0.72	0.75
mPWB95	0.05	−0.65	0.98	0.80	0.89	−0.96	−1.09	0.96	1.09	1.02	−0.72	−1.12	0.72	1.12	0.92	0.94
mPWKCIS	−0.22	−0.92	1.05	1.01	1.03	−1.17	−1.30	1.17	1.30	1.24	−0.97	−1.39	0.97	1.39	1.18	1.15
BB95	−1.08	−1.74	1.24	1.74	1.49	−2.04	−2.16	2.04	2.16	2.10	−1.83	−2.21	1.83	2.21	2.02	1.87
Hybrid GGA																
B3P86	0.70	0.06	0.83	0.37	0.60	−0.23	−0.34	0.40	0.42	0.41	−0.05	−0.44	0.33	0.44	0.38	0.46
PBE1PBE	1.01	0.33	1.03	0.55	0.79	−0.02	−0.13	0.28	0.27	0.27	0.19	−0.23	0.40	0.28	0.34	0.47
BHandHLYP	0.76	0.14	0.93	0.56	0.74	−0.23	−0.34	0.36	0.40	0.38	−0.04	−0.43	0.42	0.48	0.45	0.52
B97-1	0.75	0.10	0.94	0.71	0.82	−0.21	−0.31	0.29	0.36	0.32	−0.04	−0.43	0.45	0.45	0.45	0.53
MPW1K	0.65	−0.03	0.84	0.45	0.64	−0.40	−0.53	0.44	0.53	0.49	−0.17	−0.61	0.33	0.61	0.47	0.53
X3LYP	0.73	0.08	0.95	0.65	0.80	−0.24	−0.35	0.34	0.41	0.37	−0.05	−0.44	0.45	0.48	0.47	0.55
MPW3LYP	1.04	0.38	1.12	0.69	0.91	0.06	−0.05	0.35	0.34	0.35	0.26	−0.14	0.51	0.41	0.46	0.57
B98	0.54	−0.12	0.87	0.68	0.78	−0.44	−0.54	0.44	0.54	0.49	−0.26	−0.66	0.45	0.66	0.55	0.61
mPW1PW91	0.43	−0.25	0.75	0.54	0.65	−0.58	−0.72	0.58	0.72	0.65	−0.36	−0.79	0.39	0.79	0.59	0.63
B3LYP	0.22	−0.42	0.86	0.69	0.78	−0.73	−0.84	0.73	0.84	0.78	−0.55	−0.93	0.60	0.93	0.76	0.77
B3PW91	−0.26	−0.92	0.78	0.94	0.86	−1.23	−1.35	1.23	1.35	1.29	−1.03	−1.43	1.03	1.43	1.23	1.13
B97-2	−0.43	−1.11	0.95	1.11	1.03	−1.43	−1.55	1.43	1.55	1.49	−1.22	−1.64	1.22	1.64	1.43	1.32
O3LYP	−1.86	−2.63	1.86	2.63	2.25	−2.99	−3.10	2.99	3.10	3.04	−2.77	−3.23	2.77	3.23	3.00	2.76
Hybrid Meta GGA																
PBE1KCIS	0.59	−0.06	0.93	0.73	0.83	−0.37	−0.48	0.40	0.48	0.44	−0.20	−0.59	0.49	0.59	0.54	0.60
MPWB1K	0.52	−0.13	0.79	0.55	0.67	−0.53	−0.65	0.53	0.65	0.59	−0.31	−0.70	0.41	0.70	0.56	0.61
TPSSh	0.45	−0.24	0.76	0.56	0.66	−0.57	−0.71	0.57	0.71	0.64	−0.36	−0.80	0.41	0.80	0.60	0.63
TPSS1KCIS	0.39	−0.30	0.84	0.65	0.74	−0.63	−0.76	0.63	0.76	0.70	−0.43	−0.86	0.49	0.86	0.67	0.71
MPW1B95	0.34	−0.32	0.79	0.61	0.70	−0.69	−0.81	0.69	0.81	0.75	−0.46	−0.86	0.50	0.86	0.68	0.71
MPWKCIS1K	0.20	−0.45	0.79	0.64	0.72	−0.78	−0.91	0.78	0.91	0.85	−0.59	−1.00	0.59	1.00	0.80	0.79
MPW1KCIS	−0.11	−0.78	0.95	0.87	0.91	−1.07	−1.19	1.07	1.19	1.13	−0.87	−1.28	0.87	1.28	1.07	1.04
BB1K	−0.18	−0.82	0.72	0.82	0.77	−1.20	−1.56	1.20	1.56	1.38	−0.99	−1.37	0.99	1.37	1.18	1.11
B1B95	−0.52	−1.16	0.89	1.16	1.02	−1.52	−1.63	1.52	1.63	1.58	−1.31	−1.69	1.31	1.69	1.50	1.37
ab Initio WFT																
MP2	0.48	−1.20	0.88	1.20	1.04	0.28	−0.44	0.28	0.44	0.36	0.24	−0.93	0.26	0.93	0.60	0.66

[a] MUE denotes mean unsigned error (also called mean absolute deviation). MSE denotes mean signed error. MMUE = [MUE(no-cp) + MUE(cp)]/2. MMMUE = [MMUE(DIDZ) + MMUE(aug-cc-pVTZ) + MMUE(MG3S)]/3. [b] We use "no-cp" to denote the calculation without the counterpoise correction for the BSSE and use "cp" to denote the calculation with the counterpoise correction for the BSSE. DIDZ denotes 6-31+G(d,p) basis.

Table 5. Mean Errors (kcal/mol) for the CT7/04 Database[a,b]

	DIDZ					aug-cc-pVTZ					MG3S					
	MSE		MUE			MSE		MUE			MSE		MUE			
method	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	MMMUE
LSDA																
SPWL	6.41	6.30	6.41	6.30	6.36	7.59	6.89	7.59	6.89	7.24	6.78	6.41	6.78	6.41	6.59	6.73
SVWN5	7.59	6.89	7.59	6.89	7.24	6.42	6.30	6.42	6.30	6.36	6.78	6.41	6.78	6.41	6.60	6.73
SVWN3	7.80	7.10	7.80	7.10	7.45	6.63	6.51	6.63	6.51	6.57	6.99	6.62	6.99	6.62	6.80	6.94
GGA																
G96LYP	0.61	0.01	1.59	1.38	1.48	−0.42	−0.51	1.24	1.28	1.26	−0.18	−0.46	1.20	1.28	1.24	1.33
HCTH	1.77	1.17	2.04	1.59	1.81	0.71	0.63	1.26	1.25	1.26	0.94	0.63	1.31	1.22	1.26	1.45
BPBE	1.88	1.28	2.13	1.60	1.87	0.79	0.69	1.21	1.18	1.19	1.12	0.79	1.37	1.22	1.29	1.45
BPW91	1.92	1.32	2.15	1.63	1.89	0.83	0.73	1.22	1.19	1.20	1.16	0.83	1.39	1.23	1.31	1.47
OLYP	0.03	−0.60	1.77	1.55	1.66	−1.10	−1.18	1.56	1.59	1.58	−0.78	−1.14	1.57	1.58	1.57	1.60
BLYP	2.46	1.83	2.48	1.94	2.21	1.36	1.28	1.43	1.38	1.41	1.63	1.32	1.67	1.42	1.54	1.72
BP86	2.80	2.19	2.80	2.23	2.52	1.74	1.64	1.74	1.68	1.71	2.03	1.71	2.03	1.73	1.88	2.03
XLYP	2.98	2.33	2.98	2.33	2.66	1.85	1.77	1.85	1.77	1.81	2.14	1.82	2.14	1.82	1.98	2.15
mPWPBE	2.98	2.35	2.98	2.37	2.68	1.83	1.74	1.83	1.77	1.80	2.21	1.85	2.21	1.85	2.03	2.17
mPWPW91	3.02	2.40	3.02	2.40	2.71	2.25	1.89	2.25	1.89	2.07	2.25	1.89	2.25	1.89	2.07	2.28
mPWLYP	3.55	2.89	3.55	2.89	3.22	2.40	2.32	2.40	2.32	2.36	2.70	2.37	2.70	2.37	2.54	2.71
PBE	3.76	3.14	3.76	3.14	3.45	2.63	2.55	2.63	2.55	2.59	2.95	2.63	2.95	2.63	2.79	2.94
Meta GGA																
BB95	2.09	1.49	2.26	1.74	2.00	0.98	1.28	1.27	1.36	1.31	1.30	0.99	1.48	1.27	1.38	1.56
mPWKCIS	2.63	2.01	2.63	2.10	2.36	1.52	1.43	1.57	1.52	1.55	1.84	1.50	1.85	1.58	1.71	1.87
TPSSKCIS	2.91	2.32	2.91	2.32	2.61	1.82	1.74	1.82	1.74	1.78	2.17	1.84	2.17	1.84	2.01	2.13
TPSS	2.91	2.34	2.91	2.34	2.63	1.84	1.74	1.84	1.74	1.79	2.20	1.86	2.20	1.86	2.03	2.15
mPWB95	3.18	2.56	3.18	2.56	2.87	2.02	1.94	2.02	1.94	1.98	2.38	2.05	2.38	2.05	2.21	2.36
PBEKCIS	3.42	2.79	3.42	2.79	3.10	2.31	2.23	2.31	2.23	2.27	2.58	2.27	2.58	2.27	2.43	2.60
VSXC	3.61	3.05	3.61	3.05	3.33	2.55	2.47	2.55	2.47	2.51	2.84	2.53	2.84	2.53	2.68	2.84
Hybrid GGA																
BHandHLYP	0.31	−0.23	0.74	0.44	0.59	−0.67	−0.77	0.67	0.77	0.72	−0.42	−0.73	0.42	0.73	0.58	0.63
MPW1K	0.53	−0.03	0.94	0.69	0.81	−0.54	−0.65	0.63	0.70	0.67	−0.21	−0.56	0.44	0.66	0.55	0.68
mPW1PW91	1.26	0.68	1.38	0.91	1.15	0.19	0.08	0.51	0.52	0.51	0.53	0.18	0.65	0.51	0.58	0.75
B97-2	0.65	0.08	1.12	0.79	0.96	−0.42	−0.52	0.65	0.71	0.68	−0.10	−0.43	0.56	0.67	0.61	0.75
B3LYP	1.38	0.79	1.45	1.02	1.24	0.35	0.26	0.54	0.54	0.54	0.61	0.30	0.71	0.54	0.63	0.80
B3PW91	0.89	0.32	1.29	0.92	1.11	−0.14	−0.25	0.66	0.72	0.69	0.17	−0.17	0.64	0.69	0.66	0.82
X3LYP	1.73	1.14	1.73	1.18	1.46	0.69	0.61	0.70	0.65	0.67	0.96	0.65	0.96	0.68	0.82	0.98
B98	1.63	1.05	1.66	1.16	1.41	0.60	0.51	0.68	0.63	0.65	0.87	0.55	0.91	0.66	0.79	0.95
PBE1PBE	1.80	1.22	1.80	1.26	1.53	0.74	0.64	0.76	0.70	0.73	1.04	0.71	1.04	0.75	0.90	1.05
B3P86	1.79	1.22	1.79	1.28	1.54	0.78	0.67	0.84	0.80	0.82	1.05	0.73	1.05	0.83	0.94	1.10
B97-1	1.95	1.37	1.95	1.39	1.67	0.91	0.83	0.91	0.86	0.89	1.17	0.86	1.17	0.89	1.03	1.20
O3LYP	−0.21	−0.81	1.33	1.17	1.25	−1.30	−1.38	1.38	1.44	1.41	−0.99	−1.34	1.20	1.41	1.30	1.32
MPW3LYP	2.16	1.56	2.16	1.56	1.86	1.10	1.01	1.10	1.01	1.06	1.39	1.06	1.39	1.06	1.22	1.38
Hybrid Meta GGA																
MPWB1K	0.61	0.07	0.84	0.50	0.67	−0.45	−0.55	0.45	0.55	0.50	−0.12	−0.45	0.24	0.45	0.34	0.50
MPW1B95	1.11	0.54	1.20	0.74	0.97	0.03	−0.06	0.31	0.33	0.32	0.36	0.04	0.47	0.31	0.39	0.56
B1B95	0.46	−0.09	1.00	0.68	0.84	−0.58	−0.68	0.71	0.76	0.73	−0.27	−0.59	0.53	0.72	0.62	0.73
MPWKCIS1K	0.21	−0.34	0.79	0.55	0.67	−0.80	−0.91	0.80	0.91	0.86	−0.52	−0.85	0.52	0.85	0.69	0.74
BB1K	0.05	−0.48	0.75	0.63	0.69	−0.98	−1.09	0.98	1.09	1.04	−0.68	−1.00	0.68	1.00	0.84	0.86
PBE1KCIS	1.64	1.06	1.64	1.14	1.39	0.61	0.52	0.66	0.61	0.63	0.87	0.56	0.89	0.63	0.76	0.93
MPW1KCIS	1.46	0.87	1.59	1.12	1.36	0.39	0.30	0.68	0.68	0.68	0.70	0.37	0.85	0.68	0.77	0.93
TPSS1KCIS	1.90	1.34	1.93	1.43	1.68	0.85	0.77	0.94	0.88	0.91	1.18	0.85	1.22	0.95	1.08	1.22
TPSSh	2.13	1.57	2.13	1.64	1.89	1.08	0.98	1.14	1.08	1.11	1.43	1.09	1.44	1.16	1.30	1.43
ab Initio WFT																
MP2	0.59	−1.00	0.71	1.00	0.86	0.65	0.16	0.65	0.25	0.45	0.73	−0.21	0.73	0.26	0.49	0.60

[a] MUE denotes mean unsigned error (also called mean absolute deviation). MSE denotes mean signed error. MMUE = [MUE(no-cp) + MUE(cp)]/2. MMMUE = [MMUE(DIDZ) + MMUE(aug-cc-pVTZ) + MMUE(MG3S)]/3. [b] We use "no-cp" to denote the calculation without the counterpoise correction for the BSSE and use "cp" to denote the calculation with the counterpoise correction for the BSSE. DIDZ denotes 6-31+G(d,p) basis.

Table 6. Mean Errors (kcal/mol) for the DI6/04 Database[a,b]

	DIDZ					aug-cc-pVTZ					MG3S					
	MSE		MUE			MSE		MUE			MSE		MUE			
method	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	MMMUE
							LSDA									
SPWL	2.89	2.77	2.89	2.77	2.83	3.27	2.95	3.27	2.95	3.11	2.93	2.73	2.93	2.73	2.83	2.93
SVWN5	3.27	2.95	3.27	2.95	3.11	2.89	2.77	2.89	2.77	2.83	2.93	2.73	2.93	2.73	2.83	2.93
SVWN3	3.42	3.11	3.42	3.11	3.27	3.05	2.93	3.05	2.93	2.99	3.08	2.89	3.08	2.89	2.98	3.08
							GGA									
mPWLYP	0.17	−0.07	0.43	0.44	0.44	−0.19	−0.26	0.39	0.41	0.40	−0.11	−0.30	0.38	0.41	0.39	0.41
PBE	0.67	0.39	0.68	0.54	0.61	0.31	0.24	0.42	0.41	0.42	0.38	0.20	0.46	0.40	0.43	0.49
mPWPW91	0.01	−0.27	0.58	0.61	0.60	−0.28	−0.47	0.56	0.59	0.57	−0.28	−0.47	0.56	0.59	0.57	0.58
HCTH	−0.31	−0.53	0.44	0.53	0.49	−0.63	−0.68	0.63	0.68	0.66	−0.55	−0.71	0.55	0.71	0.63	0.59
mPWPBE	−0.01	−0.30	0.59	0.63	0.61	−0.36	−0.46	0.58	0.60	0.59	−0.30	−0.49	0.57	0.61	0.59	0.60
XLYP	−0.29	−0.53	0.55	0.58	0.56	−0.65	−0.72	0.65	0.72	0.68	−0.58	−0.76	0.58	0.76	0.67	0.64
BP86	−0.21	−0.49	0.68	0.71	0.69	−0.58	−0.67	0.68	0.74	0.71	−0.54	−0.71	0.66	0.76	0.71	0.71
BLYP	−0.70	−0.95	0.70	0.95	0.82	−1.06	−1.13	1.06	1.13	1.10	−1.00	−1.18	1.00	1.18	1.09	1.00
BPW91	−0.87	−1.15	0.88	1.15	1.02	−1.21	−1.32	1.21	1.32	1.27	−1.17	−1.35	1.17	1.35	1.26	1.18
BPBE	−0.89	−1.18	0.90	1.18	1.04	−1.24	−1.34	1.24	1.34	1.29	−1.19	−1.38	1.19	1.38	1.29	1.20
OLYP	−2.14	−2.39	2.14	2.39	2.26	−2.46	−2.51	2.46	2.51	2.48	−2.35	−2.53	2.35	2.53	2.44	2.40
G96LYP	−2.25	−2.49	2.25	2.49	2.37	−2.61	−2.69	2.61	2.69	2.65	−2.57	−2.74	2.57	2.74	2.65	2.56
							Meta GGA									
PBEKCIS	0.43	0.18	0.54	0.46	0.50	0.07	0.01	0.35	0.37	0.36	0.14	−0.03	0.37	0.37	0.37	0.41
TPSSKCIS	0.00	−0.27	0.52	0.54	0.53	−0.34	−0.40	0.49	0.51	0.50	−0.26	−0.43	0.49	0.52	0.50	0.51
mPWB95	−0.02	−0.27	0.51	0.54	0.52	−0.38	−0.45	0.52	0.56	0.54	−0.29	−0.47	0.48	0.57	0.52	0.53
TPSS	−0.03	−0.32	0.53	0.56	0.55	−0.35	−0.44	0.52	0.54	0.53	−0.29	−0.46	0.52	0.56	0.54	0.54
mPWKCIS	−0.25	−0.51	0.56	0.59	0.58	−0.60	−0.69	0.61	0.69	0.65	−0.54	−0.72	0.57	0.72	0.65	0.62
VSXC	1.21	0.98	1.21	0.98	1.10	1.14	1.08	1.15	1.11	1.13	1.10	0.95	1.10	1.02	1.06	1.10
BB95	−0.90	−1.15	0.90	1.15	1.02	−1.26	−1.33	1.26	1.33	1.29	−1.18	−1.35	1.18	1.35	1.27	1.19
							Hybrid GGA									
B97-1	0.41	0.14	0.48	0.37	0.42	0.06	0.00	0.28	0.30	0.29	0.09	−0.06	0.28	0.30	0.29	0.33
MPW3LYP	0.15	−0.10	0.36	0.36	0.36	−0.19	−0.26	0.32	0.34	0.33	−0.14	−0.30	0.31	0.36	0.34	0.34
B98	0.15	−0.12	0.41	0.39	0.40	−0.20	−0.27	0.35	0.37	0.36	−0.17	−0.32	0.34	0.40	0.37	0.38
PBE1PBE	0.35	0.05	0.47	0.42	0.45	−0.02	−0.09	0.36	0.38	0.37	0.03	−0.13	0.35	0.38	0.37	0.39
X3LYP	−0.14	−0.39	0.42	0.44	0.43	−0.48	−0.54	0.48	0.54	0.51	−0.43	−0.59	0.45	0.59	0.52	0.49
B3P86	0.03	−0.27	0.52	0.55	0.53	−0.34	−0.44	0.50	0.55	0.53	−0.32	−0.49	0.50	0.57	0.54	0.53
MPW1K	−0.15	−0.49	0.43	0.51	0.47	−0.53	−0.62	0.53	0.62	0.58	−0.51	−0.67	0.52	0.67	0.60	0.55
mPW1PW91	−0.13	−0.44	0.50	0.53	0.51	−0.49	−0.59	0.55	0.60	0.57	−0.46	−0.63	0.53	0.63	0.58	0.55
BHandHLYP	−0.44	−0.72	0.44	0.72	0.58	−0.77	−0.83	0.77	0.83	0.80	−0.75	−0.89	0.75	0.89	0.82	0.73
B3LYP	−0.48	−0.74	0.52	0.74	0.63	−0.82	−0.89	0.82	0.89	0.86	−0.78	−0.94	0.78	0.94	0.86	0.78
B97-2	−0.55	−0.83	0.58	0.83	0.71	−0.91	−0.98	0.91	0.98	0.94	−0.87	−1.02	0.87	1.02	0.94	0.86
B3PW91	−0.64	−0.94	0.69	0.94	0.82	−1.00	−1.10	1.00	1.10	1.05	−0.97	−1.14	0.97	1.14	1.06	0.97
O3LYP	−1.73	−1.99	1.73	1.99	1.86	−2.06	−2.10	2.06	2.10	2.08	−1.97	−2.13	1.97	2.13	2.05	2.00
							Hybrid Meta GGA									
PBE1KCIS	0.14	−0.14	0.40	0.37	0.39	−0.22	−0.27	0.33	0.36	0.34	−0.17	−0.32	0.32	0.38	0.35	0.36
TPSS1KCIS	−0.09	−0.37	0.47	0.50	0.49	−0.43	−0.50	0.49	0.53	0.51	−0.37	−0.53	0.46	0.55	0.50	0.50
MPWB1K	−0.18	−0.47	0.36	0.48	0.42	−0.55	−0.55	0.55	0.55	0.55	−0.50	−0.65	0.50	0.65	0.57	0.52
MPW1B95	−0.17	−0.45	0.41	0.49	0.45	−0.54	−0.60	0.54	0.60	0.57	−0.48	−0.63	0.50	0.63	0.56	0.53
TPSSh	−0.09	−0.39	0.50	0.53	0.52	−0.42	−0.51	0.51	0.56	0.54	−0.38	−0.54	0.49	0.58	0.54	0.53
MPW1KCIS	−0.35	−0.63	0.51	0.63	0.57	−0.71	−0.78	0.71	0.78	0.75	−0.66	−0.82	0.66	0.82	0.74	0.69
MPWKCIS1K	−0.42	−0.72	0.44	0.72	0.58	−0.78	−0.85	0.78	0.85	0.81	−0.75	−0.90	0.75	0.90	0.83	0.74
BB1K	−0.69	−0.98	0.69	0.98	0.84	−1.06	−1.13	1.06	1.13	1.09	−1.02	−1.16	1.02	1.16	1.09	1.01
B1B95	−0.80	−1.07	0.80	1.07	0.94	−1.16	−1.23	1.16	1.23	1.20	−1.11	−1.26	1.11	1.26	1.19	1.11
							ab Initio WFT									
MP2	−0.23	−1.41	0.23	1.41	0.82	0.74	0.22	0.74	0.22	0.48	0.45	−0.08	0.45	0.25	0.35	0.55

[a] MUE denotes mean unsigned error (also called mean absolute deviation). MSE denotes mean signed error. MMUE = [MUE(no-cp) + MUE(cp)]/2. MMMUE = [MMUE(DIDZ) + MMUE(aug-cc-pVTZ) + MMUE(MG3S)]/3. [b] We use "no-cp" to denote the calculation without the counterpoise correction for the BSSE and use "cp" to denote the calculation with the counterpoise correction for the BSSE. DIDZ denotes 6-31+G(d,p) basis.

Table 7. Mean Errors (kcal/mol) for the WI9/04 Database[a,b]

	DIDZ					aug-cc-pVTZ					MG3S					
	MSE		MUE			MSE		MUE			MSE		MUE			
method	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	MMMUE
								LSDA								
SPWL	0.42	0.38	0.42	0.38	0.40	0.47	0.33	0.47	0.33	0.40	0.44	0.36	0.44	0.36	0.40	0.40
SVWN5	0.47	0.33	0.47	0.33	0.40	0.41	0.38	0.41	0.38	0.40	0.44	0.36	0.44	0.36	0.40	0.40
SVWN3	0.51	0.38	0.51	0.38	0.44	0.46	0.42	0.46	0.42	0.44	0.49	0.40	0.49	0.40	0.44	0.44
								GGA								
PBE	−0.11	−0.26	0.30	0.29	0.30	−0.23	−0.26	0.26	0.28	0.27	−0.19	−0.26	0.26	0.28	0.27	0.28
HCTH	0.11	−0.04	0.43	0.35	0.39	0.02	−0.01	0.33	0.33	0.33	0.02	−0.05	0.39	0.34	0.36	0.36
mPWLYP	−0.08	−0.25	0.46	0.40	0.43	−0.22	−0.26	0.38	0.39	0.38	−0.17	−0.24	0.39	0.39	0.39	0.40
mPWPW91	−0.40	−0.55	0.54	0.56	0.55	−0.47	−0.55	0.49	0.56	0.52	−0.47	−0.55	0.49	0.56	0.52	0.53
mPWPBE	−0.41	−0.57	0.55	0.57	0.56	−0.55	−0.58	0.55	0.59	0.57	−0.48	−0.57	0.50	0.57	0.53	0.56
XLYP	−0.41	−0.58	0.55	0.58	0.56	−0.55	−0.58	0.55	0.58	0.57	−0.50	−0.57	0.51	0.57	0.54	0.56
BLYP	−0.67	−0.84	0.75	0.84	0.80	−0.81	−0.84	0.81	0.84	0.83	−0.77	−0.84	0.77	0.84	0.80	0.81
OLYP	−0.72	−0.91	0.94	0.95	0.94	−0.90	−0.93	0.92	0.95	0.93	−0.83	−0.92	0.88	0.94	0.91	0.93
BP86	−0.83	−0.98	0.84	0.98	0.91	−0.94	−0.97	0.94	0.97	0.95	−0.90	−0.98	0.90	0.98	0.94	0.93
BPW91	−0.98	−1.13	0.99	1.13	1.06	−1.11	−1.15	1.11	1.15	1.13	−1.05	−1.14	1.05	1.14	1.09	1.09
BPBE	−1.00	−1.15	1.01	1.15	1.08	−1.12	−1.16	1.12	1.16	1.14	−1.06	−1.15	1.06	1.15	1.11	1.11
G96LYP	−1.78	−1.94	1.78	1.94	1.86	−1.90	−1.94	1.90	1.94	1.92	−1.86	−1.95	1.86	1.95	1.90	1.89
								Meta GGA								
PBEKCIS	−0.06	−0.21	0.30	0.27	0.29	−0.18	−0.21	0.26	0.27	0.26	−0.15	−0.22	0.26	0.27	0.27	0.27
mPWB95	−0.17	−0.33	0.35	0.35	0.35	−0.29	−0.32	0.31	0.34	0.33	−0.23	−0.31	0.27	0.32	0.30	0.32
TPSSKCIS	−0.28	−0.42	0.42	0.43	0.43	−0.39	−0.42	0.41	0.43	0.42	−0.33	−0.42	0.39	0.43	0.41	0.42
TPSS	−0.35	−0.48	0.46	0.48	0.47	−0.45	−0.49	0.46	0.49	0.48	−0.39	−0.48	0.42	0.48	0.45	0.47
mPWKCIS	−0.37	−0.52	0.53	0.54	0.53	−0.50	−0.53	0.51	0.54	0.53	−0.44	−0.52	0.48	0.53	0.50	0.52
BB95	−0.76	−0.91	0.79	0.91	0.85	−0.87	−0.91	0.87	0.91	0.89	−0.82	−0.90	0.82	0.90	0.86	0.87
VSXC	1.52	1.37	1.62	1.49	1.55	1.66	1.63	1.73	1.71	1.72	1.55	1.48	1.65	1.61	1.63	1.64
								Hybrid GGA								
B97-1	−0.06	−0.19	0.20	0.20	0.20	−0.15	−0.18	0.18	0.20	0.19	−0.11	−0.18	0.18	0.20	0.19	0.19
B98	−0.17	−0.30	0.28	0.30	0.29	−0.26	−0.29	0.27	0.30	0.29	−0.22	−0.29	0.25	0.30	0.28	0.28
PBE1PBE	−0.18	−0.31	0.28	0.31	0.30	−0.27	−0.30	0.28	0.31	0.29	−0.22	−0.30	0.25	0.30	0.28	0.29
MPW3LYP	−0.07	−0.21	0.36	0.32	0.34	−0.18	−0.21	0.29	0.30	0.30	−0.12	−0.20	0.31	0.30	0.31	0.32
X3LYP	−0.28	−0.42	0.40	0.42	0.41	−0.39	−0.42	0.39	0.42	0.41	−0.34	−0.41	0.36	0.41	0.38	0.40
BHandHLYP	−0.32	−0.44	0.38	0.44	0.41	−0.40	−0.43	0.40	0.43	0.42	−0.34	−0.42	0.35	0.42	0.39	0.41
MPW1K	−0.35	−0.47	0.41	0.47	0.44	−0.44	−0.48	0.44	0.48	0.46	−0.37	−0.47	0.38	0.47	0.42	0.44
mPW1PW91	−0.37	−0.50	0.47	0.51	0.49	−0.48	−0.52	0.48	0.52	0.50	−0.12	−0.50	0.71	0.50	0.61	0.53
B97-2	−0.46	−0.58	0.53	0.58	0.56	−0.56	−0.60	0.56	0.60	0.58	−0.50	−0.59	0.50	0.59	0.54	0.56
B3LYP	−0.49	−0.63	0.56	0.63	0.60	−0.60	−0.63	0.60	0.63	0.62	−0.55	−0.62	0.55	0.62	0.59	0.60
B3P86	−0.60	−0.72	0.60	0.72	0.66	−0.67	−0.71	0.67	0.71	0.69	−0.63	−0.72	0.63	0.72	0.67	0.68
O3LYP	−0.61	−0.77	0.79	0.80	0.80	−0.77	−0.79	0.78	0.80	0.79	−0.70	−0.78	0.74	0.80	0.77	0.79
B3PW91	−0.76	−0.89	0.77	0.89	0.83	−0.86	−0.89	0.86	0.89	0.88	−0.80	−0.89	0.80	0.89	0.84	0.85
								Hybrid Meta GGA								
MPWB1K	−0.13	−0.25	0.22	0.25	0.23	−0.21	−0.24	0.22	0.25	0.23	−0.14	−0.23	0.16	0.23	0.20	0.22
MPW1B95	−0.15	−0.27	0.26	0.28	0.27	−0.24	−0.27	0.25	0.28	0.26	−0.17	−0.25	0.20	0.26	0.23	0.25
PBE1KCIS	−0.13	−0.26	0.28	0.27	0.28	−0.22	−0.25	0.25	0.27	0.26	−0.18	−0.25	0.25	0.27	0.26	0.27
TPSS1KCIS	−0.28	−0.41	0.40	0.42	0.41	−0.38	−0.41	0.40	0.42	0.41	−0.33	−0.41	0.37	0.41	0.39	0.40
MPWKCIS1K	−0.33	−0.44	0.41	0.45	0.43	−0.41	−0.45	0.42	0.45	0.43	−0.35	−0.44	0.37	0.44	0.41	0.42
TPSSh	−0.35	−0.48	0.44	0.48	0.46	−0.45	−0.48	0.45	0.48	0.47	−0.39	−0.48	0.40	0.48	0.44	0.45
MPW1KCIS	−0.36	−0.49	0.48	0.50	0.49	−0.47	−0.50	0.48	0.51	0.49	−0.41	−0.49	0.44	0.50	0.47	0.48
BB1K	−0.47	−0.59	0.48	0.59	0.54	−0.55	−0.58	0.55	0.58	0.57	−0.49	−0.57	0.49	0.57	0.53	0.54
B1B95	−0.57	−0.70	0.59	0.70	0.64	−0.66	−0.69	0.66	0.69	0.68	−0.60	−0.68	0.60	0.68	0.64	0.65
								ab Initio WFT								
MP2	−0.02	−0.38	0.12	0.38	0.25	0.12	−0.03	0.12	0.07	0.10	0.07	−0.17	0.09	0.17	0.13	0.16

[a] MUE denotes mean unsigned error (also called mean absolute deviation). MSE denotes mean signed error. MMUE = [MUE(no-cp) + MUE(cp)]/2. MMMUE = [MMUE(DIDZ) + MMUE(aug-cc-pVTZ) + MMUE(MG3S)]/3. [b] We use "no-cp" to denote the calculation without the counterpoise correction for the BSSE and use "cp" to denote the calculation with the counterpoise correction for the BSSE. DIDZ denotes 6-31+G(d,p) basis.

behavior in the large reduced-gradient region (which corresponds to the polarizable low-density outer fringe of atoms and molecules) and that seems to account for their good behavior for nonbonded interactions. Table 7 bears this out: mPWLYP and XLYP are better than BLYP, mPWB95 is better than BB95, MPW3LYP and X3LYP are better than B3LYP, MPWB1K is better than BB1K, and MPW1B95 is better than B1B95.

From Table 7, we can see that B97-1 is the best performer for the DIDZ basis set.

3.4. Overall Results. Table 8 is a summary of the performance of the tested methods for nonbonded interactions. The rank order is according to the final column, which is the average of the four database columns included in this table. Clearly the exact position in the ranking is not as meaningful as the general trends, but it provides a way to organize the discussion. The five smallest average errors for each of the individual databases and the 10 smallest average errors overall are in bold.

From Table 8, we can see that the best performers for hydrogen bonding are PBE, PBE1PBE, B3P86, MPW1K, B97-1, and X3LYP. The best performers for charge-transfer interaction are MPWB1K, MP2, MPW1B95, MPW1K, and B1B95. The best performers for dipole interaction are MPW3LYP, B97-1, PBE1KCIS, B98, and PBE1PBE. The best performers for weak interaction are MP2, B97-1, MPWB1K, MPW1B95, and PBE1KCIS.

Using MMMMUE in Table 8 as the overall, summarizing measure of quality for nonbonded interactions, we can see that MPWB1K is the best of all the tested DFT methods, and MPW1K is the best of the tested DFT methods that do not contain kinetic energy density. Both of these DFT methods have three bold entries. B97-1 has four bold entries, but it is only 27% better than average of all 44 DFT methods for charge-transfer complexes (the averages are in the last row of the table).

LSDA, PBE, and TPSS are the first three rungs of Perdew's nonempirical Jacob's ladder[51,61,71] for organizing DFT approximations. Perdew, Scuseria, and co-workers have recently made extensive comparisons of LSDA, PBE, and TPSS for atoms, molecules, solids, and surfaces.[69,71,82] In every case, TPSS was either significantly better than the other two nonempirical functionals or about the same as the better of the other. Recently we also found this pattern holds for barrier height calculations.[86] From Table 8, we can see that this pattern also holds for nonbonded interaction calculations in general, even though it does not hold for hydrogen bonding. As we climb the nonempirical ladder, the nonbonded interaction calculations improve significantly from LSDA to PBE (MMMMUE reduces from 3.67 to 1.05 kcal/mol) but only improve slightly from PBE to TPSS (MMMMUE reduces from 1.05 to 0.95 kcal/mol).

Averaging the four averages in Table 2 yields 4.13 kcal/mol for a mean binding energy characterizing the tests in this paper. The smallest error in Table 8, namely 0.46 kcal/mol, is only 11% of this. This is quite good, especially considering that DFT is often described as inaccurate for nonbonded interactions. The same calculation yield an 18% error for B3LYP, which is the most popular DFT functional, and 93% for the least accurate method (SVWN3) in Table 8.

Progress in DFT can be gauged by a statistical analysis of the 14 DFT methods that perform better than the popular B3LYP method in Table 8. We find two methods from 1993

Table 8. Overall Results (kcal/mol)[a]

		MMMUE				
rank	method	HB	CT	DI	WI	MMMMUE[b]
1	MPWB1K	0.61	**0.50**	0.52	**0.22**	**0.46**
2	MP2	0.66	**0.60**	0.55	**0.16**	**0.49**
3	MPW1B95	0.71	**0.56**	0.53	**0.25**	**0.51**
4	PBE1KCIS	0.60	0.93	**0.36**	**0.27**	**0.54**
5	MPW1K	**0.53**	**0.68**	0.55	0.44	**0.55**
6	PBE1PBE	**0.47**	1.05	**0.39**	0.29	**0.55**
7	B98	0.61	0.95	**0.38**	0.28	**0.55**
8	B97-1	**0.53**	1.20	**0.33**	**0.19**	**0.56**
9	BHandHLYP	**0.52**	**0.63**	0.73	0.41	**0.57**
10	X3LYP	0.55	0.98	0.49	0.40	**0.60**
11	mPW1PW91	0.63	0.75	0.55	0.53	0.62
12	MPW3LYP	0.57	1.38	**0.34**	0.32	0.65
13	MPWKCIS1K	0.79	0.74	0.74	0.42	0.67
14	B3P86	**0.46**	1.10	0.53	0.68	0.69
15	TPSS1KCIS	0.71	1.22	0.50	0.40	0.71
16	B3LYP	0.77	0.80	0.78	0.60	0.74
17	TPSSh	0.63	1.43	0.53	0.45	0.76
18	MPW1KCIS	1.04	0.93	0.69	0.48	0.79
19	B97-2	1.32	0.75	0.86	0.56	0.87
20	BB1K	1.11	0.86	1.01	0.54	0.88
21	B3PW91	1.13	0.82	0.97	0.85	0.94
22	TPSS	0.66	2.15	0.54	0.47	0.95
23	TPSSKCIS	0.75	2.13	0.51	0.42	0.95
24	B1B95	1.37	0.73	1.11	0.65	0.96
25	PBEKCIS	0.70	2.60	0.41	0.27	1.00
26	mPWPBE	0.82	2.17	0.60	0.56	1.03
27	mPWB95	0.94	2.36	0.53	0.32	1.04
28	HCTH	1.76	1.45	0.59	0.36	1.04
29	mPWPW91	0.77	2.28	0.58	0.53	1.04
30	mPWKCIS	1.15	1.87	0.62	0.52	1.04
31	PBE	**0.50**	2.94	0.49	0.28	1.05
32	mPWLYP	0.72	2.71	0.41	0.40	1.06
33	XLYP	0.92	2.15	0.64	0.56	1.06
34	BP86	0.88	2.03	0.71	0.93	1.14
35	BLYP	1.29	1.72	1.00	0.81	1.21
36	BPW91	1.70	1.47	1.18	1.09	1.36
37	BB95	1.87	1.56	1.19	0.87	1.37
38	BPBE	1.74	1.45	1.20	1.11	1.38
39	VSXC	0.61	2.84	1.10	1.64	1.55
40	O3LYP	2.76	1.32	2.00	0.79	1.72
41	OLYP	3.60	1.60	2.40	0.93	2.13
42	G96LYP	2.95	1.33	2.56	1.89	2.18
43	SPWL	4.62	6.73	2.93	0.40	2.75
44	SVWN5	4.63	6.73	2.93	0.40	3.67
45	SVWN3	4.87	6.94	3.08	0.44	3.83
	average	1.28	1.78	0.92	0.56	1.14

[a] MMMUE = [MMUE(DIDZ) + MMUE(aug-cc-pVTZ) + MMUE(MG-3S)]/3 and MMUE defined in the text and also in the footnote of Tables 4−7. MMMMUE is defined as MMMMUE = [MMMUE(HB) + MMMUE(CT) + MMMUE(DI) + MMMUE(WI)]/4. [b] All mean errors are computed from unrounded results, and the ranking is determined prior to rounding.

Table 9. Comparison of Mean Errors by Different Geometries[a,b]

	MC-QCISD geometries					consistently optimized geometries							
	MSE		MUE			MSE		MUE			max. error[c]		
methods	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	error[d]	complex[e]	BSSE[f]
						HB6/04							
B3LYP	−0.55	−0.93	0.60	0.93	0.76	−0.52	−0.90	0.58	0.90	0.74	−2.22	$(HCONH_2)_2$	cp
B97-1	−0.04	−0.43	0.45	0.45	0.45	−0.02	−0.41	0.44	0.44	0.44	−1.45	$(HCONH_2)_2$	cp
B97-2	−1.22	−1.64	1.22	1.64	1.43	−1.22	−1.63	1.22	1.63	1.43	−3.29	$(HCONH_2)_2$	cp
B98	−0.26	−0.66	0.45	0.66	0.55	−0.24	−0.64	0.45	0.64	0.54	−1.83	$(HCONH_2)_2$	cp
BLYP	−1.18	−1.56	1.18	1.56	1.37	−1.08	−1.46	1.08	1.46	1.27	−3.15	$(HCONH_2)_2$	cp
HCTH	−1.68	−2.12	1.68	2.12	1.90	−1.56	−1.95	1.56	1.95	1.76	−4.60	$(HCOOH)_2$	cp
MP2	0.24	−0.93	0.26	0.93	0.60	0.24	−0.91	0.27	0.91	0.59	−2.04	$(HCOOH)_2$	cp
MPW1B95	−0.46	−0.86	0.50	0.86	0.68	−0.47	−0.86	0.51	0.86	0.69	−2.12	$(HCONH_2)_2$	cp
MPW1K	−0.17	−0.61	0.33	0.61	0.47	−0.18	−0.63	0.33	0.63	0.48	−1.68	$(HCONH_2)_2$	cp
MPW3LYP	0.26	−0.14	0.51	0.41	0.46	0.30	−0.11	0.55	0.41	0.48	−1.13	$(HCONH_2)_2$	cp
MPWB1K	−0.31	−0.70	0.41	0.70	0.56	−0.33	−0.72	0.43	0.72	0.57	−1.88	$(HCONH_2)_2$	cp
PBE	0.22	−0.19	0.45	0.32	0.39	0.41	−0.03	0.59	0.28	0.44	−0.95	$(HCONH_2)_2$	cp
PBE1KCIS	−0.20	−0.59	0.49	0.59	0.54	−0.19	−0.58	0.49	0.58	0.53	−1.74	$(HCONH_2)_2$	cp
PBE1PBE	0.19	−0.23	0.40	0.28	0.34	0.24	−0.19	0.46	0.26	0.36	−1.06	$(HCONH_2)_2$	cp
SPWL	4.64	4.20	4.64	4.20	4.42	6.27	5.47	6.27	5.47	5.87	13.03	$(HCOOH)_2$	no-cp
TPSS	−0.37	−0.82	0.45	0.82	0.63	−0.22	−0.69	0.38	0.69	0.54	−1.91	$(HCONH_2)_2$	cp
TPSS1KCIS	−0.43	−0.86	0.49	0.86	0.67	−0.37	−0.81	0.48	0.81	0.65	−2.09	$(HCONH_2)_2$	cp
TPSSh	−0.36	−0.80	0.41	0.80	0.60	−0.26	−0.72	0.37	0.72	0.55	−1.91	$(HCONH_2)_2$	cp
X3LYP	−0.05	−0.44	0.45	0.48	0.47	−0.02	−0.42	0.45	0.47	0.46	−1.53	$(HCONH_2)_2$	cp
XLYP	−0.65	−1.04	0.74	1.04	0.89	−0.55	−0.94	0.70	0.94	0.82	−2.42	$(HCONH_2)_2$	cp
						CT7/04							
B3LYP	0.61	0.30	0.71	0.54	0.63	1.22	0.85	1.29	1.01	1.15	3.46	NH_3-F_2	no-cp
B97-1	1.17	0.86	1.17	0.89	1.03	1.52	1.15	1.52	1.17	1.34	3.14	NH_3-ClF	no-cp
B97-2	−0.10	−0.43	0.56	0.67	0.61	−0.06	−0.43	0.62	0.69	0.65	−1.44	HCN−ClF	cp
B98	0.87	0.55	0.91	0.66	0.79	1.13	0.76	1.16	0.86	1.01	2.76	NH_3-ClF	no-cp
BLYP	1.63	1.32	1.67	1.42	1.54	5.03	4.62	5.03	4.62	4.83	13.06	NH_3-F_2	no-cp
HCTH	0.94	0.63	1.31	1.22	1.26	2.25	1.84	2.56	2.32	2.44	8.05	NH_3-F_2	no-cp
MP2	0.73	−0.21	0.73	0.26	0.49	0.76	−0.22	0.76	0.26	0.51	1.30	NH_3-ClF	no-cp
MPW1B95	0.36	0.04	0.47	0.31	0.39	0.37	0.02	0.48	0.30	0.39	1.49	NH_3-ClF	no-cp
MPW1K	−0.21	−0.56	0.44	0.66	0.55	−0.24	−0.61	0.46	0.68	0.57	−0.96	HCN−ClF	cp
MPW3LYP	1.39	1.06	1.39	1.06	1.22	2.11	1.72	2.11	1.72	1.91	4.46	NH_3-F_2	no-cp
MPWB1K	−0.12	−0.45	0.24	0.45	0.34	−0.19	−0.52	0.27	0.52	0.40	−0.78	HCN−ClF	cp
PBE	2.95	2.63	2.95	2.63	2.79	6.34	5.85	6.34	5.85	6.10	12.59	NH_3-F_2	no-cp
PBE1KCIS	0.87	0.56	0.89	0.63	0.76	1.13	0.78	1.15	0.85	1.00	2.49	NH_3-ClF	no-cp
PBE1PBE	1.04	0.71	1.04	0.75	0.90	1.28	0.88	1.28	0.90	1.09	3.36	NH_3-ClF	no-cp
SPWL	6.78	6.41	6.78	6.41	6.59	12.79	12.15	12.79	12.15	12.47	18.43	NH_3-F_2	no-cp
TPSS	2.20	1.86	2.20	1.86	2.03	5.21	4.69	5.21	4.69	4.95	10.65	NH_3-F_2	no-cp
TPSS1KCIS	1.18	0.85	1.22	0.95	1.08	2.23	1.78	2.23	1.82	2.02	4.88	NH_3-F_2	no-cp
TPSSh	1.43	1.09	1.44	1.16	1.30	2.88	2.39	2.88	2.39	2.64	5.73	NH_3-F_2	no-cp
X3LYP	0.96	0.65	0.96	0.68	0.82	1.50	1.12	1.50	1.13	1.31	3.35	NH_3-F_2	no-cp
XLYP	2.14	1.82	2.14	1.82	1.98	5.65	5.23	5.65	5.23	5.44	13.73	NH_3-F_2	no-cp
						DI6/04							
B3LYP	−0.78	−0.94	0.78	0.94	0.86	−0.59	−0.75	0.59	0.75	0.67	−1.34	$CH_3Cl-HCl$	cp
B97-1	0.09	−0.06	0.28	0.30	0.29	0.19	0.04	0.30	0.30	0.30	0.99	$CH_3SH-HCl$	no-cp
B97-2	−0.87	−1.02	0.87	1.02	0.94	−0.67	−0.82	0.67	0.82	0.74	−1.55	$CH_3Cl-HCl$	cp
B98	−0.17	−0.32	0.34	0.40	0.37	−0.04	−0.20	0.33	0.36	0.35	−0.76	$CH_3Cl-HCl$	cp
BLYP	−1.00	−1.18	1.00	1.18	1.09	−0.72	−0.90	0.72	0.90	0.81	−1.53	$CH_3Cl-HCl$	cp
HCTH	−0.55	−0.71	0.55	0.71	0.63	−0.22	−0.37	0.30	0.37	0.33	−1.00	$CH_3Cl-HCl$	cp
MP2	0.45	−0.08	0.45	0.25	0.35	0.46	−0.05	0.46	0.23	0.35	1.30	$CH_3SH-HCl$	no-cp
MPW1B95	−0.48	−0.63	0.50	0.63	0.56	−0.42	−0.58	0.45	0.58	0.52	−1.06	$CH_3Cl-HCl$	cp
MPW1K	−0.51	−0.67	0.52	0.67	0.60	−0.40	−0.56	0.47	0.56	0.51	−1.21	$CH_3Cl-HCl$	cp
MPW3LYP	−0.14	−0.30	0.31	0.36	0.34	−0.02	−0.18	0.29	0.32	0.31	−0.66	$CH_3Cl-HCl$	cp
MPWB1K	−0.50	−0.65	0.50	0.65	0.57	−0.45	−0.60	0.45	0.60	0.53	−1.05	$CH_3Cl-HCl$	cp
PBE	0.38	0.20	0.46	0.40	0.43	0.60	0.41	0.60	0.48	0.54	1.99	$CH_3SH-HCl$	no-cp
PBE1KCIS	−0.17	−0.32	0.32	0.38	0.35	−0.06	−0.21	0.30	0.34	0.32	−0.77	$CH_3Cl-HCl$	cp
PBE1PBE	0.03	−0.13	0.35	0.38	0.37	0.15	−0.02	0.35	0.38	0.37	1.08	$CH_3SH-HCl$	no-cp
SPWL	2.93	2.73	2.93	2.73	2.83	4.20	3.92	4.20	3.92	4.06	7.11	$CH_3SH-HCl$	no-cp
TPSS	−0.29	−0.46	0.52	0.56	0.54	−0.07	−0.26	0.52	0.54	0.53	0.97	$CH_3SH-HCl$	no-cp
TPSS1KCIS	−0.37	−0.53	0.46	0.55	0.50	−0.20	−0.36	0.43	0.48	0.46	−1.02	$CH_3Cl-HCl$	cp
TPSSh	−0.38	−0.54	0.49	0.58	0.54	−0.19	−0.37	0.47	0.52	0.49	−1.03	$CH_3Cl-HCl$	cp
X3LYP	−0.43	−0.59	0.45	0.59	0.52	−0.29	−0.45	0.39	0.48	0.44	−0.97	$CH_3Cl-HCl$	cp
XLYP	−0.58	−0.76	0.58	0.76	0.67	−0.34	−0.53	0.50	0.60	0.55	−1.11	$CH_3Cl-HCl$	cp

Table 9 (Continued)

	MC-QCISD geometries					consistently optimized geometries							
	MSE		MUE			MSE		MUE			max. error[c]		
methods	no-cp	cp	no-cp	cp	MMUE	no-cp	cp	no-cp	cp	MMUE	error[d]	complex[e]	BSSE[f]
						WI9/04							
B3LYP	−0.55	−0.62	0.55	0.62	0.59	−0.41	−0.43	0.41	0.43	0.42	−1.47	$(C_2H_4)_2$	cp
B97-1	−0.11	−0.18	0.18	0.20	0.19	−0.08	−0.15	0.16	0.17	0.17	−0.67	$(C_2H_4)_2$	cp
B97-2	−0.50	−0.59	0.50	0.59	0.54	−0.31	−0.36	0.31	0.36	0.34	−1.25	$(C_2H_4)_2$	cp
B98	−0.22	−0.29	0.25	0.30	0.28	−0.16	−0.23	0.21	0.24	0.22	−0.90	$(C_2H_4)_2$	cp
BLYP	−0.77	−0.84	0.77	0.84	0.80	−0.45	−0.45	0.45	0.45	0.45	−1.43	$(C_2H_4)_2$	cp
HCTH	0.02	−0.05	0.39	0.34	0.36	0.23	0.16	0.29	0.23	0.26	0.40	CH_4−Ne	no-cp
MP2	0.07	−0.17	0.09	0.17	0.13	0.07	−0.14	0.09	0.14	0.11	−0.33	C_6H_6−Ne	cp
MPW1B95	−0.17	−0.25	0.20	0.26	0.23	−0.10	−0.18	0.22	0.19	0.20	−0.70	$(C_2H_4)_2$	cp
MPW1K	−0.37	−0.47	0.38	0.47	0.42	−0.28	−0.34	0.29	0.34	0.31	−1.23	$(C_2H_4)_2$	cp
MPW3LYP	−0.12	−0.20	0.31	0.30	0.31	−0.05	−0.13	0.30	0.26	0.28	−0.92	$(C_2H_4)_2$	cp
MPWB1K	−0.14	−0.23	0.16	0.23	0.20	−0.06	−0.14	0.20	0.17	0.19	−0.54	$(C_2H_4)_2$	cp
PBE	−0.19	−0.26	0.26	0.28	0.27	−0.14	−0.20	0.22	0.23	0.22	−0.88	$(C_2H_4)_2$	cp
PBE1KCIS	−0.18	−0.25	0.25	0.27	0.26	−0.13	−0.20	0.21	0.22	0.21	−0.84	$(C_2H_4)_2$	cp
PBE1PBE	−0.22	−0.30	0.25	0.30	0.28	−0.19	−0.25	0.22	0.25	0.24	−0.93	$(C_2H_4)_2$	cp
SPWL	0.44	0.36	0.44	0.36	0.40	0.78	0.47	0.78	0.47	0.63	1.68	$(C_2H_4)_2$	no-cp
TPSS	−0.39	−0.48	0.42	0.48	0.45	−0.24	−0.38	0.27	0.38	0.33	−1.76	$(C_2H_4)_2$	cp
TPSS1KCIS	−0.33	−0.41	0.37	0.41	0.39	−0.21	−0.28	0.25	0.28	0.27	−1.06	$(C_2H_4)_2$	cp
TPSSh	−0.39	−0.48	0.40	0.48	0.44	−0.25	−0.32	0.28	0.32	0.30	−1.16	$(C_2H_4)_2$	cp
X3LYP	−0.34	−0.41	0.36	0.41	0.38	−0.27	−0.35	0.30	0.35	0.33	−1.33	$(C_2H_4)_2$	cp
XLYP	−0.50	−0.57	0.51	0.57	0.54	−0.35	−0.41	0.36	0.41	0.38	−1.40	$(C_2H_4)_2$	cp
						Overall Results[g]							
B3LYP	−0.32	−0.55	0.66	0.76	0.71	−0.07	−0.31	0.72	0.77	0.75	3.46	NH_3−F_2	no-cp
B97-1	0.28	0.05	0.52	0.46	0.49	0.40	0.16	0.61	0.52	0.56	3.14	NH_3−ClF	no-cp
B97-2	−0.67	−0.92	0.79	0.98	0.88	−0.57	−0.81	0.71	0.88	0.79	−3.29	$(HCONH_2)_2$	cp
B98	0.06	−0.18	0.49	0.50	0.50	0.17	−0.08	0.54	0.52	0.53	2.76	NH_3−ClF	no-cp
BLYP	−0.33	−0.56	1.15	1.25	1.20	0.70	0.45	1.82	1.86	1.84	13.06	NH_3−F_2	no-cp
HCTH	−0.32	−0.56	0.98	1.10	1.04	0.17	−0.08	1.18	1.22	1.20	8.05	NH_3−F_2	no-cp
MP2	0.37	−0.35	0.38	0.40	0.39	0.38	−0.33	0.39	0.39	0.39	−2.04	$(HCOOH)_2$	cp
MPW1B95	−0.19	−0.43	0.42	0.51	0.47	−0.16	−0.40	0.41	0.48	0.45	−2.12	$(HCONH_2)_2$	cp
MPW1K	−0.32	−0.58	0.42	0.60	0.51	−0.27	−0.53	0.39	0.55	0.47	−1.68	$(HCONH_2)_2$	cp
MPW3LYP	0.35	0.11	0.63	0.53	0.58	0.58	0.32	0.81	0.68	0.74	4.46	NH_3−F_2	no-cp
MPWB1K	−0.27	−0.51	0.33	0.51	0.42	−0.26	−0.50	0.34	0.50	0.42	−1.88	$(HCONH_2)_2$	cp
PBE	0.84	0.59	1.03	0.91	0.97	1.81	1.51	1.94	1.71	1.82	12.59	NH_3−F_2	no-cp
PBE1KCIS	0.08	−0.15	0.49	0.47	0.48	0.19	−0.05	0.54	0.50	0.52	2.49	NH_3−ClF	no-cp
PBE1PBE	0.26	0.01	0.51	0.43	0.47	0.37	0.10	0.58	0.45	0.51	3.36	NH_3−ClF	no-cp
SPWL	3.70	3.43	3.70	3.43	3.56	6.01	5.50	6.01	5.50	5.76	18.43	NH_3−F_2	no-cp
TPSS	0.29	0.03	0.90	0.93	0.91	1.17	0.84	1.60	1.58	1.59	10.65	NH_3−F_2	no-cp
TPSS1KCIS	0.02	−0.24	0.64	0.69	0.66	0.36	0.08	0.85	0.85	0.85	4.88	NH_3−F_2	no-cp
TPSSh	0.08	−0.18	0.68	0.75	0.72	0.54	0.25	1.00	0.99	0.99	5.73	NH_3−F_2	no-cp
X3LYP	0.04	−0.20	0.56	0.54	0.55	0.23	−0.02	0.66	0.61	0.63	3.35	NH_3−F_2	no-cp
XLYP	0.10	−0.14	0.99	1.05	1.02	1.10	0.84	1.80	1.79	1.80	13.73	NH_3−F_2	no-cp

[a] MUE denotes mean unsigned error (also called mean absolute deviation). MSE denotes mean signed error. MMUE = [MUE(no-cp) + MUE(cp)]/2. [b] The MG3S basis set is used for all calculations in this table. [c] Maximum errors are taken from the results for the consistently optimized geometries. Although we tabulate the error that has the largest absolute value, we tabulate it as a signed quantity. [d] Error = calculation − best estimate. [e] This is the complex that gives the maximum error. [f] This column specify whether the maximum error occurs in the calculation with counterpoise corrections (cp) turned on during the optimization or without counterpoise corrections (no-cp). [g] The results in this section are 0.25 × HB + 0.25 × CT7+ 0.25 × DI6+ 0.25 × WI9, except for maximum error with the maximum over the whole nonbonded data set.

(BHandHLYP and B3P86), one from 1996 (PBE1PBE), one from 2000 (MPW1K), six from 2004 (MPWB1K, MPW1B95, X3LYP, MPW3LYP, MPW1KCIS, and TPSS1KCIS), and one from this article (PBE1KCIS). Thus, half of the best methods date from the past year. This is encouraging, and we hope that the variety of new databases and systematic tests currently being reported will lead to even more improvements soon.

3.5. Comparison of the Mean Errors by Consistently Optimized Geometries. To investigate the ability of DFT methods to predict accurate energetics when they are also used to predict the geometries of the complexes, we also performed calculations with consistently optimized geometries for the B3LYP, B97-1, B97-2, B98, BLYP, HCTH, MP2, MPW1B95, MPW1K, MPW3LYP, MPWB1K, PBE, PBE1KCIS, PBE1PBE, SPWL, TPSS, TPSS1KCIS, TPSSh,

X3LYP, and XLYP methods with MG3S basis set, and the results are given in Table 9.

Table 9 shows that the mean errors only slightly changed for hydrogen bonding, dipole interactions, and weak interactions by using consistently optimized geometries as compared to using the MC-QCISD geometries. For charge-transfer interactions, the mean errors for the B97-1, B97-2, MP2, MPWB1K, MPW1B95, MPW1K, PBE1KCIS, and PBE1PBE methods change slightly, but mean errors for B3LYP, HCTH, MPW3LYP, PBE, SPWL, TPSS, TPSS1KCIS, and TPSSh change more significantly. This is due to the systematical underestimation of the intermolecular distance by these methods, a problem for many DFT methods that was studied 10 years ago by Ruiz et al.[18] For SPWL applied to weak interactions the strong overbinding shown by the results in Table 9 is consistent with previous work.[144−146]

The HCTH, XLYP, and X3LYP results in Table 9 are particularly interesting because these methods contained some nonbonded complexes in their training set. The training set of HCTH contains nine hydrogen bonded dimers, and the training set of XLYP and X3LYP contains two van der Waals complexes (He_2 and Ne_2). However, Table 9 shows that HCTH does poorly for hydrogen bonding, and XLYP and X3LYP do poorly for weak interaction. These results show that including nonbonded interaction complexes in the training set does not guarantee that one will produce a good functional for nonbonded interaction; one needs to choose a good functional form, an appropriate training set, and a good weighting scheme to accurately parametrize a semiempirical DFT functional for nonbonded interactions.

The last three columns of Table 9 give the information about the maximum errors for each functional. For hydrogen bonding, $(HCOOH)_2$ and $(HCONH2)_2$ are difficult cases for all DFT methods. This is partly due to the fact that both complexes have two hydrogen bonds. For charge-transfer interactions, NH_3-F_2 is the most difficult case for those DFT methods that have low or zero percentage of Hartree−Fock exchange, and NH_3-ClF is the worst case for DFT methods that have a moderate or high percentage of Hartree−Fock exchange. $CH_3Cl-HCl$ and $CH_3SH-HCl$ are two difficult cases for DFT for dipole interactions. For weak interactions, $(C_2H_4)_2$ is the worst case for most DFT methods. This is because $(C_2H_4)_2$ is a $\pi\cdots\pi$ stacking complex, and it is very difficult to describe this type of weak interaction by DFT methods. It is encouraging that MPW1K and MPWB1K have smaller maximum errors than MP2, and the maximum errors in MPW1B95 and PBE1KCIS are only 4% and 22% larger, respectively, than the maximum error in MP2. If we judge the methods solely by the maximum errors, we would conclude that MPW1K, MPWB1K, and MPW1B95 (in that order) are the best DFT methods in Table 9 with PBE1KCIS ranked fourth. It is encouraging that this is very similar to the conclusion drawn form Table 8. A key difference between Tables 8 and 9 is that Table 8 averages over three basis sets, whereas Table 9 is based on a single basis set. For this reason we based on our overall evaluation in Table 8.

4. Concluding Remarks

In this paper, we developed four benchmark databases of binding energies for nonbonded interaction complexes. We tested 44 DFT methods and the 1 WFT method against the new databases.

Among the tested methods, the PBE, PBE1PBE, B3P86, MPW1K, B97-1, and BHandHLYP functionals give the best performance for hydrogen bonding, and MPWB1K, MP2, MPW1B95, MPW1K, and BHandHLYP give the best performance for charge-transfer interactions. MPW3LYP, B97-1, PBE1KCIS, B98, and PBE1PBE give the best performance for dipole interactions, and MP2, B97-1, MPWB1K, PBE1KCIS, and MPW1B95 give the best performance for weak interactions.

Overall, MPWB1K is the best of all the tested DFT methods, and MPW1K, PBE1PBE, and B98 are the best of the tested DFT methods that do not contain kinetic energy density. Interestingly, MPWB1K is found to be more accurate than MP2 for nonbonded interactions.

Moving up the rungs of Jacob's ladder for DFT, PBE improves significantly over the LSDA, and TPSS improves slightly (on average) over PBE for nonbonded interactions.

Acknowledgment. We are grateful to Jan M. L. Martin for sending us the Perl script for W1 calculations. This work was supported in part by the U.S. Department of Energy, Office of Basic Energy Science.

Supporting Information Available: The AE6 database and the calculated binding energies on nonbonded complexes with the MG3S basis and the MC-QCISD/3 geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, 1960.

(2) Kollman, P. A. In *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Politzer, P. A., Truhlar, D. G., Eds.; Plenum: New York, 1981; p 243.

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

(4) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133.

(5) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

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

(7) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.

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

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

(10) Tschinke, V.; Ziegler, T. A. *J. Chem. Phys.* **1990**, *93*, 8051.

(11) Perdew, J. P. In *Electronic Structure of Solids '91*; Ziesche, P., Eschig, H., Eds.; Akademie Verlag: Berlin, 1991; p 11.

(12) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.

(13) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(14) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

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

(16) Kristyán, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.

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

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

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

(20) Gill, P. M. W. *Mol. Phys.* **1996**, *89*, 433.

(21) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.

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

(23) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982.

(24) Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554.

(25) Filatov, M.; Thiel, W. *Mol. Phys.* **1997**, 847.

(26) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345.

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

(28) Gritsenko, O. V.; Schipper, P. R. T.; Baerends, E. J. *J. Chem. Phys.* **1997**, *107*, 5007.

(29) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.

(30) Schmider, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, *108*, 9624.

(31) Paizs, B.; Suhai, S. *J. Comput. Chem.* **1998**, *19*, 575.

(32) Voorhis, T. V.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.

(33) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264.

(34) Alkorta, I.; Rozas, I.; Elguero, J. *J. Phys. Chem. A* **1998**, *102*, 9278.

(35) Handy, N. C.; Tozer, D. J. *Mol. Phys.* **1998**, *94*, 707.

(36) Krieger, J. B.; Chen, J.; Iafrate, G. J. *Int. J. Quantum Chem.* **1998**, *69*, 255.

(37) Rey, J.; Savin, A. *Int. J. Quantum Chem.* **1998**, *69*, 581.

(38) Krieger, J. B.; Chen, J.; Iafrate, G. J.; Savin, A. In *Electron Correlations and Materials Properties*; Gonis, A., Kioussis, N., Eds.; Plenum: New York, 1999; p 463.

(39) Tuma, C.; Bosese, A. D.; Handy, N. C. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3939.

(40) Engel, E.; Dreizler, R. M. *J. Comput. Chem.* **1999**, *20*, 31.

(41) Hammer, B.; Hansen, L. B.; Norskov, J. K. *Phys. Rev. B* **1999**, *59*, 7413.

(42) Perdew, J. P.; Kurth, S.; Zupan, A.; Blaha, P. *Phys. Rev. Lett.* **1999**, *82*, 2544.

(43) Becke, A. D. *J. Chem. Phys.* **2000**, *112*, 4020.

(44) Rappe, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.

(45) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.

(46) He, Y.; Grafenstein, J.; Kraka, E.; Cremer, D. *Mol. Phys.* **2000**, *98*, 1639.

(47) Rabuck, A. D.; Scuseria, G. E. *Theor. Chem. Acc.* **2000**, *104*, 439.

(48) Proynov, E.; Chermette, H.; Salahub, D. R. *J. Chem. Phys.* **2000**, *113*, 10013.

(49) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403.

(50) Hoe, W.-M.; Cohen, A. J.; Handy, N. C. *Chem. Phys. Lett.* **2001**, *341*, 319.

(51) Perdew, J. P.; Schmidt, K. In *Density Functional Theory and Its Applications to Materials*; Doren, V., Alsenoy, C. V., Geerlings., P., Eds.; American Institute of Physics: New York, 2001.

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

(53) Wilson, P. J.; Bradley, T. J.; Tozer, D. J. *J. Chem. Phys.* **2001**, *115*, 9233.

(54) Tsuzuki, S.; Luthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949.

(55) Karpfen, A. *J. Phys. Chem. A* **2001**, *105*, 2064.

(56) Adamo, C.; Barone, V. *J. Chem. Phys.* **2002**, *116*, 5933.

(57) Menconi, G.; Tozer, D. J. *Chem. Phys. Lett.* **2002**, *360*, 38.

(58) Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2002**, *116*, 9559.

(59) Baker, J.; Pulay, P. *J. Chem. Phys.* **2002**, *117*, 1441.

(60) Toulouse, J.; Savin, A.; Adamo, C. *J. Chem. Phys.* **2002**, *117*, 10465.

(61) Mattsson, A. E. *Science* **2002**, *298*, 759.

(62) Mourik, T. V.; Gdanitz, R. J. *J. Chem. Phys.* **2002**, *116*, 9620.

(63) Coote, M. L.; Wood, G. P. F.; Radom, L. *J. Phys. Chem. A* **2002**, *106*, 12124.

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

(65) Karasiev, V. V. *J. Chem. Phys.* **2003**, *118*, 8576.

(66) Boese, A. D.; Martin, J. M. L.; Handy, N. C. *J. Chem. Phys.* **2003**, *119*, 3005.

(67) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem. A* **2003**, *107*, 1445.

(68) Boese, A. D.; Chandra, A.; Martin, J. M. L.; Marx, D. *J. Chem. Phys.* **2003**, *119*, 5965.

(69) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129.

(70) Karpfen, A. *Theor. Chem. Acc.* **2003**, *110*, 1.

(71) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(72) Hsu, J. Y. *Phys. Rev. Lett.* **2003**, *91*, 133001.

(73) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673.

(74) Zhao, Y.; Pu, J.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2004**, *6*, 673.

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

(76) Gomez-Balderas, R.; Coote, M. L.; Henry, D. J.; Radom, L. *J. Phys. Chem. A* **2004**, *108*, 2874.

(77) Li, Q. S.; Xu, X. D.; Zhang, S. *Chem. Phys. Lett.* **2004**, 20.

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

(79) Andersson, S.; Gruning, M. *J. Phys. Chem. A* **2004**, *108*, 7621.

(80) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.

(81) Ireta, J.; Neugebauer, J.; Scheffler, M. *J. Phys. Chem. A* **2004**, *108*, 5692.

(82) Perdew, J. P.; Tao, J.; Staroverov, V. N.; Scuseria, G. E. *J. Chem. Phys.* **2004**, *120*, 6898.

(83) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334.

(84) Xu, X.; William, A.; Goddard, I. *J. Phys. Chem. A* **2004**, *108*, 2305.

(85) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.

(86) Zhao, Y.; González-García, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012.

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

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

(89) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(90) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(91) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.

(92) Klopper, W.; Lüthi, H. P.; Brupbacher, T.; Bauder, A. *J. Chem. Phys.* **1994**, *101*, 9747.

(93) Yin, D.; Mackerell, A. D. J. *J. Phys. Chem.* **1996**, *100*, 2588.

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

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

(96) Tsuzuki, S.; Uchimaru, T.; Tanabe, K. *Chem. Phys. Lett.* **1998**, *287*, 202.

(97) Martin, J. M. L.; Oliveira, G. d. *J. Chem. Phys.* **1999**, *111*, 1843.

(98) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703.

(99) Halkier, A.; Klopper, W.; Helgaker, T.; Jorgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157.

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

(101) Klopper, W.; van-Duijneveldt-van-de-Rijdt, J. G. C. M.; Duijneveldt, F. B. v. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2227.

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

(103) Henry, D. J.; Parkinson, C. J.; Mayer, P. M.; Radom, L. *J. Phys. Chem. A* **2001**, *105*, 6750.

(104) Parthiban, S.; de Oliveira, G.; Martin, J. M. L. *J. Phys. Chem. A* **2001**, *105*, 895.

(105) Curtiss, L. A.; Redfern, P. C.; Rassolov, V.; Kedziora, G.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 9287.

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

(107) Martin, J. M. L.; Parthiban, S. In *Quantum Mechanical Prediction of Thermochemical Data*; Cioslowski, J., Ed.; Kluwer: Dordrecht, 2001.

(108) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.

(109) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *Chem. Phys. Lett.* **2002**, *359*, 390.

(110) Jurecka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89.

(111) Park, S. Y.; Lee, J. S. *J. Chem. Phys.* **2002**, *116*, 5389.

(112) Huh, S. B.; Lee, J. S. *Chem. Phys. Lett.* **2003**, *369*, 466.

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

(114) Karpfen, A. *J. Phys. Chem. A* **1999**, *103*, 11431.

(115) Langley, C. H.; Allinger, N. L. *J. Phys. Chem.* **2003**, *107*, 5208.

(116) Karpfen, A. *J. Phys. Chem. A* **2000**, *104*, 6871.

(117) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem. A* **2003**, *107*, 11445.

(118) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129.

(119) Dobson, J. F. *Int. J. Quantum Chem.* **1998**, *69*, 615.

(120) Patton, D. C.; Pederson, M. R. *Int. J. Quantum Chem.* **1998**, *69*, 619.

(121) Lein, M.; Dobson, J. F.; Gross, E. K. U. *J. Comput. Chem.* **1999**, *20*, 12.

(122) Kamiya, M.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2002**, *117*, 6010.

(123) Chu, X.; Dalgarno, A. *J. Chem. Phys.* **2004**, *121*, 4083.

(124) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1977**, *66*, 1977.

(125) Truhlar, D. G. *J. Chem. Phys.* **1993**, *98*, 2491.

(126) Kleinekathöfer, U.; Tang, K. T.; Toennies, J. P.; Yiu, C. L. *J. Chem. Phys.* **1997**, *107*, 9502.

(127) Peterson, K. A.; Dunning, T. H. *J. Chem. Phys.* **1995**, *102*, 2032.

(128) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem.* **1998**, *102*, 3782.

(129) Ogilvie, J. F.; Wang, F. Y. H. *J. Mol. Struct.* **1992**, *273*, 277.

(130) Ogilvie, J. F.; Wang, F. Y. H. *J. Mol. Struct.* **1993**, *291*, 313.

(131) Cappelletti, D.; Bartolomei, M.; Pirani, F.; Aquilanti, V. *J. Phys. Chem. A* **2002**, *106*, 10764.

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

(133) Slater, J. C. *Quantum Theory of Molecular and Solids. Vol. 4: The Self-Consistent Field for Molecular and Solids*; McGraw-Hill: New York, 1974.

(134) Adamo, C.; Cossi, M.; Barone, V. *Theochem.* **1999**, *493*, 145.

(135) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Wong, M. W.; Foresman, J. B.; Robb, M. A.; Head-Gordon, M.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian 92/DFT*, Revision F.2; Gaussian, Inc.: Pittsburgh, PA, 1993.

(136) Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, *45*, 566.

(137) 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.01; Gaussian, Inc.: Pittsburgh, PA, 2003.

(138) Werner, H.-J.; Knowles, P. J.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO*, 2002.6; University of Birmingham: Birmingham, 2002.

(139) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. http://comp.chem.umn.edu/database.

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

(141) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(142) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418.

(143) Kohn, W.; Meir, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.

(144) Patton, D. C.; Pederson, M. R. *Phys. Rev. A* **1997**, *56*, R2495; **2005**, *71*, 019906 (Erratum).

(145) Zhang, Y.; Pan, W.; Yang, W. *J. Chem. Phys.* **1997**, *107*, 7921.

(146) Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2005**, *122*, 114102.

CT049851D

JCTC Journal of Chemical Theory and Computation

Tetrahedral Sulfur-Based Dianions Stable in the Gas Phase: Application of the Cederbaum−Dreuw Design Principles

Carl Trindle*

Department of Chemistry, University of Virginia, Charlottesville, Virginia 22904

Ayse Yumak

Department of Physics, Marmara University, Istanbul, Turkey

Received December 14, 2004

Abstract: Cederbaum and Dreuw [*Chem. Rev.* **2002**, *102,* 181−200] developed an extensive set of examples of small dianions generally stable to autoionization and in some cases to dissociation as well. Their design principles include arranging alternating positive and negative charges in roughly spherical shells, as in the alkaline earth tetrahalides MX_4^{2-}. This principle embraces a series of surprising tetrahedral oxygen-centered dianions of $O(C_2)_4$, $O(BN)_4$, and $O(BeO)_4$. These systems are stable with respect to autoionization (their vertical ionization energy is positive) but are metastable with respect to dissociation into monoanions. We present a family of dianions with sulfur cores which are stable both with respect to autoionization and dissociation. The shell charge alternation still contributes to stability, but these systems have a positive core in contrast to the oxo systems' negative core.

Introduction

Small molecules with high negative charge may be anticipated to be unstable simply by considering the repulsions among the excess charges. There are a number of departures from this reasonable prediction however. The way toward construction of species with positive second electron affinities (i.e., which form stable dianions) was indicated by a general analysis[1] of MX_k electron affinities, where X may be F, Cl, or CN. Experimentally realized examples include dianions of heavy metal hexahalides MX_6^{2-} (M = Re, Os, Ir, and Pt; X = chloride or bromide),[2] alkaline earth tetrahalides such as BeF_4^{2-} and congeners,[3] and even octahalides of Se and Te.[4] These and other experimentally characterized systems share a structure Cederbaum and Dreuw[5] call a "Center-Ligand Sphere" which enjoys a favorable charge arrangement of concentric shells with alternating positive and negative charges.

Other means of stabilization of dianions by charge arrangement have been modeled, including linear forms XMX and XMXMX. Here X can be assigned a formal charge of −2, and M a charge of +2. SMgSMgS dianion seems to be the smallest such system which might be stable with respect to electron detachment.[6] Longer dianions such as S_n^{2-} can be stable if n is greater than 7.[7] In this case charge alternation may not be so important as simple separation in space of the two excess negative charges. Long-chain alkyl dicarboxylates[8] are examples of systems in which the separation between excess negative charges is of paramount importance. Catenated systems with metal centers such as $M_2X_4^{2-}$ enjoy both the advantages of the center-ligand sphere and charge separation modes of stabilization.[9] Druew and Cederbaum[4,10] have reviewed other dianion-stabilizing structures as well as the possibility of stabilization of a dianion by a very large dipole moment and the extension of the lifetime of a metastable system by the Coulomb barrier. The general guide to design articulated by these authors can be paraphrased: doubly negatively charged species are stable when the charges are separated from one another and each occupy an electrophilic environment, i.e., a region of very positive electrical potential.

* Corresponding author e-mail: cot@virginia.edu.

10.1021/ct049854q CCC: $30.25
Published on Web 03/30/2005

Table 1. Oxo Systems: Reported and Computed Energy Differences between Tetrahedral Oxo Dianions and Monoanions (Energies in eV)

species	Koopmans[9]	Koopmans[a]	ΔSCF[9]	ΔR(O)HF[a]	ΔMP2 [9]	ΔDFT[a]	ΔCCSD[9]
$O(CC)_4^{2-}$	2.63	1.96	1.66	1.41	2.33	0.73	2.06
$O(BN)_4^{2-}$	2.77	2.08	2.07	0.58	2.14	0.62	2.03
$O(BeO)_4^{2-}$		2.96[a]		0.90		0.74	

[a] This work: ΔE refers to monoanions' optimized structures. The DFT functional is B3LYP, basis is always 6-31G(d).

Table 2. Oxo Systems: R(O)HF/6-31G(d) and B3LYP Stability with Respect to Fragmentation[a]

species	RHF dianion	fragments	ΔE_{fr}	B3LYP dianion	fragments	ΔE_{fr}
$O(CC)_4$	−376.8727496	−376.994469	−3.31	−379.1958120	−379.313962	−3.21
$O(BN)_4$	−391.0549096	−391.148000	−2.53	−393.3837715	−393.432997	−1.34
$O(BeO)_4$	−433.0303083	−433.008572	+0.59	−435.3891784	−435.348205	+1.11

[a] Energies in Hartrees, energy differences in eV.

Table 3. Energies (in eV) Required To Remove an Electron from a SR_4 Dianion

species	Koopmans IE	$\Delta E_{R(O)HF}$ (ionization)	ΔE_{B3LYP} (ionization
$S(C_2)_4^{2-}$	2.46	0.618	0.734
$S(BN)_4^{2-}$	2.56	0.929	0.952
$S(BeO)_4^{2-}$	3.47	1.143	1.154

The presumptive instability of small dianions can be moderated and even overcome in some cases by substituents that change the electrophilicity of the excess electrons' environment. Stabilization by solvent or counterions is most common; familiar species such as sulfate and carbonate are observed in the gas phase only when solvated.[11] Electron-acceptor substituents such as CN and BO can stabilize local negative charges, as in the carboranes[12] for which the substituent is essential to the stability of $B_6(CN)_6^{2-}$.

Dianions may be unstable with respect either to autoionization or fragmentation into monoanions (or both). Familiar dianions such as CO_3^{2-},[13] SO_4^{2-},[14] and PO_4^{3-} [15] spontaneously lose an electron in the gas phase. Some dianions that appear to be stable with respect to autoionization may still be unstable with respect to dissociation, so to produce separated monoanions. This possibility has sometimes been overlooked in modeling studies, but is the case, for example, for the important class of alkaline earth tetrahalides.[16]

A distinct class of stable dianions incorporates the C_2 fragment and isoelectronic variants BN and BeO; e.g., $Be(C_2)_n^{2-}$ with $n = 2$ to 7; $Si(C_2)_3^{2-}$, $C(C_2)_3^{2-}$, and $O(C_2)_3^{2-}$ as well as the surprising $O(C_2)_4^{2-}$ and a number of isomers.[17] As the latter example proves, the center can be negatively charged if the ligand provides a positively charged shell around that core. Isoelectronic substitution of C_2 by BN and BeO produces stable tetrahedral dianions of form OX_4^{2-}. The four-coordination of oxygen in these systems is remarkable, but one may see that the $O(BN)_4^{2-}$ system is derivable from the more familiar-seeming $Be(CN)_4^{2-}$ system by a kind of isoelectronic argument. Here CN is a pseudohalide so the analogy with the established BeF_4^{2-} is plain. To obtain the novel system $O(BN)_4^{2-}$ from more familiar-seeming $Be(CN)_4^{2-}$, transfer an electron from each CN to the core atom; then increase its atomic number from 4 (Be) to 8 (O) to compensate for the charge reassignment, and finally alter C to B to balance the loss of negative charge. This example merely illustrates the chemical argument; $Be(CN)_4^{2-}$ is not stable with respect to electron loss according to Koopmans' theorem, although the larger $Mg(CN)_4^{2-}$ and the analogous $Be(CNO)_4^{2-}$ (which see below) are both stable by this criterion.

Here we characterize the thia analogies to the oxo systems, SX_4^{2-}. The thia systems might be expected to be more stable than the oxo systems since the negative charges are dispersed over a larger volume, particularly when X is larger than the well-known diatomic fragments already mentioned, but the lesser electronegativity of sulfur might oppose that apparent advantage. We also discuss the possibility of stabilizing dianionic species through coordination by Lewis acids BH_3 and BF_3.

We should make explicit that the species considered here are far from the thermodynamically most stable form of their respective collections of atoms. To take an extreme view, all species discussed in this paper are unstable with respect to their thermodynamic elements. On the other hand, unstable species can be very long-lived if there is no open channel leading toward energetically preferable products, or if (as we suspect is the case for many species considered here) a coulomb or other barrier slows passage and ensures kinetic stability. Each species described here occupies its local minimum on the potential surface, and thus we can refer to its kinetic stability. When we compare two species' energies we can refer to the relative stability of the less energetic species or to the instability or metastability of the more energetic species. A referee has pointed out that C_{3v}-symmetric dianion $OCC(C_2)_3$ is much more stable than $O(C_2)_4$ dianion. This is an important observation. The rearrangement path deserves careful study. But we can defer that study since it does not bear directly on the question of the kinetic stability of the systems under examination with respect to autoionization or dissociation to monoanionic fragments.

Methods

We have employed the Spartan software suite on PC Windows, SGI IRIX and PQS Linux systems,[18] and the Gaussian 03W[19] software suite on Windows.

It is essential to ascertain that the methods we use are capable of representing the molecules already known to be stable with respect to autoionization and sometimes dis-

Table 4. Thia Systems: R(O)HF/6-31G(d) and B3LYP/6-31G(d) Stability with Respect to Fragmentation[a]

species	dianion	fragments	ΔE_f	dianion T_d	fragments	ΔE_f
$S(CC)_4$	−699.806949	−699.771653	0.960	−702.418241	−702.403936	0.389
$S(BN)_4$	−713.747796	−713.728799	0.517	−716.399604	−716.392214	0.201
$S(BeO)_4$	−755.648911	−755.543245	2.875	−758.321603	−758.267402	1.475

[a] Energies in Hartrees, energy differences in eV.

Table 5. Bond Distances (Å) and Mulliken Atomic Charges (millielectrons) for Oxa-Kernel Systems $O(XY)_4^{2-}$

	properties				
model	R_{OX}	R_{XY}	Q_O	Q_X	Q_Y
		$O(CC)_4^{2-}$			
RHF	1.4584	1.2207	−765	−028	−280
B3LYP	1.5082	1.2524	−530	+075	−443
		$O(BN)_4^{2-}$			
RHF	1.5203	1.2272	−717	+197	−518
B3LYP	1.5271	1.2551	−472	+098	−480
		$O(BeO)_4^{2-}$			
RHF	1.6151	1.3276	−740	+214	−471
B3LYP	1.6111	1.3503	−509	+034	−407

Table 6. Bond Distances (Å) and Mulliken Atomic Charges (millielectrons) for Thia-Kernel Systems $S(XY)_4^{2-}$

	properties				
model	R_{SX}	R_{XY}	Q_S	Q_X	Q_Y
		$S(CC)_4^{2-}$			
ROHF	1.6266	1.2308	+1163	−545	−246
B3LYP	1.6819	1.2569	+623	−247	−408
		$S(BN)_4^{2-}$			
ROHF	1.8585	1.2305	+252	−008	−495
B3LYP	1.8514	1.2589	+319	−115	−464
		$S(BeO)_4^{2-}$			
ROHF	2.0747	1.3241	−623	+206	−550
B3LYP	2.0473	1.3459	−311	−011	−412

sociation. The simplest diagnosis of stability with respect to autoionization is provided by Koopmans' theorem, by which the energy of the highest occupied MO is an approximation to the ionization energy. A negative HOMO energy suggests that the dianion is stable with respect to loss of an electron. This estimate is a first-order estimate with no recognition of electronic relaxation or correlation and has failed in the important case of sulfate dianion.[20] More accurate estimates can be obtained by outer-valence Green's function methods[21] or by comparison of separate Hartree−Fock or correlation-corrected computations on the dianion and monoanion.

For initial surveys we use uncorrelated methods for which Koopmans' theorem applies. The stability with respect to ionization and dissociation for promising species can also be characterized by energy differences estimated by R(O)HF, MP2, and density functional methods. The adequacy of DFT methods must be established, particularly owing to the self-repulsion in available functionals. Galbraith and Schaefer's assurance[22] that this flaw in local density functional methods does not impede the description of strongly bound anions does not guarantee that the methods are suitable for dianions. DFT's failure to describe weak long-range interactions is not an immediate problem for our purposes, but reminds us that calibration is required. We therefore report DFT characterization of the well-defined stable dianionic species $O(CC)_4^{2-}$. As Tables 1 and 2 show, our simple methods capture major features of these systems' stability in an approximate way. It appears that compared with reported R(O)HF and correlation-corrected calculations[23] which deploy large flexible basis sets, our R(O)HF/6-31G(d) and especially the B3LYP/6-31G(d) calculations seriously underestimate the stability of the dianions relative to the monoanions (this would be consistent with an overestimate of electron repulsion in the DFT functional). Nonetheless the method affords reliable qualitative diagnoses of stability (i.e., the signs of the energy differences between dianion and monanion are correct). One detail which complicates these comparisons is that if the tetrahedral structure of the dianion is assumed for the monoanions as well, as must be the case in the report of vertical ionization energies, the monoanion's electronic state is degenerate and must be subject to Jahn−Teller spontaneous distortion. Our energy differences refer to relaxed monoanions. For the most part the optimized geometry of these monoanions is of C_{2v} or C_s-symmetry with rather modest distortions from T_d, but in a few cases the optimization leads to dissociation. $O(CC)_4^{1-}$ dissociates both in ROHF and DFT models. Some details of the optimized monoanions' structures and energies are collected in the Supporting Information.

Frequency calculations in RHF/6-31G(d) and B3LYP/6-31G(d) assure us that the dianions occupy relative minima on their potential surfaces. The detailed pathways and rates of dissociation remain to be addressed.

A Simple Extension

The most elementary extension of this family is to thia-analogues, in which the central kernel is sulfur. All species show stability with respect to autoionization but are now also stable with respect to dissociation into monoanions. Table 3 shows that Koopmans' theorem values are overestimates of the ionization energies compared with the differences in estimates by differences of either uncorrelated or correlated energies for the dianion and the associated monoanion. A part of the difference is that our monoanion structures are relaxed. Table 4 shows that the thia systems are stable with respect to dissociation into monoionic fragments. The versatility of the sulfur kernel, which can tolerate both positive and negative local charge, is a part of the story of the stability of these systems (see Tables 5 and 6 and discussion below.) We conjecture that the thia-kernel can be the core of a variety of tetracoordinate stable dianions, assuming the ligands are strong π acids, which we explore in the next section or σ pair acceptors which we describe in the final section.

Table 7. Energies (in eV) Required To Remove an Electron from a $M(XYZ)_4$ Dianion and To Dissociate into Two Monoanions XYZ^{-1} and $M(XYZ)_3{}^{-1}$

species	Koopmans IE	$\Delta E_{R(O)HF}$ (IE: vert)	$\Delta E_{R(O)HF}$ (IE: rel)	$\Delta E_{R(O)HF}$ (diss'n)	ΔE_{B3LYP} (IE: vert)	ΔE_{B3LYP} (IE: rel)	ΔE_{B3LYP} (diss'n)
$Be(CNO)_4{}^{2-}$	3.25	2.96	1.86	−0.55	0.83	0.76	−0.84
$Mg(CNO)_4{}^{2-}$	3.61	3.34	1.67	0.00	1.20	1.15	−0.23
$O(BNO)_4{}^{2-}$	1.77	1.42	diss[a]	−3.95	−0.35	diss[a]	−3.09
$S(BNO)_4{}^{2-}$	1.39	1.58	1.13	−0.47	−0.15	diss[b]	−0.82

[a] Monoanion dissociates spontaneously to $O(BNO)_2$ and $ONBBNO^{1-}$. [b] Monoanion dissociates spontaneously to $S(BNO)_2$ and $ONBBNO^{1-}$.

Table 8. Bond Distances (Å) and Mulliken Atomic Charges (millielectrons) for Oxa-Kernel Systems $O(BNO)_4{}^{2-}$ and Thia-Kernel Systems $S(BNO)_4{}^{2-}$

	properties						
model	R_{OB}	R_{BN}	R_{NO}	Q_O	Q_B	Q_N	Q_O
			$O(BNO)_4{}^{2-}$				
ROHF	1.5110	1.2024	1.2551	−695	+129	+768	−1223
B3LYP	1.5113	1.2024	1.2548	−470	+037	+181	−600
			$S(BNO)_4{}^{2-}$				
ROHF	1.8292	1.2127	1.2406	+1574	−529	+1027	−1391
B3LYP	1.8120	1.2507	1.2474	+297	−167	+155	−560

Potential New Dianionic Species: $Q(XYZ)_4{}^{2-}$

One of the first stable dianions predicted to be stable with respect to autoionization was $BeF_4{}^{2-}$. This species falls clearly into the class of center-ligand species in which stabilization arises from concentric shells of alternating charges. Simple chemical arguments suggest that all alkaline earth tetrahalides or pseudohalides would be reasonable candidates as stable dianions. We point out here that Koopmans' theorem has been a reliable guide to stability even if it systematically overestimates ionization energies. According to this criterion, $Be(CNO)_4{}^{2-}$ and $Mg(CNO)_4{}^{2-}$ are stable dianions. We can conjecture that the family derived from "isoelectronic" $O(BNO)_4{}^{2-}$ might also contain stable dianions. Our preliminary calculations of Koopmans' estimate of vertical ionization energies suggest that some of these systems will be resistant to autoionization. (See Tables 7 and 8.) Vertical energy differences between RHF dianions and ROHF monoanions confirm this view. B3LYP calculations again produce smaller values for vertical ionization energy; $O(BNO)_4{}^{2-}$ dianion in B3LYP/6-31G(d) is predicted to be unstable with respect to autoionization and also unstable with respect to distortion. B3LYP/6-31G(d) produces small imaginary e, t_1, and t_2 frequencies. Neither of these instabilities obtains in RHF, but MP2 also maintains that the system is marginally unstable (−0.35 eV) with respect to autoionization. Curiously, an outer valence Greens' function estimate[24] of the ionization energy in 6-31G(d) and at the RHF/6-31G(d) optimized geometry is unambiguously positive, ca. 0.90 eV. In view of the fact that the Green's function methods include higher order corrections than MP2, one would receive these estimates with respect. Optimization of $S(BNO)_4{}^{2-}$ in B3LYP/6-31G(d) leads to a structure of borderline instability with respect to autoionization; the vertical energy difference between the dianion and the monoanion at the geometry for the dianion optimized in T_d symmetry is −0.35 eV. This system has corresponding e, t_1, and t_2 modes with real but very small frequencies. MP2/6-31G(d) modeling of $S(BNO)_4{}^{2-}$ also produces small imaginary frequencies at the geometry optimized in T_d; the vertical energy difference between the dianion singlet in MP2 and the monoanion doublet in ROMP2 is −0.19 eV, confirming the B3LYP prediction of borderline instability with respect to autoionization. Again, an outer valence Greens' function estimate of the ionization energy in 6-31G(d) and at the RHF/6-31G(d) optimized geometry is unambiguously positive.

Table 9. Properties of BX_3-Stabilized Thia-Kernel Dianions[a]

no. of coordinated acids	Koopmans IE	R(O)HF VIE	RHF Ediss	DFT VIE	DFT Ediss
		BH_3-Stabilized Thia Dianions			
4	2.15	1.11	1.498	0.755	1.988
3	−0.18	−1.49	2.015	−1.139	2.510
2	−2.43	−3.46	2.628	−2.890	3.176
1	−3.97			−4.380	
		BF_3-Stabilized Thia Dianions			
4	4.35	3.039	0.763	+2.601	1.192
3	2.01	0.568	1.477	+0.908	1.742
2	−0.47	−1.743	2.534	−1.152	2.798
1	−3.01			−3.371	

[a] All energies in eV. KIE = Koopmans ionization energy, VIE = vertical ionization energy, and Ediss = dissociation energy of the *n* acid cluster to BX_3 and the *n*-1 acid cluster.

The monoanions $O(BNO)_4{}^{1-}$ and $S(BNO)_4{}^{1-}$ both dissociate spontaneously to dissociation to $O(BNO)_2 + (BNO)_2{}^{1-}$ and $S(BNO)_4 + (BNO)_2{}^{1-}$. These products are more stable than the parent dianion or the monoanionic fragments. We noticed analogous behavior in the B3LYP treatment of $S(BN)_4{}^{1-}$ which produced spontaneous dissociation to $S(BN)_2$ and $(BN)_2{}^{1-}$. It does not appear that the BNO-substituted oxygen or sulfur species are promising candidates as long-lived dianions, despite some promising results from Green's function and Koopmans estimates of ionization energies.

We have not established definitely that $S(BNO)_4{}^{2-}$ dianion is stable either with respect to autoionization or dissociation (be that kinetic or thermodynamic). However we consider worth exploring a family of other pseudohalogens and isoelectronic variants. Other candidates for stable dianionic species can be derived by isoelectronic substitution. XYZ examples include isoelectronic ligands NCO, BNF, NBF, NNN, and CCF for alkaline earths and BCF, CCO, BeNF, LiOF, and NCN for the chalcogens. We will describe members of this series elsewhere.

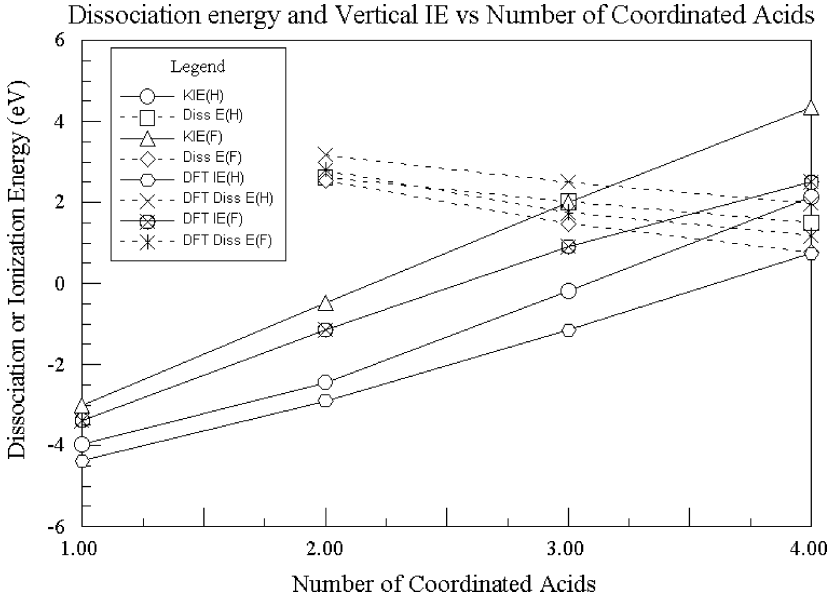


Figure 1. The dependence of vertical ionization energy and dissociation energy on the number of coordinated acids in the system $S(BX_3)_n^{2-}$. Dashed lines refer to dissociation energies, solid lines refer to vertical ionization energies, either Koopmans' estimates (KIE) or energy differences between dianion and monoanion. Positive ionization energies define stable dianions. Parenthesized (F) and (H) refer to acids BF_3 and BH_3, respectively.

Geometry and Charges of $K(XY...)_4$ Dianions

We have collected geometric parameters of the tetrahedral dianionic species described above in Tables 4 and 5 along with Mulliken charges. For the oxo systems B3LYP displays the familiar bond-length extension relative to RHF except for for the central O−Be bonds in $O(BeO)_4^{2-}$, which suggests that correlation is playing a role in that system's binding. Reports[4] of the structures of the tetrahedral dianions $O(C_2)_4^{2-}$ and $O(BN)_4^{2-}$ show that our B3LYP/6-31G(d) bond lengths agree within 0.01 Å with CCSD calculated values; in the CCSD calculation a Dunning double-ζ basis was employed, augmented by diffuse functions.

Mulliken charges are to be considered only semiquantitative, but values collected here for oxo systems seem to suggest that B3LYP favors a more uniform distribution of charge than does RHF and assigns less negative charge to the central oxygen. Charge alternation is plain, with the central and peripheral atoms unambiguously negative. The intervening atoms are positively charged except for the C in $O(CC)_4^{2-}$, for which the charge alternation is least marked.

In thia systems this central bond shortening in B3LYP is more prominent, and the central sulfur varies strikingly in charge, ranging from very positive in $S(CC)_4^{2-}$ to unambiguously negative in $S(BeO)_4^{2-}$. Again B3LYP predicts more uniform charge distributions than does RHF.

These patterns apply as well to $O(BNO)_4^{2-}$ and $S(BNO)_4^{2-}$; the kernel of the oxo system is negative, while the kernel of the thia system is positive, and the well-established charge alternation is more prominent in the RHF representation than in the B3LYP model.

BH_3 and BF_3 as Dianion-Stabilizing Shells

A remarkable T-symmetric trianion $N(BF_3)_4^{3-}$ has been characterized by computation[25] and shown to be stable with respect to autoionization though not with respect to dissociation. The less dramatic dianionic analogue, $O(BF_3)_4^{2-}$, is likewise stable with respect to ionization according to the Koopmans criterion. The thia cousin $S(BH_3)_4^{2-}$ is stable not only with respect to autoionization (by Koopmans' criterion) but also with respect to dissociation into neutral BH_3 and the $S(BH_3)_3^{2-}$ dianion. (See Table 9.) That product anion $S(BH_3)_3^{2-}$ is unstable with respect to autoionization but would not dissociate into BH_3 and $S(BH_3)_2^{2-}$. This is not due to any intrinsic stability of the dianion but reflects the penalty paid in dissociation to an even smaller product dianion; the latter dianion is still more unstable with respect to autoionization according to Koopmans' theorem. Probably another decomposition channel involving the $S(BH_3)_3^{1-}$ monoanion opens. We have not pursued this possibility.

As one expects, BF_3 produces more stable thia dianions than does BH_3; both $S(BF_3)_4^{2-}$ dianion and $S(BF_3)_3^{2-}$ dianion are stable with respect to autoionization according to Koopmans' theorem and do not dissociate into closed shell products including a less stable (because less completely coordinated) dianion. Figure 1 provides a graphic summary of this behavior.

As the degree of coordination increases and the excess negative charge is more effectively stabilized the BH and BF distances shorten. The SB distances in $S(BF_3)_n^{2-}$ increases as the bulkier BF_3 ligands crowd the dianion's central sulfur. Sulfur donates more negative charge to ligands as their number increases. Structural details are provided in Supporting Information.

Conclusions

The design principles guiding Dreuw and Cederbaum's construction of "center-ligand" dianions stable with respect to autoionization and dissociation are easily extended to systems with sulfur cores and pseudohalogen π-acid and BX_3

Lewis acid ligands. The thia family contains many species that are likely to be stable, even when the oxo analogue is not.

Acknowledgment. This work was supported by the University of Marmara, Istanbul, The Turkpetrol Foundation, and the University of Virginia. We thank the Body Foundation for equipment, travel support, and general encouragement and wish to acknowledge the key role in making this collaboration possible of Professor Dr. Zikri Altun of the Physics Department, the University of Marmara, Istanbul.

Supporting Information Available: Tables S1−S17. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Gutsev, G. L.; Boldyrev, A. I. *J. Phys. Chem.* **1990**, *94*, 2256

(2) Wang, X.-B.; Wang, L.-S. *J. Chem. Phys.* **1999,** *111*, 4497.

(3) Weikert, H. G.; Cederbaum, L. S.; Tarantelli, F.; Boldyrev, A. I. *Z. Phys. D-Atoms, Molecules, Clusters* **1991,** *18*, 229. Weikert, H. G.; Cederbaum, L. S. *J. Chem. Phys.* **1993,** *99*, 8877. Observation reported by Middleton, R.; Klein, J. *Phys. Rev. A* **1999,** *60*, 3515.

(4) Boldyrev, A. I.; Simons, J. *J. Chem. Phys.* **1992**, *97*, 2826.

(5) Cederbaum, L. S.; Dreuw, A. *Chem. Rev.* **2002**, *102*, 181−200

(6) Boldyrev, A. I.; Simons, J. *J. Chem. Phys.* **1993**, *98*, 4745.

(7) Berghof, V.; Sommerfeld, T.; Cederbaum, L. S. *J. Phys. Chem. A* **1998**, *102*, 5100.

(8) Maas, W. P. M.; Nibbering, N. M. M. *Int. J. Mass Spec. Ion Proc.* **1989**, *88*, 257.

(9) Scheller, M. K.; Cederbaum, L. S. *Chem. Phys. Lett.* **1993**, *216*, 441. Scheller, M. K.; Cederbaum, L. S. *J. Chem. Phys.* **1994**, *100*, 8934. Scheller, M. K.; Cederbaum, L. S. *J. Chem. Phys.* **1994**, *100*, 894. Scheller, M. K.; Cederbaum, L. S. *J. Chem. Phys.* **1994**, *101*, 3962.

(10) Dreuw, A.; Cederbaum, L. S. In *Theoretical Prospect of Negative Ions*; Kalcher, J., Ed.; Research Signpost: Trivandrum, Kerala, India, 2002; pp 1−28

(11) Blades, A. T.; Kebarle, P. *J. Am. Chem. Soc.* **1994**, *116*, 10761.

(12) Zint, N.; Dreuw, A.; Cederbaum, L. S. *J. Am. Chem. Soc.* **2002**, *124*, 4910.

(13) Sommerfeld, T. *J. Phys. Chem. A* **2000**, *104*, 8806.

(14) Janoschek, R. *Z. Anorg. Allg. Chem.* **1992**, *616*, 101.

(15) Boldyrev, A. I.; Simons, J. *J. Phys. Chem.* **1994**, *98*, 2298.

(16) Weikert, H. G.; Cederbaum, L. S. *J. Chem. Phys.* **1993**, *99*, 8877.

(17) Dreuw, A.; Schweinsberg, H.; Cederbaum, L. S. *J Phys. Chem. A* **2002**, *106*, 1406.

(18) Spartan 2002, Wavefunction, Inc., Irvine, CA. See methods described in the following: Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, C. D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rasselov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. A.; Dombrowski, J. P.; Daschel, H.; Zhang, W.; Korambath, P. P.; Baker, J.; Boyd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C.-P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532.

(19) Gaussian 03, Revision B.2, Frisch, M. G.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc., Pittsburgh, PA, 2003.

(20) Boldyrev, A.; Simons, J. *J. Phys. Chem.* **1994**, *98*, 2298.

(21) These agree closely with the vertical ionization energies estimated by MP2 according to Dreuw, A.; Sommerfeld, T.; Cederbaum, L. S. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1889. Dreuw, A.; Sommerfeld, T.; Cederbaum, L. S. *J. Chem. Phys.* **1998**, *109*, 2727.

(22) Galbraith, J. M.; Schaefer, H. F. *J. Chem. Phys.* **1996**, *105*, 862; see also Tschumper, G. S.; Schaefer, H. F. *J. Chem. Phys.* **1997**, *107*, 2529.

(23) Dreuw, A.; Schweinsberg, H.; Cederbaum, L. *J Phys. Chem. A* **2002**, *106*, 1406.

(24) Cederbaum, L. S.; Domcke, W. *Adv. Chem. Phys.* **1977**, *36*, 205. Implementation in Gaussian according to Zakrzewski, V. G.; Ortiz, J. V. *Int. J. Quantum Chem.* **1995**, *53*, 583. Zakrzewski, V. G.; Ortiz, J. V. *J. Chem. Phys.* **1995**, *102*, 294. Dolgounitcheva, O.; Zakrzewski, V. G.; Ortiz, J. V. *J. Chem. Phys.* **1998**, *109*, 87.

(25) Dreuw, A.; Cederbaum, L. S. *J. Phys. Chem. A* **2001**, *105*, 10577.

CT049854Q

JCTC Journal of Chemical Theory and Computation

Variational Iterative Time Dependent Method for Eigenvalues and Eigenfunctions of the Hamiltonian

Markus Saltzer and Eli Pollak*

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

Received January 13, 2005

Abstract: Short time information on the time evolution of wave packets is combined with the variational theorem to determine eigenvalues and eigenfunctions. As in the Filter Diagonalization Method the input that is needed is a correlation function and its time derivative. The method is iterative and convergent. The time interval needed is short, for example, the determination of tunneling splitting energies ΔE is obtained in a time interval which is substantially shorter than the Fourier time $2\pi\hbar/\Delta E$. The method is applied to some model problems including determining the ground tunneling state in a quartic double well potential using numerically exact short time results obtained from the semiclassical initial value representation series of the exact propagator. This is another example in which tunneling is obtained using only coherent classical paths. Implications of the method for ab initio computation of molecular electronic energies is discussed.

A central challenge facing theoretical Physicists and Chemists today is the invention of convergent methods for the computation of real time quantum dynamics in "large" systems. The past decade has seen significant progress. Makri and co-workers,[1,2] using the QUAPI method,[3] have managed to compute accurate quantum reaction rates in dissipative systems. Mak, Stockburger, and Grabert[4,5] have derived and applied exact stochastic Schroedinger equations. The Bohmian formulation of quantum mechanics is also used to compute real time quantum dynamics.[6] Classical propagation of coherent states is being implemented to study time dependent quantum phenomena.[7,8] Another approach is the use of multiconfigurational time dependent Hartree methods.[9] We have recently developed a SemiClassical Initial Value Representation (SCIVR) series expression for the exact propagator and used it to compute numerically exact quantum dynamics[10] as well as "deep" tunneling probabilities.[11]

In all of these methods, a serious limitation is the extent in time for which accurate results can be obtained. One way of overcoming this limitation is by use of the Filter Diagonalization Method (FDM)[12,13] where the characteristic time needed is $2\pi\hbar/\bar{E}$ where $\bar{E}$ is the local average level spacing. The central object in the FDM is the correlation function of an initial state $|\Psi\rangle$ with its time evolved form

$$c(t) = \langle\Psi|K(t)|\Psi\rangle \tag{1}$$

where $K(t) = \exp(-iHt/\hbar)$. Harmonic inversion is then used to represent the correlation function in terms of the eigenvalues of the Hamiltonian. The FDM method is not foolproof, as noted by Mandelshtam,[13] "the degree of convergence will always be a delicate issue".

A related topic is what may be broadly termed as ab initio chemistry. Present day technology is impressive. A variety of methods, such as density functional theory,[14] quantum Monte Carlo methods,[15] and basis set methods,[16] are used for the computation of molecular energies, but any improvement would be invaluable. One may thus pose the question, how can one use the information obtained from a short time computation of the correlation function to improve upon existing estimates of the energy. These two themes are the central topics of this letter.

A fundamental building block of ab initio chemistry is the Rayleigh-Ritz functional

$$E[\Psi] = \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} \geq \epsilon_0 \tag{2}$$

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

10.1021/ct0500085 CCC: $30.25
Published on Web 03/22/2005

where ϵ_0 is the ground-state energy of the Hamiltonian H. Inserting the time evolved wave function $K(t)|\Psi\rangle$ as a trial function into the functional will not improve anything, since the Hamiltonian commutes with the propagator. However, if one chooses the linear combination $|\Phi(t)\rangle = |\Psi\rangle + K(t)|\Psi\rangle$, the functional becomes time dependent, provided that the initial state $|\Psi\rangle$ is not an eigenstate of the Hamiltonian.[17]

The initial wave packet $|\Psi\rangle$ may be expanded in terms of the eigenfunctions (ψ_n) of the Hamiltonian as $|\Psi\rangle = \sum_{j=0}^{\infty} a_j|\psi_j\rangle$ and $a_j = \langle\psi_j|\Psi\rangle$. The time dependent energy functional then becomes (the eigenvalues of H are denoted as ϵ_n)

$$E[\Phi(t)] = \frac{\sum_{j=0}^{\infty}|a_j|^2\epsilon_j(1+\cos(\epsilon_j t/\hbar))}{\sum_{j=0}^{\infty}|a_j|^2(1+\cos(\epsilon_j t/\hbar))} \tag{3}$$

showing explicitly, how the time evolution changes the contribution of different states at different times to the energy functional.

Noting that $i\hbar(\mathrm{d}c(t)/\mathrm{d}t) = \langle\Psi|HK(t)|\Psi\rangle$, assuming that the initial state is normalized ($\langle\Psi|\Psi\rangle = 1$), allows us then to write the energy functional explicitly in terms of the correlation function as

$$E[\Phi(t)] = \frac{E[\Psi] - \hbar\dfrac{\mathrm{d}\ \mathrm{Im}\ c(t)}{\mathrm{d}t}}{1 + \mathrm{Re}\ c(t)} \tag{4}$$

Minimization of this functional with respect to the time will lead to an improved estimate of the ground-state energy provided that the ground-state energy is positive. As proof, we consider its short time dependence. Using the notation $E_j[\Psi] = \langle\Psi|H^j|\Psi\rangle$, one has that $\mathrm{Re}\ c(t) = 1 - (t^2/2\hbar^2)E_2[\Psi] + o(t^4)$ and similarly $-\hbar(\mathrm{d}\ \mathrm{Im}\ c(t)/\mathrm{d}t) = E_1[\Psi] - (t^2/2\hbar^2)E_3[\Psi] + o(t^4)$. It follows that

$$E[\Phi(t)] = E[\Psi] - \frac{t^2}{4\hbar^2}(E_3[\Psi] - E[\Psi]E_2[\Psi]) + o(t^4) \tag{5}$$

If the spectrum of the Hamiltonian is positive, then one readily finds that $E_3[\Psi] - E[\Psi]E_2[\Psi] \geqslant 0$ so that for short times, the time dependent functional $E[\Phi(t)]$ can only lead to a lower estimate for the ground-state energy than the original estimate $E[\Psi]$. Since the spectrum of the Hamiltonian is typically bounded from below, one can shift the Hamiltonian by a constant value to ensure that the initial time dependence is convex.

One may now use the variational theorem iteratively. One propagates $c(t)$ until the functional reaches a minimum energy $E[\Phi(t_1)]$ at the time t_1. One then resets the initial state to be $|\Psi_1\rangle = (1/\sqrt{\langle\Phi(t_1)|\Phi(t_1)\rangle})|\Phi(t_1)\rangle$ and the new time dependent function to be $|\Phi_1(t)\rangle = |\Psi_1\rangle + K(t)|\Psi_1\rangle$.

The new correlation function however can be expressed solely in terms of the previous correlation function. That is

$$c_1(t) \equiv \langle\Psi_1|K(t)|\Psi_1\rangle = \frac{c(t) + \frac{1}{2}[c(t-t_1) + c(t+t_1)]}{1 + \mathrm{Re}\ c(t_1)} \tag{6}$$

while the new functional becomes

$$E[\Phi_1(t)] = \frac{E[\Phi(t_1)] - \hbar\dfrac{\mathrm{d}\ \mathrm{Im}\ c_1(t)}{\mathrm{d}t}}{1 + \mathrm{Re}\ c_1(t)} \tag{7}$$

In other words, to carry out the iterative procedure, there is no need to know the wave function, all that is needed is the initial correlation function $c(t)$ albeit for ever increasing times, as one proceeds with the iteration. Due to the convex short time property of the functional, this iterative procedure will usually converge to an eigenstate of the Hamiltonian.

Let us assume for the moment, that we have converged to the ground-state eigenvalue at the time t_0. The ground-state wave function $|\psi_0\rangle$ is then known in principle by following the time evolution. It is a linear combination of the initial wave function $|\Psi\rangle$ propagated to the set of known discrete propagation times. Thus, its overlap with the initial wave function is given as the corresponding linear combination of the correlation function $c(t)$. To obtain the excited-state energy, one projects out the (normalized) ground-state wave function from the initial state, that is one defines $|\bar{\Psi}\rangle = |\Psi\rangle - \langle\psi_0|\Psi\rangle|\psi_0\rangle$. The correlation function $c(t) = \langle\bar{\Psi}|\hat{K}(t)|\bar{\Psi}\rangle = c(t) - e^{-i\epsilon_0 t/\hbar}|\langle\Psi|\psi_0\rangle|^2$ can be expressed in terms of the original correlation function $c(t)$ and the ground-state results only. Moreover, it is known for all the times for which the original correlation function $c(t)$ is known accurately. Therefore one may now repeat the iterative process to obtain the first excited-state eigenvalue and eigenfunction. This can then be continued for the second excited state etc.

We will first consider an almost trivial example, to demonstrate that only short time information is needed. Consider a two-state system with eigenvalues $\epsilon_1 < \epsilon_2$ and respective eigenfunctions $|\psi_1\rangle$ and $|\psi_2\rangle$. The "tunneling splitting" is by definition $\Delta E = \epsilon_2 - \epsilon_1 \ll \epsilon_1, \epsilon_2$. We choose the initial wave packet to be localized in one of the wells, that is $|\Psi\rangle = (1/\sqrt{2})(|\psi_1\rangle + |\psi_2\rangle)$. The associated energy $E[\Psi] = (1/2)(\epsilon_1 + \epsilon_2)$. The correlation function is $c(t) = \exp[-i(E[\Psi]t/\hbar)]\cos[(\Delta E/2\hbar)t]$. The time dependent energy functional is then

$$E[\Phi(t)] = \frac{\epsilon_1 + \epsilon_2 + \epsilon_1\cos\left(\frac{\epsilon_1 t}{\hbar}\right) + \epsilon_2\cos\left(\frac{\epsilon_2 t}{\hbar}\right)}{2 + \cos\left(\frac{\epsilon_1}{\hbar}t\right) + \cos\left(\frac{\epsilon_2}{\hbar}t\right)} \tag{8}$$

It has its minimum value at the time $t_1 = \pi\hbar/\epsilon_2$ at which $E[\Phi(t_1)] = \epsilon_1$ immediately giving the ground-state energy. To obtain the excited-state energy, one projects out the ground-state wave function from the initial state, that is one defines $|\Psi_1\rangle = |\Psi\rangle - [1 + c^*(t_1)]/2[1 + \mathrm{Re}\ c(t_1)]|\Phi(t_1)\rangle = 1/\sqrt{2}|\psi_2\rangle$. Clearly then $E[\Psi_1] = \epsilon_2$, demonstrating that the tunneling splitting is obtained by propagation up to time $t_1 \ll 2\pi\hbar/\Delta E$.

Table 1. Iterative Convergence toward the Ground State of a Model System

j	E_j	$\|a_j(0)\|^2$	$\|a_j(t_1^a)\|^2$	$\|a_j(t_2^a)\|^2$	$\|a_j(t_3^a)\|^2$
1	1.00000	0.98	0.9978	0.9997	0.99996
2	2.00000	0.01	0.0019	0.0002	$4\cdot10^{-5}$
3	2.52360	0.005	$1\cdot10^{-5}$	$2\cdot10^{-7}$	$1\cdot10^{-8}$
4	2.81828	0.003	0.0001	$1\cdot10^{-5}$	$2\cdot10^{-7}$
5	3.04159	0.002	0.0002	$6\cdot10^{-5}$	$6\cdot10^{-6}$

[a] The values of the times t_1, t_2, and t_3 at which one finds successive minima of the time dependent energy are 0.605, 0.66, and 0.60, respectively.

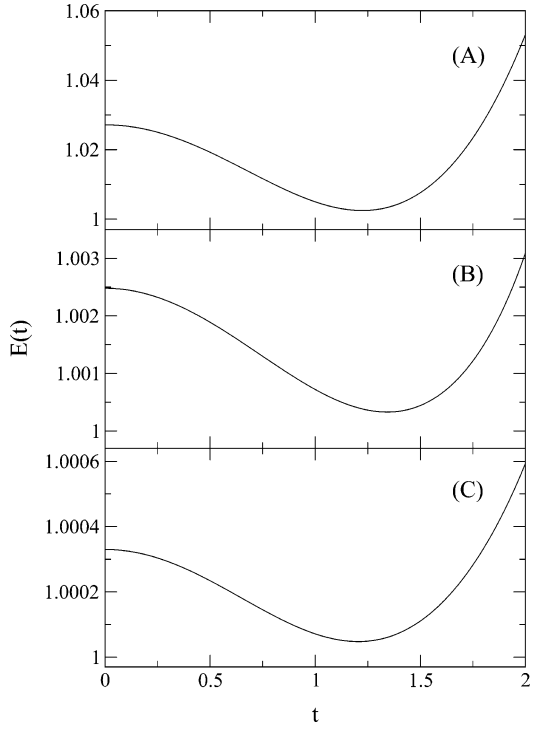


Figure 1. The time dependence of the energy functional for a model ab initio computation. The ground-state energy is 1. The initial value of the average energy, using the initial state as defined in Table 1 was 1.0271. The successive estimates for the energy, as obtained from panels a−c are 1.0025, 1.00033, and 1.000048, respectively.

To demonstrate the utility of the method for ab initio computations, we consider a model spectrum and an initial wave packet which is close to the ground state but slightly contaminated by the excited states. (The eigenvalues and coefficients of the wave packet are given in Table 1.) This would be the typical case of a "good" ab initio computation of the ground-state potential of an atom, which due to computational limitations cannot be converged precisely to the ground state. In Figure 1 we plot the first three iterations for the average energy. Each iteration provides an additional order of magnitude in the accuracy of the ground state. The total integration time needed for this type of accuracy is very short compared to using Fourier transforms.

As a third example, we consider the tunneling split ground state in a symmetric quartic double well potential $V(x) = -1.4x^2 + (1/4)x^4 + 4$. This potential has a pair of bound states whose energies 3.052830 and 3.171736 are below the barrier energy (4). The next bound state is found above the barrier at $E = 4.536449$. The initial wave packet is taken to be a Gaussian ($\langle x|\Psi\rangle = (\alpha/\pi)^{-1/4}\exp[(-\alpha/2)(x - x_0)^2]$ with $x_0 = -1.501$ and $\alpha = 2.806$), centered in the left wall, with average energy $E[\Psi] = 3.1928$.

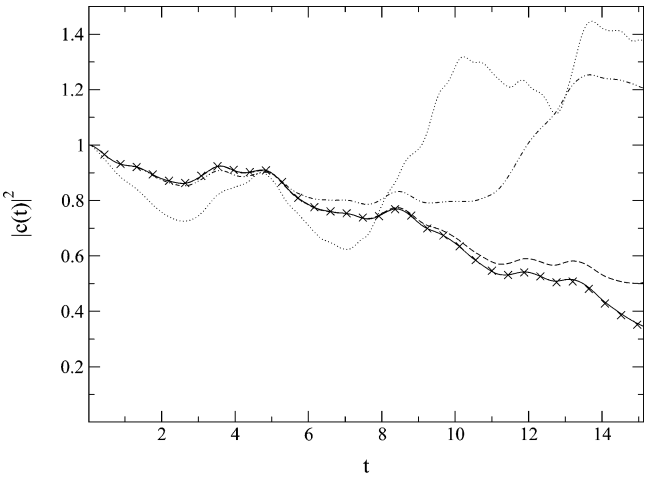


Figure 2. SCIVR series determination of the correlation function $c(t)$. The dotted, dashed−dotted, dashed, and solid lines correspond to inclusion of the 0, 0+1, 0+1+2, 0+1+2+3 terms in the SCIVR series expansion. The crosses show the numerically exact values as obtained from a basis set diagonalization computation.

This one-dimensional quantum problem can be solved numerically exactly using basis set methods. However, here we choose to employ the SCIVR series method,[10] to show explicitly that (a) even if one obtains the correlation function numerically exactly only for short times as is the case when using the SCIVR series method, one may still determine rather accurately the ground tunneling state. The systematic error of truncation of the series is not serious, and (b) the ground-state energy and wave function may be obtained by considering only the real time coherent classical paths.[11]

Briefly reviewing the SCIVR series method, the quantum propagator is approximated by the Herman-Kluk SCIVR propagator ($K_0(t)$)[18] which obeys the time evolution equation[19]

$$i\hbar\frac{\partial K_0(t)}{\partial t} = HK_0(t) + C(t) \tag{9}$$

where $C(t)$ is a known "correction operator".[10] The exact quantum propagator may then be represented exactly in terms of a power series in the correction operator $K(t) = \sum_{j=0}^{\infty} K_j(t)$ such that each successive term in the series is obtained from the recursion relation

$$K_{j+1}(t) = \frac{i}{\hbar}\int_0^t \mathrm{d}s\, K_j(t-s)C(s) \tag{10}$$

The width parameter chosen for the coherent state appearing in K_0 is $\gamma = 2.5$ chosen by minimizing the expectation value of the correction operator, as described in ref 10c. The absolute value squared of the resulting correlation function $c(t) = \langle\Psi|\ K(t)|\Psi\rangle$ obtained for increasing order in the SCIVR series is compared in Figure 2 with the numerically exact results obtained using matrix diagonalization. As shall be also shown below, it suffices to go only up to the third

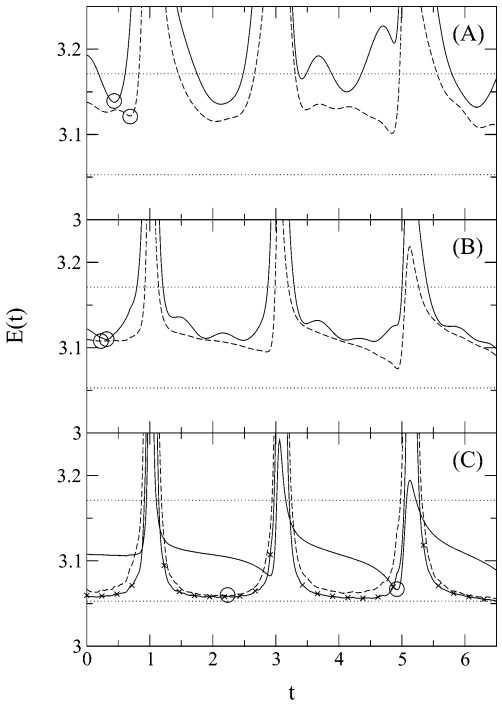


Figure 3. The iterative variational method for computing the ground-state tunneling energy in a double well quartic potential. Panel a shows the zeroth and first iteration. The small circles denote the respective time dependent minima, $t_1 = 0.435$ for the zeroth order functional and then $t_2 = 0.69$ for the first iterated functional. Panel b shows the same for the next two iterations ($t_3 = 0.315$, $t_4 = 0.225$), and panel c shows the final two iterations ($t_5 = 4.92$, $t_6 = 2.235$). The horizontal dotted lines denote the ground and first excited-state eigenvalues as determined by matrix diagonalization of the Hamiltonian. The line with crosses in panel (c) shows the energy after the sixth iteration.

order in the series to obtain results sufficiently accurate for determining the ground-state energy.

In Figure 3 we then show how the correlation function may be used to estimate the ground-state energy via the energy functional. Six iterations suffice to determine that at $t_7 = 2.07$ the ground-state eigenvalue is 3.057352 (a seventh iteration gives 3.054693), to be compared with the numerically exact value 3.052830. The total time interval needed for this determination is $t = 10.89$ which is substantially shorter than the Fourier time $t = 2\pi/\Delta E = 52.84$. One notes, that when convergence is obtained, the time dependent energy becomes rather insensitive to the time, except for short time intervals in which the ground-state contribution to the energy functional vanishes (as may be inferred from eq 3 these times are $\sim(2n + 1)\pi/\epsilon_0$, $n = 0, 1, ...$), and then the functional becomes large compared to the ground-state energy. The energy interval between successive maxima thus also provides information on the ground-state energy. The resulting ground-state wave function, obtained by following the time evolution through the iteration scheme $(N_0|\psi_0\rangle = (1 + \sum_{i=j}^{6} K(t_j) + \sum_{i<j}^{6} K(t_i + t_j) + ... + K(t_1 + t_2 + ... + t_6))|\Psi\rangle$ where the t_i's are the successive iteration times) is shown in Figure 4.

In summary, we have shown how the variational theorem may be used to determine eigenvalues of Hamiltonian

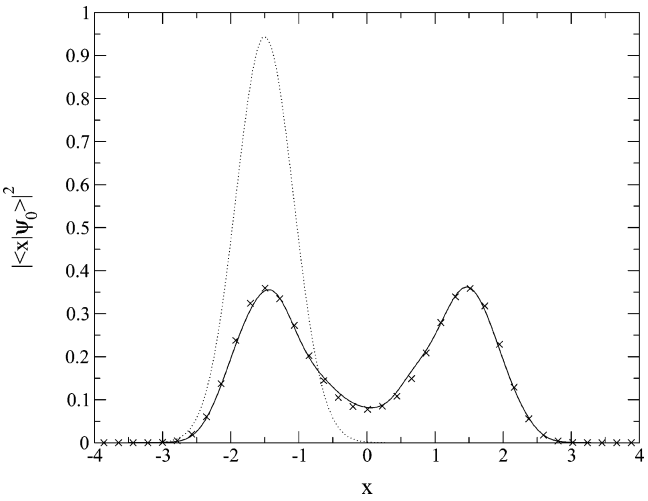


Figure 4. The ground-state tunneling wave function as determined from the iterative variational method. The crosses denote the numerically exact wave function obtained using the standard basis set method. The dotted line shows the initial Gaussian wave function $|\langle x|\Psi\rangle|^2$.

systems using short time quantum dynamics data. We have also shown that use of the SCIVR series method leads to convergent eigenvalues and that one does not need to go to a very high order in the series to obtain convergence. This methodology should be useful for extending the accuracy of ab initio computations of electronic energies, using the "best" available ab initio wave function combined with a short time quantum propagation. We have shown that this strategy ensures an even better estimate of the ground-state energy. The same methodology can then be used also to estimate the first excited-state energy. This would be a significant step forward in the implementation of ab initio technology for spectroscopy. Although in principle one could obtain the whole spectrum, each additional excited state demands an increasingly more accurate determination of the previous eigenvalues which in turn will demand longer time propagation which at some point will become prohibitive. Finally, we note that use of the variational method presented here does not suffer from the ambiguities of the nonlinear harmonic inversion method which are found in the FDM method.

Acknowledgment. M.S. gratefully acknowledges a Minerva Foundation (Munich) fellowship. This work has also been supported by grants of the U.S.−Israel Binational Science Foundation and the Israel Science Foundation. The PACS classification is (03.65.Sq,03.65.Xp,31.10.+z,31.15.-p,31.15.Gy).

References

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

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

(3) Makri, N. *Chem. Phys. Lett.* **1992**, *193*, 435.

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

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

(6) Wyatt, R. E. *J. Chem. Phys.* **2002**, *117*, 9569. Babyuk, D.; Wyatt, R. E.; Frederick, J. H. *J. Chem. Phys.* **2003**, *119*, 6482. Hughes, K. H.; Wyatt, R. E. *J. Chem. Phys.* **2004**, *120*, 4089.

(7) Shalashilin, D. V.; Child, M. S. *J. Chem. Phys.* **2000**, *113*, 10028. Shalashilin, D. V.; Child, M. S.; Clary, D. C. *J. Chem. Phys.* **2004**, *120*, 5608.

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

(9) Beck, M. H.; Jaeckle, A.; Worth, G. A.; Meyer, H.-D.; Kuehn, O. *Phys. Rep.* **2000**, *324*, 1. Wang, H.; Thoss, M. *J. Chem. Phys.* **1289**, *119*, 2003.

(10) Pollak, E.; Shao, J. *J. Phys. Chem. A* **2003**, *107*, 7112. Zhang, S.; Pollak, E. *Phys. Rev. Lett.* **2003**, *91*, 190201; *J. Chem. Phys.* **2003**, *119*, 11058; *J. Chem. Phys.* **2004**, *3384*, 121.

(11) Zhang, D.-H.; Pollak, E. *Phys. Rev. Lett.* **2004**, *93*, 140401.

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

(13) Mandelshtam, V. A. *Prof. Nuc. Magn. Res. Spectrosc.* **2001**, *38*, 159.

(14) Geerling, P.; De Proft, F.; Langenaeker, W. *Density Functional Theory, A bridge Between Chemistry and Physics*; VUB University Press: Brussels, 1999.

(15) Bernu, B.; Ceperley, D. M.; Lester, W. A. *J. Phys. Chem.* **1990**, *93*, 552; *J. Phys. Chem.* **1991**, *95*, 7782.

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

(17) A more general choice would be $|\Phi(t)\rangle = |\Psi\rangle + \alpha K(t)|\Psi\rangle$. Varying the functional with respect to α shows that it is stationary for $\alpha = 1$.

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

(19) Ankerhold, J.; Saltzer, M.; Pollak, E. *J. Chem. Phys.* **2002**, *116*, 5925.

CT0500085

JCTC Journal of Chemical Theory and Computation

TDHF Evaluation of the Dipole−Quadrupole Polarizability and Its Geometrical Derivatives

Olivier Quinet, Vincent Liégeois, and Benoît Champagne*

Laboratoire de Chimie Théorique Appliquée, Facultés Universitaires Notre-Dame de la Paix, B-5000 Namur, Belgium

Received November 15, 2004

Abstract: Analytical procedures based on the time-dependent Hartree−Fock (TDHF) scheme are elaborated to evaluate the frequency-dependent electric dipole−electric quadrupole polarizability and its derivatives with respect to atomic Cartesian coordinates. On one hand, the mixed second-order TDHF equations are solved iteratively to obtain the second-order derivatives of the linear combination of atomic orbitals coefficients, once with respect to atomic Cartesian coordinates and once with respect to external dynamic electric fields or electric field gradients. On the other hand, taking advantage of the $2n + 1$ rule, the first-order derivatives of A are expressed with respect to atomic Cartesian coordinates in terms of lower-order derivatives. These procedures have been implemented in the GAMESS quantum chemistry package and have been illustrated in the case of several small molecules as well as adamantane.

1. Introduction

The goal of this paper is to present a set of new analytical methods to compute the first-order geometrical derivatives of the electric dipole−electric quadrupole polarizability, denoted A in the following. The electric dipole−electric quadrupole polarizability and its derivatives with respect to atomic Cartesian coordinates or, more specifically, to vibrational normal mode coordinates, Q, are properties governing the interactions between molecules and electromagnetic fields, including thereof the intermolecular interactions.[1,2] Thus, A and $(\partial A/\partial Q)$ enter in the description of several optical phenomena as well as spectroscopic signatures.

A is related to both the molecular electric dipole, μ, and electric quadrupole, θ, moments, multipole moments which describe a system in weak interactions with external electric fields or with other molecules. The Hamiltonian, which accounts for these interactions, reads[1]

$$\mathscr{H} = \mathscr{H}^0 - \mu_\xi E_\xi - \frac{1}{3}\theta_{\zeta\eta} E_{\zeta\eta} - \cdots \quad (1)$$

where E_ξ and $E_{\zeta\eta}$ are components of the electric field and electric field gradient, respectively, and summations over the repeated indices is implicit. This equation assumes that the quadrupole moment operator is given in the form of a traceless quantity: $\theta_{\zeta\eta} = 3/2[q_i(r_{i\zeta}r_{i\eta} - r_i^2/3\delta_{\zeta\eta})]$. However, in the methodological part, the 'standard', Cartesian form, operator, $\theta_{\zeta\eta} = 1/2 q_i r_{i\zeta} r_{i\eta}$, is adopted, both expressions being related by a simple transformation.

The relations between, on the one hand, the electric dipole and quadrupole moments, and, on the other hand, the components of the electric field and electric field gradient are given by[1]

$$\mu_\xi = \mu_\xi^{\,0} + \alpha_{\xi\zeta} E_\zeta + 1/3\, A_{\xi,\zeta\eta} E_{\zeta\eta} + \cdots \quad (2)$$

$$\theta_{\zeta\eta} = \theta_{\zeta\eta}^{\,0} + A_{\xi,\zeta\eta} E_\xi + \cdots \quad (3)$$

where $\alpha_{\xi\zeta}$ and $A_{\xi,\zeta\eta}$ are tensor elements of the electric dipole−electric dipole and electric dipole−electric quadrupole polarizabilities, respectively; μ_ξ^0 and $\theta_{\zeta\eta}^0$ are the electric moments in the absence of external fields.

A as well as α are properties relevant to the description of intermolecular interactions and of dielectric properties.[1,2] $\partial\alpha/\partial Q$ governs the Raman scattering intensities of fundamental transitions in the harmonic approximation.[3,4] A and $\partial A/\partial Q$ are involved in collision-induced Rayleigh and Raman

* Corresponding author e-mail: benoit.champagne@fundp.ac.be.

10.1021/ct049888y CCC: $30.25
Published on Web 04/02/2005

scatterings, respectively.[5−11] For instance, Elliasmine et al.[8] obtained a value of 1 Å^4 (= 12.75 au) for the absolute value of A in CF_4. Its bond length derivative, $|\partial A/\partial R|$, was later deduced from collision-induced Raman experiment by Bancewicz et al.[9] and attains 5.3 Å^3 or 35.8 au. These experimental $|A|$ and $|\partial A/\partial R|$ values have been compared to high-level ab initio calculations performed by Maroulis.[12,13] Combining a numerical finite field approximation with the Møller−Plesset scheme, the best/recommended A value amounts to 12.4 au,[13] while the estimated $\partial A/\partial R$ is 27.7 au,[12] in good agreement with experiment.

The amplitude of the dipole−quadrupole polarizability can also be addressed from high-resolution infrared absorption spectra and in particular from the analysis of the rotational energy levels of the molecule under study interacting with surrounding molecules. Thus, using CH_4 and CD_4 trapped in solid parahydrogen, a $A(CD_4)/A(CH_4)$ ratio of 0.875 was determined by Momose and co-workers.[14,15]

Together with the electric dipole−magnetic dipole polarizability (G'), A describes the electric field-induced differential scattering effect.[16,17] It is also involved in the description of the electric-field-gradient-induced birefringence.[18] In ref 19, a coupled-perturbed Hartree−Fock (CPHF) scheme is presented to evaluate the static A value. As a preliminary step toward evaluating the derivatives of A wrt Q, the present paper also describes a method for evaluating dynamic A values.

The first-order geometrical derivatives of A enter directly in the determination of the vibrational Raman Optical activity (VROA) intensities together with the first-order geometrical derivatives of α and G'.[20−26] VROA spectroscopy, which provides molecular structure information on chiral species, is focusing an increasing interest due to its broad range of applications including the determination of absolute configurations[27] and the characterization of conformational dynamics in proteins.[28] Simulating VROA spectra by means of analytical procedures taking into account the frequency of the incident light brings indeed the main motivation for this work because, so far, the evaluation of the VROA intensities is carried out by using finite difference procedures for part or for all the response properties.

In this work, an analytical procedure is elaborated and implemented for calculating the frequency-dependent $\partial A/\partial Q$ quantities. A similar procedure was previously worked out for the $\partial\alpha/\partial Q$ which are—in addition to the vibrational frequencies—the sole required quantities for simulating Raman spectra in Placzek's approximation. A procedure for evaluating the more cumbersome $\partial G'/\partial Q$ is under development. In addition, within the evaluation of $\partial A/\partial Q$, the Time-dependent Hartree−Fock (TDHF) procedure is applied here, for the first time, to the evaluation of a response property with respect to three different perturbations: the electric fields, electric field gradients, and geometrical distortions. Section 2 presents the methodology and discusses the iterative and less (or partially) iterative schemes. The implementation and the illustration of the methods are described in section 3 before conclusions are drawn in section 4.

2. Methodology

Considering eqs 2 and 3, the elements of the dipole−quadrupole polarizability tensor can be defined either with respect to the quadrupole moment or to the dipole moment

$$A_{\xi,\zeta\eta}(-\omega;\omega) = \frac{\partial\theta_{\zeta\eta}}{\partial E_\xi(-\omega)} \tag{4a}$$

$$= \frac{\partial\mu_\xi}{\partial\nabla_\eta E_\zeta(\omega)} \tag{4b}$$

In the first equation, the quadrupole moment, i.e. $\theta_{\zeta\eta} = -Tr[D^0H^{\zeta\eta}]$, is differentiated with respect to the dynamic electric field, whereas in the second, the dipole moment, i.e. $\mu_\xi = -[TrD^0H^\xi]$, is differentiated with respect to the dynamic electric field gradient. However, both expressions lead to the same results. Subsequently, the matrix analogues of eqs 4a and 4b read

$$A_{\xi,\zeta\eta}(-\omega;\omega) = -\mathrm{Tr}[D^\xi(-\omega)H^{\zeta\eta}] \tag{5a}$$

$$= -\mathrm{Tr}[H^\xi D^{\zeta\eta}(\omega)] \tag{5b}$$

H^ξ and $H^{\zeta\eta}$ are the matrices of the ξ and $\zeta\eta$ components of the dipole and quadrupole moments in the atomic orbitals (AO) basis representation, respectively. $D^\xi(-\omega)$ and $D^{\zeta\eta}(\omega)$ are the first-order derivatives of the density matrix with respect to the electric field component directed along the ξ axis and to the electric field gradient component along the ζ and η axis, both oscillating at frequencies $-\omega$ and ω, respectively. They can be evaluated following the TDHF scheme described by Sekino and Bartlett[29] and by Karna and Dupuis[30,31] to determine electric field derivatives or its generalization to electric field gradient perturbation.

Subsequently, the derivatives of these two expressions with respect to the a Cartesian coordinate are given by

$$A^a_{\xi,\zeta\eta}(-\omega;\omega) = \frac{\partial}{\partial a}[A_{\xi,\zeta\eta}(-\omega;\omega)]$$

$$= -\mathrm{Tr}[D^{\xi,a}(-\omega)H^{\zeta\eta} + D^\xi(-\omega)H^{\zeta\eta,a}] \tag{6a}$$

$$= -\mathrm{Tr}[H^{\xi,a}D^{\zeta\eta}(\omega) + H^\xi D^{\zeta\eta,a}(\omega)] \tag{6b}$$

$H^{\xi,a}$ and $H^{\zeta\eta,a}$ are the a derivatives of the H^ξ and $H^{\zeta\eta}$ matrices, respectively. $H^{\xi,a}$ is also required to determine the IR intensities. $D^{\xi,a}(-\omega)$ and $D^{\zeta\eta,a}(\omega)$ are mixed second-order derivatives of the density matrix. The first can be obtained via the iterative procedure described in ref 32, which provides the frequency-dependent Raman intensities, whereas the procedure to evaluate the second quantities is presented below.

2.1. Basics of the Mixed TDHF Approach. The strategy of refs 32−35 is followed. It combines procedures for evaluating derivatives with respect to oscillating electric fields[29−31] and with respect to atomic Cartesian coordinates.[36,37] Therefore, it extends to dynamic perturbations the successive derivative schemes developed in the book of Schaefer and collaborators.[38] Several theoretical and methodological aspects related to these mixed dynamic derivatives have recently been reviewed by one of us.[39] The starting

point consists of the TDHF equation

$$FC - i\frac{\partial SC}{\partial t} = SC\epsilon \tag{7}$$

the normalization condition of the wave function

$$C^{\dagger}SC = 1 \tag{8}$$

and the definition of the density matrix

$$D = CnC^{\dagger} \tag{9}$$

where F, ϵ, S, and C are the Fock matrix, the energy matrix, the overlap matrix, and the linear combination of atomic orbitals (LCAO) matrix, respectively. n is the diagonal matrix of occupation numbers (here, for a closed-shell system, $n_{ii} = 2$ for occupied molecular orbitals and zero otherwise). In addition to the geometrical distortions associated with each of the $3N$ atomic Cartesian coordinates (a)

$$\lambda^{a} = x_{a} \tag{10}$$

and the application of an external oscillating electric field oriented along the ξ axis

$$\lambda^{\xi} = E_{\xi}e^{i\omega_{\xi}t} = (\lambda^{-\xi})^{*} \tag{11}$$

gradient along the η axis of an electric field oriented along the ζ axis is also considered

$$\lambda^{\zeta\eta} = \nabla_{\eta}E_{\zeta}e^{i\omega_{\zeta\eta}t} = (\lambda^{-\zeta\eta})^{*} \tag{12}$$

Each matrix is developed with respect to the three kinds of perturbation. For example, the Fock matrix

$$F = F^{0} + \sum_{i}^{\xi,\zeta\eta,a}\lambda^{i}F^{i} + \frac{1}{2}\sum_{ij}^{\xi,\zeta\eta,a}\lambda^{i}\lambda^{j}F^{ij} + ... \tag{13}$$

where the summations run not only over all $3N$ coordinates, over all the field directions (x, y, z), and over all the electric field gradient directions (xx, xy, ..., zz) but also over all opposite directions ($-x$, $-y$, $-z$ and $-xx$, $-xy$, ..., $-zz$) in order to obtain, after combination, real electric field and electric field gradient components. For first order in the electric field gradient as well as for mixed second order in the electric field gradient and geometrical distortion, the two sets of equations are

$\lambda^{\zeta\eta}$:

$$F^{0}C^{\zeta\eta} + F^{\zeta\eta}C^{0} + \omega_{\zeta\eta}S^{0}C^{\zeta\eta} = S^{0}C^{0}\epsilon^{\zeta\eta} + S^{0}C^{\zeta\eta}\epsilon^{0} \tag{14}$$

$$C^{(-\zeta\eta)\dagger}S^{0}C^{0} + C^{0\dagger}S^{0}C^{\zeta\eta} = 0 \tag{15}$$

$$D^{\zeta\eta} = C^{\zeta\eta}nC^{0\dagger} + C^{0}nC^{(-\zeta\eta)\dagger} \tag{16}$$

$\lambda^{\zeta\eta}\lambda^{a}$:

$$F^{0}C^{\zeta\eta,a} + F^{\zeta\eta}C^{a} + F^{a}C^{\zeta\eta} + F^{\zeta\eta,a}C^{0} + \omega_{\zeta\eta}S^{0}C^{\zeta\eta,a} + \omega_{\zeta\eta}S^{a}C^{\zeta\eta} = S^{0}C^{0}\epsilon^{\zeta\eta,a} + S^{0}C^{a}\epsilon^{\zeta\eta} + S^{a}C^{0}\epsilon^{\zeta\eta} + S^{a}C^{\zeta\eta}\epsilon^{0} + S^{0}C^{\zeta\eta,a}\epsilon^{0} + S^{0}C^{\zeta\eta}\epsilon^{a} \tag{17}$$

$$C^{(-\zeta\eta),a\dagger}S^{0}C^{0} + C^{(-\zeta\eta)\dagger}S^{a}C^{0} + C^{(-\zeta\eta)\dagger}S^{0}C^{a} + C^{a\dagger}S^{0}C^{\zeta\eta} + C^{0\dagger}S^{a}C^{\zeta\eta} + C^{0\dagger}S^{0}C^{\zeta\eta,a} = 0 \tag{18}$$

$$D^{\zeta\eta,a} = C^{\zeta\eta,a}nC^{0\dagger} + C^{\zeta\eta}nC^{a\dagger} + C^{a}nC^{(-\zeta\eta)\dagger} + C^{0}nC^{(-\zeta\eta),a\dagger} \tag{19}$$

2.2. Solution of the TDHF Equations. Following the usual strategy, the (mixed) derivatives of the LCAO matrix, $C^{\zeta\eta}$ and $C^{\zeta\eta,a}$, which enter into the definition of the $D^{\zeta\eta}$ and $D^{\zeta\eta,a}$ matrices, are written as products of the unperturbed LCAO matrix and a matrix denoted $U^{\zeta\eta}$ and $U^{\zeta\eta,a}$, respectively,

$$C^{x}_{ri}(\pm\omega) = \sum_{j}^{\mathrm{MO}}C^{0}_{rj}U^{x}_{ji}(\pm\omega) \tag{20}$$

where the x subscript stands for any of these derivatives. The derivatives of the Fock matrix are expressed in the MO basis representation, e.g.

$$G^{x}_{ij}(\pm\omega) = \sum_{rs}^{AO}C^{0\dagger}_{ir}F^{x}_{rs}(\pm\omega)C_{sj}{}^{0} \tag{21}$$

Then, using the definitions of the U matrices (eq 20), the derivatives of the normalization condition (eqs 15 and 18) are rewritten as follows

$$U^{x\dagger}(\mp\omega) + U^{x}(\pm\omega) = X^{x}(\pm\omega) \tag{22}$$

where $X^{x}(\pm\omega)$ gathers the suborder terms, if any. The latter are defined in the Appendix.

Using eq 22 and following ref 32, it can be shown that only the individual occupied-virtual and virtual-occupied elements of the $U^{x}(\pm\omega)$ matrices are required to evaluate the derivative of the density matrix. These off-diagonal block elements are obtained by solving the corresponding TDHF eqs 14 and 17. This is done by left multiplying these two equations by $C^{0\dagger}$ and by considering the noncanonical approximation, in which the (mixed) derivatives of the energy matrix is considered block diagonal, i.e., $\epsilon^{x}_{ij}(\pm\omega) = 0$ with $(i, j) = (j, i) = (\mathrm{occ}, \mathrm{virt})$. Using eqs 20 and 21 as well as related relations, the off-diagonal block elements of the $U^{x}(\pm\omega)$ are given by

$$U^{x}_{ij}(\pm\omega) = \frac{G^{x}_{ij}(\pm\omega) + Q^{x}_{ij}(\pm\omega)}{\epsilon_{j}{}^{0} - \epsilon_{i}{}^{0} \mp \omega} \tag{23}$$

with $Q^{x}_{ij}(\pm\omega)$ gathering all the suborder terms from eqs 14 and 17. They are defined in the Appendix. $G^{x}_{ij}(\pm\omega)$ depends on the $U^{x}_{ij}(\pm\omega)$ matrix elements via eqs 20 and 21 as well as the definition of $F^{x}_{rs}(\pm\omega)$. Equation 23 is thus solved iteratively by using $1/2\ X^{x}_{ij}(\pm\omega)$ for initializing $U^{x}_{ij}(\pm\omega)$.

2.3. First-Order Iterative Procedure To Compute $A_{\xi,\zeta\eta}{}^{a}(\mp\omega;\pm\omega)$. In this section, the $2n + 1$ rule[40,41] is used to evaluate the $\mathrm{Tr}[D^{\xi,a}(-\omega)H^{\zeta\eta}]$ term, avoiding the explicit determination of the mixed derivatives of the density matrix. The procedure leading to the less iterative expression is decomposed into the following steps:

1. Left multiply the equivalent eq 19 of ref 32 for the ξ electric field component by $C^{\zeta\eta\dagger}(-\omega)$.

2. Right multiply the adjoint of eq 14 for the $-(\zeta\eta)$ component by $C^{\xi,a}(-\omega)$.

3. Subtract the result of step 2 from the result of step 1 and take the sum over all diagonal occupied elements.

4. Take the adjoint of the equations resulting from steps 1 and 2, follow step 3, and replace $\pm\omega$ by $\mp\omega$.

5. Sum the equations resulting from steps 3 and 4. The terms multiplying the $F_{rs}^{\zeta\eta}(\omega)$ matrix, i.e. $C_{si}^{0}C_{ir}^{\xi,a\dagger}(\omega) + C_{si}^{\xi,a}(-\omega)C_{ir}^{0\dagger}$, can be rewritten as, $1/2D_{sr}^{\xi,a}(-\omega) - 1/2\mathscr{D}_{sr}^{\xi,a}(-\omega)$ with $D^{\xi,a}$ and $\mathscr{D}^{\xi,a}$ defined in the Appendix.

6. By developing the derivatives of the Fock matrix ($F_{rs}^{\zeta\eta}(\omega)$ and $F_{rs}^{\xi,a}(-\omega)$) and by taking advantage of the superproduct between density matrices and two-electron integrals $(D_{sr}^{\xi,a}(-\omega)D_{qp}^{\zeta\eta}(\omega)[2J - K]_{pq,rs}^{0} = D_{sr}^{\zeta\eta}(\omega)\text{-}D_{qp}^{\xi,a}(-\omega)[2J - K]_{pq,rs}^{0})$, one finally obtains

$$
\begin{aligned}
&A_{\xi,\zeta\eta}^{a}(-\omega;\omega)\\
&= -\sum_{r,s}^{\mathrm{AO}} D_{sr}^{\xi}(-\omega)H_{rs}^{\zeta\eta,a} + D_{sr}^{\zeta\eta}(\omega)H_{rs}^{\xi,a}\\
&\quad - \sum_{p,q,r,s}^{\mathrm{AO}} D_{sr}^{\zeta\eta}(\omega)D_{qp}^{\xi}(-\omega)[2J - K]_{pq,rs}^{a}\\
&\quad -2\sum_{i,j}^{\mathrm{MO}} \mathscr{U}_{ij}^{\xi,a}(-\omega)G_{ji}^{\zeta\eta}(\omega) + \mathscr{U}_{ij}^{\zeta\eta,a}(\omega)G_{ji}^{\xi}(-\omega) +\\
&\qquad\qquad \mathscr{U}_{ij}^{\xi,\zeta\eta}(-\omega,\omega)G_{ji}^{a}\\
&\quad -2\sum_{i}^{\mathrm{occ}} \omega \mathscr{S}_{i}^{\xi,\zeta\eta,a}(-\omega,\omega) + \mathscr{X}_{i}^{\xi,\zeta\eta,a}(-\omega,\omega)\epsilon_{i}^{0}\\
&\quad -2\sum_{i,j}^{\mathrm{occ}} X_{ij}^{\xi,a}(-\omega)\epsilon_{ji}^{\zeta\eta}(\omega) + X_{ij}^{\zeta\eta,a}(\omega)\epsilon_{ji}^{\xi}(-\omega) +\\
&\qquad\qquad X_{ij}^{\xi,\zeta\eta}(-\omega,\omega)\epsilon_{ji}^{a} \quad (24)
\end{aligned}
$$

Intermediate quantities are defined in the Appendix. The last expression requires only first-order derivatives of the wave function with respect to the atomic Cartesian coordinates, the electric field components, or their gradients.

The procedure to evaluate the $Tr[H^{\xi}D^{\zeta\eta,a}(\omega)]$ term, avoiding the explicit determination of $D^{\zeta\eta,a}(\omega)$, leads to the same expression of $A_{\xi,\zeta\eta}^{a}(-\omega;\omega)$ (eq 24). This is achieved by adopting the same procedure as described above, besides the interchange of the field and field gradient indices and of their associated pulsations. In other words, this means that eq 24 is symmetric with respect to the field indices and the field gradient indices.

3. Implementation and Results

All these procedures have been implemented into the GAMESS quantum chemistry package[42] within the TDHFX module and have been checked against a numerical finite distortion scheme. In the later, the dipole−quadrupole polarizability is evaluated for distorted structures along the different vibrational normal coordinates, and the derivatives are evaluated using finite difference expressions. The amplitude of the Cartesian displacements amounts to ±0.01, ±0.02, ±0.04, and ±0.08 au, while the Romberg quadrature[43] is employed in order to increase the accuracy on the numerical derivatives. Within the given precision of 10^{-5} au both the numerical and analytical derivative schemes provide the same values. In addition, the iterative and the less iterative procedures give the same results within the numerical threshold (10^{-7}). The CPU time ratios between the two types of analytical procedures are greater than one in favor of the less iterative one (Table 1). It increases with the size of the system. As expected, the numerical procedure is, computation-wise, the less advantageous one, except when considering a reduced set of modes of interest.

Table 1. Relative CPU Time for Evaluating the $\partial A/\partial Q$ Quantities Using the Analytical Iterative and Numerical Differentiation Schemes[a]

	method		
	analytical/	numerical	
molecule	iterative	one mode	all modes
methane	3	$0.23 \times 8 = 1.8$	$0.23 \times 8 \times 9 = 17$
adamanthane	48	$0.16 \times 8 = 1.3$	$0.16 \times 8 \times 72 = 92$

[a] The CPU ratios are given with respect to the CPU times for the less-iterative $(2n + 1)$ procedure. The calculations were performed on V60x of SUN Microsystems. The less-iterative procedure took 5.4 and 261.4 min for methane and adamantane, respectively.

For illustration purpose, these methods are applied to the determination of the A and $(\partial A/\partial Q)_0$ quantities of three reference systems of T_d symmetry, methane (CH_4), tetrafluoromethane (CF_4), and adamantane ($C_{10}H_{16}$) as well as for three other molecules of other symmetry: water (H_2O, C_{2v}), ammonia, (NH_3, C_{3v}), and benzene (C_6H_6, D_{6h}). For the molecules of T_d symmetry, the electric dipole−electric quadrupole polarizability tensor possesses only one independent component, $A_{x,yz}$, provided the standard orientation is adopted. Moreover, the derivatives of $A_{x,yz}$ with respect to the vibrational normal coordinates present a nonzero value, independent of the permutation of the x, y, and z indices, only for the totally symmetric modes. These two quantities, $A_{x,yz}$ and $\partial A_{x,yz}/\partial Q$, are determined within collision-induced light scattering and high-resolution infrared absorption spectroscopies.[8,9,14,15] However, for modes of T_2 symmetry, the independent/invariant quantity is the norm of the vector part of the $\partial A/\partial Q$ tensor. Following the expression of Zyss[44] for tensors of rank 3, it reads

$$X_{\mathrm{vec}} = \sqrt{X_x^{\,2} + X_y^{\,2} + X_z^{\,2}} \quad (25)$$

with

$$X_i = \frac{1}{3}\sum_{j}^{x,y,z} X_{i,jj} + X_{j,ij} + X_{j,ji} \quad (26)$$

where X stands for A or $\partial A/\partial Q$. These results are listed in Tables 2−3, 4−5, and 6−7 for methane, tetrafluoromethane, and adamantane, respectively.

The calculations on CH_4 and CF_4 are carried out using the cc-pvdz and aug-cc-pvdz basis sets.[45] For adamantane, the DV0 basis set used in ref 46 has been chosen. For comparison purposes, the electric dipole−electric dipole polarizability and its vibrational normal coordinate deriva-

Table 2. Frequency Dispersion of the Dipole−Dipole and Dipole−Quadrupole Polarizabilities and Their Derivatives with Respect to Normal Coordinates of A_1 Symmetry for the Methane Molecule (cc-pvdz and aug-cc-pvdz Basis Sets)[a]

	cc-pvdz		aug-cc-pvdz	
ω	$\bar{\alpha}$	$\partial\bar{\alpha}/\partial Q$ $\bar{\nu}=3165$	$\bar{\alpha}$	$\partial\bar{\alpha}/\partial Q$ $\bar{\nu}=3152$
0.00	13.012774	−0.159903	15.998164	0.185984
0.02	13.028341	−0.160293	16.018242	0.186443
0.04	13.075335	−0.161472	16.078860	0.187832
0.06	13.154650	−0.163473	16.181196	0.190192
0.08	13.267828	−0.166352	16.327290	0.193592
0.10	13.417141	−0.170190	16.520168	0.198139
A_α, $A_{\partial\alpha/\partial Q}$	2.98	6.07	3.13	6.14
ω	$A_{x,yz}$	$\partial A_{x,yz}/\partial Q$ $\bar{\nu}=3165$	$A_{x,yz}$	$\partial A_{x,yz}/\partial Q$ $\bar{\nu}=3152$
0.00	14.405079	−0.327439	9.476620	0.246624
0.02	14.431805	−0.328285	9.498117	0.247390
0.04	14.512530	−0.330848	9.563175	0.249713
0.06	14.648930	−0.335194	9.673527	0.253670
0.08	14.843897	−0.341444	9.832198	0.259398
0.10	15.101692	−0.349775	10.043725	0.267104
A_A, $A_{\partial A/\partial Q}$	4.63	6.44	5.65	7.73

[a] All the values are given in atomic units (1 au of $\alpha = 1.6488 \times 10^{-41}$ C^2 m^2 J^{-1} and 1 au of $A = 8.724958 \times 10^{-52}$ C^2 m^3 J^{-1}), besides the vibrational pulsation in cm^{-1}.

Table 3. Frequency Dispersion of the Derivatives of the Dipole−Quadrupole Polarizabilities with Respect to Normal Coordinates of T_2 Symmetry for the Methane Molecule (cc-pvdz and aug-cc-pvdz Basis Sets)[a]

	cc-pvdz $(\partial A/\partial Q)_{vec}$		aug-cc-pvdz $(\partial A/\partial Q)_{vec}$	
ω	$\bar{\nu}=1434$	$\bar{\nu}=3285$	$\bar{\nu}=1425$	$\bar{\nu}=3265$
0.00	0.215050	0.599220	0.247610	0.787832
0.02	0.215558	0.600470	0.248196	0.789531
0.04	0.217097	0.604252	0.249971	0.794670
0.06	0.219709	0.610662	0.252993	0.803389
0.08	0.223471	0.619866	0.257362	0.815931
0.10	0.228497	0.632115	0.263231	0.832658
$A_{\partial A/\partial Q}$	5.88	5.20	5.88	5.37

[a] All the values are given in atomic units, besides the vibrational pulsation in cm^{-1}.

Table 4. Frequency Dispersion of the Dipole−Dipole and Dipole−Quadrupole Polarizabilities and Their Derivatives with Respect to Normal Coordinates of A_1 Symmetry for the Tetrafluoromethane Molecule (cc-pvdz and aug-cc-pvdz Basis Sets)[a]

	cc-pvdz		aug-cc-pvdz	
ω	$\bar{\alpha}$	$\partial\bar{\alpha}/\partial Q$ $\nu=996$	$\bar{\alpha}$	$\partial\bar{\alpha}/\partial Q$ $\bar{\nu}=982$
0.00	10.989920	−0.023478	16.162236	0.031226
0.02	10.994104	−0.023507	16.169889	0.031263
0.04	11.006682	−0.023592	16.192911	0.031373
0.06	11.027736	−0.023735	16.231501	0.031559
0.08	11.057400	−0.023938	16.285995	0.031821
0.10	11.095869	−0.024202	16.356875	0.032165
A_α, $A_{\partial\alpha/\partial Q}$	0.95	3.02	1.18	2.93
ω	$A_{x,yz}$	$\partial A_{x,yz}/\partial Q$ $\bar{\nu}=996$	$A_{x,yz}$	$\partial A_{x,yz}/\partial Q$ $\bar{\nu}=982$
0.00	12.838830	−0.077774	10.943045	0.073150
0.02	12.845476	−0.077855	10.950699	0.073243
0.04	12.865457	−0.078096	10.973717	0.073522
0.06	12.898904	−0.078502	11.012271	0.073991
0.08	12.946040	−0.079075	11.066651	0.074656
0.10	13.007180	−0.079822	11.137268	0.075525
A_A, $A_{\partial A/\partial Q}$	1.29	2.58	1.75	3.16

[a] All the values are given in atomic units, besides the vibrational pulsation in cm^{-1}.

Table 5. Frequency Dispersion of the Derivatives of the Dipole−Quadrupole Polarizabilities with Respect to Normal Coordinates of T_2 Symmetry for the Tetrafluoromethane Molecule (aug-cc-pvdz Basis Sets)[a]

	cc-pvdz $(\partial A/\partial Q)_{vec}$		aug-cc-pvdz $(\partial A/\partial Q)_{vec}$	
ω	$\bar{\nu}=686$	$\bar{\nu}=1448$	$\bar{\nu}=671$	$\bar{\nu}=1414$
0.00	0.034040	0.076902	0.044925	0.183403
0.02	0.034073	0.077016	0.044974	0.183637
0.04	0.034171	0.077358	0.045122	0.184342
0.06	0.034337	0.077933	0.045371	0.185528
0.08	0.034571	0.078747	0.045724	0.187211
0.10	0.034876	0.079809	0.046185	0.189416
$A_{\partial A/\partial Q}$	2.40	3.69	2.73	3.18

[a] All the values are given in atomic units, besides the vibrational pulsation in cm^{-1}.

tives are also evaluated following the procedure described in ref 32. Only the average value—corresponding to one-third of the trace of the polarizability tensor—and its derivatives with respect to totally symmetric modes are considered.

For the water and ammonia molecules, the computations have been carried out with the aug-cc-pvdz basis set, whereas for the benzene, the DV0 basis set has been used. Only the nonzero vector components, or the mean value for degenerate modes, are reported. In water, the nonzero quantities are A_z, $(\partial A/\partial Q)_z$ for the A_1 bending and symmetric stretching modes, and $(\partial A/\partial Q)_y$ for the B_2 asymmetric stretching mode. In the case of ammonia, A_z is different from zero as well as $(\partial A/\partial Q)_z$ for the A_1 modes and $(\partial A/\partial Q)_x = (\partial A/\partial Q)_y$, denoted hereafter $(\partial A/\partial Q)_{(x=y)}$, for the degenerate E modes. Finally, A of benzene is zero. The nonzero $(\partial A/\partial Q)$ terms are associated with the E_{1u} and A_{2u} modes. These results are given in Tables 8−10 for water, ammonia, and benzene, respectively.

The effects of the frequency have been addressed by performing calculations with a pulsation of the external electric field (or electric field gradient) ranging between 0.00 and 0.10 au. This corresponds to a photon energy ranging between 0.0 and 2.71 eV or a wavelength between ∞ and 456 nm. The frequency effects are characterized by the coefficients (A, B, ...) of the power series expansion in ω^2:

$$\alpha(-\omega;\omega) = \alpha(0;0)[1 + A_\alpha\omega^2 + B_\alpha\omega^4 + \cdots] \quad (27)$$

$$A(-\omega;\omega) = A(0;0)[1 + A_A\omega^2 + B_A\omega^4 + \cdots] \quad (28)$$

Table 6. Frequency Dispersion of the Dipole−Dipole and Dipole−Quadrupole Polarizabilities and Their Derivatives with Respect to Normal Coordinates of A_1 Symmetry for the Adamantane Molecule (DV0 Basis Set)[a]

		$\partial\bar{\alpha}/\partial Q$				
ω	$\bar{\alpha}$	$\bar{\nu}=805$	$\bar{\nu}=1149$	$\bar{\nu}=1679$	$\bar{\nu}=3057$	$\bar{\nu}=3114$
0.00	89.821106	0.092692	0.002022	0.009827	0.045178	0.354509
0.02	89.909246	0.092874	0.002019	0.009846	0.045243	0.355170
0.04	90.174820	0.093423	0.002011	0.009904	0.045438	0.357163
0.06	90.621335	0.094352	0.001998	0.010002	0.045769	0.360529
0.08	91.254776	0.095676	0.001977	0.010140	0.046241	0.365331
0.10	92.083824	0.097425	0.001950	0.010320	0.046866	0.371668
A_α, $A_{\partial\alpha/\partial Q}$	2.45	4.90	−3.26	4.90	3.57	4.65
		$\partial A_{x,yz}/\partial Q$				
ω	$A_{x,yz}$	$\bar{\nu}=805$	$\bar{\nu}=1149$	$\bar{\nu}=1679$	$\bar{\nu}=3057$	$\bar{\nu}=3114$
0.00	9.736843	−0.074647	−0.080733	0.026879	1.774280	0.035720
0.02	9.741735	−0.074856	−0.080809	0.027157	1.778279	0.035569
0.04	9.756504	−0.075487	−0.081038	0.027999	1.790362	0.035112
0.06	9.781434	−0.076557	−0.081424	0.029425	1.810798	0.034332
0.08	9.817037	−0.078091	−0.081974	0.031475	1.840048	0.033203
0.10	9.864102	−0.080128	−0.082696	0.034204	1.878795	0.031684
A_A, $A_{\partial A/\partial Q}$	1.25	6.97	2.35	25.78	5.62	−10.51

[a] All the values are given in atomic units, besides the vibrational pulsation in cm^{-1}.

Table 7. Frequency Dispersion of the Derivatives of the Dipole−Quadrupole Polarizabilities with Respect to Normal Coordinates of T_2 Symmetry for the Adamantane Molecule (DV0 Basis Set)[a]

	$(\partial A/\partial Q)_{vec}$				
ω	$\bar{\nu}=508$	$\bar{\nu}=722$	$\bar{\nu}=866$	$\bar{\nu}=1083$	$\bar{\nu}=1248$
0.00	0.000084	0.076755	0.418069	0.166360	0.046304
0.02	0.000039	0.076967	0.418903	0.166730	0.046355
0.04	0.000095	0.077607	0.421422	0.167849	0.046509
0.06	0.000319	0.078689	0.425680	0.169741	0.046768
0.08	0.000634	0.080238	0.431766	0.172452	0.047132
0.10	0.001044	0.082287	0.439814	0.176048	0.047605
$A_{\partial A/\partial Q}$	/	6.88	4.97	5.44	2.77
	$(\partial A/\partial Q)_{vec}$				
ω	$\bar{\nu}=1505$	$\bar{\nu}=1555$	$\bar{\nu}=1659$	$\bar{\nu}=3062$	$\bar{\nu}=3089$	$\bar{\nu}=3123$
0.00	0.203037	0.628231	0.358723	0.557977	3.347433	1.345444
0.02	0.203447	0.629240	0.359324	0.558745	3.353880	1.347922
0.04	0.204684	0.632285	0.361137	0.561059	3.373355	1.355409
0.06	0.206770	0.637421	0.364192	0.564955	3.406265	1.368063
0.08	0.209744	0.644743	0.368541	0.570492	3.453312	1.386158
0.10	0.213660	0.654390	0.374259	0.577756	3.515531	1.410099
$A_{\partial A/\partial Q}$	5.04	4.01	4.18	3.43	4.80	4.59

[a] All the values are given in atomic units, besides the vibrational pulsation in cm^{-1}.

which are determined by least-squares fittings. In all cases, the resulting correlation coefficient (R^2) is larger than 0.999. The results are listed in Tables 2 and 3, 4 and 5, and 6 and 7 for CH_4, CF_4, and adamantane, respectively, for modes having A_1 (T_2) symmetry. The results for water, ammonia, and benzene are listed in Tables 8−10, respectively. A brief discussion of these results is given below.

Methane. The static (CPHF) aug-cc-pvdz A ($=A_{x,yz}$) value (9.48 au) is in good agreement with the SCF results of Maroulis,[13,47] who pointed out that electron correlation effects are small and mostly accounted for at the MP2 level: the SCF, MP2, and CCSD(T) values from ref 13 are 9.48, 8.98, and 9.03 au, respectively. A increases with the frequency faster than $\bar{\alpha}$ as demonstrated by the larger value of A_A with respect to A_α. The frequency dispersion is also stronger for $\partial A/\partial Q$ than for $\partial\alpha/\partial Q$. Although it is not the main focus of the present investigation, the experimental A value determined by Buck et al.[48] is a bit larger and attains 11.3 au.

Tetrafluoromethane. The static aug-cc-pvdz A value (10.94 au) of CF_4 is slightly larger than for CH_4, but its frequency dispersion is smaller. At the SCF level, Maroulis obtained a value of 11.15 au, while including electron correlation at the MP2 and MP4 levels[12] (different basis sets are used for evaluating the SDQ and T fourth-order corrections) leads to A values of 12.00 and 12.27 au, respectively. The agreement with the experimentally derived values of

Table 8. Frequency Dispersion of the Nonzero Vector Component of the Dipole−Quadrupole Polarizability and Its Derivatives with Respect to Normal Coordinates for the Water Molecule (aug-cc-pvdz Basis Set)[a]

ω	A_z	$(\partial A/\partial Q)_z$ $\bar{\nu} = 1742$ (A_1)	$(\partial A/\partial Q)_z$ $\nu = 4133$ (A_1)	$(\partial A/\partial Q)_y$ $\bar{\nu} = 4236$ (B_2)
0.00	−6.158025	0.109973	−0.310840	0.344829
0.02	−6.168362	0.110137	−0.311397	0.345375
0.04	−6.199663	0.110632	−0.313083	0.347027
0.06	−6.252817	0.111469	−0.315945	0.349825
0.08	−6.329399	0.112668	−0.320066	0.353841
0.10	−6.431829	0.114261	−0.325575	0.359187
A,$A_{\partial A/\partial Q}$	4.18	3.71	4.46	3.94

[a] All the values are given in atomic units, besides the vibrational pulsation in cm^{-1}.

Table 9. Frequency Dispersion of the Nonzero Vector Component of the Dipole−Quadrupole Polarizability and Its Derivatives with Respect to Normal Coordinates for the Ammonia Molecule (aug-cc-pvdz Basis Set)[a]

ω	A_z	$(\partial A/\partial Q)_z$ $\bar{\nu} = 1105$ (A_1)	$(\partial A/\partial Q)_{(x=y)}$ $\bar{\nu} = 1765$ (E)	$(\partial A/\partial Q)_z$ $\bar{\nu} = 3685$ (A_1)	$(\partial A/\partial Q)_{(x=y)}$ $\bar{\nu} = 3815$ (E)
0.00	4.967627	0.148364	0.150261	0.344611	0.495260
0.02	4.978129	0.148619	0.150576	0.345387	0.496125
0.04	5.010083	0.149386	0.151533	0.347739	0.498741
0.06	5.064903	0.150680	0.153164	0.351737	0.503179
0.08	5.145182	0.152521	0.155527	0.357502	0.509565
0.10	5.255131	0.154933	0.158718	0.365222	0.518092
A, $A_{\partial A/\partial Q}$	5.24	4.29	5.22	5.61	4.35

[a] All the values are given in atomic units, besides the vibrational pulsation in cm^{-1}.

Table 10. Frequency Dispersion of the Nonzero Vector Components of the Derivatives with Respect to Normal Coordinates of the Dipole−Quadrupole Polarizabilities for the Benzene Molecule (DV0 Basis Set)[a]

ω	$(\partial A/\partial Q)_z$ $\bar{\nu} = 781$ (A_{2u})	$(\partial A/\partial Q)_{(-x=y)}$ $\bar{\nu} = 1138$ (E_{1u})	$(\partial A/\partial Q)_{(x=y)}$ $\bar{\nu} = 1636$ (E_{1u})	$(\partial A/\partial Q)_{(x=y)}$ $\bar{\nu} = 3407$ (E_{1u})
0.00	−0.088044	0.346603	0.492332	2.355252
0.02	−0.089535	0.348050	0.494174	2.359332
0.04	−0.094116	0.352482	0.499819	2.371645
0.06	−0.102127	0.360183	0.509648	2.392409
0.08	−0.114196	0.371684	0.524369	2.421996
0.10	−0.131353	0.387857	0.545154	2.460949
$A_{\partial A/\partial Q}$	41.7	10.3	9.21	4.32

[a] All the values are given in atomic units, besides the vibrational pulsation in cm^{-1}.

Elliasmine et al.[8] (12.75 au) is very good. In what concerns $\partial A/\partial Q$ of CF_4, both its static value and its frequency dispersion are smaller than in the case of methane. On the other hand, similar to methane, the frequency dispersions of A and $\partial A/\partial Q$ are stronger than for α and $\partial\alpha/\partial Q$, respectively. The corresponding $\partial A/\partial R$, which can be deduced from $\partial A/\partial Q$, where Q is the normal coordinate of the totally symmetric vibration, amounts to the same value of 27.7 au as in ref 12 in the static limit.

Basis Set Effects. The inclusion of diffuse and polarized functions, characterizing the passage from the cc-pvdz to aug-cc-pvdz basis sets, produces opposite effects upon the dipole−dipole and dipole−quadrupole polarizability responses and their derivatives with respect to totally symmetric modes. On the one hand, an increase of the $\bar{\alpha}$ and $\partial\bar{\alpha}/\partial Q$ quantities is observed, and this effect is much important for tetrafluoromethane (+47% and +33%) than for methane (+22% and +16%). On the other hand, a much important decrease of the $A_{x,yz}$ and $\partial A_{x,yz}/\partial Q$ quantities is ascertained for methane (−34% and −25%) than for tetrafluoromethane (−15% and −6%). Moreover, a general increase of the frequency dispersion coefficients is also observed with the exception of $A_{\bar{\alpha}}$ of CF_4 which decreases by going from cc-pvdz to aug-cc-pvdz. For the modes belonging to the T_2 symmetry, the $(\partial A/\partial Q)_{vec}$ quantity is more sensitive to the choice of the basis set than the (x, yz) tensor component, and this effect is much important for CF_4 (+32% and +138%) than for CH_4 (+15% and +31%). However, the corresponding frequency dispersion coefficients vary less. Since expressions for geometrical derivatives of the two-electron integrals between atomic orbitals are not implemented in the GAMESS package for f, g, ... type functions, further systematic investigation of the basis set effects (i.e. considering aug-cc-pvtz, ...) is presently nonfeasible.

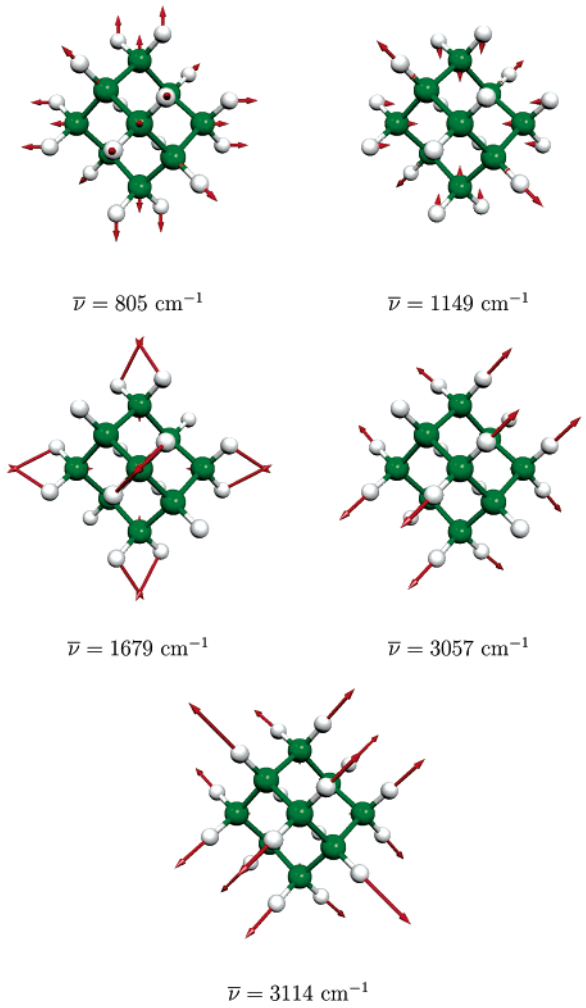


Figure 1. Schematic representations of the totally symmetric vibrational normal modes of adamantane.

Adamantane. The value of the electric dipole−electric quadrupole polarizability of adamantane as calculated at the

TDHF level of approximation is of similar magnitude as for the CH_4 and CF_4 molecules. The electric-dipole−electric-dipole polarizability—which displays a closer relationship to the molecular size or volume—of adamantane is however much larger than for CH_4 and CF_4, as expected from the larger size of the former. Besides the sign, A values of similar magnitude have been evaluated in ref 46, whereas the value determined from collision-induced light scattering experiment is much larger (-102.0 ± 7.8 au). Again, like for CH_4 and CF_4, the variations in the frequency of A are smaller than for α. The situation is different for the $\partial A/\partial Q$ and $\partial\alpha/\partial Q$ quantities. In addition, for the 1149 and 3144 cm^{-1} modes, the $\partial\alpha/\partial Q$ and $\partial A/\partial Q$ quantities decrease with the frequency, respectively. The different totally symmetric vibrational modes of adamantane are sketched in Figure 1. From Table 6, it also turns out that the larger $\partial A/\partial Q$ value is associated with the 3057 cm^{-1} mode, while for $\partial\alpha/\partial Q$, it is the 3114 cm^{-1} mode. With the exception of the 508 cm^{-1} mode which shows a nonmonotonic behavior with respect to ω, all the other T_2 symmetry modes present frequency dispersion coefficients of similar amplitude to those for $\partial A/\partial Q_{A_1}$.

Water, Ammonia, and Benzene. Numerical illustrations are also provided for the A and $(\partial A/\partial Q)$ quantities of water, ammonia, and benzene. The A_z amplitude of H_2O is larger than for NH_3, but its frequency dispersion is smaller. The $\partial A/\partial Q$ values and their evolution with frequency are similar for both H_2O and NH_3. For benzene, the A frequency dispersion coefficients are particularly large. To our knowledge, no experimental data are available for these three systems.

4. Conclusions

Analytical procedures based on the time-dependent Hartree−Fock (TDHF) scheme are elaborated to evaluate the frequency-dependent electric dipole−electric quadrupole polarizability (A) and its derivatives with respect to atomic Cartesian coordinates and vibrational normal coordinates ($\partial A/\partial Q$). On one hand, the mixed second-order TDHF equations are solved iteratively to obtain the second-order derivatives of the linear combination of atomic orbitals coefficients, once with respect to atomic Cartesian coordinates and once with respect to external dynamic electric fields or electric field gradients. On the other hand, taking advantage of the $2n + 1$ rule, the first-order derivatives of A with respect to atomic Cartesian coordinates are expressed in terms of lower-order derivatives. These procedures are implemented in the GAMESS quantum chemistry package and are illustrated in the case of several small molecules as well as adamantane.

The frequency dispersion of $\partial A/\partial Q$ (and A) is shown to be of the same order of magnitude as for $\partial\alpha/\partial Q$ (and α). Thus, in addition to the importance of taking into account explicitly the pulsation of the incident light in Raman simulations,[32,49,50] the evaluation of $\partial A/\partial Q$, which enters in the determination of the VROA intensities, should also consider explicitly the pulsation. A comparison between the cc-pvdz and aug-cc-pvdz basis sets further demonstrates that diffuse functions may have opposite effects on the dipole−dipole (α, $\partial\alpha/\partial Q$) and dipole−quadrupole (A, $\partial A/\partial Q$) responses. To our knowledge, this is the first time that the TDHF scheme is applied to the evaluation of a response property with respect to three different perturbations and thus to the analytical evaluation of $\partial A/\partial Q$.

Acknowledgment. The authors thank Profs. Werner Hug and Kenneth Ruud for fruitful discussions and Dr. Maxime Guillaume for pertinent comments. O.Q., V.L., and B.C. thank the Belgian National Fund for Scientific Research for their Postdoctoral Researcher, Research Fellow, and Senior Research Associate positions, respectively. The calculations have been performed on the Interuniversity Scientific Computing Facility (ISCF), installed at the Facultés Universitaires Notre-Dame de la Paix (Namur, Belgium), for which the authors gratefully acknowledge the financial support of the FNRS-FRFC and the "Loterie Nationale" for the convention no. 2.4578.02, and of the FUNDP.

Appendix: Definitions of Quantities

In the following expressions, the Einstein convention is assumed throughout, and, unless specified, the sums run over (occupied and unoccupied) molecular orbitals.

$$X_{ij}^{\zeta\eta}(\omega) = 0 \tag{29}$$

$$Q_{ij}^{\zeta\eta}(\omega) = 0 \tag{30}$$

$$X_{ij}^{\zeta\eta,a}(\omega) = U_{ik}^{a} U_{kj}^{\zeta\eta}(\omega) - U_{ik}^{\zeta\eta}(\omega) U_{kj}^{a\dagger} \tag{31}$$

$$Q_{ij}^{\zeta\eta,a}(\omega) = G_{ik}^{\zeta\eta}(\omega) U_{kj}^{a} + U_{ik}^{a\dagger} \epsilon_{kj}^{\zeta\eta}(\omega) + G_{ik}^{a} U_{kj}^{\zeta\eta}(\omega) - U_{ik}^{\zeta\eta}(\omega) \epsilon_{kj}^{a} + I_{ik}^{a} U_{kj}^{\zeta\eta}(\omega)[\omega - \epsilon_{j}^{0}] \tag{32}$$

where $I_{ij}^{a} = \sum_{rs}^{AO} C_{ir}^{0\dagger} S_{rs}^{a} C_{sj}^{0}$ is the derivative of the overlap matrix in the MO basis representation. The U^a matrix elements are evaluated using the CPHF scheme of refs 36 and 37.

$$\mathcal{S}_{i}^{\xi,\zeta\eta,a}(-\omega,\omega) = U_{ik}^{\zeta\eta\dagger}(-\omega) I_{kl}^{a} U_{li}^{\xi}(-\omega) - U_{ik}^{\xi\dagger}(\omega) I_{kl}^{a} U_{li}^{\zeta\eta}(\omega) \tag{33}$$

$$\mathcal{X}_{ij}^{\xi,\zeta\eta,a}(-\omega,\omega) = -U_{ik}^{\zeta\eta\dagger}(-\omega) I_{kl}^{a} U_{li}^{\xi}(-\omega) - U_{ik}^{\xi\dagger}(\omega) I_{kl}^{a} U_{lj}^{\zeta\eta}(\omega) \tag{34}$$

$$X_{ij}^{\xi,\zeta\eta}(-\omega,\omega) = -U_{ik}^{\zeta\eta\dagger}(-\omega) U_{kj}^{\xi}(-\omega) - U_{ik}^{\xi\dagger}(\omega) U_{kj}^{\zeta\eta}(\omega) \tag{35}$$

$$\mathcal{U}_{ij}^{\zeta\eta,a}(\omega) = \sum_{k}^{\mathrm{occ}} U_{ik}^{a} U_{kj}^{\zeta\eta\dagger}(-\omega) + U_{ik}^{\zeta\eta}(\omega) U_{kj}^{a\dagger} \tag{36}$$

$$\mathcal{U}_{ij}^{\xi,\zeta\eta}(-\omega,\omega) = \sum_{k}^{\mathrm{occ}} U_{ik}^{\xi}(-\omega) U_{kj}^{\zeta\eta\dagger}(-\omega) + U_{ik}^{\zeta\eta}(\omega) U_{kj}^{\xi\dagger}(-\omega) \tag{37}$$

$$\mathcal{U}_{ij}^{\xi,a}(-\omega) = \sum_{k}^{\mathrm{occ}} U_{ik}^{a} U_{kj}^{\xi\dagger}(-\omega) + U_{ik}^{\xi}(\omega) U_{kj}^{a\dagger} \tag{38}$$

$$X_{ij}^{\xi,a}(-\omega) = U_{ik}^{a} U_{kj}^{\xi}(\omega) - U_{ik}^{\xi}(\omega) U_{kj}^{a\dagger} \tag{39}$$

$$D^{\xi,a}(\omega) = C^{\xi,a}(\omega) n C^{0\dagger} + C^{0} n C^{(-\xi),a\dagger} + \mathcal{D}^{\xi,a}(-\omega) \tag{40}$$

$$\mathcal{D}^{\xi,a}(\omega) = C^{\xi}(\omega) n C^{a\dagger} + C^{a} n C^{(-\xi)\dagger}(-\omega) \tag{41}$$

The $U^{\zeta}(\omega)$ matrix elements are evaluated using the TDHF approach of refs 29−31.

References

(1) Buckingham, A. D. *Adv. Chem. Phys.* **1967**, *12*, 107.

(2) Bancewicz, T.; Le Duff, Y.; Godet, J.-L. Adv. Chem. Phys. In *Contemporary Optics and Electrodynamics*; Evans, M., Ed.; Wiley: London, 2001; Vol. 110, p 267.

(3) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations − The Theory of Infrared and Raman Vibrational Spectra*; Dover: New York, 1980.

(4) Long, D. A. *The Raman Effect;* John Wiley & Sons: New York, 2002.

(5) Posch, H. A. *Mol. Phys.* **1982**, *46*, 1213.

(6) Bancewicz, T.; Teboul, V.; Le Duff, Y. *Phys. Rev. A* **1992**, *46*, 1349.

(7) Bancewicz, T. *Chem. Phys. Lett.* **1993**, *213*, 363.

(8) Elliasmine, A.; Godet, J.-L.; Le Duff, Y.; Bancewicz, T. *Phys. Rev. A* **1997**, *55*, 4230.

(9) Bancewicz, T.; Elliasmine, A.; Godet, J.-L.; Le Duff, Y. *J. Chem. Phys.* **1998**, *108*, 8084.

(10) Hohm, U. *Chem. Phys. Lett.* **1999**, *311*, 117.

(11) Hohm, U.; Loose, A.; Maroulis, G.; Xenides, D. *Phys. Rev. A* **2000**, *61*, 053202.

(12) Maroulis, G. *Chem. Phys. Lett.* **1996**, *259*, 654.

(13) Maroulis, G. *J. Chem. Phys.* **1996**, *105*, 8467.

(14) Hoshina, H.; Wakabayashi, T.; Momose, T.; Shida, T. *J. Chem. Phys.* **1999**, *110*, 5728.

(15) Momose, T.; Miki, M. Wakabayashi, T.; Shida, T.; Chan, M.-C.; Lee, S. S.; Oka, T. *J. Chem. Phys.* **1997**, *107*, 7707.

(16) Buckingham, A. D.; Raab, R. E. *Proc. Roy. Soc. A* **1975**, *345*, 365.

(17) Buckingham, A. D.; Shatwell, R. A. *Phys. Rev. Lett.* **1980**, *45*, 21.

(18) Buckingham, A. D.; Disch, R. L.; Dunmur, D. A. *J. Am. Chem. Soc.* **1968**, *90*, 3104.

(19) Amos, R. D. *Chem. Phys. Lett.* **1982**, *87*, 23.

(20) Barron, L. D.; Buckingham, A. D. *Mol. Phys.* **1971**, *20*, 1111.

(21) Barron, L. D.; Buckingham, A. D. *Annu. Rev. Phys. Chem.* **1975**, *26*, 381.

(22) Polavarapu, P. L. *J. Chem. Phys.* **1990**, *94*, 8106.

(23) Hug, W. *Chem. Phys.* **2001**, *264*, 53.

(24) Ruud, K.; Helgaker, T.; Bour, P. *J. Phys. Chem. A* **2002**, *106*, 7448.

(25) Pecul, M.; Rizzo, A. *Mol. Phys.* **2003**, *101*, 2073.

(26) Zuber, G.; Hug, W. *J. Phys. Chem. A* **2004**, *108*, 2108.

(27) Zuber, G.; Hug, W. *Helv. Chim. Acta* **2004**, *87*, 2208.

(28) McColl, I. H.; Blanch, E. W.; Gill, A. C.; Rhie, A. G. O.; Ritchie, M. A.; Hecht, L.; Nielsen, K.; Barron, L. D. *J. Am. Chem. Soc.* **2003**, *125*, 10019.

(29) Sekino, H.; Bartlett, R. J. *J. Chem. Phys.* **1986**, *85*, 976.

(30) Karna, S. P.; Dupuis, M.; Perrin, E.; Prasad, P. N. *J. Chem. Phys.* **1990**, *92*, 7418.

(31) Karna, S. P.; Dupuis, M. *J. Comput. Chem.* **1991**, *12*, 487.

(32) Quinet, O.; Champagne, B. *J. Chem. Phys.* **2001**, *115*, 6293.

(33) Quinet, O.; Champagne, B.; Kirtman, B. *J. Comput. Chem.* **2001**, *22*, 1920 erratum: 23, 1495 (2002).

(34) Quinet, O.; Champagne, B. *J. Chem. Phys.* **2002**, *117*, 2481 publisher's note: **2003**, *118*, 5692 .

(35) Quinet, O.; Champagne, B.; Kirtman, B. *J. Chem. Phys.* **2003**, *118*, 505.

(36) Pulay, P. *Mol. Phys.* **1969**, *17*, 197.

(37) *Geometrical Derivatives of Energy Surfaces and Molecular Properties;* volume 166 of *Series C: Mathematical and Physical Sciences;* Jørgensen, P., Simons, J., Eds.; D. Reidel Publishing Company: Dordrecht, Holland, 1986.

(38) Yamaguchi, Y.; Osamura, Y.; Goddard, J. D.; Schaefer, H. F., III *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Struture Theory;* Oxford University Press: Oxford, 1994.

(39) Quinet, O. *Int. J. Quantum Chem.* Special issue on vibrational spectroscopies, in press.

(40) Silverman, L.; van Leuven, J. L. *Phys. Rev.* **1967**, *162*, 1175.

(41) Nee, T. S.; Parr, R. G.; Bartlett, R. J. *J. Chem. Phys.* **1976**, *64*, 2216.

(42) 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.

(43) Davis, P. J.; Rabinowitz, P. In *Numerical Integration*; Blaisdell Publishing Company: London, 1967.

(44) Zyss, J. *J. Chem. Phys.* **1993**, *98*, 6583.

(45) Dunning, T. *J. Chem. Phys.* **1989**, *90*, 1007.

(46) Maroulis, G.; Xenides, D.; Hohm, U.; Loose, A. *J. Chem. Phys.* **2001**, *115*, 7957.

(47) Maroulis, G. *Chem. Phys. Lett.* **1994**, *226*, 420.

(48) Buck, U.; Schleusener, J.; Malik, D. J.; Secrest, D. *J. Chem. Phys.* **1981**, *74*, 1707.

(49) Van Caillie, C.; Amos, R. D. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2123.

(50) Quinet, O.; Champagne, B. *Int. J. Quant. Chem.* **2002**, *89*, 341.

CT049888Y

JCTC Journal of Chemical Theory and Computation

Monte Carlo Simulations of Water Adsorption Isotherms in Silicalite and Dealuminated Zeolite Y

Matthieu Fleys[†] and Robert W. Thompson*

Department of Chemical Engineering, Worcester Polytechnic Institute, 100 Institute Road 01609, Worcester, Massachusetts 01609-2280

Received November 8, 2004

Abstract: The affinity that adsorbents have for water can influence their effectiveness in organics removal from drinking water due to competitive adsorption. The extent of the affinity of microporous zeolites for water is determined, in part, by AlO_4^- tetrahedral sites (T-sites) in the crystal lattice and lattice defects in the form of silanol nests. In this study, water adsorption isotherms in silicalite and in dealuminated zeolite Y (DAY) were simulated using the Compass force-field. The results show that the simulations can predict the shape of water adsorption isotherms and predict adsorption levels comparable to literature results. Moreover, simulations revealed that the results are influenced significantly by the presence of AlO_4^- T-sites. The results confirm the capacity of the Compass force-field to predict water sorption properties in silicalite and in DAY.

Introduction

Numerous investigators have studied water adsorption in confined media, such as activated carbons and molecular sieve zeolites. Among these works, some are related to the competitive adsorption of water with other organic components,[1−4] while others deal with the single-component adsorption of water in zeolites.[4−13] Moreover, water can be used as a probe of zeolite properties such as the evaluation of the dealumination rate,[14−16] acid strength,[17] product quality, and micropore volume.[18] For these reasons, water is a particular component whose specific properties in confined media need to be elucidated.

Numerous simulations of the behavior of water in zeolites pore systems have been carried out over the past decade. In silicalite, the diffusion and the behavior of water was studied using MD simulations by Demontis et al.[19] Similarly, Bussai et al.[20−22] used both MD simulations and quantum calculations with ab initio potentials. Water behavior in zeolite NaA also was studied by Faux et al.[23,24] who were interested in diffusion of water and preferred adsorption sites in the α- and β-cages. A study of water in SiO_2 pores with different diameters was reported by Hartnig et al.[25] in order to elucidate the hydration process and the nature of the hydrogen-bond network.

Recently, we investigated the behavior of water in silicalite and dealuminated zeolite Y (DAY) by Molecular Dynamics (MD) simulations.[26] The self-diffusion coefficients and the number of hydrogen bonds per confined water molecule were computed at temperatures in the range 100−600 K, and at different loadings, using the Compass force-field. This force-field proved to be appropriate to explain experimental findings such as the observation that water is vaporlike in silicalite but liquidlike in DAY at room temperature and atmospheric pressure.[12,27] As a result of the previous successes, the Compass force-field was used in the present study where adsorption isotherms of water were computed at different temperatures and compared with experimental results.

Computational Details

The simulations were done with a Silicon Graphics workstation using the Compass force-field and the Sorption module developed by Accelrys in the Cerius2 environment. Electrostatic interactions modeled by a Coulomb potential and the van der Waals, London, and hydrogen-bonding interac-

* Corresponding author phone: (508)831-5525; e-mail: rwt@wpi.edu.

[†] Current address: Département de Chimie-Physique des Réactions, UMR 7630 CNRS, INPL-ENSIC, 1, rue Grandville - BP 451-54001 NANCY Cedex - France.

10.1021/ct049896e CCC: $30.25
Published on Web 03/17/2005

tions modeled by a 6−9 Lennard-Jones potential were taken into account.

The starting crystal structures were perfect, i.e., no defects such as bond breakages or silanol groups were considered. Initially, both zeolite frameworks were all-silica which is a hypothetical limit especially for DAY, as that assumes that the dealumination procedure was 100% effective. The consequences of this hypothesis will be discussed later. Simulations were carried out at different temperatures in two different pressure ranges; a low range (0−2.7 kPa) and a high range (0−500 kPa) for both crystal structures. After considering the all-silica end-member crystal types, additional simulations were done with Al-ZSM-5 where Si/Al = 47 and Na/Al = 1 so that more realistic zeolite structures could be studied.

After creating the zeolite structures and the adsorbent molecules, the next step consisted of calculating a Connolly surface in order to estimate the accessible micropore volume. The Connolly surface is overlaid on the zeolite structure, so that it is possible to visualize the accessible volume in the different channels. Because of the small size of water molecules, the accessible volumes were in the zigzag and straight pores in silicalite and in the sodalite cages and supercages in DAY. No blocking or dummy atoms were used in the simulations. The computed micropore volumes were in agreement with experimental and theoretical values and were found to be 0.18 cm^3/g for silicalite and 0.36 cm^3/g for DAY. It is important to note that these volumes do not correspond to the effective occupied volumes by water in the crystals but are an indication of the possible accessible volumes estimated on the basis of energy and steric hindrance considerations.

The value for the bad contact rejection factor used was 0.5, which was the default value. This means that if any water atoms and framework atoms were closer to each other than half of their van der Waals radii, the configuration was rejected and no energetic calculations were done. As the calculations were time-consuming, the purpose of this criterion was to save time.

The simulations were done in the grand-canonical ensemble where the number of adsorbed molecules was not fixed, but their chemical potentials were. Ten million steps were required before both energy and the loaded amount converged to their final values. Each simulation point obtained at a given temperature and pressure followed these steps.

The same parameters as the ones used in our last simulation work[26] were used. The interaction cutoff distance, which is the distance beyond which interactions between atoms are not taken into account, was about half the size of the smallest simulation cell, that is 6 Å for silicalite and 12 Å for DAY. Electrostatic interactions were considered by using the Ewald method which accelerated the long-range Coulomb calculation.

Similations with Al-Free Silicalite and Zeolite Y

Low Pressures. The adsorption isotherms for water in silicalite were simulated at five temperatures from 300 to 370 K, shown in Figure 1a. The isotherms look very similar in shape to previous reported experimental works; however, the predicted amount adsorbed was substantially less than observed experimentally, discussed further below.[12,16] It is interesting to note that at 300 K, the isotherm shape is different from the others, as an increase of the adsorbed amount of water appears at about 2300 Pa (17 Torr). This phenomenon was observed by Olson et al.,[16] Giaya et al.,[12] and Sano et al.,[14] and its origin is not clearly understood. Olson et al.[16] suggested that a framework defect, such as silanol groups, could explain the phenomenon. But, since the phenomenon was observed in our simulations for the crystallographically perfect silicalite structure, the explanation may be different.

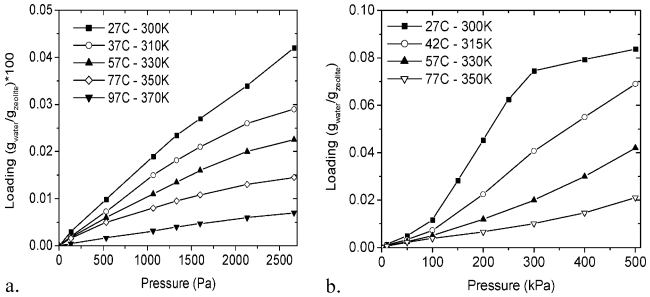


Figure 1. a. Water adsorption isotherms in defect-free silicalite from 300 to 370 K at low pressures and b. from 300 to 350 K at high pressures.

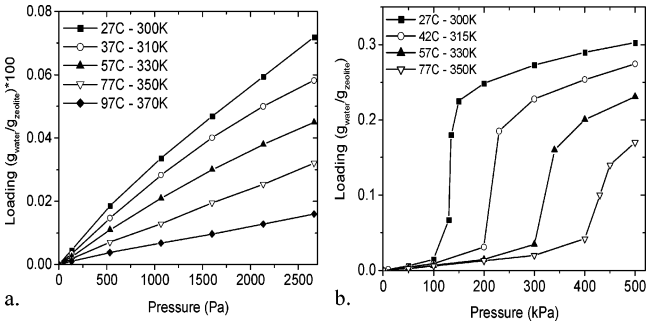


Figure 2. a. Water adsorption isotherms in defect-free DAY from 300 to 370 K at low pressures and b. from 300 to 350 K at high pressures.

The major difference between hypothetical and real silicalite is the adsorption capacity. Indeed, the adsorbed amount of water is 50−100 times lower in the simulated isotherms compared to the experimental values.[12,14,16] Hence, this might suggest that the Compass force-field fails to estimate the adsorption capacity of water in silicate, despite its ability to predict isotherms shapes. However, another hypothesis needs to be considered before concluding this, i.e., that real silicalite is not perfect; it is a high-silica zeolite but not a pure-silica zeolite. Thus, it may contain aluminum T-atoms and associated cations which could modify the adsorption capacity of water.[14,16] This hypothesis will be developed and discussed in the last section.

Similarly, low-pressure water adsorption isotherms in DAY were simulated from 300 to 370 K and are shown in Figure 2a. The adsorbed amount increases linearly with pressure, and the adsorbed amount in this pressure range is twice as much as that predicted for silicalite. This result can

be related to the pore volume which is also twice larger in DAY than in silicalite. As the dealumination procedure does not yield all-silica zeolite Y, real zeolite DAY also contains defects such as aluminum T-atoms. Hence, the simulated adsorbed amounts of water are expected to be underpredicted compared to experimentally measured values.

High Pressures. The simulations performed in the lower pressure range clearly show that silicalite and DAY are very hydrophobic. Only a small fraction of the accessible micropore volume is occupied by water molecules.[6] Thus, simulations at high pressures, from 0 to 500 kPa (5 bar), were carried out in order to characterize the increase of the adsorbed amount of water. The results are given in Figures 1b and 2b for silicalite and DAY, respectively.

In silicalite, the amount of adsorbed water increases slowly with pressure increases. The isotherms can be divided into three parts; first, the loading increases very slowly until a transition pressure is reached. Then, the adsorbed amount increases faster with pressure before reaching a plateau. The lower the temperature, the lower the transition pressure, and the faster the plateau is reached. Only the simulation at 300 K reached the plateau before 500 kPa. From these results, it can be said that the adsorbed amount of water increases so slowly with pressure that the maximum adsorbed amount will be reached at very high pressures. This result is in agreement with previous experimental work[28] where several "strongly hydrophobic zeolite-water" systems were submitted to increasing hydrostatic pressure from 0 to 120 MPa. It was shown that water molecules penetrated the pores when the capillary pressure was reached, that is, at about 90 MPa for the most hydrophobic zeolite, silicalite.

In DAY, the adsorption isotherms are different from those obtained in silicalite. First, the loading increases slightly with pressure but then, after reaching a critical pressure, the amount adsorbed increases sharply before reaching a plateau. Moreover, for simulations performed at higher temperatures, the step is smaller and less pronounced. The origin of this step can be explained by considering the micropore structure of DAY and the micropore filling mechanism. Zeolite DAY is composed of supercages (diameter 1.2 nm) interconnected by β-cages (diameter 0.8 nm). It was found that 4 water molecules can be accommodated per β-cage compared to 26−28 water molecules in the supercage.[10] Water adsorption proceeds by filling the β-cages first and subsequently by filling the supercages.[18] Hence, the step in adsorbed amount could be due to filling of the supercages. Moreover, it was reported that about 250 water molecules per unit cell can be accommodated at complete pore filling at 300 K.[10] In our simulations, at 500 kPa and 300 K, the loading is equivalent to 194 water molecules per unit cell. Hence, at 300 K and 500 kPa, a large proportion of the accessible micropore volume is occupied by water molecules. This means that no step is expected to occur after 500 kPa and that the step which is observed is significant and supports the pore filling mechanism proposed by Boddenberg et al.[18]

The filling process of the DAY cages is not straightforward. While the supercages are more accessible than the β-cages, the potential energy in the center of the supercages is higher than in the vicinity of the walls.[29] Hence, at low loadings, water molecules locate close to the walls indicating that the supercages are not the preferential adsorption sites. This is shown by Faux's results[23,24] who also noted that water molecules located at preferred sites in the α-cages of zeolite NaA, while water located within the center of the α-cages only at higher loadings. It is important to note that the discussion presented by Faux about the self-diffusion coefficients took account only the mobile water molecules contained in the NaA α-cage volumes. Moreover, Faux considered high loadings (from 56 to 224 water molecules in the simulation cell), and at these loadings β-cages were already occupied by water molecules and represented about 15% of the total amount of adsorbed water. In this work, we considered rather low loadings, whereas Faux's paper dealt with diffusion at higher loadings.

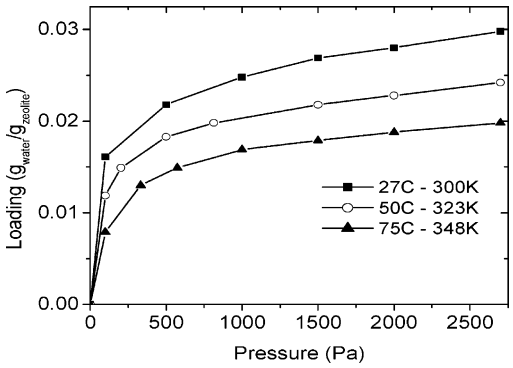


Figure 3. Water adsorption isotherms in Al-ZSM-5, with Si/Al = 47 and Na/Al = 1 from 300 to 348 K at low pressures.

Simulations with Al-ZSM-5 Where Si/Al = 47 and Na/Al = 1

After considering all-silica zeolites, simulations using Al-ZSM-5 were conducted in order to assess the influence of aluminum T-atoms with associated Na^+ cations in the crystal structure on the loading.[30] Simulations were carried out at different AlO_4^- loadings, between 0.5 and 2.5 Al atoms per unit cell. All the adsorption isotherms exhibited a type I shape in the IUPAC classification. Only the simulations with 2 aluminum atoms per unit cell (Si/Al = 47) are given in Figure 3 at three temperatures, 300, 323, and 348 K, in the low-pressure range from 0 to 2700 Pa, since the results were all similar in nature.

In these simulations, all the calculation parameters were conserved. It is noteworthy that the amounts of adsorbed water are in agreement with experimental values.[12,14,16] Nevertheless, the isotherm shapes are somewhat different from those shown in Figure 1a. In Figure 3, the loading increases very quickly at very low pressures before increasing slowly. This result is inherent to the presence of aluminum T-atoms with the associated Na^+ cations. The strong low-pressure sorption is attributed to the hydration of the Na^+ cation.[31] That is why, in Figures 1a and 2a corresponding to all-silica zeolites, the adsorption increases very slightly with pressure in the very low-pressure range, i.e., since no AlO_4^- T-sites and cations were present.

It is important to note that the purpose of the simulation reported in Figure 3 is to illustrate that the presence of defects

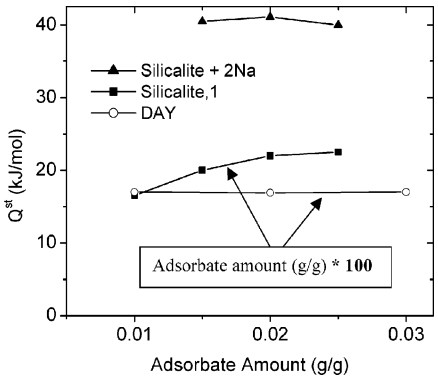


Figure 4. Isosteric heat of adsorption of water in defect-free silicalite and DAY (lower curves) and in Al-ZSM-5, with Si/Al = 47 and Na/Al = 1 (upper curve).

influences the predicted adsorption isotherms. But, it does not provide the exact nature of the defect. Different defects have different influences on the adsorption isotherms. For example, the particular cation is important. It was reported by Olson et al.[16] that the smaller proton, H^+, interacts more strongly than the large Cs^+ cation. Moïse et al.[8] reported that zeolite Y adsorbs more water with Mg^{2+} than with Na^+ as the compensating cation. The presence of internal or external silanol groups also has an influence on water adsorption capacities.[14] Silanol groups are adsorption sites for water molecules whose strength is less than a cation such as H^+.[14] All these different defects modify the adsorption isotherms in terms of shape and loading.

Isosteric Heat of Adsorption

The isosteric heats of adsorption, Q^{st}, were calculated from the isotherm data in Figures 1a, 2a, and 3 using

$$Q^{st} = -R\left[\frac{\partial(\ln P)}{\partial(1/T)}\right]_{loading} \quad (1)$$

The values are reported in Figure 4 for Al-free silicalite, Al-free DAY, and Al-ZSM-5 with Si/Al = 47. For DAY, the isosteric heat is constant with loading and equal to 17 kJ/mol. This value is low compared to the experimentally measured heat of vaporization for water at 25 °C, 43.99 kJ/mol,[32] and the value computed in this study for bulk liquid water, 46.1 kJ/mol. Experiments conducted with NaY, BaY, or CsY zeolites showed that the isosteric heat of adsorption was around 80 kJ/mol, and that Q^{st} decreased with loading.[8,18] Keeping in mind that our simulations were done with hypothetical all-silica zeolites, the comparison is not surprising. Indeed, cations have a significant influence on the isosteric heat. At low loadings, cations act as pumps, water molecules adsorb at cation sites, and the isosteric heat is high. Then, at intermediate loadings, a monolayer, and then multilayers, of water molecules cover the walls of zeolite DAY, leading to a decrease of the isosteric heat as the stronger sites were already occupied. In our simulation where no cations were present, there was no pumping effect to enhance adsorption, and the adsorbed amount was extremely low, as was the isosteric heat.

For silicalite-1, the isosteric heat increased very gently with loading, but the value was still low, equal to about 20 kJ/mol, which is close to the value obtained for DAY. The reason the isosteric heat of adsorption was slightly higher in silicalite-1 than in all-silica zeolite DAY was probably due to the higher adsorbate−wall interaction energies in silicalite-1 compared to DAY. In silicalite, the pore diameters are smaller, and so the confinement effect is more important compared to zeolite DAY. The fact that the isosteric heat obtained for silicalite-1 and all-silica DAY are close reveals that the adsorption process is similar in both all-silica crystals where hydrophobicity is very high and loadings are very low.

The explanation for the low heat of adsorption is related to the low adsorbed amount of water. The behavior of water in silicalite and DAY pore systems, including radial distribution functions, was detailed in our previous paper.[26] The main results are the following: at ambient temperature, the number of hydrogen bonds per water molecule is ca. 1.6 in silicalite and about 3.3 in DAY. In the present paper, the simulated adsorbed amounts of water in Al-free silicalite and DAY are about 100 times lower than the measured values. Hence, in both cases the number of hydrogen bonds is extremely low and clusters do not form in the zeolite pores. Thus adsorbate−adsorbate interactions are not significant, which leads to a low value of the isosteric heat for both zeolites. Comparing the isosteric heats between silicalite-1 and Al-free DAY shows that the isosteric heat in silicalite-1 is slightly higher than in DAY. This is due to the fact that the adsorbate−wall interactions are higher in silicalite compared to DAY, because water molecules are more confined in the smaller pore silicalite compared to DAY.

For Al-ZSM-5 with Si/Al = 47 and Na/Al = 1, the isosteric heat is much higher than for silicalite-1 and equal to about 41 kJ/mol. This value is in agreement with experimental data,[12,16] which confirms the ability of the simulations to predict water sorption properties.

In silicalite and DAY the adsorbed amount is quite low. At these low loadings, water does not cluster, and water molecules are dispersed in the zeolite pore system. Only the water−wall interactions are important, since the water−water interactions are negligible at these low loadings. These interactions are slightly stronger in silicalite due to the stronger confinement, i.e., smaller pore size, compared to DAY. That is why the isosteric heat increases slightly with loading in Al-free silicalite. Water−water interactions become more important at higher loadings, as seen in the case if Al-ZSM-5 with Si/Al = 47.

Conclusions

Adsorption isotherms were simulated in all-silica ZSM-5 and DAY. The results show that the simulations predict the isotherm shapes at low and high pressures and that the results are in agreement with experimental measurements. However, it was observed that at low pressures the amount of adsorbed water was underpredicted in the all-silica zeolites. The reason for this is not related to algorithm issues or parametrization flaws but was due to the intrinsic properties of real zeolites which can contain impurities and defects. These observations can be used to compare the adsorption isotherms for ideal zeolites containing no defects, zeolites with T-atoms or defects included, and experimental isotherms to give at least

qualitative insights on the possible structures of the real zeolites. Moreover, if the influence of the different defects on the adsorption isotherms can be identified, it should be possible to determine the defect types qualitatively. In this case, the experimental adsorption isotherm would be one characteristic feature of the zeolite, as noted previously.[14]

Acknowledgment. The authors acknowledge the support of the National Science Foundation through award NSF NIRT DMI-0210258. The authors also acknowledge the assistance of Professor John C. MacDonald, Chemistry & Biochemistry Department at WPI for use of the MD software.

References

(1) Farrell, J.; Manspeaker, C; Luo, J. Understanding competitive adsorption of water and trichloroethylene in a high-silica Y zeolite. *Microporous Mesoporous Mater.* **2003**, *59*, 205−210.

(2) Anderson, M. A. Removal of MTBE and Other Organic Contaminants from Water by Sorption to High Silica Zeolites. *Environ. Sci. Technol.* **2000**, *34*, 725−727.

(3) Halasz, I.; Kim, S.; Marcus, B. Uncommon Adsorption Isotherm of Methanol on a Hydrophobic Y-zeolite. *J. Phys. Chem. B* **2001**, *105*, 10788−10796.

(4) Khalid, M.; Joly, G.; Renaud, A.; Magnoux, P. Removal of Phenol from Water by Adsorption Using Zeolites. *Ind. Eng. Chem. Res.* **2004**, *43*, 5275−5280.

(5) Serbezov, A. Adsorption Equilibrium of Water Vapor on F-200 Activated Alumina. *J. Chem. Eng. Data* **2003**, *48*, 421−425.

(6) Turov, V. V.; Brei, V. V.; Khomenko, K. N.; Leboda, R. H NMR studies of the adsorption of water on silicalite. *Microporous Mesoporous Mater.* **1998**, *23*, 189−196.

(7) Inagaki, S.; Fukushima, Y. Adsorption of water vapor and hydrophobicity of ordered mesoporous silica, FSM-16. *Microporous Mesoporous Mater.* **1998**, *21*, 667−672.

(8) Moïse, J. C.; Bellat, J. P.; Méthivier, A. Adsorption of water vapor on X and Y zeolites exchanged with barium. *Microporous Mesoporous Mater.* **2001**, *43*, 91−101.

(9) Oh, S. O.; Shim, W. G.; Lee, J. W.; Kim, J. H.; Moon, H.; Seo, G. Adsorption Equilibrium of Water Vapor on Mesoporous Materials. *J. Chem. Eng. Data* **2003**, *48*, 1458−1462.

(10) Dubinin, M. M.; Isirikyan, A. A.; Regent, N. I.; Baier, K. K.; Belenkaya, I. M. Adsorption heats of water and benzene vapors in high silica zeolite of faujasite type. *Izs. Akad. Nauk SSSR Ser. Khim.* **1987**, *3*, 484−492.

(11) Knez, J.; Novak, Z. Adsorption of Water Vapor on Silica, Alumina, and Their Mixed Oxide Aerogels. *J. Chem. Eng. Data* **2001**, *46*, 858−860.

(12) Giaya, A.; Thompson, R. W. Single-component gas-phase adsorption and desorption studies using a tapered element oscillating microbalance. *Microporous Mesoporous Mater.* **2002**, *55*, 265−274.

(13) Frunza, L.; Kosslick, H.; Frunza, S.; Schönhals, A. Unusual Relaxation Behavior of Water Inside the Sodalite Cages of Faujasite-Type Molecular Sieves. *J. Phys. Chem. B* **2002**, *106*, 9191−9194.

(14) Sano, T.; Yamashita, N.; Iwami, Y.; Takeda, K.; Kawakami, Y. Estimation of dealumination rate of ZSM-5 zeolite by adsorption of water vapor. *Zeolites* **1996**, *16*, 258−264.

(15) Simonot-Grange, M.-H.; Elm'Chaouri, A.; Weber, G.; Dufresne, P.; Raatz, F.; Joly, J. F. Characterization of the dealumination effect into H faujasites by adsorption: Part1. The water molecule as a structural aluminium ion selective probe. *Zeolites* **1992**, *12*, 155−159.

(16) Olson, D. H.; Haag, W. O.; Borghard, W. S. Use of water as a probe of zeolitic properties: interaction of water with HZSM-5. *Microporous Mesoporous Mater.* **2000**, *35*−*36*, 435−446.

(17) Semmer-Herlédan, V.; Heeribout, L.; Batamack, P.; Dorémieux-Morin, C.; Fraissard, J.; Gola, A.; Benazzi, E. Comparison of the acid strength of Dealuminated H-faujasites determined by H NMR after water adsorption. *Microporous Mesoporous Mater.* **2000**, *34*, 157−169.

(18) Boddenberg, B.; Rakhmatkariev, S.; Hufnagel, S.; Salimov, Z. A calorimetric and statistical mechanics study of water adsorption in zeolite NaY. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4172−4180.

(19) Demontis, P.; Stara, G.; Suffritti, G. B. The Behavior of Water in the Hydrophobic Zeolite Silicalite at Different Temperatures. A Molecular Dynamics Study. *J. Phys. Chem. B* **2003**, *107*, 4426−4436.

(20) Bussai, C.; Hannongbua, S.; Haberlandt, R. Understanding the Movement, Encapsulation, and Energy Barrier of Water Molecule Diffusion into and in Silicalites Using Ab Initio Calculations. *J. Phys. Chem. B* **2001**, *105*, 3409−3414.

(21) Bussai, C.; Vasenkov, S.; Liu, H.; Böhlmann, W; Fritzsche, S; Hannongbua, S; Haberlandt, R; Kärger, J. On the diffusion of water in silicalite-1: MD simulations using ab initio fitted potential and PFG NMR measurements. *Appl. Catal. A* **2002**, *General 232*, 59−66.

(22) Bussai, C.; Hannongbua, S; Fritzsche, S; Haberlandt, R. Ab initio potential energy surface and molecular dynamics simulations for the determination of the diffusion coefficient of water in silicalite-1. *Chem. Phys. Lett.* **2002**, *354*, 310−315.

(23) Faux, D. A.; Smith, W.; Forester, T. R. Molecular Dynamics Studies of Hydrated and Dehydrated Na^+-Zeolite-4A. *J. Phys. Chem. B* **1997**, *101*, 1762−1768.

(24) Faux, D. A. Molecular Dynamics Studies of Hydrated Zeolite 4A. *J. Phys. Chem. B* **1999**, *103*, 7803−7808.

(25) Hartnig, C.; Witschel, W.; Spohr, E.; Gallo, P.; Ricci, M. A.; Rovere, M. Modifications of the hydrogen bond network of liquid water in a cylindrical SiO2 pore. *J. Mol. Liq.* **2000**, *85*, 127−137.

(26) Fleys, M.; Thompson, R. W.; MacDonald, J. C. Comparison of the Behavior of water in Silicalite and Dealuminated Zeolite Y at Different Temperatures by Molecular Dynamic Simulations. *J. Phys. Chem. B* **2004**, *108*, 12197−12203.

(27) Giaya, A.; Thompson, R. W.; Denkewicz, R., Jr. Liquid and vapor phase adsorption of chlorinated volatile organic compounds on hydrophobic molecular sieves. *Microporous Mesoporous Mater.* **2000**, *40* (1−3), 205−218.

(28) Eroshenko, V.; Regis, R.-C.; Soulard, M.; Patarin, J. Energetics: A new Field of Applications for Hydrophobic Zeolites. *J. Am. Chem. Soc.* **2001**, *123*, 8129−8130.

(29) Fritzsche, S.; Haberlandt, R.; Kärger, J.; Pfeifer, H.; Heinzinger, K. An MD Simulation on the Applicability of the Diffusion Equation for Molecules Adsorbed in a Zeolite. *Chem. Phys. Lett.* **1992**, *198*, 283−287.

(30) The locations of the Na^+ cations were determined by Al substitution for Si tetrahedra. The Cerius 2 software allowed us to specify the Si/Al ratio, and then the software substituted the appropriate number of Si atoms by Al tetrahedra in random crystallographic locations but in keeping with Löwenstein's rule. Moreover, only the Si atoms in the walls of the channels in the zeolite framework were substituted by Al atoms. Simulations with 2 aluminum atoms per unit cell (Si/Al = 47) were carried out by choosing different locations in the channels (both in straight channels, both in zigzag channels, and one in each channel type, one at the intersection and the other in the straight channel); the computed adsorbed amount was the same as long as the length of the simulation was sufficient for equilibration (at least 5 millions steps). The Na^+ cations were located close to AlO_4^- sites. The charge neutrality was provided by sodium ions Na^+ which compensated the negative charge inherent in AlO_4^- sites.

(31) Chen N. Y. Hydrophobic properties of zeolites. *J. Phys. Chem.* **1976**, *80*, 60−64.

(32) In *Recommended Reference Materials for the Realization of Physicochemical Properties*; Marsh, K. N., Ed.; Blackwell: Oxford, 1987.

CT049896E

JCTC Journal of Chemical Theory and Computation

Effects of the Aluminum Content of a Zeolite Framework: A DFT/MM Hybrid Approach Based on Cluster Models Embedded in an Elastic Polarizable Environment

Elena A. Ivanova Shor,[†] Alexei M. Shor,[†] Vladimir A. Nasluzov,*[,†] Georgi N. Vayssilov,*[,‡] and Notker Rösch*[,§]

Institute of Chemistry and Chemical Technology, Russian Academy of Sciences, 660049 Krasnoyarsk, Russian Federation, Faculty of Chemistry, University of Sofia, 1126 Sofia, Bulgaria, and Department Chemie, Technische Universität München, 85747 Garching, Germany

Received October 26, 2004

Abstract: We report the first computational study with a sophisticated quantum mechanics/molecular mechanics (QM/MM) technique that addresses the effect of the aluminum content on the properties of acidic zeolites. To account for both electrostatic and mechanical interaction between the QM cluster and its MM environment, we used cluster models embedded in the covalent variant of the elastic polarizable environment (covEPE) [Nasluzov, V. A.; Ivanova, E. A.; Shor, A. M.; Vayssilov, G. N.; Birkenheuer, U.; Rösch, N. *J. Phys. Chem. B* **2003**, *107*, 2228]. For the practical application of the covEPE method, it was necessary to develop a new force field for Al containing zeolites. Two types of zeolite materials, FAU and MFI, were employed as examples. We modeled the variation of the Al content both in the MM environment and in the QM cluster, and we studied pertinent properties of bridging OH groups of the zeolite frameworks, OH vibrational frequencies, and deprotonation energies. The computational results suggest that the local structure and the location of the OH groups exert a stronger effect than the variation of the Al content of the framework.

1. Introduction

The contribution of computational methods for investigating structure and catalytic properties of zeolite materials is nowadays commonly acknowledged.[1] Still, large unit cells of realistic zeolite systems present a major obstacle that restricts the wide-spread application of simulation methods. Ab initio calculations that account for the periodicity of such systems are, as a rule, computationally expensive and, therefore, available mainly for highly symmetric frameworks with unit cell of modest size.[2,3] The symmetry of practically important zeolite systems is often lowered by the presence of defects (e.g. interstitials or adsorbed molecules) which further complicate calculations. A well-recognized approach for studying such local phenomena employs cluster models where active sites are represented by a chemically relevant finite part of the lattice.[4−7] A significant drawback of this approach is that it totally neglects the influence of the surrounding crystal lattice on the reaction center. Therefore, such models are not suitable for a variety of pertinent problems of zeolite chemistry where one has to discriminate active sites in structurally or chemically different environments, e.g. Brønsted acidic sites in aluminosilicates of different framework structures and varying Al content.

Hybrid quantum mechanics/molecular mechanics (QM/MM) methods[8,9] are known to afford an accurate treatment of such problems by taking into account thousands of atoms

* Corresponding authors e-mail: nv@krsk.info (V.A.N.), gnv@chem.uni-sofia.bg (G.N.V.), roesch@ch.tum.de (N.R.).
[†] Russian Academy of Sciences.
[‡] University of Sofia.
[§] Technische Universität München.

10.1021/ct049910n CCC: $30.25
Published on Web 04/19/2005

in the system at the cost of dozens. This efficiency is achieved by describing only the immediate neighborhood of the active site (cluster) with a high precision QM method and by representing its environment at an more economic level, e.g. with a MM method. Such QM/MM schemes inherit the capability to model easily point defects (including charged ones) from strategies based on traditional (isolated) cluster models but overcome their deficiencies connected with neglecting steric constraints and the electrostatic field of the environment. However, to properly construct a QM/MM method, one has to account both for long-range electrostatic effects of the environment and the mechanical coupling of the QM cluster with its immediate surrounding. In addition, it is highly desirable to preserve the variationality of the method as well as to reduce and control the influence of the QM/MM border region.

A specific complication of high-level, accurate QM/MM schemes is connected with the requirement to use a properly parametrized MM force field (FF) for describing both mechanical and electrostatic interactions between the QM region and its MM environment. In view of the polar covalent nature of siliceous and zeolite materials, it is crucial to model correctly the electrostatic properties of the framework. A practical and efficient way to achieve this goal assigns so-called potential derived point charges (PDC)[10] to lattice centers and represents the polarizability of lattice oxygen centers by a "shell model" scheme.[11] However, to the best of our knowledge, so far no force field has been developed for Al-containing zeolites that simultaneously uses such charges and allows for the polarization of O centers. The PDC-based rigid ion model of Blake et al.[12] is applicable to Na-forms of aluminosilicates only, whereas the model of van Santen et al.[13] for H-forms of aluminoslicates relies on a rigid-ion parametrization with atomic charges that are relatively large compared to those derived from the electrostatic potential. The valence force field of Hill and Sauer[14] utilizes rather small charges; at the other extreme, shell models[15,16] employ too large (absolute) formal charges. These force fields aim at reproducing structural features of zeolites rather than their electrostatic field. Indeed, the latter parametrizations[16] are invoked in the QM-Pot cluster embedding which, at the quantum mechanical level, neglects long-range interactions across the QM/MM border.[9] Available PDC-based force fields for siliceous frameworks[12,17] are also limited in number and scope; in particular, when used in embedded cluster schemes, they are suitable for modeling pure silica systems or systems with an extremely low Al content where it is sufficient to locate Al atoms only in the QM part of the system.[8,17]

The present work is the first QM/MM study of the Al content effect on the properties of Brønsted active centers in zeolites at the level of a combined electrostatic and mechanical embedding. We used our covEPE embedding scheme[17] but developed a new FF for Al containing zeolites that provides both a shell model treatment of oxygen centers and assigns PDC to all centers of the zeolite framework, including protons of bridging OH groups. This FF is an extension of our previously reported PDC-based shell model FF for pure siliceous materials.[17] With these tools, we investigated to which degree the structure, OH frequencies, and deprotonation energies of Brønsted acidic sites in zeolites are affected by changes of its Al loading. Our goal was to derive a general correlation between chemical composition and properties of the zeolite.

We studied a faujasite (FAU) zeolite with a wide range of Si/Al ratios (∞, 47, 23, 11). Such highly dealuminated lattices are not quite typical for faujasite, the Si/Al ratio of which is usually below 3.[18,19] To model structures with high local Al concentration, we considered two interacting (or "paired") acidic sites, i.e., a situation where Al centers are located at closest possible distance. We compared the effects of the Al content with that of the local structure around the acidic OH group in a regular framework. For that purpose, we modeled Brønsted centers located at O1(H) and O3(H) crystallographic positions of a faujasite lattice as well as at Al7−O17(H)−Si4 sites of zeolite HZSM-5 (with MFI structure). Moreover, we discussed the accuracy of our results with regard to size limitations of QM cluster models as well as to the proximity of active Brønsted centers and the QM/MM boundary.

2. Method and Models

2.1. CovEPE Embedding Approach. The covEPE embedding scheme[17] is specially set up to deal with systems that feature covalent bonding. Dangling bonds of the model QM cluster are saturated with monovalent atomic pseudopotentials (*pp*) representing crystal lattice oxygen centers. The parameters of these border atoms are adjusted to mimic the behavior of real oxygen atoms as well as to minimize the perturbation of the electrostatic field at the cluster boundary. The method completely and explicitly includes both mechanical and electrostatic interactions between the QM cluster and its MM lattice (*lat*) environment. For a correct description of the cluster surrounding, a force field is required that is based on potential-derived charges and accounts for the polarization of lattice centers.

The entire system under study is divided into an active area, region I, treated at the QM level, and its MM surrounding, region II, described with a force field. region II is further partitioned into areas of explicit optimization (IIa), with Mott-Littleton polarization (IIb), and an external area (IIc). The total energy E_{tot} of the QM/MM system comprises three contributions which describe the interactions in the QM cluster, E_{cl}, and the coupling between the QM cluster and its environment, E_{int}, as well as the interactions within the MM lattice, E_{lat}.[17]

The environment affects the QM cluster via its Madelung field and short-range forces. The last contribution includes the FF interaction between regions I and IIa as well as forces acting on the QM/MM boundary to restore the structure of the frontier region of the cluster model as in a defect-free lattice. The environmental lattice is subjected to a displacement polarization, and the oxygen atoms of the MM framework are also allowed to be polarized in shell-model fashion due to changes of the charge distribution in the QM cluster.

A covEPE treatment consists of the following steps: (i) optimization of the structure of the ideal (regular − *reg*)

periodic lattice with an atomistic force field; (ii) optimization of the QM cluster (*cl*) modeling region I as *reference* (*ref*) embedded in a "frozen" environment; (iii) energy minimization of the *defect* site under study (e.g. an adsorption complex or defect center) created within the QM part; and (iv) the relaxation of region II while accounting for Coulomb and short-range coupling with region I. The structure of the whole system, comprising QM and classical regions, is determined variationally by minimizing the total energy in steps (iii) and (iv). One may expect artificial changes of structure and charge distribution of an embedded system relative to the system treated completely classically—even without any guest species, impurities, or defects. To eliminate these artifacts caused by the mismatch of descriptions inherent to any hybrid method, one applies correction terms to the charge density and the short-range interaction. These corrections are estimated as difference between the total charge densities and the short-range interactions for the regular MM and the reference QM structures of the unperturbed lattice, calculated in two preparatory steps 1 and 2, respectively:

$$E_{int}^{corr} = (\rho_{cl}^{tot,MM}(R_{cl}^{reg}, R_{pp}^{reg}) - \rho_{cl}^{tot,QM}(\rho_{cl}^{ref}, R_{cl}^{ref}, R_{pp}^{ref}))||\rho_{lat}(R_{lat}) + V_{short}(R_{cl}^{reg}, R_{pp}^{reg}/R_{lat}) - V_{short}(R_{cl}^{ref}, R_{pp}^{ref}/R_{lat}) \quad (1)$$

Here, R_{cl}, R_{pp}, and R_{lat} collectively refer to atomic positions in the cluster, the boundary region, and the lattice environment, respectively, of the regular (*reg*) and the reference (*ref*) structures. The double bar || represents the classical Coulomb interaction between two charge densities. By construction, the environment is assured to retain its MM equilibrium structure if a cluster representing a regular site is embedded.[17] Steps (i) and (ii) need to be performed only once, before an active region, modified by the presence of any type of the defect, can be described in steps (iii) and (iv). The covEPE computational procedure allows one to carry out a completely variational treatment of the total energy with respect to all degrees of freedom of regions I and II.

At variance with most of the other QM/MM schemes, the covEPE method does not use H atoms to cap dandling bonds of the QM cluster but specially adjusted monovalent oxygen pseudoatoms, O*. The interactions of these pseudoatoms with both the rest of the QM cluster and the MM lattice were modified by short-range interaction terms and partial charges $\Delta q_{pp} = -0.3\ e$ located at the position of the O* nuclei. Different from our previous work,[17] in the present study the coupling via the QM/MM border does not only include two-body O*−Si(MM) terms but also three-body interactions of the types O*−Si(MM)−O(MM) and O*−Si(MM)−O*. These extra terms ensure that the geometry of SiO_4 units at the QM/MM boundary remain closer to an ideal tetrahedral configuration. The parameters of these three-body terms were set equal to those for the O−Si−O interaction in our PDC-based force field[17] ($k_\theta = 0.89844$; $\theta_0 = 109.47°$). The parametrization of the pseudopotential and the additional interaction terms were specific for O* centers bound to Si atoms both of the QM cluster and the MM environment; Al atoms were not admitted as neighbors of O* centers to restrict the number of parameters of the current covEPE scheme, but restrictions result when one constructs QM cluster models of Al-containing zeolites (see below). Further details of the covEPE method and the parametrization can be found in ref 17.

2.2. Calculation of the QM Region. All embedded cluster calculations were carried out with the covEPE scheme as implemented in parallel density functional program ParaGauss.[20,21] To treat the QM part of the systems, we used the gradient-corrected exchange-correlation functional as suggested by Becke (exchange) and Perdew (correlation);[22] all calculations were performed in spin-restricted fashion. We described Kohn−Sham orbitals with the following Gaussian-type basis sets, contracted in generalized fashion: (6s1p) → [4s1p] for H, (9s5p1d) → [5s4p1d] for O, (12s9p2d) → [6s4p2d] for Al and Si.[23] The Hartree contribution of the electron−electron interaction was calculated with an auxiliary Gaussian-type representation of the electronic charge density.[24] The corresponding exponents were constructed by scaling the exponents of the orbital basis; a standard set of p and d polarization exponents was added on each atomic center.[25] The structure of the QM clusters as well as the location of the environmental EPE centers were optimized without imposing any symmetry restriction.

As a measure of the acid strength of a bridging OH groups, we calculated their deprotonation energies (DE) by subtracting the total energy of the relaxed neutral (protonic) form of the zeolite and the relaxed anionic (deprotonated) form.

The analysis of OH vibrational frequencies was carried out on optimized structures. The force constant for the vibrational mode was calculated numerically, using finite differences of analytical energy gradients; we invoked the approximation of one independent harmonic oscillator, i.e., only the O−H internal coordinate was varied during the frequency calculation.

2.3. Force Field for Protonic Forms of Aluminosilicates. When we introduced the covEPE approach, we described a strategy for deriving a PDC-based shell-model force field, and we successfully derived a FF for pure siliceous frameworks.[17] Here, we report the extension of this FF to protonic forms of aluminosilicates based on the same parametrization procedure. The following general expression for the total energy is used for the aluminosilicate version of the PDC-based shell model FF:

$$V = \sum_{i,j}\frac{q_i q_j}{r_{ij}} + \sum_{i,j}[A_{ij}\exp(-r_{ij}/\rho_{ij}) - C_{ij}r_{ij}^{-6}] + \sum_i k_i^{sh}\Delta r_i^2 + \sum_{j=T,O} k_\theta(\theta_{ijk} - \theta_0)^2 \quad (2)$$

The charges q_i, describing the Coulomb interaction, are estimated for optimized gas-phase clusters, invoking the CHELPG procedure[10] as implemented in the software package Gaussian98.[26] PDCs were determined from BP-DF calculations with a TZV(d,p) basis set, using seven model clusters of 2 to 5 T atoms which were terminated by OH groups (Table 1; see also the Supporting Information). In addition to the PDCs of Si, Al, and H, three different PDC values were assigned to O centers, depending on their position in the framework, O_I and O_{II} for Si−O−Si and Al−

Table 1. Potential-Derived Charges (in *e*) for Aluminosilicate Clusters[a] Optimized without Embedding

	potential-derived charges[b]					
cluster	O_I	O_{II}	O_{III}	Si_I	Al_I	H_I
$Si_3O_{10}H_8$ [c]	−0.48			1.20		
$Si_4O_{12}H_8$ [d]	−0.48			1.11		
$AlSiO_7H_7$ [e]			−0.15		1.14	0.35
$AlSi_2O_{10}H_9$ [f]		−0.56	−0.28	1.19	1.13	0.33
$AlSi_2O_{10}H_9$ [g]	−0.43		−0.29	1.14	1.09	0.33
$AlSi_3O_{12}H_9$ [d]	−0.33	−0.47	−0.20	1.07	0.96	0.27
	−0.48					
$AlSi_4O_{16}H_{13}$ [h]		−0.62	−0.34	1.21	1.22	0.35
		−0.67		1.29		
final assignment	−0.6	−0.75	−0.4	1.2	1.1	0.35

[a] For sketches of the model clusters, see Figure 4 in the Supporting Information. [b] For the designation of atom types, see Figure 1. [c] 3T chainlike cluster (Figure 4a). [d] Four-membered ring (Figure 4b,f). [e] 2T model (Figure 4c). [f] 3T Si−Al−Si chainlike cluster (Figure 4d). [g] 3T Al−Si−Si chainlike cluster (Figure 4e). [h] 5T cluster with central Al atom surrounded by O and Si shells (Figure 4g).

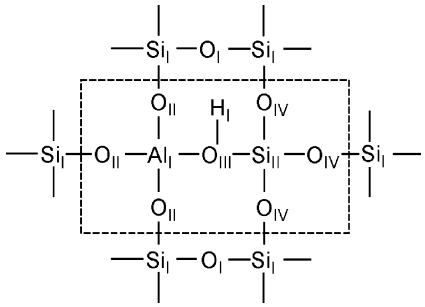


Figure 1. Designation of atomic centers used in the PDC-based shell-model force field for H-zeolites. Centers outside the dashed rectangle are common to the force fields for pure silica and aluminosilicates; centers inside the rectangle are specific to Al-containing lattices.

O−Si bridges, respectively, and O_{III} for a bridging OH group (Table 1 and Figure 1). We took the charges $q(Si_I) = 1.2\ e$ and $q(O_I) = -0.6\ e$ from the FF for pure silica.[17] The other charges, $q(O_{II})$, $q(O_{III})$, $q(Al_I)$, and $q(H_I)$, were restricted by the relation

$$q(Si_I) + 4q(O_I) = q(Al_I) + q(H_I) + 3q(O_{II}) + q(O_{III}) \quad (3)$$

to preserve the neutrality of the zeolite lattice. In other words, the incorporation of Brønsted sites in the framework is assumed to redistribute charge among the centers Al_I, H_I, O_{II}, and O_{III} near the acidic center but leave the charges of the remaining system unchanged (Table 1). Note that the PDC of aluminum, $q(Al_I) = 1.1\ e$, is only 0.1 *e* smaller than that of silicon centers—at variance with the rigid ion force fields of van Santen et al.[13] where the Al charge was fixed to be 1.0 *e* smaller than that of a Si center. Incidentally, our PDC values are similar to the Mulliken charges obtained in a periodic HF calculation of H-chabazite where a small STO-3G basis set had been used:[27] $q(Al) = 1.21\ e$, $q(Si) = 1.42\ e$, $q(O) = -0.71\ e$, $q(H) = 0.21\ e$.

To derive a uniform FF suitable for modeling both siliceous and aluminosiliceous materials, we kept the parameters for stretching (Si_I-O_I, O_I-O_I) and bending interactions ($Si_I-O_I-Si_I$ and $O_I-Si_I-O_I$) as reported earlier for a SiO_2 framework.[17] However, these parameters were found to be unsuitable for describing short-range interactions within a tetrahedron centered on Si in the immediate vicinity of Al−O(H)−Si bridges. To overcome this problem, we defined a further type of oxygen centers, O_{IV}, saturating three of the four bonds of an Si center at an acidic site Al−O(H)−Si; the latter silicon center was reassigned to type Si_{II} (Figure 1). The atomic parameters of Si_{II} and O_{IV} centers (PDCs and core−shell splitting of O_{IV}) were taken from the corresponding centers Si_I and O_I.

Because there is no precise experimental information on the local structure of a zeolite framework in the proximity of Al centers and bridging OH groups, we based the parametrization of the interactions specific to aluminosilica on computational data. For this purpose, we carried out reference QM calculations at the same level as in the subsequent hybrid covEPE calculations, to provide a coherent description of the QM and MM parts of the system. The parametrization procedure consisted of three steps, repeated iteratively until convergence was reached. In the first step, parameters for two-body O−O and O−H interactions [second term in eq 2] and three-body O−T−O interactions [T = Si, Al; fourth term in eq 2] were established, using potential energy curves obtained for the corresponding reference gas-phase cluster from DF calculations (Table 2). Then the pair-potential interactions Si−O and Al−O were determined in the second step. In the third step, the core−shell spring constants, k^{sh}, and the charges of core and shell of polarizable oxygen centers were adjusted to reproduce changes in the dipole moment of the free 3T cluster $(OH)_3Si-O(H)-Al-O-Si(OH)_3$ as induced by point charge of 0.5 *e* located at about 10 Å from the center of the model cluster. Finally, once the values of these iteratively determined parameters had converged, the force constants of three-body interactions, Al−O−Si and H−O−Si(Al), were fitted in a single step. At variance with the Si−O−Si interaction in the silica FF,[17] they were not fitted to experimental data but determined to yield structure and unit cell parameters of H-chabazite. In fact, these results turned out to agree closely with those obtained from the DF-based FF of Sierka and Sauer (Table 3).[16b]

The resulting force field was found to reproduce successfully the structure of isolated acidic sites (also of systems with several such centers) but failed to predict structures where acidic centers are separated by one Si atom only, e.g. where two hydroxyl groups are located in the same 4-ring. Apparently, at a high local Al concentration, oxygen pairs, e.g. $O_{II}-O_{II}$ or $O_{II}-O_{III}$, interact not only via an Al atom but also via a Si center. In such special cases, global short-range parameters determined for all interactions within the common cutoff of 8 Å do not seem to work well. To avoid this problem, we added a local two-body interaction of O−O type to the force field (with a smaller cutoff of 4 Å) which acts only *within* certain tetrahedra; see Table 2. Thus, only pairs of O centers bound to the *same* T atom interact via this additional term.

In Table 4, we provide several examples that demonstrate the performance of our new PDC-based FF for alumino-

Table 2. Parameters[a] of a PDC-Based Shell-Model Force Field for the Protonic Form of Aluminosilicates

	short-range parameters		
FF interaction	A, eV	ρ, Å	C, eV Å^6
	Two-Body		
Al_I-O_{II} [b]	29617.225	0.189495	122.39
Al_I-O_{III} [b]	40768.575	0.197235	181.04
$Si_I-O_I = Si_I-O_{IV}$	51431.799	0.174872	131.11
Si_I-O_{II} [c]	35091.342	0.190375	174.57
$Si_{II}-O_{II}$	35291.445	0.189265	187.52
$Si_{II}-O_{III}$ [d]	40184.694	0.183955	166.99
$Si_{II}-O_{IV}$ [d]	55400.776	0.172102	144.27
$O_I-O_I = O_I-O_{IV}$	95169.354	0.199100	43.636
O_I-O_{II} [c] $= O_{II}-O_{IV}$	17471.593	0.268175	413.39
$O_{II}-O_{IV}$ ($T = Si_{II}$)[e]	30471.574	0.259285	159.59
$O_I-O_{III} = O_{III}-O_{IV}$ [d]	38115.954	0.239040	125.58
$O_{II}-O_{II}$ [b]	93947.963	0.223816	57.488
$O_{II}-O_{II}$ ($T = Si_I$ or Si_{II})[e]	17471.593	0.268175	413.39
$O_{II}-O_{III}$ [b]	9418.570	0.240302	6.4444
$O_{II}-O_{III}$ ($T = Si_{II}$)[e]	37214.853	0.238941	120.98
$O_{IV}-O_{IV}$ [d]	95169.354	0.199100	43.636
$O_{IV}-O_{IV}$ ($T = Si_{II}$)[e]	59194.280	0.238835	147.22
$O_{III}-H_I$ [b]	7518.521	0.117834	2.0915
FF interaction	k_θ, eV rad^{-1}		θ_0, degree
	Three−Body		
$O_{II}-Al_I-O_{II}$ [b]	0.87410		109.47
$O_{II}-Al_I-O_{III}$ [b]	0.84520		109.47
$Om-Si_I-On$	0.89844		109.47
$O_{II}(O_{IV})-Si_{II}-O_{II}(O_{IV})$[d]	0.86564		109.47
$O_{II}-Si_{II}-O_{III}$ [d] $= O_{IV}-Si_{II}-O_{III}$	0.87080		109.47
$Si_I-O_I-Si_I = Si_I-O_{IV}-Si_{II}$	5.62		163.4
$Al_I-O_{II}-Si_I$	6.72		162.4
$Al_I-O_{II}-Si_{II}$	6.57		163.4
$Al_I-O_{III}-Si_{II}$	6.44		150.3
$Al_I-O_{III}-H_I$	9.65		129.8
$Si_{II}-O_{III}-H_I$	9.75		141.0
$Si_{II}-O_{IV}-Si_{II}$	6.70		166.2

[a] For the definition, see eq 2; see also Figure 1 for the designation of the atom types. [b] From 5T model $(OH)_3Si-O(H)-Al(OSi(OH)_3)_2-OSi(OH)_3$ with central AlO_4 tetrahedron. [c] From 6T model $(OH)_3Si-O(H)-Al(OH)_2-O-Si(OSi(OH)_3)_3$. [d] From 5T model $(OH)_3Al-O(H)-Si(OSi(OH)3)_2-OSi(OH)_3$ with central SiO_4 tetrahedron. [e] Local interaction, valid only within the tetrahedron centered on atom T (see text).

silicates (PDC in Table 4). We compare the force field results on structural parameters of faujasite (Si/Al = 47, 23, 11) with those for zeolite lattices optimized with other first-principles force fields and free QM reference models. In agreement with the B3LYP-based FF which utilizes formal charges (FC in Table 4),[16b] our force field demonstrates no significant structural changes for isolated acidic sites when the Al loading of a unit cell increases. Bond lengths calculated with both FFs agree within 0.03 Å. Except for $Si_{II}-O_{III}$ bonds, the B3LYP-derived FF gives shorter bond distances; that feature seems to be inherited from the B3LYP approach used to generate set the QM reference data. Compared to the B3LYP-based FF, results from our FF turned out to be less sensitive to the presence of neighboring Al centers; in comparison to isolated sites, bond distances of paired acidic sites changed by 0.025 Å in the B3LYP-based FF, whereas such changes do not exceed 0.004 Å in our FF. As expected, our parametrization fits very well the set of reference structure parameters of isolated QM clusters (Table 4). Al_I-O_{II}, $Si_{II}-O_{III}$, and $O_{III}-H_I$ bond lengths agree perfectly between periodic and QM cluster structures; $Si_{II}-O_{IV}$ and Si_I-O_{II} bond distances are reproduced within 0.02−0.03 Å. The largest discrepancy, 0.04 Å, was obtained for the Al_I-O_{III} bond. Note that $Al_I-O_{III}-Si_{II}$ angles from FF and QM calculations agree within 2° although this interaction was not parametrized on QM data.

Table 3. Comparison of Unit Cell Parameters (Å, deg), Elastic Constants (10^{10} Nm^{-2}), and Bulk Modulus (GPa) of H-Chabazite Calculated with Our PDC-Based (PDC) Force Field and a B3LYP-Derived Force Field Based on Formal Charges (FC)

	chabazite (Si/Al = 11)		chabazite (Si/Al = 5)	
	PDC	FC[a]	PDC	FC[a]
a	9.415	9.438	9.359	9.456
b	9.333	9.458	9.513	9.518
c	9.335	9.386	9.246	9.504
α	93.987	94.776	92.924	95.451
β	94.691	93.951	92.541	93.905
γ	93.766	95.363	95.416	94.932
C_{11}	11.08	9.56	8.01	9.54
C_{12}	5.18	5.13	4.18	5.76
C_{13}	6.17	5.79	2.56	5.71
C_{22}	9.94	8.88	8.60	9.35
C_{23}	4.95	5.41	2.37	5.54
C_{33}	11.65	9.88	6.46	9.25
C_{44}	1.51	1.26	1.47	1.16
C_{55}	1.72	1.10	1.41	1.15
C_{66}	1.55	1.08	1.34	1.08
ϵ_{11}	1.61	2.78	1.69	2.85
ϵ_{33}	1.62	2.92	1.65	2.88
ϵ_{11}^∞	1.53	1.43	1.56	1.43
ϵ_{33}^∞	1.53	1.44	1.56	1.43
bulk modulus	69.94	60.73	43.09	62.53

[a] Reference 16b.

Very recently, EXAFS studies[28] on the surrounding of Al centers in zeolites with FAU and MFI structure provided experimental interatomic Al−O and Al−Si distances as well as O−Al−O angles which are suitable for comparison with our results. In line with these data, our force field predicts an Al−O bond in a Al−O(H)−Si fragment 0.2 Å longer (Al_I-O_{III} in Table 4) than in an Al−O−Si bridge (Al_I-O_{II} in Table 4). Our force field is parametrized to reproduce structural data obtained at the BP level, an approach which has a tendency to overestimate bond lengths. This known trend is reflected by our calculated Al−O bond distances in faujasite which are 0.02−0.04 Å longer than the experimental values determined with EXAFS. Similarly, the average measured Al−Si distance, 3.11,[28] is shorter than the corresponding average of calculated distances, by 0.08 Å (Table 4).

In summary, for the first time, we generated a shell-model force field for H-forms of aluminosilicates. This new FF uses realistic charges for the lattice centers and thus is suitable to create an adequate electrostatic field in the vicinity of hydroxyl acidic sites.[17]

Table 4. Selected Structural Parameters (Distances in Å, Angles in deg) for H-Faujasite with 1, 2, and 4 Al Centers in the Unit Cell (Si/Al = 47, 23, 11) Obtained with Our PDC-Based Force Field (PDC) and a B3LYP-Derived Shell Model Force Field Based on Formal Charges (FC)[a,g]

				Si/Al = 23		
parameter[b]	QM model[c]		Si/Al = 47	isolated	paired	Si/Al = 11
Al_I-O_{III} [d]	1.954	PDC	1.910	1.914	1.905, 1.920	1.913
		FC	1.897	1.902	1.869, 1.875	1.897
$Si_{II}-O_{III}$	1.696	PDC	1.686	1.687	1.687, 1.694	1.686
		FC	1.707	1.707	1.731, 1.732	1.704
$O_{III}-H_I$	0.980	PDC	0.985	0.985	0.984, 0.985	0.985
		FC	0.971	0.971	0.968, 0.969	0.971
Al_I-O_{II} [e]	1.709	PDC	1.706	1.705	1.705, 1.712	1.704
		FC	1.714	1.709	1.722, 1.730	1.717
$Si_{II}-O_{IV}$	1.607	PDC	1.632	1.632	1.630, 1.635	1.631
		FC	1.616	1.616	1.608, 1.615	1.613
Si_I-O_{II}	1.604	PDC	1.628	1.625	1.619, 1.626	1.622
		FC	1.598	1.602	1.594, 1.597	1.601
$\langle Al_I-Si_{II}\rangle$[f]		PDC	3.188	3.192	3.192, 3.189	3.189
$Al_I-O_{III}-Si_{II}$	131.6	PDC	129.3	129.7	129.8	129.5

[a] Reference 16b. [b] For the definition of the types of atoms, see Figure 1. [c] Values used for establishing FF parameters (with the exception of the angle $Al_I-O_{III}-Si_{II}$). [d] Exp. 1.87 ± 0.01 Å in ref 28a; 1.89 ± 0.025 Å in ref 28b. [e] Exp. 1.68 Å in ref 28a; 1.66 Å in ref 28b. [f] Averaged value; exp. 3.11 Å in ref 28a. [g] In all structures, the acidic OH groups were located at O1(H) crystallographic positions. Also shown are QM (DFT-BP) results for the QM ref 5T and 6T cluster models in the gas phase; for the description of the clusters, see the footnotes to Table 2 and the Supporting Information.

2.4. Structures Modeled. The QM/MM systems were constructed so that Al-containing QM clusters (exactly, their protonic form) represent *regular* structures of zeolite lattices, i.e., the Al-containing structures were first optimized at the MM level and the nuclei of the QM cluster were placed at the positions of the corresponding MM centers. Structures with Si/Al = ∞ form an exception: there, Al centers were constructed as defects in a pure silica lattice. We denote QM/MM models with x T-atoms and n Al centers in the QM cluster by **xT-nAl−Oz/N**; z refers to the crystallographic position of the oxygen center of the bridging OH group, and N is the Si/Al ratio in the unit cell. The relative position of the Al centers (in the case of more than one Al atom in a unit cell) was chosen to obey Löwenstein's rule, i.e., Al atoms are separated by at least one Si center. Additional conditions were imposed by the specifics how we treat the QM/MM boundary region in our covEPE scheme. As mentioned in section 2.1, we avoided Al(QM)−O* and O*−Al(MM) bonds in QM/MM calculations because they would have required further short-range parameters and an additional type of O* pseudopotential.

2.4.1. Faujasite. The faujasite structure consists of sodalite cages connected via hexagonal prisms which form so-called supercages.[29] The lattice features four- and six-membered rings as parts of these building blocks and twelve-membered windows between the supercages. The pure silica form of faujasite exhibits only one crystallographic T atom position and four oxygen positions in a unit cell. The primitive unit cell of an Al-free structure contains 144 lattice atoms (48 SiO_2 units). The H-forms of faujasite were constructed by replacing Si atoms with Al and adding a charge compensating proton to one of the four neighboring oxygen centers. Structures with Si/Al ratios of 47, 23, and 11 were modeled by inserting 1, 2, and 4 Al atoms per unit cell, respectively. A structure with infinitely small Al content (1 Al center in the entire lattice, Si/Al = ∞) was modeled by introducing the Al center only at the QM level with the surrounding MM lattice being pure silica.

Brønsted acidic sites were represented by various embedded QM cluster models. The smallest, **5T-1Al−O1/∞** and **5T-1Al−O1/47**, have five tetrahedral (T) atoms in the QM cluster with one central Al atom and an acidic proton attached to an O1 crystallographic position (Figure 2a). With the systems **8T-1Al−O1/N** (N = ∞, 47, 23, 11), we studied how the Al content affects the properties of acidic sites; in these model series, chemically identical QM parts represented two coupled 4-rings with a central Al−O(H)−Si bridge terminated by an $OSiO^*_3$ groups at each T atom of the bridge (Figure 2b). We also used an alternative 8T QM model, **8T(4R)-1Al−O1/23**; there, one 4-ring with two opposite T-atoms terminated by $OSiO^*_3$ moieties was embedded in a framework with Si/Al = 23 (not shown in the figure). A pair of acidic sites, located in one four-membered ring, was described with the models **8T-2Al−O1/23** and **10T-2Al−O1/23** (Figure 2c,d). The cluster models **8T-1Al−O3/47** and **10T-1Al−O3/47** (Figure 2e,f) were used to describe Brønsted acidic sites with hydrogen connected to oxygen in an O3 crystallographic position.

2.4.2. ZSM-5. Zeolite ZSM-5 features an MFI lattice which consists of five- and six-membered rings whose edges constitute 10-membered rings that border the largest cavities. The latter rings form straight channels in [010] direction, perpendicularly intersected by zigzag channels in [100] direction. The Al-free form of ZSM-5 exhibits 12 crystallographic positions for Si atoms and 26 for O atoms. The primitive unit cell of pure silica is orthorhombic and comprises 288 atoms in 96 SiO_2 units. We modeled a ZSM-5 structure with Si/Al = 11, a value typical for this material.[19] This Si/Al ratio corresponds to 8 Al atoms and thus 8 bridging OH groups in a unit cell. The eight Al−O(H)−Si bridges in the initial classical optimization of the structure were located in four T7-O17-T4 positions, three T1-O15-

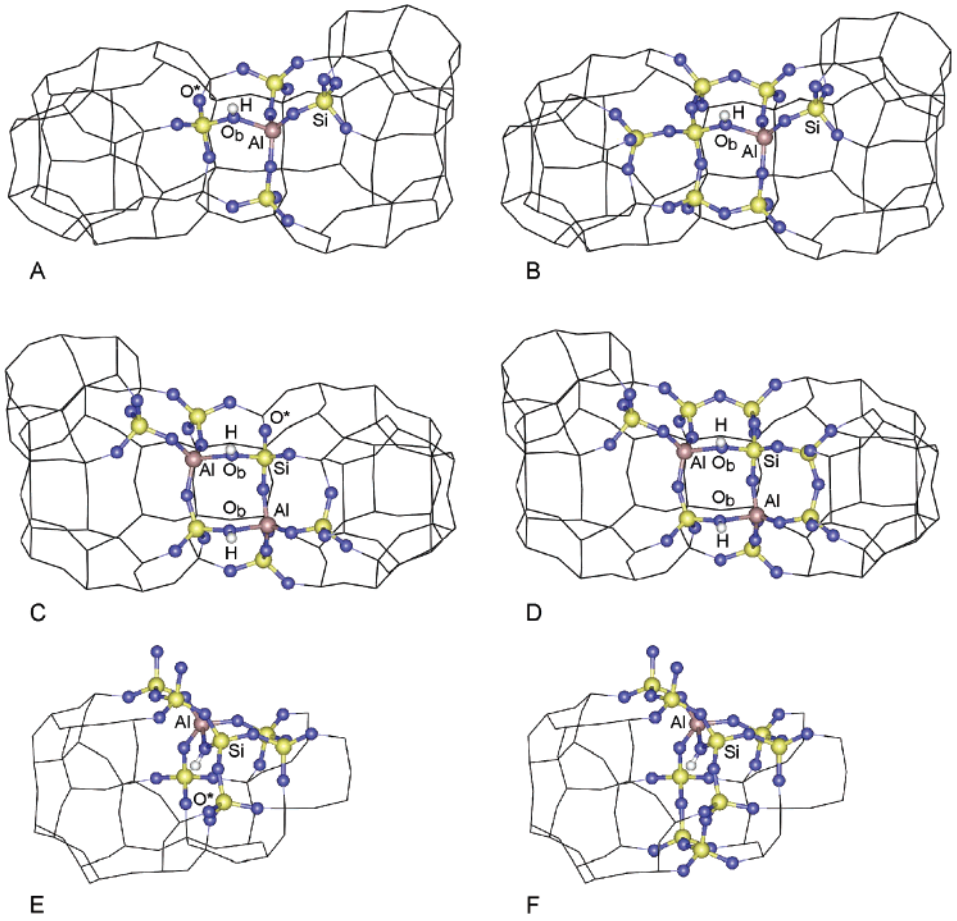


Figure 2. Embedded QM cluster models of various sites of faujasite: a − **5T-1Al**−**O1**, b − **8T-1Al**−**O1**, c − **8T-2Al**−**O1**, d − **10T-2Al**−**O1**, e − **8T-1Al**−**O3**, and f − **10T-1Al**−**O3**.

T10 positions, and one T8-O8-T9 crystallographic positions of the pure silica unit cell. For the QM/MM calculations, we used the cluster model **9T-1Al/11** with a bridging OH group in position O17, e.g. to yield the acidic site Al7−O17(H)−Si4 (Figure 3a) with the OH group directed into a zigzag channel. As experimental information about preferred substitution of T-sites by Al is lacking, we chose the Al positions based on theoretical predictions. In an atomic simulation using an empirical potential, the highest probability for an Al atom in the ZSM-5 lattice was attributed to position T7.[30] Using a DFT-based shell model, Al substitution at T7 was calculated only 5.4 kJ mol^{-1} less stable than the energetically most preferred site T1.[31] The possibility to compare our results with those of the QM-pot cluster embedding approach[32] further motivated us to choose this site.

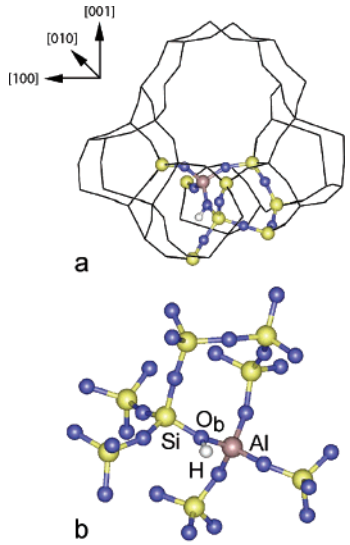


Figure 3. 9T QM model containing the acidic site Al7−O17(H)−Si4 of an HZSM-5 framework: a − view along a straight channel (pseudoatoms O* terminating the QM cluster represented as lines) and b − view from a zigzag channel (surrounding lattice omitted).

3. Results and Discussion

3.1. Effect of QM Cluster Size. Before discussing the framework effects on the characteristics of Brønsted acid sites in QM clusters, we considered computational uncertainties connected with the choice of QM cluster models. With covEPE embedded clusters, a major source of uncertainties could be the QM/MM border and its interaction with the active center (an acidic proton in the present study). The effective Mulliken charge of border oxygen pseudoatoms O* (−0.5 *e*) is less (in absolute value) than the Mulliken charge of real oxygen QM centers[17] (−0.8 *e*), despite the PDC increment of −0.3 *e* assigned to the core of O*. As noted previously,[17] the electrostatic potential close to the O* consequently is 0.02 au more positive than the potential of

Table 5. Selected Structural Parameters (Distances in Å, Angles in deg), Harmonic OH Frequencies ν(OH) (in cm^{-1}), and Deprotonation Energies DE (in kJ mol^{-1}) for Acidic Al−O_b(H)−Si Sites Modeled within QM Clusters Embedded in a Faujasite Lattice of Varying Al Content[l]

Si/Al	∞			47			23				11
cluster	**5T-1Al**[a]	**8T-1Al**[b]	Δ[c]	**5T-1Al**[a]	**8T-1Al**[b]	Δ[c]	**8T-1Al**[b]	**8T-1Al**[d]	**8T-2Al**[d]	**10T-2Al**[e]	**8T-1Al**[b]
Al−O_b	1.954	1.945	−0.009	1.958	1.954	−0.004	1.958	1.949	1.953; 1.923	1.942; 1.916	1.956
Al−O	1.711	1.710	−0.001	1.711	1.712	0.001	1.712	1.706	1.715; 1.700	1.718; 1.701	1.710
Al−O	1.714	1.724	0.010	1.724	1.722	−0.002	1.724	1.723	1.723; 1.719	1.722; 1.724	1.722
Al−O	1.723	1.728	0.005	1.728	1.731	0.003	1.731	1.732	1.739; 1.759	1.739; 1.759	1.733
Si−O_b	1.716	1.718	0.002	1.718	1.722	0.004	1.724	1.717	1.753; 1.742	1.754; 1.742	1.724
O_b-H	0.978	0.978	0.000	0.976	0.978	0.002	0.978	0.978	0.978; 0.976	0.979; 0.976	0.978
H−O*[f]	2.480	4.484		2.618	4.569		4.561	2.751	2.791; 2.773	4.584; 2.756	4.528
H−O_{Al}[g]	2.88	2.62		2.80	2.70		2.72	2.68	2.67; 2.77	2.65; 2.78	2.74
⟨Al−Si⟩[h]	3.19	3.21		3.21	3.21		3.22	3.21	3.21; 3.22	3.21; 3.21	3.21
Al−Al[i]				17.44	17.44		8.65	8.65	4.56	4.56	7.48
O−Al−O_b	98.1;	97.1;		98.1;	99.1;		99.5;	98.8;	99.3; 100.9;	98.9; 104.0;	100.0;
	98.9;	102.;		100.;	102.3;		102.4;	100.4;	108.0/99.9;	107.0/100.7;	102.1;
	101.2	103.9		101.9	102.8		102.8	102.2	101.5; 103.6	101.7; 102.7	102.2
O−Al−O	116.8;	114.4;		116.7;	115.0;		114.4;	115.2;	113.1; 115.8;	112.5; 115.2;	114.6;
	117.6;	115.4;		117.0;	116.0;		115.7;	117.0;	117.0/111.2;	117.0/111.8;	116.2;
	117.8	119.0		117.4	117.6		118.2	118.1	118.3; 118. 5	117.7; 118.5	117.6
Al−O_b−Si	127.5	132.0	4.5	128.7	131.3	2.6	131.3	130.3	127.7; 128.2	128.2; 127.5	131.0
H−O_b−Al	111.3	110.0	−1.3	114.0	111.2	−2.8	111.3	110.3	108.6; 112.1	108.4; 112.7	111.9
H−O_b−Si	112.4	116.3	3.9	117.1	115.8	−1.3	115.6	116.8	116.0; 114.7	114.8; 114.2	115.6
H-opl[k]	8.8	1.8	−7.0	0.2	1.7	1.5	1.8	2.6	7.7; 5.0	8.6; 5.6	1.5
ν(OH)	3747	3714	−33	3754	3720	−34	3721	3730	3727; 3749	3723; 3746	3725
DE	1276	1271	−5	1285	1270	−15	1266	1264	1270	1270	1265

[a] Branched pentameric QM model (Figure 2a). [b] QM cluster in the form of two fused 4-rings (Figure 2b). [c] Changes in parameters of the 8T QM cluster with respect to the 5T QM cluster. [d] QM cluster based on one 4-ring with branches: systems **8T-1Al(4R)**−**O1/23** (not shown) and **8T-2Al**−**O1/23** (Figure 2c). [e] See Figure 2d (DE value for removal of the proton at 4.58 Å from O*). [f] Distance between acidic H and nearest pseudoatom O*. [g] Distance between acidic H and oxygen connected to Al atom. [h] Average Al−Si distance. [i] Shortest distance between Al centers in QM cluster and MM environment. [k] Bending angle of acidic H out of the Al−O_b−Si plane. [l] Bridging oxygen centers O_b of all models are located at O1 crystallographic positions.

a reference MM system comprising an array of potential-derived charges at the lattice positions. Also the Mulliken charge of the Si center close to the border pseudoatom is 0.2 *e* less positive than charge on Si atoms farther from QM/MM border. Therefore, upon embedding of small QM clusters, the electrostatic field in the vicinity of an acidic proton might be perturbed, affecting the properties of the proton. To estimate this effect, we compared results for structure, charge distribution, OH vibrational frequency ν(OH), and deprotonation energy DE when the QM region is extended from a 5T to an 8T cluster model for systems with Si/Al ratio of infinity and 47 (Figure 2a,b).

Comparing calculated structural features of isolated Brønsted sites as obtained with the smaller 5T and the larger 8T QM models, one finds only minor differences in bond lengths (Table 5). The maximum change occurs for the Al−O_b bond: it is about 0.01 Å longer in the smaller models. Other bond distances change at most 0.004 Å. In the 5T models, the H center of the Al−O_b(H)−Si bridge is rather close to terminating pseudoatoms (Figure 2a); thus, bond angles around the O_b center are distorted as compared to the extended 8T QM clusters (Figure 2b), where the acidic site and border centers O* are separated by oxygen and silicon shells. In the 5T model with Si/Al = 47 in the environment, the angle Al−O_b−Si was reduced by 3° and the angle H−O_b−Al increased by 3°. For Si/Al = ∞, larger distortions occur in the 5T model: angles H−O_b−Si and Al−O_b−Si decrease by 4−5° and concomitantly the proton bends 7° more out of the Al−O_b−Si plane. The large perturbations in the latter system may be due to a shorter H−O* distance, 2.48 Å in **5T-1Al**−**O1/**∞ vs 2.62 Å in **5T-1Al**−**O1/47**. The nonbonding distance H−O_{Al} between the proton of the Al−O(H)−Si site and oxygen centers connected to the Al atom is also sensitive to the distance between the active site and the nearest pseudoatom O* (Table 5). Larger H−O* separations in 8T QM models (4.5−4.6 Å vs 2.5−2.6 Å in 5T models) correlate with shorter H−O_{Al} distances, 2.6−2.7 Å, compared to 2.80−2.88 Å in 5T models. Thus, a weaker interaction of H with a pseudoatom O* results in stronger H−O_{Al} interaction. O_{Al} centers are known to exhibit a high basicity;[33] thus, when the interaction of such a center with the acidic proton increases, one expects an elongation of the O−H bond, as is indeed the case (cf. 0.976 Å in **5T-1Al**−**O1/47** vs 0.978 Å in **8T-1Al**−**O1/47**). These deliberations on the proximity of O_{Al} and acidic hydrogen is able to rationalize the difference of the O−H bond lengths in the **8T-2Al**−**O1/23** model; indeed, the OH group at 2.77 from the O_{Al} center is 0.976 Å, while the OH group at 2.67 Å is 0.978 Å (section 3.2).

As a consequence of the above-mentioned structural changes, the charge distribution at the active OH site of small 5T QM clusters differed very slightly from that of larger 8T models. With the larger separation of the active site from border atoms O*, the Mulliken charge of the O_b oxygen

center decreased (by absolute value) from −0.78 to −0.76 *e* and the charge of the proton increased from 0.38 to 0.39 *e*.

In line with the elongation of the O−H bond length by 0.002 Å, the OH vibrational frequency decreases by 33−34 cm^{-1} when the 5T cluster is expanded to an 8T model. The proximity of the QM/MM boundary to the acidic proton also affects the deprotonation energy DE: it decreases by 5−15 kJ mol^{-1} when the size of the QM model increases (Table 5). The QM-Pot embedding scheme also yielded small QM cluster size effects on DE values of faujasite and ZSM-5 lattices, 1−6 kJ mol^{-1}.[9]

The OH frequencies of larger 8T QM clusters are not affected by pseudoatoms, probably because their H−O* distance is larger than in 5T models (see above). This is corroborated by further extending the QM cluster in the 10T models. For instance, in the model **10T-2Al−O1/23** the OH frequency decreases by 4 cm^{-1} (section 3.2) and ν(OH) increases by 6 cm^{-1} (section 3.3). Concomitantly, DE values change at most 12 kJ mol^{-1}. Thus, the features of acidic centers under discussion are already converged for the 8T models.

3.2. Effect of Al Content. In our study we tried to distinguish two ways how the Al content can affect properties of acidic sites in zeolites. First, we considered long-range effects of the overall Al content of the zeolite as monitored by the Si/Al ratio of the MM framework. Second, we studied the direct influence of two Al centers close to each other, by including a second Al center in the QM cluster.

In Table 5 we present the results for QM clusters embedded in a faujasite structure with different Si/Al ratio in the MM region, namely 11, 23, 47, and ∞. As can be seen, the structure of the Al−O_b(H)−Si fragment within the 8T QM cluster is only weakly affected by the Al content of the environment. The Al−O_b and Si−O_b bond lengths involving the protonated oxygen center O_b are elongated only by 0.012 and 0.008 Å, respectively, when the Si/Al ratio decreases from infinity to 11. The other three Al−O bond lengths remain essentially constant; changes do not exceed 0.005 Å. The O_b−H bond distances in the models **8T-1Al−O1/N** with N = 47, 23, 11 are all calculated equal, and only 0.001 Å shorter than for the system **8T-1Al−O1/∞** with a pure-silica environment. Along this series, the average Al−Si distance varies within 0.01 Å only. Pertinent bond angles of all clusters agree within 2°. As the structure of the MM lattice is rather tolerant to changes of the Si/Al ratio from 47 to 11 (Table 4), the MM surrounding imposes very similar steric constraints on all these QM models. This may explain why clusters embedded in the same framework, but with different Al content, yield nearly identical structural parameters. Based on atomistic simulation with the valence force field of Hill and Sauer,[14] faujasite was found to retain its local structure over an even wider range of Si/Al ratios, from 191 to 2.43.

Reduction of the Si/Al ratio in a faujasite lattice from infinity to 11 results in a small increase of the OH vibrational frequency, by 11 cm^{-1} (Table 5), which correlates with minor changes in O−H bond length in this row. Sierka et al. reported[34] a similarly small effect of the chemical composition of the FAU lattice on the O−H stretching frequency in a QM-Pot study. Using 4T embedded QM clusters at the B3LYP level, they determined an increase of the OH frequency by 11 cm^{-1} for the O1(H) position when the Si/Al ratio decreased from 47 to 3; at the HF level, they obtained a positive shift of the same order, 25 cm^{-1}. The calculated trend of increasing frequencies with increasing Al content is in line with experimental observation: for HY zeolite (Si/Al = 20.7 to 2.5), the O1(H) band was found to shift by 17 cm^{-1} to higher frequencies.[35] OH frequencies from fully periodic B3LYP calculations on H-chabazite scattered more.[2] An increase of the Al content from Si/Al = 11 to 3 caused the vibrational frequency of an O1(H) acidic site to increase by 21 cm^{-1}. Even larger Al loading, Si/Al = 1, resulted in a sharp jump of ν(OH) by 24 cm^{-1}. However, this could be an artifact of the computational procedure[2] because in that work zeolite structures (including OH groups) with Si/Al > 1 were optimized at the MM level, at variance with the case Si/Al = 1 where everything was done at the DFT level.

Isolated acidic sites modeled within 8T QM clusters have very similar deprotonation energies, irrespective of the Si/Al ratio in the surrounding MM part. Variations of the DE values, 1265−1271 kJ mol^{-1}, are comparable to the accuracy of the computational modeling; for instance, errors due to the limited size of the QM model are estimated to 5−15 kJ mol^{-1}, see section 3.1. Nearly constant DE values, 1242−1250 kJ mol^{-1} (HF) and 1196−1198 kJ mol^{-1} (B3LYP), were also obtained with 4T QM clusters embedded in FAU environments with Si/Al ratios ranging from 47 to 3.[34]

The present QM/MM study, carried out at the level of mechanical and electrostatic embedding, does not find any strong effect of the Al concentration in the zeolite lattice on structural parameters, spectroscopic features, and deprotonation energies of isolated Brønsted sites. This result can be rationalized by the large spatial separation among pertinent Al−O(H)−Si sites. Indeed, the distance from the Al atom in the QM clusters to the nearest Al center in the surrounding exceeds 7 Å for lattices with the largest Al loading, Si/Al = 11 (Table 5). At such distances, a minor charge redistribution within the MM lattice, induced by incorporating acidic sites (0.1−0.2e for the substitutions $Si_I \rightarrow Al_I$, $O_I \rightarrow O_{II}$, $O_I \rightarrow O_{III}$ and 0.35e for proton addition), is hardly noticeable in the QM region. As corroboration, note that the electrostatic potential in the vicinity of acidic hydrogen centers is perturbed less than 10 kJ mol^{-1} when the Si/Al ratio in the MM environment decreases from 47 to 11.

Al centers in paired acidic sites are only 4.5 Å apart. For such a configuration, **8T-2Al−O1/23** (Figure 2c), the calculated DE value is 6 kJ mol^{-1} higher than for an isolated acidic site, **8T(4R)-1Al−O1/23**: (1270 and 1264 kJ mol^{-1}, respectively; Table 5). We also considered a larger cluster model with two Al centers, **10T-2Al−O1/23**, where, upon removal of an OH group far from the border, we obtained a DE value that was 4 kJ mol^{-1} higher than the corresponding model with one Al center, **8T-1Al−O1/23.** This finding corroborates the assertion that the deprotonation energy grows with the number of Al neighbors of the silicon atom in the Si−O(H)−Al bridge as observed experimentally by

Datka et al.[36] A somewhat larger increase of DE for paired acidic sites was reported in a QM-pot study based on smaller 4T QM cluster models[34] that were embedded in a faujasite lattice; there, the DE values for models with two Al centers were calculated 24−28 kJ mol^{-1} higher for paired 4T-2Al acidic sites compared to isolated 4T-1Al acid sites. A recent DF study based on isolated cluster models of a faujasite 6-ring (containing also two Al centers in next-nearest neighbor positions) also suggested an increase of DE from 1203 kJ/mol for a monoprotonated to 1251 kJ/mol for a diprotonated cluster.[37] Note, however, that our calculations were carried out on larger 8T and 10T QM clusters, at variance with 4T[34] and 6T QM models[37] used in the two studies just mentioned above; also, the present computational approach is more sophisticated.

At variance with deprotonation energies, OH vibrational frequencies were calculated to be sensitive to the presence of an additional Al atom as nearest neighbor of the Si center of the bridging OH group. One OH frequency increased from 3721 cm^{-1} for **8T-1Al−O1/23** to 3749 cm^{-1} for **8T-2Al−O1/23** and to 3746 cm^{-1} for **10T-2Al−O1/23** (Table 5). However, the second OH band (3727 cm^{-1} in **2Al-8T** and 3723 cm^{-1} in **2Al-10T**) remained almost unchanged, exhibiting shifts of only 2−6 cm^{-1} with respect to the frequency of the isolated acidic site. The differentiation of OH frequencies in "paired" acidic sites can be explained by their different crystallographic surroundings. Indeed, the Al center of one Al−O(H)−Si bridge connects to the Si atom of another acidic site via oxygen at an O2 crystallographic position, whereas the Al atom of the other acidic site has an oxygen neighbor at position O3, located in a four-ring with two Al centers. Therefore, the changes due to the close location of two Al−O(H)−Si bridges in the local structures of these acidic sites are different (Table 5). Note that the higher OH frequency corresponds to the OH group with a shorter O−H bond (by 0.002−0.003 Å). Such bond contraction, in turn, weakens the attraction of the proton to the corresponding basic center O_{Al}; cf. the H−O_{Al} distances of 2.77 Å and 2.67 Å for the two sites of **2Al-8T**.

In summary, Al loading of a zeolite lattice affects some of the properties of acidic sites in a noticeable fashion, but only when Al centers are located at closest possible distances, i.e., separated by only one SiO_4 tetrahedron. Such a neighborhood of Al centers results in significant geometry changes of the corresponding acidic sites and a notable increase of the frequency of the nearby OH group. Variation of the Al content far from a bridging OH group does not lead to significant changes in the properties of the active center. A similar conclusion was drawn from an experimental IR study on N_2 sorption at HY samples with Si/Al ratios between 2 and 17.[38]

3.3. Effect of Crystallographic Oxygen Position and Framework Structure. In contrast to the minor influence of the Al loading of the lattice on structure, spectroscopic features, and deprotonation energies of acidic centers, we observed in our modeling that structural constraints of the surrounding framework on the geometry of an Al−O(H)−Si bridge exert a clear effect. In particular, the OH vibrational frequency is sensitive to the crystallographic position of the bridging oxygen center to which the acidic proton is attached. Frequencies corresponding to different crystallographic hydroxyl positions in a faujasite lattice can differ by as much as 85−91 cm^{-1}; see the results for the cluster **8T-1Al−O1/47** modeling an group O1(H) which points inside a 12-membered ring and the clusters **8T-1Al−O3/47** and **10T-1Al−O3/47** modeling O3(H) groups which point inside six-membered rings (Table 6). These results are consistent with the experimentally measured splitting of 73 cm^{-1} between high O1(H) and low-frequency O3(H) bands in IR spectra of faujasites.[35]

Also this frequency difference has a structural origin: the O1−H distance of a hydroxyl group pointing into the larger 12-ring is 0.004 Å shorter than the distance O3−H. This difference may be rationalized by the formation of a weak hydrogen bond across the smaller 6-ring, with oxygen centers at about 2.7 Å from the proton attached to O3. The 12-membered ring is too large for the formation of a hydrogen bond involving the O1(H) group. Because border pseudoatoms O* as a part of the 6-ring of the **8T-1Al−O3/47** model (Figure 2e) are still too close to the active center, they may distort the electrostatic field in the neighborhood of the acidic proton and consequently affect its properties. To exclude this possibility, we included the 6-ring completely in the QM cluster in the **10T-1Al−O3/47** model. Yet, we found the geometrical features and the OH frequency of the 8T and 10T models to be very similar, and we calculate only a very small shift of ν(OH), from 3629 cm^{-1} in the 8T model to 3635 cm^{-1} in the 10T model (Table 6). Therefore, these results demonstrate that the properties under discussion are already converged for the 8T QM cluster and that this model should also be adequate for describing the formation of a H-bond involving the 6-ring.

Despite significant structure and frequency changes, deprotonation energies of O1(H) and O3(H) sites agree within 12 kJ mol^{-1}, using DE values for the **10T-1Al−O3/47** model; this corroborates an earlier finding with QM-Pot modeling.[32,34] The local geometry of an acidic site, its OH vibrational frequency, and energy of proton detachment change drastically with the structure of the framework. The Brønsted site located in the Al7−O17(H)−Si4 position of the HZSM-5 framework (Figure 3) features Al−O_b and Si−O_b bonds that are 0.01−0.02 Å shorter than those of the acidic site O1(H) in faujasite (Table 6). Our results are at variance with structure data of recent EXAFS studies which reported a notably shorter Al−O_b bond, by 0.11 Å, in FAU than in HZSM-5.[28] The deviation in the latter case may be related to the fact that in the experimental sample the OH group is located in a different crystallographic position. However, the discrepancy may also be caused by the limited accuracy of interatomic distances from EXAFS studies as can be seen from the large difference between the experimental values for MFI and FAU, 0.11 Å. Our calculated Al−O_b distances of both types of zeolites, 1.94−1.96 Å, fall between the EXAFS values for FAU and MFI. Moreover, the other calculated Al−O distances not involving the bridging oxygen center O_b (with the proton) have been found almost identical in both materials. Our theoretically predicted shortening of the average Al−Si distance ⟨Al−Si⟩ from 3.21

Table 6. Selected Structural Parameters (Distances in Å, Angles in deg), Harmonic OH Frequencies ν(OH) (in cm^{-1}), and Deprotonation Energies DE (in kJ mol^{-1}) Calculated for Embedded Cluster Models Representing Various Crystallographic Positions of Bridging Oxygen Centers O_b in Faujasite (FAU) and HZSM-5 (Si/Al Ratio Also Shown)

lattice	FAU				HZSM-5	
O(H) site	Al−O1−Si	Al−O3−Si			Al7−O17−Si4	
cluster/Si:Al	**8T-1Al/47**[a]	**8T-1Al/47**[b]	**10T-1Al/47**[c]	exp.	**9T-1Al/11**[d]	exp.[e]
Al−O_b	1.954	1.959	1.955	1.87 ± 0.01,[e] 1.89 ± 0.025 [f]	1.936	1.98 ± 0.01
Al−O	1.712	1.717	1.717	1.68,[e] 1.69[f]	1.721	1.66
Al−O	1.722	1.717	1.718		1.722	1.71
Al−O	1.731	1.718	1.721		1.730	1.74
Si−O_b	1.722	1.725	1.723		1.710	
O_b−H	0.978	0.982	0.982		0.977	
H−O*	4.569	2.763	4.109		3.748	
H−O_{Al}	2.70	2.76	2.78		2.67	
<Al−Si>	3.21	3.23	3.23	3.11[e]	3.17	3.09
$(Al-Si)_{max}$ [g]					3.24	3.28
Al−Al	17.44	17.42	17.42		7.28	
O−Al−O_b	99, 102, 102	99, 102, 102	100, 102, 102		94, 98, 105	97 ± 6
O−Al−O	115, 116, 117	115, 116, 118	114, 115, 119		112, 119, 121	119 ± 6
Al−O_b−Si	131.3	129.3	128.9		125.1	
Al−O−Si	137, 140, 151				128, 143, 150	
H−O_b−Al	111.2	113.4	114.2		109.8	
H−O_b−Si	115.8	117.1	116.6		117.2	
H−opl[h]	1.7	0.2	0.3		7.9	
ν(OH)	3720[i]	3629[j]	3635[j]		3715[k]	
DE	1270	1268	1256		1237	

[a] Figure 2b. [b] Figure 2e. [c] Figure 2f. [d] Figure 3. [e] Reference 28a. [f] Reference 28b. [g] Longest Al−Si distance. [h] Bending angle of acidic H out of the Al−O_b−Si plane. [i] Exp. 3623 cm^{-1}, Si/Al = 20.7, ref 35. [j] Exp. 3550 cm^{-1}, Si/Al = 20.7, ref 35. [k] Exp. 3610−3617 cm^{-1}, Si/Al = 20−13.6; see refs 39 and 40.

Å in FAU to 3.17 Å in HZSM-5 agrees with the experimental observation that the average ⟨Al−Si⟩ contracts from 3.11 (H−FAU, Si/Al = 2.5) to 3.09 Å (HZSM-5, Si/Al = 18).[28a] Likely, the reduction of Al−O−Si bond angles in HZSM-5 contributes in a significant way to this shortening (Table 6); for instance, we calculated the Al−O_b−Si angle in HZSM-5 6° smaller than in H−FAU.

The O_b−H bond in HZSM-5 was determined 0.001 Å shorter, but the OH frequency was calculated at 3715 cm^{-1}, i.e., 5 cm^{-1} lower than for O1(H) in **8T-1Al−O1/47** but 80 cm^{-1} higher than for O3(H) in **10T-1Al−O3/47** (Table 6). This result agrees with IR measurements[39,40] which show the OH band of bridging hydroxyls in HZSM-5 zeolite, 3610−3617 cm^{-1}, only slightly lower than the band of bridging O1(H) groups in high-silica faujasite, 3623 cm^{-1}, but significantly higher than the vibrational frequency of O3(H) sites in faujasite, 3550 cm^{-1}.

The calculated Mulliken charge, 0.39 *e*, of the acidic proton is essentially independent of both the Al content and the position of the OH group; a slightly larger value, 0.40 *e*, was calculated for the cluster O3H cluster due to formation of a weak hydrogen bond.

We calculated the deprotonation of the Al7−O17−Si4 site to proceed easier, DE = 1237 kJ mol^{-1}, compared to the sites O1(H) and O3(H) of a faujasite lattice with DE = 1270 and 1256 kJ mol^{-1}, respectively; see the results for the models **8T-1Al−O1/47** and **10T-1Al−O3/47** (Table 6). The larger energy release accompanying the protonation of HZSM-5 correlates with the stronger distortion of O−Al−O angles in the AlO_4 tetrahedron induced by attaching a proton to one of these oxygen centers.[28a] Indeed, the O−Al−O angles at the Al7−O17(H)−Si4 site of HZSM-5 vary in the range 94−121°, whereas the corresponding angles for an O1(H) site of faujasite vary in a much smaller interval, 99−117° (Table 6). The calculated DE for HZSM-5 is about 10−20 kJ mol^{-1} smaller compared to values calculated previously for four acidic sites in chabazite, 1250−1258 kJ mol^{-1}, using our covEPE scheme.[17] Thus, our covEPE scheme correctly predicts the experimentally observed decrease of the deprotonation energy in the series FAU > CHA > HZSM-5, tentatively estimated from measured shifts of the OH frequency upon adsorption of a probe CO molecule.[41] The lower DE value of HZSM-5 zeolite calculated with our covEPE model is at variance with QM-Pot results[32,42] where the energy for proton removal from site Al7−O17(H)−Si4 in HZSM-5 (Si/Al = 95) was calculated 29−36 kJ mol^{-1} higher than from O1(H) sites in H−Y zeolite (Si/Al = 47). The OM-Pot deprotonation energy of faujasite O1(H) sites remains lower than that of HZSM-5 also for higher Al content (Si/Al = 2.43);[34] even an increase of the number of Al centers near the Si center of an Al−O(H)−Si bridge, which affected the DE values stronger (24 kJ mol^{-1}) than the overall Al content (2−8 kJ mol^{-1}), did not render FAU less acidic than HZSM-5. Thus, in that aspect, QM-Pot results[32,42] did not agree with experiment.[41]

Our comparative study on the effect of two factors—the overall Al content and the structure of the zeolite framework—on OH frequencies and deprotonation energies of Brønsted

centers in FAU and HZSM-5 zeolites shows that the zeolite structure exerts a more pronounced influence on acidic sites. Overall Al loading exerts an effect that is five to six times smaller and hence unable to change the ordering of deprotonation energies between faujasite and HZSM-5.

4. Conclusions

The present study demonstrated that the DFT based covEPE embedding scheme is able to describe aluminosilicate systems. We investigated industrially important zeolites of faujasite and HZSM-5 type. We studied computationally how the Al content affects the local structure, the OH vibrational frequency, and the deprotonation energy of Brønsted acidic sites in zeolite frameworks. For the first time, this problem was considered by fully accounting for both mechanical and electrostatic coupling between QM and MM parts of a hybrid QM/MM system. The newly parametrized shell-model force field for protonic forms of zeolites, assigning potential-derived charges to lattice centers, is an important aspect of the present study. In this way, an adequate electrostatic field[17] is produced in the vicinity of hydroxyl acidic sites in zeolite cavities. We also discussed guidelines for constructing satisfactory QM cluster models that minimize undesirable effects of the QM/MM border region on calculated properties.

Based on the computational results, we concluded that structural features, spectroscopic characteristics, and deprotonation energies of isolated acidic sites in zeolite frameworks with low to medium Al content (Si/Al = 47−11) are similar to those in lattices with infinite Al dilution (Si/Al = ∞). For QM clusters embedded in a faujasite framework with Si/Al ratio from infinity to 11, bond distances and angles change at most by 0.01 Å and 2°, respectively, OH frequencies by 11 cm^{-1}, and deprotonation energies by 6 kJ mol^{-1}. It is crucial to evaluate the effect of the Al content with sufficiently large QM cluster models. Otherwise, computational uncertainties of ∼30 cm^{-1} for OH frequencies and ∼5−15 kJ mol^{-1} for deprotonation energies due to the proximity of the QM/MM border may occur. The influence of *other* Al centers is somewhat more noticeable if their concentration is high enough to generate Al centers at minimum separation from (i.e. as nearest neighbors of) the Si center of the Al−O(H)−Si site. One OH frequency of such "paired" acidic sites is 25−28 cm^{-1} higher than that of an isolated Al−O(H)−Si bridge, whereas the other one remains unchanged within 2−6 cm^{-1}. Observed frequency shifts reflect this nonequivalence of the immediate crystallographic surroundings of the acidic sites which appears when they are close to each other. However, the proton detachment energy is only weakly affected by such structural differences; it was calculated 4−6 kJ mol^{-1} larger than for an isolated OH group.

Effects due to a different local structure of the zeolite framework are larger than those due to changes of the Al content. This stronger influence of the local framework structure was demonstrated for OH frequencies which differ by 85 cm^{-1} between O1(H) and O3(H) sites of a faujasite lattice (Figure 2), although the change in the DE values is rather small, 14 kJ mol^{-1}. With our covEPE embedding strategy, we were also able to reproduce the trend of decreasing proton removal energies (by about 20−25 kJ mol^{-1}) along the series FAU > CHA > HZSM-5, in agreement with experiment. In addition, ordering of and differences between simulated OH frequencies of the three bridging hydroxyl groups, O1(H) (3720 cm^{-1}) and O3(H) (3635 cm^{-1}) of faujasite and the OH group of HZSM-5 (3715 cm^{-1}), very well fit the experimental data: 3623, 3550, 3610−3617 cm^{-1}, respectively.

Acknowledgment. We thank J. D. Gale for providing the latest version of his program GULP. E.A.I.S. gratefully acknowledges an individual grant of the Krasnoyarsk Regional Scientific Foundation (Grant 14G241). This work was supported by Volkswagen-Stiftung (Grant I/73653), INTAS/RFBR (Grant IR-97-1071/RFBR 97-03-71057), Alexander von Humboldt Foundation, Deutsche Forschungsgemeinschaft, Fonds der Chemischen Industrie, and Bulgarian National Science Council.

Supporting Information Available: Sketches of the seven QM clusters used to estimate PDCs, figures of four reference gas-phase QM models utilized for establishing short-range and core−shell parameters of the aluminosilicate force field, table representing core−shell parameters for polarizable oxygen centers, and Cartesian coordinates of all optimized embedded QM clusters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) *Computer Modelling of Microporous Materials*; Catlow C. R. A., van Santen, R. A., Smit. B., Eds.; Academic Press: London, 2004; p 320.

(2) Ugliengo, P.; Civalleri, B.; Zicovich-Wilson, C. M.; Dovesi, R. *Chem. Phys. Lett.* **2000**, *318*, 247−255.

(3) White, J. C.; Hess, A. C. *J. Phys. Chem.* **1993**, *97*, 8703−8706.

(4) Strodel, P.; Neyman, K. M.; Knözinger, H.; Rösch, N. *Chem. Phys. Lett.* **1995**, *240*, 547−552.

(5) Rösch, N.; Vayssilov, G. N.; Neyman, K. M. In *Host−Guest-Systems Based on Nanoporous Crystals*; Laeri, F., Schüth, F., Simon, U., Wark, M., Eds.; Wiley-VCH: Weinheim, 2003; pp 339−357. (b) Rösch, N.; Nasluzov, V. A.; Neyman, K. M.; Pacchioni, G.; Vayssilov, G. N. In *Computational Material Sciences*; Leszczynski, J., Ed.; Theoretical and Computational Chemistry Series, Elsevier: Amsterdam, 2004; Vol. 15, pp 367−438.

(6) Bates, S. P.; van Santen, R. A. *Adv. Catal.* **1998**, *42*, 1−114.

(7) Gonzales, N. O.; Chakraborty, A. K.; Bell, A. T. *Catal. Lett.* **1998**, *50*, 135−139.

(8) Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79−92.

(9) Sauer J.; Sierka, M. *J. Comput. Chem.* **2000**, *21*, 1470−1493.

(10) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894−905. (b) Breneman, C. M. Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.

(11) Catlow, C. R. A.; Mackrodt, W. C. In *Computer Simulations of Solids, Lecture Notes in Physics*; Catlow, C. R. A., Mackrodt W. C., Eds.; Springer: Berlin, 1982; Vol. 166, p 3.

(12) Blake, N. P.; Weakliem P. C.; Metiu, H. *J. Phys. Chem.* **1998**, *102*, 67−74.

(13) Kramer, G. J.; de Man, A. J. M.; van Santen, R. A. *J. Am. Chem. Soc.* **1991**, *113*, 6435−6441. (b) van Beest, B. W. H.; Kramer, G. J.; van Santen, R. A. *Phys. Rev. Lett.* **1990**, *64*, 1955−1958. (c) Kramer, G. J.; Farragher, N. P.; van Beest, B. W. H.; van Santen, R. A. *Phys. Rev. B* **1991**, *43*, 5068−5080.

(14) Hill, J.-R.; Sauer, J. *J. Phys. Chem.* **1995**, *99*, 9536−9550.

(15) Schröder, K.-P.; Sauer, J.; Leslie, M.; Catlow, C. R. A.; Thomas, J. M. *Chem. Phys. Lett.* **1992**, *188*, 320−325.

(16) Schröder, K.-P.; Sauer, J. *J. Phys. Chem.* **1996**, *100*, 11043−11049. (b) Sierka, M.; Sauer, J. *Faraday Discuss.* **1997**, *106*, 41−62.

(17) Nasluzov, V. A.; Ivanova, E. A.; Shor, A. M.; Vayssilov, G. N.; Birkenheuer, U.; Rösch, N. *J. Phys. Chem. B* **2003**, *107*, 2228−2241.

(18) Czjzek, M.; Jobic, H.; Fitch, A. N.; Vogt, T. *J. Phys. Chem.* **1992**, *96*, 1535−1540.

(19) Barthomeuf D. *Catal. Rev. Sci. Eng.* **1996**, *38*, 521−612.

(20) Belling, T.; Grauschopf, T.; Krüger, S.; Nörtemann, F.; Staufer, M.; Mayer, M.; Nasluzov, V. A.; Birkenheuer, U.; Hu, A.; Matveev, A. V.; Shor, A. M.; Fuchs-Rohr, M. S. K.; Neyman, K. M.; Ganyushin, D. I.; Kerdcharoen, T.; Woiterski, A.; Gordienko, A. B.; Majumder, S.; Rösch, N. ParaGauss version 3.0, Technische Universität München, 2004.

(21) Belling, T.; Grauschopf, T.; Krüger, S.; Mayer, M.; Nörtemann, F.; Staufer, M.; Zenger, C.; Rösch, N. In *High Performance Scientific and Engineering Computing*; Bungartz, H.-J., Durst, F., Zenger, C., Eds.; Lecture Notes in Computational Science and Engineering, Springer: Heidelberg, 1999; Vol. 8, pp 441−455.

(22) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824; **1986**, *34*, 7406−7406.

(23) Van Duijneveld, F. B. IBM Res. Report RJ 945, 1971. (b) *Gaussian Basis Sets for Molecular Calculations*; Huzinaga, S., Ed.; Elsevier: Amsterdam, 1984. (c) Veillard, A. *Theor. Chim. Acta* **1968**, *12*, 405−411. (d) Bär, M. R.; Sauer, J. *Chem. Phys. Lett.* **1994**, *226*, 405−412.

(24) Dunlap, B. I.; Rösch, N. *Adv. Quantum Chem.* **1990**, *21*, 317−339.

(25) Chung, S. C.; Krüger, S.; Pacchioni, G.; Rösch, N. *J. Chem. Phys.* **1995**, *102*, 3695−3702.

(26) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian98, Revision A.7, Gaussian, Inc., Pittsburgh, PA, 1998.

(27) Vollmer, J. M.; Stefanovich, E. V.; Truong, T. N. *J. Phys. Chem. B* **1999**, *103*, 9415−9422.

(28) Joyner, R. W.; Smith, A. D.; Stockenhuber, M.; van den Berg, M. W. E. *Phys. Chem. Chem. Phys.* **2004**, *6*, 5435−5439. (b) van Bokhoven, J. A.; van der Eerden, A. M. J.; Prins R. *J. Am. Chem. Soc.* **2004**, *126*, 4506−4507.

(29) http://www.iza-structure.org/databases

(30) Schröder, K.-P.; Sauer, J.; Leslie, M.; Catlow, C. R. A. *Zeolites* **1992**, *12*, 20−23.

(31) Nachtigallova, D.; Nachtigall, P.; Sierka, M.; Sauer, J. *Phys. Chem. Chem. Phys.* **1999**, *1*, 2019−2026.

(32) Eichler, U.; Brändle M.; Sauer, J. *J. Phys. Chem. B* **1997**, *101*, 10035−10050.

(33) Vayssilov, G. N.; Staufer, M.; Belling, T.; Neyman K. M.; Knözinger, H.; Rösch, N. *J. Phys. Chem. B* **1999**, *103*, 7920−7928. (b) Vayssilov, G. N.; Rösch, N. *J. Catal.* **1999**, *186*, 423−432.

(34) Sierka, M.; Eichler, U.; Datka, J.; Sauer, J. *J. Phys. Chem. B* **1998**, *102*, 6397−6404.

(35) Anderson, M. W.; Klinowski, J. *Zeolites* **1986**, *6*, 455.

(36) Datka, J.; Broclawik, E.; Gil, B. *J. Phys. Chem.* **1994**, *98*, 5622−5626. (b) Datka, J.; Gil, B.; Domagala, T.; Gora-Marek, K. *Microporous Mesoporous Mater.* **2001**, *47*, 61−66. (c) Datka, J.; Gil, B.; Baran, P. *J. Mol. Struct.* **2003**, *645*, 45−49.

(37) Vayssilov, G. N.; Rösch, N. *J. Phys. Chem. B* **2001**, *105*, 4277−4284.

(38) Lonyi, F.; Valyin, J.; Pal-Borbely, G. *Microporous Mesoporous Mater.* **2003**, *66*, 272−282.

(39) Zholobenko, V. L.; Kustov, L. M.; Borovkov, V. Yu.; Kazansky, V. B. *Zeolites* **1988**, *8*, 175−178.

(40) Kubelkova, L.; Beran, S.; Lercher, J. A. *Zeolites* **1989**, *9*, 539−543.

(41) Coliccia, S.; Marchese, L.; Martra, G. *Microporous Mesoporous Mater.* **1999**, *30*, 43−56.

(42) Brändle, M.; Sauer, J. *J. Am. Chem. Soc.* **1998**, *120*, 1556−1570.

CT049910N

JCTC Journal of Chemical Theory and Computation

Histidine−Aromatic Interactions in Proteins and Protein−Ligand Complexes: Quantum Chemical Study of X-ray and Model Structures

Emilie Cauët,[†] Marianne Rooman,[‡] René Wintjens,[§] Jacques Liévin,[†] and Christophe Biot*,[‡]

Service de Chimie quantique et Photophysique, Université Libre de Bruxelles, CP 160/09, 50 Avenue F.D. Roosevelt, B-1050 Bruxelles, Belgium, Unité de Bioinformatique génomique et structurale, Université Libre de Bruxelles, CP 165/61, 50 Avenue F.D. Roosevelt, Institut de Pharmacie, B-1050 Bruxelles, Belgium, and Service de Chimie générale, Université Libre de Bruxelles, CP 206/04, Bld du Triomphe, B-1050 Bruxelles, Belgium

Received November 23, 2004

Abstract: His−aromatic complexes, with the His located above the aromatic plane, are stabilized by $\pi-\pi$, $\delta^+-\pi$ and/or cation$-\pi$ interactions according to whether the His is neutral or protonated and the partners are in stacked or T-shape conformations. Here we attempt to probe the relative strength of these interactions as a function of the geometry and protonation state, in gas phase, in water and protein-like environments (acetone, THF and CCl_4), by means of quantum chemistry calculations performed up to second order of the Møller−Plesset pertubation theory. Two sets of conformations are considered for that purpose. The first set contains 89 interactions between His and Phe, Tyr, Trp, or Ade, observed in X-ray structures of proteins and protein−ligand complexes. The second set contains model structures obtained by moving an imidazolium/imidazole moiety above a benzene ring or an adenine moiety. We found that the protonated complexes are much more stable than the neutral ones in gas phase. This higher stability is due to the electrostatic contributions, the electron correlation contributions being equally important in the two forms. Thus, $\pi-\pi$ and $\delta^+-\pi$ interactions present essentially favorable electron correlation energy terms, whereas cation$-\pi$ interactions feature in addition favorable electrostatic energies. The protonated complexes remain more stable than the neutral ones in protein-like environments, but the difference is drastically reduced. Furthermore, the T-shape conformation is undoubtedly more favorable than the stacked one in gas phase. This advantage decreases in the solvents, and the stacked conformation becomes even slightly more favorable in water. The frequent occurrence of His−aromatic interactions in catalytic sites, at protein−DNA or protein−ligand interfaces and in 3D domain swapping proteins emphasize their importance in biological processes.

Introduction

Cation$-\pi$ interactions were shown over the past 15 years to play important roles in many chemical and biological processes.[1−3] In proteins, for example, cation$-\pi$ interactions between an aromatic π ring (i.e. Phe, Tyr or Trp) and an organic cation (i.e. Lys or Arg) are frequently observed.[4−9] Their stabilizing nature was demonstrated by both in silico energy calculations and experimental means. In particular,

* Corresponding author e-mail: christophe.biot@ensc-lille.fr. Current address: Laboratoire de Catalyse de Lille − UMR CNRS 8010, ENSCL, Bâtiment C7, Université des Sciences et Technologies, B.P. 90108, 59652, Villeneuve d'Ascq Cedex, France.

[†] Service de Chimie quantique et Photophysique.

[‡] Unité de Bioinformatique génomique et structurale.

[§] Service de Chimie générale.

10.1021/ct049875k CCC: $30.25
Published on Web 03/09/2005

a specific cation−π interaction between a phenyl ring and a protonated amine in an α-helix was measured as stabilizing the protein structure by 0.4 kcal/mol.[10]

Several biomolecular association processes such as ligand-antibody binding and receptor−ligand interactions also feature cation−π interactions.[11,12] Cation−π interactions have also been identified at the protein−DNA[13,14] and protein−ligand interface,[15,16] where the nucleic acid bases play the role of aromatic partners. The existence of cation−π interactions between aromatic side chains and metal ions is more debated; note however the interaction between a Na^+ and a Trp ring experimentally evidenced in hen egg white lysozyme.[17]

His side chains are also commonly observed as potential cation−π partners in protein structures but have much less been investigated, due to their ubiquitous nature. Indeed, the imidazole cycle of His has a pK_a value of 6.1 and can therefore occur as neutral or protonated under physiological conditions. Two cases can be distinguished according to whether the imidazole ring is located above a positively charged or aromatic side chain. In the former, the His is obviously neutral and plays the role of aromatic moiety of the cation−π pair, whereas in the latter, the His may be either neutral or protonated. When neutral, it forms aromatic−aromatic interactions with its aromatic partner as well as δ^+−π interactions, due to polarization. When protonated, it plays the role of cation in the cation−π pair. We focus in the present paper on His residues interacting with aromatic side chains and analyze the effect of protonation on the strength of the interaction.

In barnase, experimental evidence indicates that protonated His18 interacts with Trp94 more strongly than the neutral form, by about 1 kcal/mol, and increases the pK_a value of His18; this suggests that the His−aromatic cation−π interactions play an important role in the stabilization of the structure.[18] The stabilizing role of protonated His interacting with aromatic side chains has also been experimentally evidenced in α-helices[19] and in apoflavodoxin.[20] Moreover, His-involving cation−π interactions have been described as having a functional role, by affecting ligand binding[21,22] or catalysis in ion channels,[23−25] in G-protein-coupled receptors,[26] in transporters[27] and in enzymes.[28]

To our knowledge, no in silico characterization of His−aromatic interactions have been performed up to now. To fill this lacuna, we searched X-ray structures of proteins and protein−ligand complexes for His−aromatic interactions and probed their interaction free energy contribution in water and in different solvents mimicking protein environments (CCl_4, THF and acetone) by means of quantum chemistry computations. The free energies calculated for the protonated and neutral forms of His were compared so as to evaluate the relative weight of the cationic and aromatic natures of His in stabilizing the His−aromatic interaction (δ^+−π or cation−π). To complete the analysis, we also considered model structures to allow a more thorough investigation of the stacked and T-shape geometries of His−Phe and His−Ade pairs. We end by discussing the potential functional role of these interactions.

Methods

1. Sets of X-ray Cation−π Complexes. Two sets of high-resolution X-ray structures with less than 25% sequence identity were considered. One comprehends 141 protein chains[30] and the other 188 complexes of proteins and nucleobase containing ligands.[15] The first set was searched for cation−π interactions between an His and an aromatic side chain (Trp, Tyr, Phe) and the second for His−nucleic acid base (Ade, Gua, Cyt, Thy).

Cation−π interactions were identified according to a distance and an angle criterion.[13] The distance criterion requires that at least one of the atoms of the aromatic ring is located no further than 4.5 Å from one of the atoms carrying the positive charge of the His, considered to the $N_{\delta 1}$, $N_{\epsilon 2}$, and $C_{\epsilon 1}$-atoms in the absence of H-atoms. The angle criterion demands the latter atom to be situated above the plane defined by the aromatic ring, more precisely, inside a cylinder of height 4.5 Å, whose base includes the ring and has a radius equal to twice the radius of the ring.

The cation−π pairs were simplified for computational study. Ligands were reduced to the nucleobase(s) they contain. Phe, Tyr and Trp were represented as benzene, phenol and indole moieties, respectively, and His as an imidazole ring (protonated and unprotonated). The H-atoms were added by construction. For neutral His, the two tautomeric forms are considered, with an H-atom linked to either $N_{\delta 1}$ or $N_{\epsilon 2}$.

The table containing the list of His−aromatic interactions is given as Supporting Information.

2. Geometrical Parameters Characterizing the Model Imidazole−Aromatic Complexes. Figure 1 illustrates the geometrical parameters used to investigate how interaction energies evolve with the relative positions of the molecular fragments. Stacked and T-shape conformations of the imidazole−benzene complexes are represented in parts (a) and (b) of the figure respectively, and part (c) illustrates a stacked imidazole−adenine complex. Parameters α, β, d and θ have the same meaning in all complexes: α and β are the molecular planes of both interacting partners, d is the distance from the imidazole ring center to the β plane and θ defines the orientation of imidazole within α with respect to benzene or adenine. More details about the parameters are given in the legend of Figure 1 and in the next sections detailing the calculations. Note that all calculations were performed with the geometrical parameters of the fragments frozen to their HF/6-31G(d, p) equilibrium values.

3. Quantum Chemistry Energy Calculations. All energy calculations were carried out using the Gaussian 03 suite of programs.[31] In a first step, the nucleic acid bases and the benzene, phenol, indole and neutral and protonated imidazole rings were optimized separately, using the Hartree−Fock method and the 6-31G(d, p) basis set. No further geometry optimizations of the internal degrees of freedom of the individual partners of the complexes were carried out. When dealing with X-ray structures, the neutral or protonated imidazole and the aromatic moiety were replaced by the individually optimized partners using the U3BEST algorithm.[32]

The gas-phase interaction energies ΔE_{MP2} of the His−aromatic systems were calculated at second-order level of

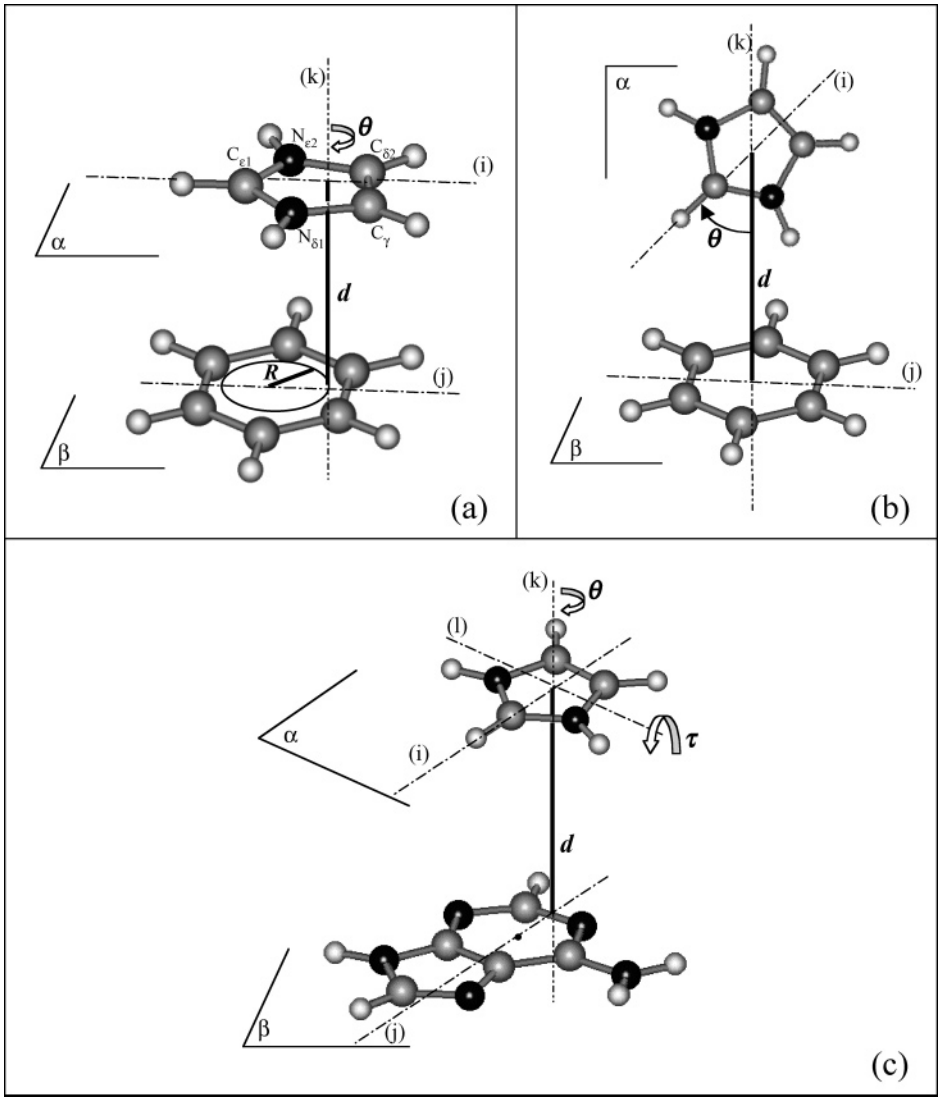


Figure 1. Geometrical parameters defining imidazole−aromatic (benzene or adenine) complexes. Only the protonated form of imidazole is illustrated, but the same parameters hold for the neutral species. According to standard labeling, one defines the nitrogen atoms of imidazole as $N_{\epsilon 2}$ and $N_{\delta 1}$, with carbon $C_{\epsilon 1}$ between them. $N_{\epsilon 2}$ is unprotonated in neutral imidazole. Hydrogens are labeled with a subscript following the heavy atom to which they are bonded ($H_{\epsilon 1}$ bonded to $C_{\epsilon 1}$ for instance). Carbon, nitrogen and hydrogen atoms are drawn in gray, black and white, respectively. The molecular planes of the imidazole and of the aromatic species are α and β respectively. (a) stacked imidazole−benzene complex with $\alpha//\beta$. (b) T-shape imidazole−benzene complex $\alpha\perp\beta$. (c) stacked imidazole−adenine complex. *d* is the distance between β and the center of the imidazole ring. *R* is the radius of a circle in β centered on the benzene ring center and containing the orthogonal projection in β of the imidazole center. θ defines the rotation of (i), the line passing through the $C_{\epsilon 1}-H_{\epsilon 1}$ bond, around the imidazole ring center. Parts a and c show the orientation of imidazole with $\theta = 0°$. τ is the tilt angle, which is equal to zero for a parallel conformation.

Møller−Plesset pertubation theory (MP2),[33] as the sum of the Hartree−Fock (HF) energy contribution ΔE_{HF} and a correlation interaction energy ΔE_{cor}, evaluated as the second-order term of the Moller−Plesset perturbation expansion:

$$\Delta E_{MP2} = \Delta E_{HF} + \Delta E_{cor} \quad (1)$$

The standard counterpoise (CP) method was applied to correct the interaction energies for the basis set superposition error (BSSE).[34] The basis set used is 6-31G(2d (0.8,0.2),p), which corresponds to the standard 6-31G(d,p) basis set where the Gaussian α_d-exponent of the d-polarization functions on the heavy atoms C, N and O is equal to 0.8, with an additional α_d-exponent equal to 0.2. It has indeed been shown that this extended description of the d-polarization functions allows a more accurate description of cation−π energies.[35]

Finally, the interaction free energy of complexes in the presence of a solvent, noted $\Delta\mathbf{G}$, was evaluated as

$$\Delta\mathbf{G} = \Delta E_{MP2} + \Delta\Delta G_{solv} \quad (2)$$

where $\Delta\Delta G_{solv}$ is the solvation free energy contribution of the complex. The latter were estimated using the integral equation formalism version (IEF) of the polarized continuum model (PCM) implemented in the Gaussian 03 program.[36] It is a continuum solvation model in a quantum mechanical framework, where the solvent is mimicked by a polarizable continuum surrounding a cavity having the shape and

Table 1. Vacuum Interaction Energy (in kcal/mol) of the 89 X-ray His−Aromatic Pairs, Calculated at HF and MP2 Levels of Theory, with 6-31G(2d (0.8,0.2), p) Basis Set[a]

		protonated His			neutral His		
	N	ΔE_{HF}	ΔE_{cor}	ΔE_{MP2}	ΔE_{HF}	ΔE_{cor}	ΔE_{MP2}
His−Phe	34	−3.3 ± 2.4 (−8.5)	−3.3 ± 2.1 (−8.5)	−6.7 ± 1.7 (−11.0)	1.6 ± 2.9 (−1.8)	−3.8 ± 2.1 (−9.4)	−2.2 ± 1.4 (−4.7)
His−Tyr	28	−3.4 ± 2.3 (−10.0)	−3.8 ± 1.7 (−6.9)	−7.2 ± 2.5 (−13.9)	2.2 ± 1.4 (0.2)	−4.2 ± 1.7 (−7.6)	−1.9 ± 1.3 (−5.4)
His−Trp	13	−4.7 ±3.5 (−11.2)	−4.0 ± 2.0 (−8.1)	−8.7 ± 3.1 (−13.4)	1.0 ± 1.9 (−1.6)	−4.5 ± 1.9 (−8.9)	−3.4 ± 0.5 (−4.2)
His−Ade	14	−3.5 ± 2.5 (−7.0)	−4.2 ± 2.2 (−8.1)	−7.7 ± 4.0 (−12.0)	0.9 ± 1.8 (−2.0)	−4.6 ± 2.4 (−8.6)	−3.6 ± 0.9 (−5.1)

[a] N denotes the number of occurrences in the sets of proteins and protein−ligand complexes. The mean energy values ± the standard deviation are given, with the minimum energy values in parentheses.

dimension of the solute molecule. The cavity is described by interlocking spheres centered on solute atoms; we used the default values for the atomic radii (UATM), multiplied by a default factor that accounts for the fact that the distance between the solvent and solute atoms is normally somewhat larger than the van der Waals radii.

By use of a PCM, the molecular free energy G of the solute embedded in a continuum medium can be expressed as the sum:

$$G = G_{el} + G_{dis-rep} + G_{cav} + G_{mm} \quad (3)$$

G_{el} collects the components of electrostatic origin, $G_{dis-rep}$ is a term related to the solute−solvent dispersion energy, and G_{cav} is the work needed to form the cavity where the solute must be accommodated. G_{mm} includes all the contributions due to molecular motions. Neglecting the entropic contributions due to molecular motions, the solvation free energy is obtained by subtracting the gas-phase energy E_{gas}

$$\Delta G_{solv} = \Delta G_{el} + G_{nel} \quad (4)$$

with

$$\Delta G_{el} = G_{el} - E_{gas} \text{ and } G_{nel} = G_{dis-rep} + G_{cav}$$

We thus find

$$\Delta\Delta G_{solv} = \Delta\Delta G_{el} + \Delta G_{nel} \quad (5)$$

where $\Delta\Delta G_{el}$ and ΔG_{nel} are the differences of electrostatic and nonelectrostatic contributions between the complex and its isolated partners.

The IEF−PCM calculations were performed at the HF/6-31G(2d (0.8,0.2), p) level, and some tests were performed at the MP2/6-31G(2d (0.8,0.2), p) level. Note that the quantity that emerges from the MP2/PCM calculations is G_{el}. The nonelectrostatic terms must be then added to it.

Results

1. Geometries of Molecular Fragments. To probe the His−aromatic complexes using quantum chemistry energy calculations, Phe, Tyr, Trp and protonated/neutral His were reduced to benzene, phenol, indole, and imidazolium/imidazole, respectively; Ade was kept as such. The geometries of the molecular fragments were optimized at the HF/6-31G(d, p) level of approximation. Both imidazole species were found to be planar, but some in plane geometry changes occur upon protonation. In particular, changes of +0.02 and −0.04 Å are observed for the $C_{\epsilon1}-N_{\epsilon2}$ and $C_{\epsilon1}-N_{\delta1}$ bond lengths respectively (see Figure 1a for the atom labeling), and changes of 1° to 4° for the aromatic ring bond angles.

The natural bond orbital (NBO) atomic charges calculated at the MP2/6-31G(2d (0.8,0.2), p) level of theory for the protonated (neutral) imidazole species are as follows: $C_{\epsilon1}$: 0.26 (0.17); $H_{\epsilon1}$: 0.28 (0.22); $N_{\delta1}$: −0.49 (−0.57); $H_{\delta1}$: 0.49 (0.44); $N_{\epsilon2}$: −0.49 (−0.50); C_{γ}: −0.05 (−0.11); H_{γ}: 0.29 (0.23); $C_{\delta2}$: −0.05 (−0.10); $H_{\delta2}$: 0.29 (0.23); $H_{\epsilon2}$: 0.49. In the case of protonated imidazole, a delocalization of the positive charge is observed over the $C_{\epsilon1}$, $H_{\epsilon1}$, $H_{\epsilon2}$ and $H_{\delta1}$ atoms. These atoms are bearing together a total charge of 0.69, which means that the positive charge of the imidazole cation is globally located in the vicinity of the $C_{\epsilon1}-H_{\epsilon1}$ bond. In the case of neutral imidazole, a partial positive charge of 0.51 is found over the same $C_{\epsilon1}-H_{\epsilon1}$ bond. Hence, when the imidazole is protonated, it can form a cation−π interaction with an aromatic ring, and when neutral, it can form interactions of the type C−H−π[29] or $\delta^+-\pi$.[37−39]

2. His−Aromatic Interactions in X-ray Structures. A total of 75 His−aromatic interactions were identified in the set of protein X-ray structures: 34 His−Phe, 28 His−Tyr and 13 His−Trp. In the set of protein−ligand complexes, 14 His−Ade complexes were found. No His−Gua, His−Cyt and His−Thy were observed, probably because of the infrequency of these nucleic acid bases in ligands. The average angle between the aromatic plane is equal to 42 degrees, reflecting a slight preference for stacked compared to T-shape conformations (see Supporting Information).

The vacuum interaction energy of these complexes, with the separate molecular fragments replaced by the optimized ones (see Methods), were computed at HF and MP2 theory levels with the 6-31G(2d (0.8,0.2), p) basis set. Three species of His were considered: the protonated imidazole cycle (imidazolium) and the two neutral ring forms resulting from the tautomeric equilibrium of His. The results are summarized in Table 1. For the protonated His form, the minimal HF interaction energies were found to be comprised between −7 kcal/mol for His−Ade pairs and −11 kcal/mol for His−Trp. The electronic correlation contributions are of the same order of magnitude: from −7 kcal/mol for His−Tyr up to −9 kcal/mol for His−Phe. The most favorable HF interaction energies, known to be dominated by electrostatic forces, appears thus in the Trp double cycle, whereas the most favorable electronic correlation is found with the benzene ring of Phe, the only aromatic moiety that does not carry heteroatoms.

The relative strengths of the ΔE_{HF} and ΔE_{cor} contributions are different in the protonated and neutral complexes. Indeed,

Table 2. HF/6-31G(2d (0.8,0.2), p) Solvation Free Energies $\Delta\Delta G_{solv}$ and Total Interaction Free Energies ΔG (in kcal/mol) with the Vacuum Contributions Evaluated at the MP2/6-31G(2d (0.8,0.2), p) Level, in Four Solvents Characterized by Their Dielectric Constant ϵ (IEF−PCM Model), for the 89 X-ray His−Aromatic Pairs[a]

solvent ϵ	*His−X*	$\Delta\Delta G_{solv}$	ΔG
water	*F*	8.1 ± 1.7 (4.8)	1.4 ± 1.6 (−1.4)
78.4	*Y*	8.3 ± 2.0 (4.8)	1.2 ± 1.2 (−1.6)
	W	10.2 ± 3.3 (5.2)[b]	1.5 ± 1.2 (−1.3)[b]
	Ade	7.5 ± 2.5 (4.1)	2.9 ± 0.5 (1.9)
acetone	*F*	7.2 ± 1.4 (4.6)	0.6 ± 1.7 (−2.3)
20.7		[7.0 ± 1.3 (4.7)]	
	Y	7.6 ± 1.5 (5.2)	0.5 ± 1.6 (−3.0)
		[7.6 ± 1.4 (5.3)]	
	W	9.2 ± 2.4 (5.2)	0.5 ± 1.7 (−3.4)
		[9.0 ± 2.3 (5.4)]	
	Ade	8.9 ± 2.7 (3.2)	1.2 ± 1.7 (−1.4)
		[8.9 ± 2.6 (3.4)]	
THF	*F*	7.1 ± 1.3 (4.9)	0.5 ± 1.8 (−2.6)
7.6	*Y*	7.6 ± 1.6 (5.0)	0.5 ± 2.5 (−6.4)
	W	9.3 ± 2.9 (5.1)	0.6 ± 1.7 (−3.2)
	Ade	9.0 ± 2.9 (3.4)	1.3 ± 1.6 (−1.0)
CCl_4	*F*	5.5 ± 1.2 (2.7)	−1.2 ± 2.1 (−8.3)
2.2	*Y*	6.0 ± 0.7 (4.4)	−1.2 ± 2.1 (−6.6)
	W	6.5 ± 2.1 (0.6)	−2.3 ± 2.2 (−5.5)
	Ade	7.3 ± 1.9 (3.6)	−0.4 ± 2.5 (−3.6)
gas phase	*F*		−6.7 ± 1.7 (−11.0)
MP2	*Y*		−7.2 ± 2.5 (−13.9)
1.0	*W*		−8.7 ± 3.1 (−13.4)
	Ade		−7.7 ± 4.0 (−12.0)

[a] All His are considered to be protonated. Results for neutral His are given in the Supporting Information. The mean energy values ± the standard deviation are given, with the minimum values in parentheses. For acetone, $\Delta\Delta G_{solv}$ values calculated at MP2 level are given in brackets. The last row displays the gas phase $\Delta G_{gas} = \Delta E_{MP2}$ values. [b] The complex between W131 and H88 in 2gcr was removed from the analysis due to abnormal $\Delta\Delta G_{solv}$ value in water.

the ΔE_{cor} values are roughly the same, while the ΔE_{HF} energies stabilize the protonated species and slightly destabilize the neutral ones. This is not surprising, as the electrostatic contributions almost vanish in the neutral complexes and HF contributions are mainly of electrostatic nature.

For both neutral and charged complexes, the conformations presenting the most favorable electronic correlation interaction energies have a rather stacked geometry, stabilized by significative dispersion energy contributions.

The solvation free energies $\Delta\Delta G_{solv}$ of the complexes immersed in water, acetone, THF, and CCl_4 were estimated using the IEF−PCM method (Table 2 and Supporting Information). In the first stage, the calculations for acetone were performed at both HF and MP2 levels. For the charged complexes, the $\Delta\Delta G_{solv}$ values are found to be almost identical within 0.2 kcal/mol (see Table 2), as previously reported for Arg−Ade complexes in water.[16] The same agreement (within 0.1 kcal/mol) is observed for the neutral form (see Supporting Information). On the basis of this result, and in order to save computer time, we limited further solvent calculations to the HF level, unless otherwise stated. The total interaction free energy values were thus calculated as $\Delta\mathbf{G} = \Delta E_{MP2} + \Delta\Delta G_{solv(HF)}$.

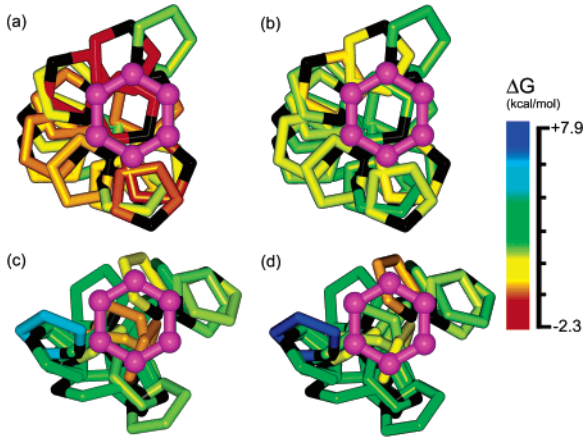


Figure 2. Distribution of the imidazole rings of His residues around the phenyl cycle of Phe amino acids, observed in the X-ray protein set, as a function of their calculated interaction free energy values $\Delta\mathbf{G}$ in acetone. The phenyl groups are superimposed, and all the imidazole rings have been located "under" the Phe planes by suitable projection, to facilitate comparison. The calculated free energy values are colored according to a rainbow ramp where the most favorable conformations are symbolized in red and the less favorable in blue. Carbon atoms of the Phe rings are colored in purple and Cγ atoms of the imidazole cycle are colored in black. The images were generated using Insight II (Accelrys, Inc.) Stacked conformation of the protonated (a) and neutral (b) imidazole rings. T-shape conformations of the protonated (c) and neutral (d) imidazole cycles.

The penalty upon immersing His−aromatic complexes in the solvent increases with the dielectric constant (Table 2), especially from $\epsilon = 2.2$ to 7.6. It is also much larger for the protonated than for the neutral form. Indeed, the $\Delta\Delta G_{solv}$ values for the former are about 4 kcal/mol larger than those for the latter, on the average (see the Supporting Information). As a consequence, the larger stability of protonated His−aromatic complexes in a vacuum relative to neutral ones is reduced by the solvents. Nevertheless, they remain more stable in the four solvents considered. In summary, the interaction free energy $\Delta\mathbf{G}$ increases from gas to water, with minimal values remaining negative except for the Ade−His complexes in water, and remaining slightly more negative for protonated than for neutral form.

We now analyze in more detail the His−Phe and His−Ade complexes by means of X-ray and model structures.

3. Detailed Analysis of His−Phe Interactions. (a) X-ray Structures. The 34 His−Phe X-ray pairs were divided into stacked and T-shape conformations, according to whether the angle between the His and benzene planes is lower or higher than 45°. The stacked and T-shape groups contains 19 and 15 His−Phe pairs, respectively. For each of these two groups, the interaction free energies $\Delta\mathbf{G}$ in acetone as a function of the relative position and orientation of the His and Phe moieties are depicted in Figure 2.

As clearly seen in Figure 2, the stacked conformations have more favorable interaction free energies than T-shape conformations on average, the protonated forms have more

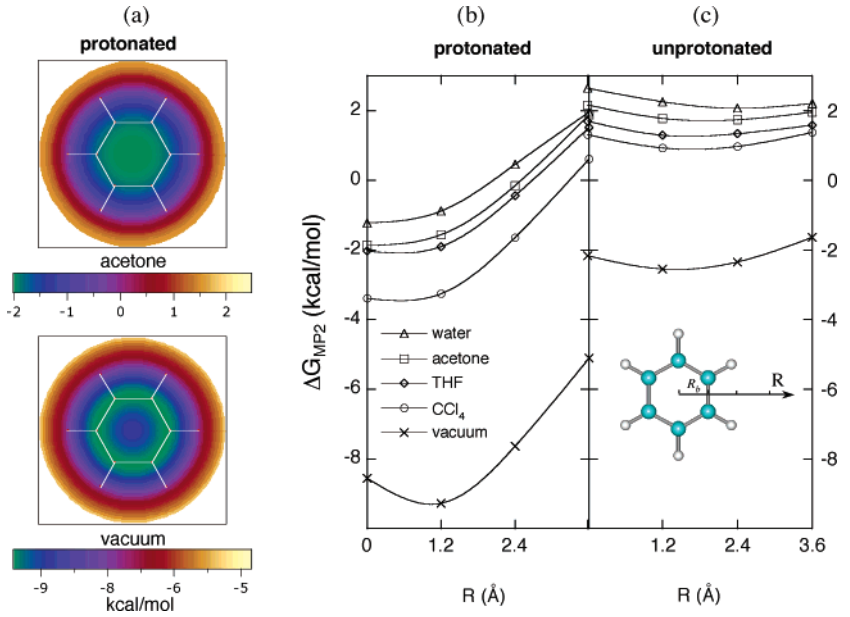


Figure 3. Model structures of the stacked imidazole−benzene complex: maps of $\Delta\mathbf{G}$ in acetone as a function of the position of imidazole in the α plane. All calculations are performed at the MP2/6-31G(2d (0.8,0.2), p) level and with the IEF/PCM approach for solvents. Imidazole is located in a plane parallel to the aromatic ring of benzene at an optimal distance of $d = 3.5$ Å (3.7 Å) for the protonated (neutral) species. The maps presented in panel (a) for the protonated complex in the vacuum and in acetone are obtained by moving the projection of the imidazole center in β on circles of radii R = 0 to 3.6 Å (3 times the radius R_b of the inscribed circle to the benzene hexagon). See Figure 1a for a visual definition of the parameters. Note that the colored scale covers the same energy range in both maps, but the minimum is shifted. Panels (b) and (c) show one-dimensional cuts ΔG(R) in the 2D surfaces (see figure inset).

favorable interaction free energies than the neutral forms, and the most favorable interaction free energies are observed above the benzene ring.

For several His−Phe pairs, the $\Delta\mathbf{G}$ value of the neutral form is equal to or even slightly more favorable than that of the protonated form. Considering that the protein X-ray structures contain a mixture of protonated and neutral His residues, it can be argued that the complexes computed as more favorable in neutral form actually contain a neutral His. This could be a way to determine which His residues in a folded structure are charged and which are not. This possibility should however be considered with care, because of the residual flexibility in folded proteins and the limited resolution of the structures, which do not exclude local inaccuracies in atomic positioning.

(b) Model Conformations: Stacked. To probe in detail the evolution of the interaction free energy as a function of the position and orientation of imidazolium/imidazole with respect to benzene, we analyzed ensembles of model structures in which the molecular planes α and β were kept parallel (perfectly stacked conformations). Three parameters define the relative positions of both species (Figure 1a): the interplane distance d, the radius R of a circle centered on the benzene ring center and containing the orthogonal projection in β of the imidazole center, and the angle θ defining the rotation of the $C_{\epsilon 1}-H_{\epsilon 1}$ bond around the imidazolium/imidazole ring center.

In the first step, we determined the minimum of the vacuum interaction energy as a function of the distance d, by approaching both species while keeping their centers superimposed (R=0). The optimal distances, calculated at the MP2/6-31G(2d (0.8,0.2), p) level of theory, are of 3.5 and 3.7 Å for protonated and neutral imidazoles, respectively.

In the second step, having fixed the interplane distances d to its optimal value, we studied the effect of moving the imidazole with respect to the benzene in α. The projection in β of the imidazole center was displaced on circles of radius R (see Figure 1a), R being varied from 0 to 3 times the radius R_b of the inscribed circle to the benzene hexagon ($R_b = 1.2$ Å). The energetic effect of the imidazole orientation parameter θ was investigated at different relative positions of the fragments, but its effect on the interaction energies was found to be small ($\Delta E < 0.5$ kcal/mol), and the preferred orientation of $\theta = 0°$ was adopted in all further calculations.

At this stage, we added the solvation contributions in water, acetone, THF or CCl_4 and evaluated the evolution of the interaction free energy $\Delta\mathbf{G}$ as a function of the position of the stacked imidazole with respect to benzene, when the imidazole moves away from the benzene in a linear displacement parallel to the benzene plane (see Figure 3).

In agreement with the results on the X-ray complexes, the protonated complex is found significantly more stable than the neutral one (see Figure 3). This can be explained by the enhancement of the interaction induced by the positive charge carried by imidazole. The difference in interaction energy between protonated and neutral species thus provides a direct measure of this charge contribution. It amounts to −7 kcal/mol in a vacuum and is significantly weakened in solution: −4 kcal/mol in CCl_4 and −3 kcal/mol in THF, acetone and water, due to the screening of the electrostatic interactions by the solvent. Another finding is the existence of pronounced minima over the benzene ring in the protonated case,

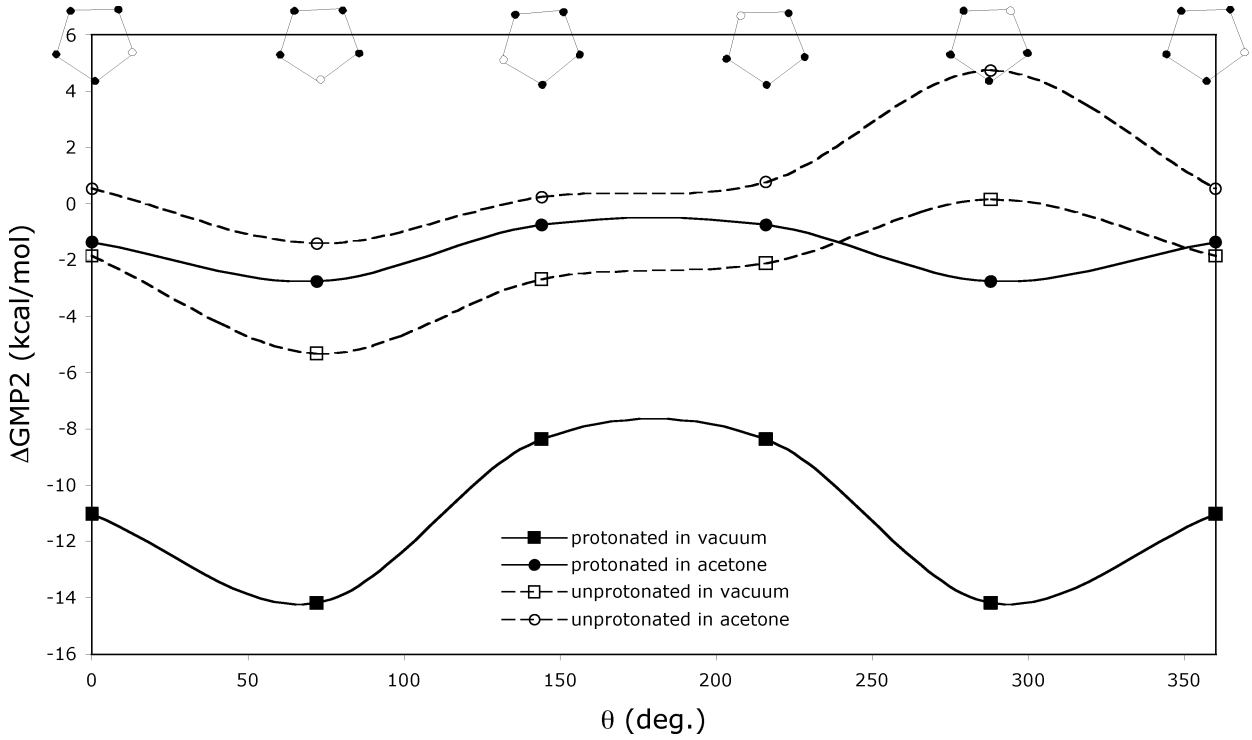


Figure 4. Model conformations of T-shape imidazole−benzene complex: evolution of $\Delta\mathbf{G}$ in acetone as a function of angle θ (see Figure 1b). All calculations are performed at the MP2/6-31G(2d (0.8,0.2), p) level and with the IEF/PCM approach for representing the acetone solvent environment. Drawings of the imidazole kernel (hydrogens are omitted) show its orientation in the T-shape structure, the lowest atom pointing toward the benzene ring. Color code for atoms is gray for carbon, white for protonated nitrogen and black for unprotonated nitrogen.

while very flat profiles characterize the neutral complex. One can again invoke the cation$-\pi$ interaction, when the positively charged imidazolium approaches the π-electron benzene cloud, inducing a deepening of the potential well.

The (free) energy maps in a vacuum and acetone, shown in Figure 3, exhibit a central symmetry in which the 6-fold symmetry of the benzene electronic structure does not appear, as a consequence of the aromatic π-electron delocalization. The energy changes along a given circle are indeed negligible (<0.02 kcal/mol). A comparison of both maps shows that the effect of the solvent on the shape of the potential energy surface is minor. Indeed, the two maps exhibit a flat minimum covering all the benzene ring and a destabilization of the same amplitude when the imidazole moves away from the benzene ring (R > 1.2 Å). However, the vacuum energy surface is shifted by about +7 kcal/mol in the presence of the solvent. Furthermore, the existence of a very shallow minimum at a value of R=R_b distinguishes the gas-phase map from the solvated one. This minimum corresponds to a position of the imidazole center just above the C−C bonds of the benzene. The positively charged atoms of imidazole ($C_{\epsilon 1}$, $H_{\epsilon 1}$, $H_{\epsilon 2}$ and $H_{\delta 1}$) are then located above the aromatic ring of the benzene, thus in a favorable position for cation$-\pi$ interactions. Similarly shifted optimal conformations have already been reported in benzene[40,41] and naphthalene dimers.[42]

(c) Model Conformations: T-Shape. In the case of the perpendicular conformation, two parameters, illustrated in Figure 1b, were explored: the distance d between the β plane and the center of the imidazole ring and the angle θ defining the orientation of imidazole within the α plane perpendicular to β.

Optimal values of the parameters were obtained in a vacuum at the MP2 level of theory, as in the case of the stacked conformation. The following three step procedure has been adopted: i) optimization of θ for a d distance fixed to 5 Å; ii) optimization of d at the previous optimal θ-value; and iii) reoptimization of θ at the previous d-value. The optimal θ-value is found to be the same ($\theta \sim 72°$) in the first and last step, indicating the decoupling of the parameters d and θ. The optimal d values are equal to 4.2 and 4.5 Å for the protonated and neutral complexes, respectively. The interaction (free) energy profiles in a vacuum and acetone as a function of θ are shown in Figure 4.

Let us first remark that all the interaction (free) energy curves present a minimum when the N−H bond points toward the π-electron cloud of the benzene, leading thus to a couple of symmetric minima ($\theta \sim 72°$ and 288°) for the protonated species and a single minimum ($\theta \sim 72°$) for the neutral one. At these minima, the MP2 binding energies reach −14 and −5 kcal/mol in a vacuum for the protonated and neutral complexes, respectively. The corresponding values are −6 and −2 kcal/mol respectively in CCl_4, −3 and −1 kcal/mol in THF, −3 and −1 kcal/mol in acetone and 0.8 and 0.9 kcal/mol in water. The contribution of the positive charge to the complex stability, obtained from differences between the above cited energies for protonated and neutral complexes, are as follows: −9 kcal/mol in a vacuum, −4 kcal/mol in CCl_4, −2 kcal/mol in THF, −2 kcal/mol in

Table 3. Interaction Energies (in kcal/mol) for the Modeled Benzene−Imidazole Complexes in Protonated and Neutral Forms[a]

solvent	stacked				T-shape			
	protonated		neutral		protonated		neutral	
ϵ	ΔG	$\Delta\Delta G_{solv}$	ΔG	$\Delta\Delta G_{solv}$	ΔG	$\Delta\Delta G_{solv}$	ΔG	$\Delta\Delta G_{solv}$
water	−0.9	8.4	2.3	4.8	0.8	15.0	0.9	6.2
78.4	(4.9)	(8.9)	(7.5)	(5.1)	(6.0)	(15.3)	(5.3)	(6.3)
	[−5.8]	[−0.5]	[−5.2]	[−0.3]	[−5.2]	[−0.3]	[−4.4]	[−0.1]
acetone	−1.9	7.4	1.3	3.8	−2.8	11.4	−1.4	3.9
20.7	(3.6)	(7.6)	(6.4)	(4.0)	(2.3)	(11.6)	(2.9)	(3.9)
	[−5.5]	[−0.2]	[−5.1]	[−0.2]	[−5.1]	[−0.2]	[−4.3]	[0.0]
THF	−1.6	7.7	1.8	4.3	−3.2	11.0	−1.2	4.1
7.6	(3.9)	(7.9)	(6.9)	(4.5)	(1.9)	(11.2)	(3.1)	(4.1)
	[−5.5]	[−0.2]	[−5.1]	[−0.1]	[−5.1]	[−0.2]	[−4.3]	[0.0]
CCl_4	−3.3	6.0	0.9	3.5	−6.5	7.7	−2.2	3.1
2.2	(2.1)	(6.1)	(5.9)	(3.5)	(−1.5)	(7.8)	(2.1)	(3.1)
	[−5.4]	[−0.1]	[−5.0]	[0.0]	[−5.0]	[−0.1]	[−4.3]	[0.0]
gas phase	−9.3		−2.5		−14.2		−5.3	
1.0	(−4.0)		(2.4)		(−9.3)		(−1.0)	
	[−5.3]		[−4.9]		[−4.9]		[−4.4]	

[a] Results obtained at the MP2/6-31G(2d (0.8,0.2), p) level are given for the gas phase and in solvents of increasing ϵ dielectric constants (IEF/PCM model). Values in parentheses are from HF/6-31G(2d (0.8,0.2), p) calculations. Values in brackets are the corresponding MP2 correlation energy contributions. ΔG values are corrected for the BSSE calculated at the corresponding level of theory.

acetone and −0.1 kcal/mol in water. These values confirm the trends already discussed for the stacked complex. Eventually note the striking parallelism between the gas phase and acetone energy curves.

(d) Comparison of Stacked and T-Shape Complexes. Table 3 collects the HF and MP2 interaction (free) energies in a vacuum and in the different solvents for the most stable stacked and T-shape conformations of the model structures. The results show the importance of the correlation energy contributions to the stabilities of both the stacked and T-shape complexes, for both the protonated and neutral species, in agreement with the results obtained with X-ray structures. The balance between electrostatic and dispersion energy contributions differs however. Indeed, the correlation effect is slightly larger (∼0.5 kcal/mol) for the stacked than the T-shape geometries and is also larger (∼0.8 kcal/mol) for the protonated than for the neutral species.

The calculations seem to indicate that the perpendicular arrangement of imidazole−benzene complexes is energetically favored as well in a vacuum as in a condensed phase, except for water. This result seems in contradiction with those obtained on the basis of the X-ray structures, where the stacked conformations displayed less favorable interaction energies in a vacuum but more favorable interaction free energies in the solvents, on the average. Note that the $\Delta\mathbf{G}$ value of the most favorable of all stacked X-ray structures is lower that that of all stacked model structures; it is however higher than that of the most favorable T-shape model structure.

These results lead us to conclude that the restriction to model or X-ray structures defines two different, but incomplete, samplings of the conformational space, the first because of the neglect of degrees of freedom and the second due to the limited number of "randomly" chosen structures. The latter point also suggests that the T-shaped free energy minima are more narrow than the stacked ones, thereby explaining the higher average interaction free energies of the random (X-ray) structures.

Another point is the minor importance of the electron correlation effect in the IEF−PCM treatment (<0.5 kcal/mol). Nevertheless, this effect increases systematically with ϵ (see values in brackets), which means that HF IEF−PCM calculations tend to overestimate the solvation free energies. Recalling that $\Delta\Delta G_{solv}$ is the sum of an electrostatic $\Delta\Delta G_{el}$ and nonelectrostatic ΔG_{nel} contribution (see Methods), we would also like to stress that the relative magnitude of these contributions is far from being the same for the different complexes. Taking results in acetone as an example, the following values (in kcal/mol) of ($\Delta\Delta G_{el}$, ΔG_{nel}) are found: (9.2, 2.2), (4.7, 2.7), (1.3, 2.6) and (0.7, 3.1) for the protonated T-shape, protonated stacked, neutral stacked and neutral T-shape complexes, respectively. The corresponding ratio $\Delta\Delta G_{el}/\Delta G_{nel}$ thus exhibits the following drastic changes: 4.2, 1.7, 0.2 and 0.5, respectively. These changes essentially arise from $\Delta\Delta G_{el}$ and reflect the strength of the screening of the solute electrostatic interactions by the solvent.

4. Detailed Analysis of His−Ade Interactions. (a) X-ray Structures. Among the 14 His−Ade X-ray pairs, 6 are rather T-shaped and 8 stacked. The positioning of the imidazole moieties above the Ade plane are depicted in Figure 5 and colored according to their interaction free energies $\Delta\mathbf{G}$ in acetone. As in the case of His−Phe, the $\Delta\mathbf{G}$ values are on the average more favorable for the complexes with protonated than with neutral His. Moreover, the imidazole is more favorable above the 6-cycle than above the 5-cycle. Note also that the minimal values of ΔG_{HF}, ΔG_{cor} and $\Delta\mathbf{G}$ are slightly less favorable for the His−Ade pairs than for His−Phe (Tables 1 and 2), but the number of His−Ade pairs in our set is too low to allow firm conclusions.

(b) Model Structures. We focused on stacked His−Ade conformations, slightly more frequent among the X-ray

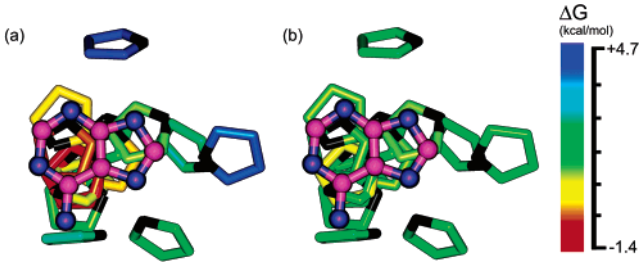


Figure 5. Distribution of the imidazole rings of His residues around the Ade ring systems, observed in the protein−ligand set, as a function of their calculated free energy values in acetone. The Ade bases are superimposed and the imidazole rings have been located "under" the Ade planes, for a better comparison, by projection. The calculated free energy values are colored according to a rainbow ramp where the most favorable conformations are symbolized in red and the less favorable in blue. Carbon and nitrogen atoms of the Ade rings are colored in purple and blue respectively, and Cγ atoms of the imidazole cycle are colored in black. The images were generated using Insight II (Accelrys, Inc.) Structures of the protonated (a) and neutral (b) imidazole rings.

structures. The relative positions of both species were specified by three parameters, the distance d between the β plane and the center of imidazole ring, the orientation angle θ defining the rotation of the $C_{\epsilon 1}-H_{\epsilon 1}$ bond around the imidazole center, and the tilt angle τ defining the rotation of this bond around the perpendicular line passing through the imidazole center (Figure 1c). On the basis of the protein−ligand X-ray structures, we chose to fix d=3.5 Å and explored the other two parameters.

In the first step, we kept both species parallel (τ=0°) and studied the effect of moving the imidazolium parallel with respect to Ade. To explore the entire plane above the Ade system, the orthogonal projection of the imidazole center in the Ade plane was placed at 12 positions, corresponding to the centers of the two Ade rings and to the middles of all bonds constituting these rings. At each position, the imidazole orientation parameter θ was optimized at the vacuum MP2/6-31G(2d (0.8,0.2), p) level of theory. All interaction energies ΔE_{MP2} were found to be comprised between −7 and −12 kcal/mol. Whereas ΔE_{MP2} is rather insensitive to θ (0.9 kcal/mol on the average), as found for the imidazole−benzene complex, it strongly depends on the position of the imidazole center. According to whether the imidazole center is situated above the 5- or 6-cycle of Ade, ΔE_{MP2} is in the range of −7 to −10 kcal/mol or −10 to −12 kcal/mol, respectively. The minimum ΔE_{MP2} corresponds to a position of the imidazole center above the bond between the two cycles. At this position, ΔE_{MP2} reaches −12 kcal/mol for the protonated complex.

In a second step, having fixed the orientation angle θ to its position-dependent optimal value, we have studied the effect of tilting the imidazole. Therefore, we varied the τ angle from 0° to ± 45°, the positive values of τ corresponding to the situations where the positive charge carried by the His approaches the Ade ring (Figure 1c). The calculated ΔE_{MP2} values show that, except for one position, the protonated complex is more stable when τ is comprised between +15 and +30°, which can be explained by the enhancement of the cation−π interaction induced by the approach of the positively charged atoms. One observes that this stabilizing effect is on average 3 times larger when the projection of the imidazole center is situated on the 5-cycle, reaching then −1.8 kcal/mol.

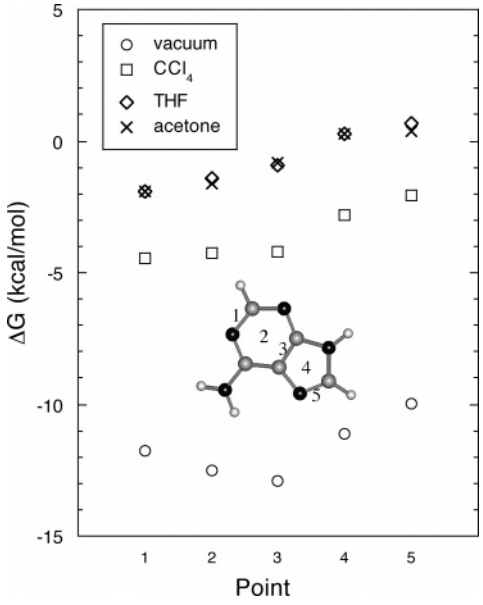


Figure 6. Stacked imidazole−adenine complex: evolution of ΔG in various solvents as a function of the position of imidazole with respect to adenine. The five points considered (see inset of the figure) correspond to the projection of the imidazole center in the adenine aromatic plane α. Parameters θ and τ (see Figure 1c) defining the orientation of imidazole in α and the tilted orientation of α have been optimized at each point.

At this stage, we evaluated the contributions of solvation to the interaction energies by using the IEF−PCM model. Figure 6 shows the interaction energies ΔE_{MP2} and the interaction free energies $\Delta\mathbf{G}$ in CCl_4, THF, acetone and water, when the orthogonal projection of the imidazole center into the aromatic plane is situated at some characteristic positions located on a line across the two aromatic ring centers. At each position, the preferred parameters of θ and τ were adopted. The positions with the lowest free energies are located in the 6-cycle. The interaction (free) energies become less favorable when moving away above the 5-cycle. The existence of a pronounced minimum above the bond common to the 2 cycles distinguishes the gas phase and solvated energy profiles. At this minimum, ΔE_{MP2} reaches −13 kcal/mol. The corresponding $\Delta\mathbf{G}$ values are −4. kcal/mol in CCl_4, −1 kcal/mol in THF and acetone and 1 kcal/mol in water. The contribution of the positive charge to the stability of the complex are −10 kcal/mol in a vacuum, −5 kcal/mol in CCl_4, −2. kcal/mol in THF and acetone, and −0.5 kcal/mol in water.

The correlation effects contribute significantly to the stability of the complex, like for His−Phe. The ΔE_{cor} value in the vacuum amounts to −7 and −9 kcal/mol for points 1 and 3 respectively, which corresponds to 56 and 68% of the gas phase $\Delta\mathbf{E}$ values, respectively (−12 and −13 kcal/mol).

The excess of correlation at point 3 clearly induces the observed gas-phase minimum. The large values of the solvation energies (10 and 12 kcal/mol for points 1 and 3 in acetone) compensate however the correlation energy contributions, with as a result a shift of the minimum from points 3 to 1. As in the case of the benzene, one finds that the balance between the $\Delta\Delta G_{el}$ and ΔG_{nel} solvation terms follow the changes occurring in the electronic structure of the solute. At positions 1 and 3 ($\Delta\Delta G_{el}$, ΔG_{nel}) are (in kcal/mol) (8.6, 1.7) and (10.7, 1.4) respectively, the corresponding ratio $\Delta\Delta G_{el}/\Delta G_{nel}$ being 5 and 8.

Note also the systematic overestimation of the solvation energies by HF IEF−PCM calculations (0.2 and 0.5 kcal/mol at point 1 and 3 respectively), as in the Phe−His case. The larger discrepancy for point 3 probably reflects the enhancement of the correlation effect pointed out above for this geometry

5. Biological Role of His−Aromatic Interactions. There are several arguments in favor of the biological role of His−aromatic interactions in proteins. First, these interactions are often found in catalytic sites. This is the case, among the proteins of our data set, for His48-Tyr52 in phospholipase A2 (4p2p),[43] His54-Trp16 in D-xylose isomerase (7xia),[44] His195-Tyr25 in chloramphenicol acetyltransferase (3cla),[45] and His57-Phe54 in serine proteinases (2apl and 2sga).[46] All these His−aromatic pairs were determined as essential to the enzymatic reactions, and the aromatic partner was proven to participate indirectly in the stabilization of the catalytic site.[43,47,48] This suggests that His−aromatic interactions are key components of the catalytic mechanism of these enzymes, either in providing critical constraints on the precise conformation of the catalytic site or in participating in transition-state stabilization.

Second, His−aromatic cation−π interactions are recurrently observed at crucial positions in 3D domain swapping proteins.[49] Indeed, using the data set of domain swapping proteins of Dehouck et al.,[49] we find several His−π pairs located in the hinge loop of either the closed or swapped form, for example, His116-Tyr102 in the closed form of Spo0A-N (1qmp) and His384-Phe530 in swapped diphtheria toxin (1ddt). In other proteins, the same interaction is observed in the two forms, involving residues that are part of the same chain in the closed form but of two different chains in the swapped form and connecting the chain pieces across the closed interface. Such interactions seem thus to zip up the closed interface of domain swapped oligomers. Examples of this type are His86-Tyr42 in cystatine C (1g96), His90-Trp49 in cyanovirin-N (2ezm, 3ezm) and His18-Trp94 in barnase (1brn, 1yvs). We computed the free energy in acetone for the His18-Trp94 pair of the closed form of barnase. As usual, the protonated form appears more stable ($\Delta\mathbf{G}=-1.6$ kcal/mol in acetone) than the two neutral forms ($\Delta\mathbf{G}=+0.9$ and $+1.7$ kcal/mol). So, a positive charge on the His ring stabilizes barnase by about 2.5 to 3.3 kcal/mol, in agreement, though slightly overestimated, with the experimental values of 1 to 2 kcal/mol.[50,51]

Hence, because of their frequent occurrence at specifically important positions for 3D domain swapping and their stabilizing properties, His−aromatic interactions can be expected to contribute in a pH-dependent way to the stability of the closed and/or swapped forms and to modulate the conversion between the two forms.

Third, at the interface between protein and DNA, stair motifs involving simultaneously cation−π, H-bond and base stacking interactions are often observed.[52] They consist of two successive bases along the DNA stack, one in cation−π interaction with an amino acid side chain that carries a net or partial positive charge, and the other H-bonded with the same side chain. In some proteins, like zinc fingers, several successive stair motifs are observed. When His is considered as potential cation/δ^+−π partner in addition to Arg, Lys, Asn and Gln, the number of succesive stairs is sometimes impressive. In the zinc finger 1ag1, in particular, we find the multiple stair: Thy(B5)$\therefore$His(A149)$\vee$Gua(B6)$\therefore$Arg(A146)$\vee$Gua(B7)$\therefore$Arg(A124)$\vee$Gua(B8)$\therefore$Asn(A121)$\vee$Cyt(B9), where $\therefore$ denotes cation−π and $\vee$ H-bond. The exact structural or functional role of these motifs is not known, but their recurrent occurrence suggests that they play a role in protein−DNA recognition.

Conclusions

Our quantum chemistry analysis of both X-ray and model structures of His−aromatic interactions concur to indicate that the positively charged complex is more stable than the neutral one by about -6 to -9 kcal/mol in gas phase, -4 to -5 kcal/mol in CCl_4, -2 to -3 kcal/mol in THF, -1 to -3 kcal/mol in acetone and 0 to -2 kcal/mol in water (Tables 1 and 3 and Supporting Information). Our results are thus in agreement with the experimental finding that in protein environments, best mimicked by THF and acetone, the protonation of His has a stabilizing effect of about 1 to 2 kcal/mol.[18]

The substantial difference in interaction free energy between the protonated and neutral forms of His−aromatic complexes, and the property of His to become protonated near physiological conditions, makes this type of interactions likely to be involved in the biological function. This hypothesis is supported by the frequent occurrence of His−aromatic interactions in 3D domain swapping proteins, at protein−DNA interfaces, and among catalytic residues.

Our results do not allow to clearly settle whether His−aromatic interactions in protein environments are more favorable in T-shape or stacked arrangement. Both conformations are roughly equally frequently observed in X-ray structures, and their computed relative stabilities are similar, with some differences according to the precise geometry, the type of partners and the solvent. However, some trends can be distinguished. In particular, T-shaped structures are much more stable in gas phase, and this advantage is reduced in the solvents considered. In water, the stacked conformation appears even slightly more favorable. We can thus tentatively conclude that T-shaped conformations are more favorable in gas phase and nonpolar solvents and that the stacked conformations become more favorable as the polarity of the solvent increases, in agreement with molecular dynamics studies[53] and the more frequent observation of stacked conformations at the protein surface. Finally note that the

favorable nature of stacked conformations in protein environments must not be related with the higher accessibility of H-bond forming groups as sometimes stated,[54] at least for His−aromatic complexes, since the two groups ($N_{\delta 1}$ and $N_{\epsilon 2}$) in the His cycle that are able to form H-bonds are roughly evenly accessible in the stacked and T-shape conformations. We verified indeed that the number of His involving H-bonds in the complexes studied are independent of their conformation.

Acknowledgment. The Communauté Française de Belgique (Action de Recherche Concertée), the Cost P9 Action, the European Community and The Belgian National Fund for Scientific Research are acknowledged for support. Eric Buisine is thanked for stimulating discussions.

Supporting Information Available: The set of 89 His−aromatic interactions identified (Table S1) and solvation free energies in acetone for the 89 X-ray His−aromatic cation−π pairs (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dougherty, D. A. *Science* **1996**, *271*, 163−168.

(2) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303−1324.

(3) Zacharias, N.; Dougherty, D. A. *Trends Pharmacol. Sci.* **2002**, *23*, 281−287.

(4) Burley, S. K.; Petsko, G. A. *FEBS Lett.* **1986**, *203*, 139−143.

(5) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709−717.

(6) Gallivan, J. P.; Doughert, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459−9464.

(7) Minoux H.; Chipot C. *J. Am. Chem. Soc.* **1999**, *121*, 10366−10372.

(8) Gromiha, M. M.; Thomas, S.; Santhosh, C. *Prep. Biochem. Biotechnol.* **2002**, *32*, 355−362.

(9) Gromiha, M. M. *Biophys. Chem.* **2003**, *103*, 251−258.

(10) Tsou, L. K.; Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2002**, *124*, 14917−14921.

(11) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12088−12093.

(12) Peterson, E. J.; Choi, A.; Dahan, D. S.; Lester, H. A.; Dougherty, D. A. *J. Am. Chem. Soc.* **2002**, *124*, 12662−12663.

(13) Wintjens, R.; Liévin, J.; Rooman, M.; Buisine, E. *J. Mol. Biol.* **2000**, *302*, 395−410.

(14) Biot, C.; Wintjens, R.; Rooman, M. *J. Am. Chem. Soc.* **2004**, *126*, 6220−6221.

(15) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40816−40822.

(16) Biot, C.; Buisine, E.; Rooman, M. *J. Am. Chem. Soc.* **2003**, *125*, 13988−13994.

(17) Wouters, J. *Protein Sci.* **1998**, *7*, 2472−2475.

(18) Loewenthal, R.; Sancho, J.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 759−770.

(19) Fernández-Recio, J.; Vazquez, A.; Civera, C.; Sevilla, P.; Sancho J. *J. Mol. Biol.* **1997**, *267*, 184−197.

(20) Fernández-Recio, J.; Romero, A.; Sanjo, J. *J. Mol. Biol.* **1999**, 290, 319−330.

(21) Feinberg, H.; Torgersen, D.; Drickamer, K.; Weis, W. I. *J. Biol. Chem.* **2000**, *275*, 35176−35184.

(22) Williams, S.; Bledsoe, R. K.; Collins, J. L.; Boggs, S.; Lambert, M. H.; Miller, A. B.; Moore, J.; McKee, D. D.; Moore, L.; Nichols, J.; Parks, D.; Watson, M.; Wisely, B.; Willson, T. M. *J. Biol. Chem.* **2003**, *278*, 27138−27143.

(23) Li, H. L.; Galue, A.; Meadows L.; Ragsdale D. S. *Mol. Pharmacol.* **1999**, *55*, 134−141.

(24) Okada, A.; Miura, T.; Takeuchi, H. *Biochemistry* **2001**, *40*, 6053−6060.

(25) Takeuchi, H.; Okada, A.; Miura, T. *FEBS Lett.* **2003**, *552*, 35−38.

(26) Inoue, Y.; Nakamura, N.; Inagami, T. *J. Hypertens.* **1997**, *15*, 703−714.

(27) Chen, X. Z.; Steel, A.; Hediger, M. A. *Biochem. Biophys. Res. Commun.* **2000**, *272*, 726−730.

(28) Argiriadi, M. A.; Morisseau, C.; Hammock, B. D.; Christianson, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10637−10642.

(29) Spiwok, V.; Lipovova, P.; Skalova, T.; Buchtelova, E.; Hasek, J.; Kralova, B. *Carbohydr. Res.* **2004**, *339*, 2275−2280.

(30) Wintjens, R.; Rooman, M.; Wodak, S. J. *J. Mol. Biol.* **1996**, *255*, 235−253.

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

(32) Kabsch W. *Acta Crystallogr. A* **1978**, *34*, 827−828.

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

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

(35) Wintjens, R.; Biot, C.; Rooman, M.; Liévin, J. *J. Phys. Chem. A* **2003**, *107*, 6249−6258.

(36) Tomasi, J.; Mennucci, B.; Cancès, E. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 211−226.

(37) Meyer E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 1210−1250.

(38) Meyer E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 4120.

(39) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J. Mol. Biol.* **1984**, *180*, 301−329.

(40) Hunter, C. A. *Chem. Soc. Rev.* Meldola Lecture, **1994**, 101−109.

(41) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104−112.

(42) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *J. Chem. Phys.* **2004**, *120*, 647−659.

(43) Soares, A. M.; Giglio, J. R. *Toxicon* **2003**, *42*, 855−868.

(44) Whitaker, R. D.; Cho, Y.; Cha, J.; Carrell, H. L.; Glusker, J. P.; Karplus, P. A.; Batt, C. A. *J. Biol. Chem.* **1995**, *270*, 22895−22906.

(45) Lewendon, M. R.; Murray, I. A.; Shaw, W. V. *Biochemistry* **1994**, *33*, 1944−1950.

(46) Fuhrmann, C. N.; Kelch, B. A.; Ota, N.; Agard, D. A. *J. Mol. Biol.* **2004**, *338*, 999−1013.

(47) Lambeir, A. M.; Lauwereys, M.; Stanssens, P.; Mrabet, N. T.; Snauwaert, J.; van Tilbeurgh, H.; Matthyssens, G.; Lasters, I.; De Maeyer, M.; Wodak, S. J.; Jenkins, J.; Chiadmi, M.; Janin, J. *Biochemistry* **1992**, *31*, 5459−5466.

(48) Murray, I. A.; Lewendon, A.; Shaw, W. V. *J. Biol. Chem.* **1991**, *266*, 11695−11698.

(49) Dehouck, Y.; Biot, C.; Gilis, D.; Kwasigroch, J. M.; Rooman, M. *J. Mol. Biol.* **2003**, *330*, 1215−1225.

(50) Sali, D.; Bycroft, M.; Fersht, A. R. *Nature* **1988**, *335*, 740−743.

(51) Loewenthal, R.; Sancho, J.; Fersht, A. R. *Biochemistry* **1991**, *30*, 6775−6779.

(52) Rooman, M.; Liévin, J.; Buisine, E.; Wintjens, R. *J. Mol. Biol.* **2002**, *319*, 67−76.

(53) Gervasio, F. L., Chelli, R., Marchi, M., Procacci, P., Scettino, V. *J. Phys. Chem. B* **2001**, *105*, 7835−7846.

(54) Mitchell, J. B.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315−331.

CT049875K

JCTC Journal of Chemical Theory and Computation

Optimized Radii for Poisson−Boltzmann Calculations with the AMBER Force Field

Jessica M. J. Swanson,*[,†,‡,§] Stewart A. Adcock,[†,‡,§] and J. Andrew McCammon[†,‡,§,||]

Howard Hughes Medical Institute, Center for Theoretical Biological Physics, Department of Chemistry and Biochemistry, and Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365

Received December 29, 2004

Abstract: Implicit solvent models are a standard tool for assessing the electrostatics of biomolecular systems. The accuracy of quantitative predictions, such as pK_a values, transfer free energies, binding energies, and solvation forces, is strongly dependent on one's choice of continuum parameters: the solute charges, dielectric coefficient, and radii, which define the dielectric boundary. To ensure quantitative accuracy, these parameters can be benchmarked against explicit solvent simulations. Here we present two sets of optimized radii to define either abrupt or cubic-spline smoothed dielectric boundaries in Poisson−Boltzmann calculations of protein systems with AMBER (parm99) charges. Spline smoothing stabilizes the electrostatic potential at the molecular surface, allowing for continuum force calculations. Most implementations, however, require significantly different radii than the abrupt boundary surfaces. The optimal continuum radii are initially approximated from the solvent radial charge distribution surrounding each atom type. A genetic algorithm is then used to fine-tune the starting values to reproduce charging free energies measured from explicit solvent simulations. The optimized radii are tested on four protein-like polypeptides. The results show increased accuracy of molecular solvation energies and atomic forces relative to commonly used continuum parameter sets. These radii are suitable for Poisson−Boltzmann calculations with the AMBER force field and offer energetic congruence to any model that combines molecular mechanics and Poisson−Boltzmann solvation energies.

I. Introduction

Continuum solvation models are a useful link between the microscopic and macroscopic realms of theoretical biochemistry and are widely used to evaluate the electrostatics of biomolecular systems. One of the most common models solves the Poisson Equation (PE)[1] or the Poisson−Boltzmann Equation (PBE)[2,3] as a function of the solute's charge density, a spatially dependent dielectric coefficient, and, in the case of the PBE, ion concentration. The resulting electrostatic potential can be used, for example, to predict electrostatic complementarity of molecular surfaces, to estimate electrostatic binding and transfer free energies, to approximate pK_a values, and to supply electrostatic forces for use in molecular dynamics.

The accuracy of continuum models is dependent on the parameters that are used to define the solute charges, the solvent and solute dielectric coefficients, and the atomic radii which define the dielectric boundary. It has been shown that the average solvent charge density in the continuum model is a sharply peaked function in the region of dielectric discontinuity.[4] The solvent charge density gives rise to the solvent reaction field. Thus, quantitative results are especially sensitive to the location of the solvent−solute dielectric

* Corresponding author phone: (858)822-2771; fax: (858)534-4974; e-mail: jswanson@ucsd.edu.
† Howard Hughes Medical Institute.
‡ Center for Theoretical Biological Physics.
§ Department of Chemistry and Biochemistry.
|| Department of Pharmacology.

10.1021/ct049834o CCC: $30.25
Published on Web 03/31/2005

boundary, i.e., the molecular surface. To appreciate where the dielectric boundary should be, it is helpful to relate its macroscopic and microscopic descriptions. Early studies comparing atomistic simulations of aqueous solutions of monatomic ions to the Born model were helpful in elucidating this connection.[5−7] These studies revealed significant differences in the solvent structure surrounding anions and cations with the same ionic radius but different solvation energies. The Born model correctly predicts disparate anionic and cationic solvation energies if, and only if, the radii used to define the continuum dielectric boundary correspond with the first peak in the microscopic solvent density. Thus, the optimal Born radii are not the ionic radii, a function of the ion alone, but those that trace out solvent excluded cavities, which are a function of both the ion and the surrounding solvent structure.

Optimal biomolecular radii could, presumably, be measured in a similar manner. Biomolecules are, however, more complicated because the solvent structure surrounding each atom is influenced by neighboring atoms and the solute's conformation. Continuum radii for biomolecules must, therefore, balance accuracy for a given conformation with robustness across multiple conformations. Though the solvent structure can still be used to approximate the location of the dielectric boundary, the radii that define that boundary must ultimately be benchmarked against quantitative observables such as solvation energies or forces. Several continuum parameter sets, consisting of solute charges and radii that complement a specified protein dielectric constant and molecular surface definition, have been optimized to reproduce either experimental solvation energies[8,9] or explicit solvent simulations.[10,11] These two methods will be considered in turn.

Referencing continuum models against experimentally determined solvation free energies is an appealing approach because there are many small molecule data to draw upon and because computational models should ultimately be benchmarked by experiment. There are, however, two disadvantages to this approach. First, it necessitates estimating the nonpolar contribution to the solvation free energy since electrostatic and nonelectrostatic contributions cannot be clearly distinguished in experiment. Solvation energies, rigorously defined as the reversible work involved in transferring a molecule from gas phase to bulk solvent, are typically separated into electrostatic and nonelectrostatic contributions via a thermodynamic cycle. This cycle involves cavity formation, the introduction of solute−solvent vdW interactions, and the introduction of solute−solvent electrostatic interactions; $\Delta G_{solv} = \Delta G_{cavity} + \Delta G_{vdW} + \Delta G_{elec}$. PB calculations only account for the electrostatic contribution, ΔG_{elec}. The remaining nonpolar contribution is commonly approximated by the Solvent Accessible Surface Area (SASA) model, $\Delta G_{np} = \Delta G_{cavity} + \Delta G_{vdw} \cong \gamma\ \Delta SASA$, despite its inadequacy as discussed by Gallicchio et al.[12] The second disadvantage is that one is limited to the small, neutral molecules for which experimental results are available. Accordingly, the transferability of these parameters to biomolecular systems such as proteins, where backbone hydrogen bonding and charged residues likely affect solvent−solute interactions, is unknown.

Benchmarking continuum parameters against explicit solvent simulations avoids the aforementioned challenges; simulations can be applied to charged molecules that mimic biomolecular systems, and thermodynamic cycles can be used to separate solvation energies into electrostatic and nonelectrostatic contributions. In the first leg of the thermodynamic cycle, the 'growth' phase, the nonpolar contributions are accounted for by growing a neutral solute cavity into bulk solvent and introducing solute−solvent vdW interactions. In the next leg, the 'charging' phase, the electrostatics contributions are measured by introducing solvent−solute charge−charge interactions. The main disadvantage to using explicit solvent simulations is the error inherent in modern force fields, limited sampling, and simulations techniques. It is encouraging that attempts to minimize these errors have become increasingly successful.[13,14] Simulations additionally offer energetic congruence to methods that combine continuum solvation and molecular mechanics energies such as end-point free energy calculations,[15,16] pK_a calculations[17,18] continuum dynamics,[19] and constant pH molecular dynamics.[20]

It is also possible to compare the forces acting on specific atoms from explicit solvent simulations and continuum models.[21] Continuum force evaluations require a smooth dielectric boundary because abrupt boundaries often result in numerical instability in the electrostatic potential. Techniques such as Gaussian[22] and cubic spline[23] based volume functions have been introduced to smooth the dielectric boundary, making continuum force calculations and continuum dynamics possible. Spline smoothed surfaces are now standard options in several Poisson−Boltzmann solvers.[24,25] To maintain quantitative accuracy, however, these implementations require significant rescaling of the radii used to define the dielectric boundary. If the radii are not rescaled, solvation energies and forces tend to be overestimated by 10 to 40% (results not shown). Nina et al. used explicit solvent simulations to optimize of a set of radii for the CHARMM22 force field for abrupt boundary definitions[10] and later rescaled these radii for spline smoothed boundary definitions.[10,11] Since the location of the dielectric boundary is highly dependent on the solute charge distribution, the Nina et al. radii are not transferable to other solute charge definitions. Thus, a similar effort is needed for other force fields as has been suggested by a number of authors.[19,21,26]

We present two sets of optimized radii for the AMBER (parm99) force field for either abrupt or cubic spline smoothed dielectric boundary definitions. The radii are initially approximated from the solvent charge distributions measured during explicit solvent simulations and then optimized with a genetic algorithm (GA) to reproduce explicit solvent charging energies. We present the relative performance of several commonly used continuum parameter sets on the model compounds and four protein-like polypeptide chains. The latter were included to demonstrate the transferability of these parameters to protein systems. Both sets of optimized radii improve the accuracy of continuum solvation energies and the smooth boundary radii improve

Table 1. 14 Polyalanine Peptides Used To Optimize the Backbone Radii

polypeptide[a]	residues	description[b]	H-bonding
l-beta1	17−20	type II	yes
l-beta2	36−39	type I′	yes
l-beta3	39−42	type I	yes
l-beta4	59−62	type IV	no
l-beta5	60−63	type I	no
l-beta6	85−88	type VIII	no
l-helix1	5−14	type H	yes
l-helix2	109−114	type H	yes
l-hairpin1	42−53	class 3:5 IG	yes
l-hairpin2	51−59	class 4:4	yes
c-beta1	17−20	type I	yes
c-beta2	42−45	type IV	no
c-helix1	7−20	type H	yes
c-helix2	23−30	type H	yes

[a] Peptides taken from lysozyme (1aki) and crambin (1ejg). [b] Secondary structure descriptions provided by PDBsum.[29]

the correlation between explicit and implicit forces. These radii are suitable for PB quantitative measurements with AMBER partial charges and are recommended for methods that combine AMBER molecular mechanics and PB solvation energies.

The next section describes the model systems and the methodology used in the explicit solvent simulations, the continuum calculations, and the genetic algorithm optimizations. Section 3 defines the radius groups, discusses the explicit solvent charging energies, compares continuum solvation energies from different parameter sets, verifies radii transferability to proteins with protein test cases, and presents the quantitative affects of these radii on atomic forces. Section 4 summarizes the work and possible future directions.

2. Methods

2.1. Model System. The explicit solvent simulations, continuum calculations, and genetic algorithm optimizations were divided into two stages. First, the protein backbone atoms were optimized using 14 polyalanine peptides of varying lengths in common secondary structure conformations. Each conformation (see Table 1) was modeled from fragments of either lysozyme (pdb code 1ati) or crambin (pdb code 1ejg). The fragments were mutated to polyalanine and terminated with neutral blocking groups with the MMTSB Tool Set.[27] Second, the side chain radii were optimized using 20 nonzwiterionic *N*-acetyl-X-*N*′-methylamide dipeptides where X represents one of the twenty standard amino acids. Two conformations of each side chain dipeptide were used. The first conformation, chosen for the sake of comparison with previous optimizations,[10] used extended backbone phi and psi angles (180°, 180°) and the most frequent side chain dihedral angles from a Dunbrack backbone-*independent* rotamer library. The second conformation used a much more common backbone conformation (−60°, −40°) and the most frequent rotamers from a Dunbrack backbone-*dependent* rotamer library.[28] All model conformations are provided in the Supporting Information.

2.2. Explicit Solvent Simulations. The AMBER parm99 force field converted to CHARMM format was used in all simulations. Hydrogen atoms were first energy minimized in vacuum with 50 steps of steepest descent followed by 1000 steps of the Adopted Basis Newton Raphson (ABNR) method. All solute atoms were then fixed for the duration of the simulation. Each model compound was solvated in a sphere of explicit TIP3P water molecules that extended 6.5 Å beyond the dipeptides and 10.0 Å beyond the polyalanine peptides. This resulted in 3−4 hydration shells around every solute atom. Running simulations with larger and smaller solvent shells verified that the chosen dimensions were sufficient for energetic convergence. The spherical solvent boundary potential (SSBP) model including Kirkwood's mulitpolar expansion reaction field was used to approximate the influence of bulk water beyond the explicit water sphere.[30] This model alleviates many of the difficulties that result from perturbing charged systems with periodic boundary conditions and has been shown to give reliable results for proteins, nucleic acids, and, most recently, small molecules.[4,14,31] The solvent was first energy minimized with 50 steps of steepest descent followed by 1000 steps of the ABNR method and then equilibrated for 100 ps.

All simulations employed Langevin dynamics at constant temperature (300 K) using SHAKE enabled 2 fs time steps, infinite cutoffs for nonbonded interactions, and a friction constant corresponding to a relaxation time of 5 ps applied to water oxygen atoms. The preequilibrated fully charged systems were simulated for 200 ps to obtain the solvent charge distribution surrounding each solute atom. The solvent charge distributions were used to verify that the continuum radii were properly grouped and to estimate their starting values.

The charging free energies were measured with free energy perturbation (FEP) simulations run in the PERT module of CHARMM.[25] Each simulation consisted of 10 windows in which the solute's charge was scaled by a thermodynamic coupling parameter λ varying by $\pm$ 0.1 from 0 to 1 and then from 1 to 0 according to, $q(\lambda) = \lambda q_{final}$. The weighted histogram analysis method (WHAM) was used to combine the results of the individual windows to calculate the total charging free energy. WHAM is a self-consistent iterative procedure that optimizes the distribution of data from separate simulations and thus decreases the amount of sampling required for convergence.[32] Two tests were used to ensure that the simulations were converged. First, the standard error, calculated as half the difference between the forward and reverse WHAM postprocessed results and shown in Table 3, was required to be less than 3% of the total energy. Second, simulations of twice the length were required to be within 2% of the original simulations for the largest and most charged model compounds (results not shown). Different window lengths were required for convergence by the backbone and side chain model compounds; the polyalanine peptides were equilibrated for 10 ps followed by 40 ps of collection, while the dipeptides showed convergence in 5 ps of equilibration followed by 20 ps of collection.

2.3. Continuum Calculations. All continuum calculations were performed with the Adaptive Poisson−Boltzmann Solver (APBS)[24] using zero bulk ionic strength, a temperature of 300 K, a solvent dielectric of 78.4, and a solute dielectric

Table 2. Radius Groups with GA Starting Values and the Final Optimized Values

atom name[a]	residues	start value	final value[b]	final value[c]
	Backbone			
C	all	2.30	1.903	2.170
O	all	1.58	1.454	1.742
N	all	2.50	1.835	2.262
CA	all except G	2.80	1.591	2.339
CA	G	2.60	1.654	2.133
CAY/CAT	ACE,NME	2.51	2.595	2.375
	Side Chains			
CB	D,E,C,H,M,F,S,T,W,Y	2.55	2.087	2.370
CB	A,R,N,Q,I,L,K,V	2.75	1.829	2.063
CG*	R,Q,I,L,K,M,T,V	2.49	2.039	2.329
CG	H,F,W,Y	2.10	1.651	2.290
CG/CD	N,Q,D	2.19	1.995	2.257
CG	E	2.45	1.942	2.432
CB/CG/CD	P	2.70	2.008	2.157
CD	R,K	2.81	2.034	2.303
CD*	I,L	2.45	1.897	2.103
CD*/CE*/CZ	H,F,W,Y	2.05	1.837	2.122
CE	M	2.40	1.902	2.157
CZ/CE	R,K	2.66	2.020	2.414
OD*/OE*	N,Q,D,E	1.55	1.516	1.727
OG*	S,T	1.65	1.562	1.832
OH	Y	1.72	1.738	2.022
NE,NH*,NZ	R,K	2.48	1.523	1.861
ND2/NE2	N,Q	2.12	2.222	2.453
ND1,NE2	H	1.90	1.436	1.782
NE1	W	2.11	1.898	2.147
SG/SD	C,M	2.00	1.978	2.169
	Hydrogens[d]			
type H	bb HN	1.20	1.600	1.967
type H	bound to N	1.20	1.119	1.379
type HO/HS	bound to O/S	1.00	1.201	1.406
type H1/HP	polar	1.31	1.914	2.033
type HC/HA	nonpolar	1.30	0.840	1.321

[a] Radius groups are distinguished by AMBER atom names for all heavy atoms and by atom type for hydrogen atoms. [b] Final radii for abrupt dielectric surfaces. [c] Final radii for spline smoothed surfaces. [d] Hydrogens specified by atom type with type 'H' divided into two groups; amide backbone 'HN' and all other N-bound hydrogens.

Table 3. FEP Charging and Continuum Solvation Energies for the 14 Polyalanine Peptides Using AMBER and Optimized Radii[c]

polypep	WHAM		AMBER	opt[a]	opt[b]
l-beta1	−34.17	(0.06)	−31.87	−35.44	−33.53
l-beta2	−32.69	(0.05)	−25.67	−30.68	−31.85
l-beta3	−30.89	(0.48)	−27.88	−31.40	−30.79
l-beta4	−31.42	(0.31)	−28.89	−32.26	−32.88
l-beta5	−33.69	(0.03)	−32.18	−34.83	−32.87
l-beta6	−31.09	(0.06)	−27.27	−31.53	−32.34
l-helix1	−65.29	(0.45)	−53.78	−64.15	−65.31
l-helix2	−49.18	(0.42)	−43.36	−48.18	−51.49
l-hairpin1	−66.50	(1.33)	−55.70	−66.14	−66.63
l-hairpin2	−81.45	(1.06)	−69.17	−81.60	−81.43
c-beta1	−34.72	(0.24)	−33.64	−34.54	−33.90
c-beta2	−40.90	(0.24)	−36.47	−40.88	−41.06
c-helix1	−68.91	(0.44)	−54.99	−66.39	−67.15
c-helix2	−51.91	(0.04)	−44.01	−51.05	−51.77

[a] Abrupt smoothed dielectric boundary definitions. [b] Spline smoothed dielectric boundary definitions. [c] All energies are in kcal/mol. Standard errors are reported as half the difference between the forward and reverse WHAM energies.

of 1. The PB grid was centered on each solute and extended at least 20 Å beyond its dimensions. A grid resolution of 0.25 Å was proven sufficient for energetic convergence by comparing calculations with 0.15 Å grid resolution, which resulted in a correlation coefficient of 0.9995 and an average absolute error (AAE) of 0.97 kcal/mol. Solute charges were distributed onto grid points using a cubic B-spline discretization. Electrostatic solvation free energies were calculated from the energetic difference in the solvated ($\epsilon_{bulk} = 78.4$) and gas-phase systems ($\epsilon_{bulk} = 1$). The molecular surface was defined by the interface of a 1.4 Å solvent probe and the solute radii.

Some of the choices made in the continuum calculations protocol warrant explanation. First, the molecular surface traced by rolling a solvent probe, $r = 1.4$ Å, around the solute atoms was used to define the location of the abrupt dielectric boundary. Previous PB radii optimizations[10] have used the van der Waals (vdW) surface made up of overlapping solute atoms which excludes the interstitial spaces that lie between solute crevices from the molecular volume. Although the vdW surface definition works well for small molecules, it often results in buried high dielectric pockets in larger molecules such as proteins.[33] These pockets can change quantitative results significantly, making the radii less robust across similar conformations. The vdW surface is particularly problematic in continuum dynamics as high dielectric pockets can appear and disappear as frequently as every time step, resulting in numerical instability.[19] Second, for the spline smoothed surfaces, cubic B-spline charge discretization was used instead of trilinear interpolation to avoid large orientational artifacts.[22] Finally, a solute dielectric of 1 was chosen for consistency with the nonpolarizable force field and fixed solute conformation.

2.4. GA Optimization. The GA was used to optimize the radii to their final values. The GA is an efficient stochastic optimization method that has been widely applied to minimization problems because it is ideally suited for multiple-dimensional global search problems where the search space contains multiple local minima and the search variables may or may not be correlated. The GA begins with the generation of an initial population of a given number of solutions. The fitness of each solution is evaluated, and a new population is generated via selection, crossover, and mutation. This process is repeated until a desired fitness is reached or the maximum number of generations exceeded.

In the initial evolutions, populations of 50 solutions were run for 100 generations. In subsequent evolutions, populations of 100 solutions run for 50 generations. Solutions in the initial population of a given evolution were generated from a uniform distribution ± 0.10 Å around the starting radii. In subsequent generations, solutions were generated via a process of selection followed by crossover or mutation. Selections were performed with the Stochastic Universal Sampling algorithm,[34] which is designed to give zero bias

in the selection and minimal spread. It selects solutions with a probability proportional to their fitness

$$p_i = \frac{F_i}{\sum_{j=1}^{N} F_j} \quad (1)$$

where p_i is the probability that solution i will be selected and F_i is its fitness. The fitness function was normalized to ensure that fitness scores remained between 0 (poor) and 1 (perfect)

$$F_i = \frac{1}{(1 - AAE)_i} \quad (2)$$

where AAE_i is the average absolute error of solution i. A uniform crossover process was applied to 90% of the population. Each crossover event consisted of randomly distributing the radii from two previous-generation solutions to two new-generation solutions. Mutations were applied to 20% of the population and consisted of perturbing a solution's radii using a Gaussian distribution with a standard deviation of ±0.05 Å. The evolution was terminated if the total fitness and the best solution converged to the same value for 6 consecutive generations. In the absence of convergence a new evolution was started with the best radii from the previous evolution.

3. Results and Discussion

3.1. Radii Grouping and Starting Values. In optimization procedures, the number of parameters that can be meaningfully optimized is generally limited by the number of reference values. Although there is likely an ideal radius for every atom in every conformation, it is desirable to find a set of radii that are robust across multiple conformations. It is generally accepted that atoms in similar chemical environments have comparable surrounding solvent structures and thus similar optimal continuum radii. The factors directly influencing solvent structure are the atom's charge, vdW parameters, and structural neighbors. It may seem appealing to use AMBER atom types to define a set of continuum radius groups. This does not work, however, since AMBER atom types are distinguished by vdW parameters and often have significantly different charges in different residues. Instead, atoms were initially grouped according to similar chemical environments. The groups were then tested for similar surrounding solvent structure.

The solvent charge distribution surrounding each atom type was measured from 100 ps of Langevin dynamics of solvent plus rigid, fully charged solute molecules. The first peak in the solvent charge distribution corresponds to the closest explicit water molecules and can be used to approximate the optimal continuum radius. Figure 1a shows almost identical solvent charge distributions for the carboxyl oxygen in both conformations of the aspartate and glutamate residues. As expected, the first peak is positive due to electropositive water hydrogens crowding around the negatively charged carboxyl oxygen. Conversely, the first peak in the solvent charge distribution surrounding positively charged atoms,

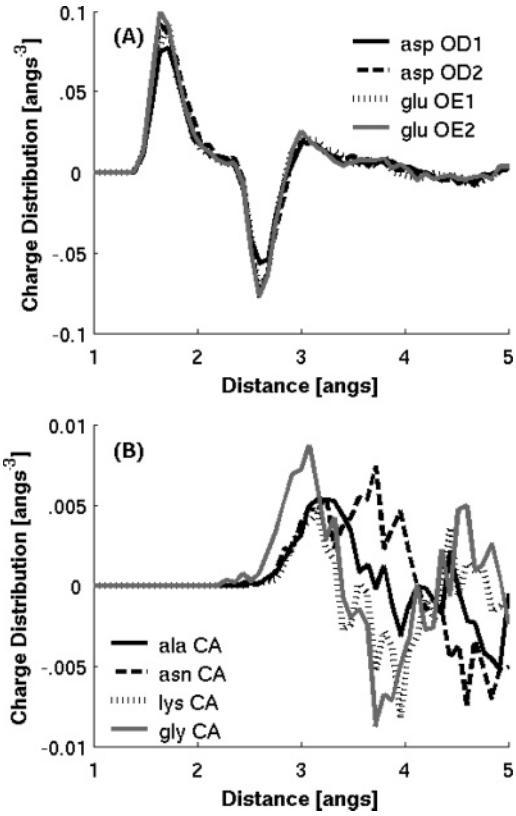


Figure 1. Radial solvent charge distributions show that similar chemical environments result in similar solvent structures around (A) the electronegative carboxyl oxygen in asp and glu as well as (B) the alpha carbons of ala, asn, and lys. The alpha carbon of gly reveals a unique solvent structure.

such as the amine nitrogen in arginine and lysine (results not shown), are negative due to electronegative water oxygen atoms. Figure 1b shows the solvent charge distribution for alpha-carbons in alanine, asparagine, lysine, and glycine. It demonstrates a slightly smaller solvent excluded volume for the alpha-carbon in glycine, likely due to the absence of a side chain. Thus, glycine's alpha carbon was put in a separate radius group from the rest of the alpha carbons. Similar plots were used to classify all of the radius groups which are shown in Table 2 along with their starting values.

We highlight two distinctions from the Nina et al.[10] optimizations, which also used the solvent charge distribution to approximate starting radii. First, we used nonzero hydrogen radii. This was deemed important because FEP charging free energies deviated 2−13% depending upon hydrogen placement (results not shown). GB models have also shown sensitivity to hydrogen radii.[35] Although the GB and PB models are fundamentally different, both are hopeful methods for DNA dynamics and both will likely need to fine-tuned their hydrogen radii. Second, although the starting radii work well with vdW surface definitions, as used by Nina et al., they clearly underestimated solvation effects with molecular surface definitions. We found that decreasing the starting values by 10% reduced the number of evolutions necessary to converge the radii to their optimal values.

3.2. FEP Simulations. The approximate starting radii were fine-tuned to reproduce explicit solvent FEP charging free energies. These energies, shown in Tables 3 and 4, were used

Table 4. Explicit Solvent Charging Free Energies (WHAM) and Continuum Solvation Energies for the 20 Amino Acid Dipeptides Using AMBER and Optimized Radii[c]

res	WHAM		AMBER	opt[a]	opt[b]
		Nonpolar Groups			
Gly	−13.91	(0.10)	−14.35	−14.43	−13.93
Gly2	−18.45	(0.15)	−18.51	−18.36	−18.11
Ala	−14.01	(0.18)	−13.77	−14.26	−13.82
Ala2	−17.86	(0.02)	−16.97	−16.92	−16.90
Val	−12.48	(0.17)	−11.22	−12.58	−12.45
Val2	−18.79	(0.45)	−16.99	−17.36	−17.73
Leu	−12.58	(0.02)	−11.72	−12.68	−12.40
Leu2	−18.45	(0.06)	−16.77	−17.97	−17.94
Ile	−11.99	(0.06)	−10.87	−12.01	−12.00
Ile2	−19.29	(0.06)	−17.25	−17.52	−17.85
Pro	−14.75	(0.01)	−13.29	−14.76	−14.74
Pro2	−16.84	(0.16)	−14.66	−16.75	−16.98
Phe	−16.42	(0.46)	−14.68	−16.45	−16.43
Phe2	−21.95	(0.01)	−20.16	−21.66	−20.97
Trp	−19.76	(0.04)	−19.50	−19.73	−19.73
Trp2	−22.23	(0.45)	−23.25	−23.14	−21.93
Met	−12.98	(0.02)	−12.76	−13.12	−13.09
Met2	−19.53	(0.23)	−18.40	−18.85	−19.04
		Neutral Polar Groups			
Ser	−17.34	(0.03)	−18.28	−17.33	−17.29
Ser2	−19.60	(0.18)	−21.03	−19.68	−19.82
Thr	−16.25	(0.18)	−17.06	−16.69	−16.52
Thr2	−19.58	(0.30)	−19.68	−19.11	−19.24
Cys	−18.18	(0.36)	−18.69	−18.17	−18.11
Cys2	−17.97	(0.07)	−18.39	−17.96	−17.95
Tyr	−19.78	(0.13)	−20.49	−20.42	−19.77
Tyr2	−25.30	(0.23)	−25.32	−25.08	−24.26
Asn	−20.88	(0.10)	−23.09	−20.89	−20.86
Asn2	−23.38	(0.47)	−24.96	−23.36	−23.12
Gln	−18.40	(0.47)	−20.16	−18.43	−18.68
Gln2	−27.63	(0.24)	−29.29	−26.40	−26.35
His	−23.63	(0.07)	−24.24	−24.77	−24.26
His2	−22.12	(0.16)	−21.13	−22.52	−21.56
		Charged Polar Groups			
Arg	−60.99	(0.44)	−69.19	−63.89	−63.05
Arg2	−77.28	(0.70)	−81.88	−76.10	−74.50
Lys	−65.99	(0.06)	−71.66	−67.44	−68.71
Lys2	−82.65	(1.14)	−85.93	−81.69	−80.33
Asp	−77.86	(0.47)	−72.05	−77.90	−78.02
Asp2	−79.38	(0.18)	−71.06	−77.35	−77.33
Glu	−79.00	(0.51)	−72.09	−79.09	−78.99
Glu2	−76.31	(0.24)	−69.47	−76.30	−76.30

[a] Abrupt smoothed dielectric boundary definitions. [b] Spline smoothed dielectric boundary definitions. [c] All energies are in kcal/mol. Conformation 1 and 2 have phi/psi angles of (180°,180°) and (−40°,-60°), respectively. Standard errors are reported as half the difference between the forward and reverse WHAM energies.

as the target values in the GA optimizations. All simulations were well converged as previously described and as indicated by the reported standard errors. Values for the polyalanine peptides ranged from −31 to −81 kcal/mol depending on their length and conformation, while those of the side chain dipeptides ranged from −12 to −83 kcal/mol. The neutral nonpolar residues had the lowest magnitudes, −12 to −16 kcal/mol; the polar residues were slightly higher, −16 to −28 kcal/mol; and the charged residues were the highest, −61 to −83 kcal/mol. Comparing the charging free energies between dipeptide conformations demonstrates a moderate range of conformational sensitivity. Lysine demonstrated the largest range with a charging free energy of −61 kcal/mol for one conformation and −77 kcal/mol for the other.

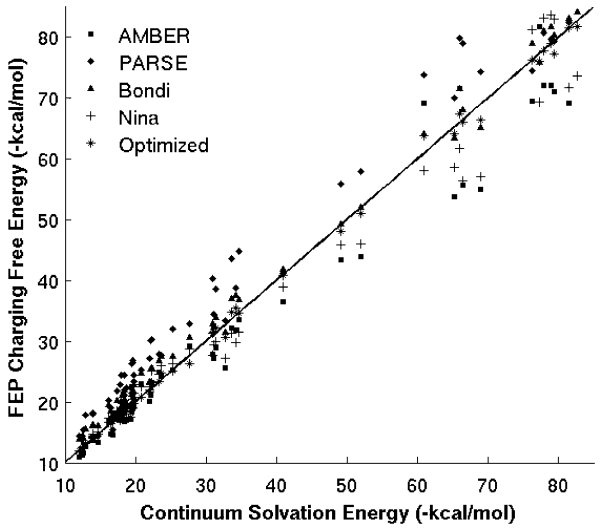


Figure 2. Explicit solvent charging free energies versus continuum solvation free energies calculated with the AMBER charges and different radii.

3.3. GA Optimizations. Results for the GA optimizations were typical for a highly dimensional rough energy landscapes. Specifically, different radius sets with similar fitness values were often encountered. After 6 evolutions the optimal abrupt and smooth boundary radii, as presented in Table 1, had AAEs of 0.54 and 0.64, respectively.

3.4. Continuum Calculations Testing Parameter Sets. To gauge the importance of these optimized radii, we tested several commonly used continuum parameter sets: AMBER (parm99) charges combined with parm99 vdW radii, Bondi radii, and the previously optimized Nina et al. radii in addition to PARSE charges combined with PARSE radii. Figure 2 shows the resulting continuum solvation energies compared to explicit solvent charging energies. The relative performance between parameter sets is more clearly illustrated in Figure 3a where the continuum deviation for each model compound is represented as a single point.

Each parameter set was tested for the appropriate solute dielectric and surface definitions for the comparisons made in Figures 2 and 3a. Failure to use the appropriate solute dielectric and surface definitions results in overestimating or underestimating solvation effects. For example, the PARSE parameters, originally optimized with a solute dielectric of 2, consistently overestimate solvation effects when used with a solute dielectric of 1. Likewise the Nina et al. radii, which were optimized with a vdW surface, underestimate solvation effects with a molecular surface definition. Although changing either the solute dielectric or the surface representation generally increases or decreases solvation energies, it has little effect on the relative solvation energies between different solutes. This is demonstrated in Figure 3b with the Bondi radii where using a vdW surface and solute dielectric of 1 overestimates solvation effects and

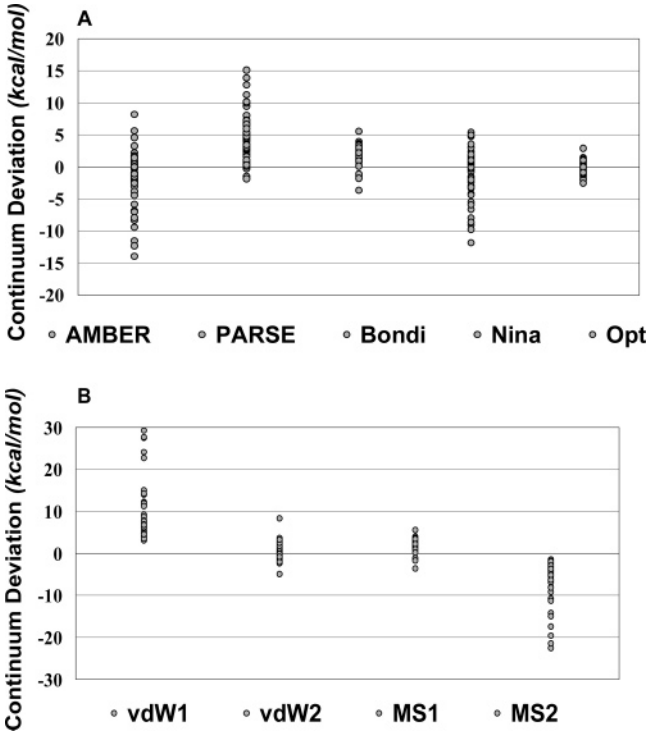


Figure 3. Continuum solvation energy deviations from explicit solvent reference values. Each circle represents one model compound. (A) Five different parameter sets are compared with their optimal surface definitions and solute dielectrics: PARSE charges with PARSE radii, a molecular surface, and solute dielectric of 2; AMBER charges with AMBER radii a molecular surface and solute dielectric of 1; AMBER charges with Bondi radii a molecular surface and solute dielectric of 1; AMBER charges with Nina radii a vdW surface and solute dielectric of 1; and AMBER charges with the newly optimized radii a molecular surface and solute dielectric of 1. (B) The effect of different surface and solute dielectric definitions on solvation energies calculated with the Bondi radii.

a molecular surface with a solute dielectric of 2 underestimates them.

The AMBER parm99 vdW radii were tested because they are frequently chosen for use with parm99 charges despite the fact that they have never been optimized for continuum models. Optimal performance, found with a molecular surface definition and a solute dielectric of 1, yielded an AAE of 3.18 kcal/mol. The Bondi radii[36] were tested because they are the most common choice for the intrinsic radii used in generalized Born models.[37] Using a solute dielectric of 1 and a molecular surface representation they performed the best out of the unoptimized parameters, with an AAE of 2.27 kcal/mol.

The Nina et al. radii were tested to query the transferability of radii between the AMBER and CHARMM force fields. They were not expected to work perfectly with the AMBER charges because continuum charge and radius definitions are strongly interdependent. It has been postulated, however, that the two force fields should have similar solute−solvent interactions, similar charging free energies, and thus similar continuum radii.[38] To test this, FEP simulations were run with the CHARMM force field. They yielded substantially different charging free energies and solvent radial distributions. In continuum calculations using a solute dielectric of 1 and a vdW surface, the Nina et al. radii combined with AMBER charges resulted in an AAE of 2.67 kcal/mol. This is an improvement over parm99 vdW radii but worse than the Bondi radii. It is far worse than their performance with the CHARMM charges when compared to FEP simulations run with the CHARMM force field (results not shown), which had an AAE of 0.69 kcal/mol. Differences in the FEP charging free energies, the solvent radial distribution functions, and the optimal continuum radii for AMBER and CHARMM are likely due to differences in partial charges.

Finally, the PARSE parameters were tested as they have likely been the most respected and frequently used continuum charge and radius definitions since their development in 1994.[8] Their performance relative to the explicit solvent model was worse than expected. Regardless of the surface and solute dielectric definitions, they consistently overestimated solvation effects on the model systems. Optimal performance was found with a molecular surface definition and a solute dielectric of 2 resulting in an AAE of 5.22 kcal/mol.

3.5. Protein Test Cases. The optimized radii clearly perform better on the model compounds for which they were optimized, but this is a biased test. As an unbiased test and to evaluate their transferability to proteins, FEP simulations were run on four short polypeptide chains: Trpcage (1l2y),[39] the C-terminal fragment (41−56) of protein G (2gb1),[40] the C-peptide of ribonuclease A (8rat),[41] and the first helix taken from lysozyme (residues 5−14) with all of the native side chains present. The proteins were prepared and simulated with the same procedures used on the model compounds. Simulations were held to the same requirements for energetic convergence. The resulting charging free energies are shown in Table 5 in addition to continuum deviations with the AMBER, PARSE, Bondi, and newly optimized radii. The surface definitions and solute dielectrics found optimal for the model compounds, as described in the previous section, were used for consistency. The optimized radii perform the best on abrupt boundary surfaces with an AAE of 5.96 kcal/mol. All abrupt boundary radius sets overestimated solvation energies by 20−40% when used with the spline smoothed surfaces (results not shown). The smooth boundary radii, presented in the 9th column, are a significant improvement but still overestimate solvation effects slightly. The source of this systematic overestimation is unknown. It can be compensated for by using a higher solute dielectric, which, as previously mentioned, shifts all solvation energies down but has little effect relative continuum deviations. As shown in the final column, the AAE can be decreased to 4.42 kcal/mol by using a solute dielectric of 1.15 instead of 1.0.

3.6. Spline Smoothed Radii and Atomic Forces. If abrupt boundary radii are used with a spline smoothed dielectric boundary, then solvation forces, similar to solvation energies, tend to be overestimated by 10 to 40% (results not shown). The width of the spline smoothed dielectric transition region, often referred to as the *spline window*, has been shown to effect solvation energies.[11] Thus, the optimal spline smoothed radii will be dependent on the spline window width. Previous optimizations have shown that reasonable quantitative ac-

Table 5. FEP Charging Energies and Continuum Deviations for 4 Protein-like Polypeptides[d]

protein	WHAM-40ps	WHAM-60ps	AMBER	PARSE	Bondi	Nina	opt[a]	opt[b]	opt[c]
trpcage	−309.85 (2.00)	−308.63 (1.89)	−20.47	−27.41	1.10	−2.97	0.51	19.57	−3.32
protein G	−314.37 (0.51)	−314.60 (0.25)	−17.78	14.58	21.55	70.55	15.74	19.72	4.86
ribonucleaseA	−167.77 (1.55)	−165.25 (1.21)	−6.21	38.76	11.69	45.41	3.82	17.37	3.83
lysozyme	257.88 (3.89)	−259.79 (2.53)	4.99	−3.09	1.93	31.16	3.77	5.99	−5.69
AAE			12.36	20.96	9.07	37.52	5.96	15.66	4.43

[a] Abrupt and smoothed dielectric boundary definitions with solute dielectric of 1. [b] Spline smoothed dielectric boundary definitions with solute dielectric of 1. [c] Spline smoothed boundary with solute dielectric of 1.15. [d] All energies in kcal/mol. FEP standard errors are reported as half the difference between the forward and reverse WHAM energies.

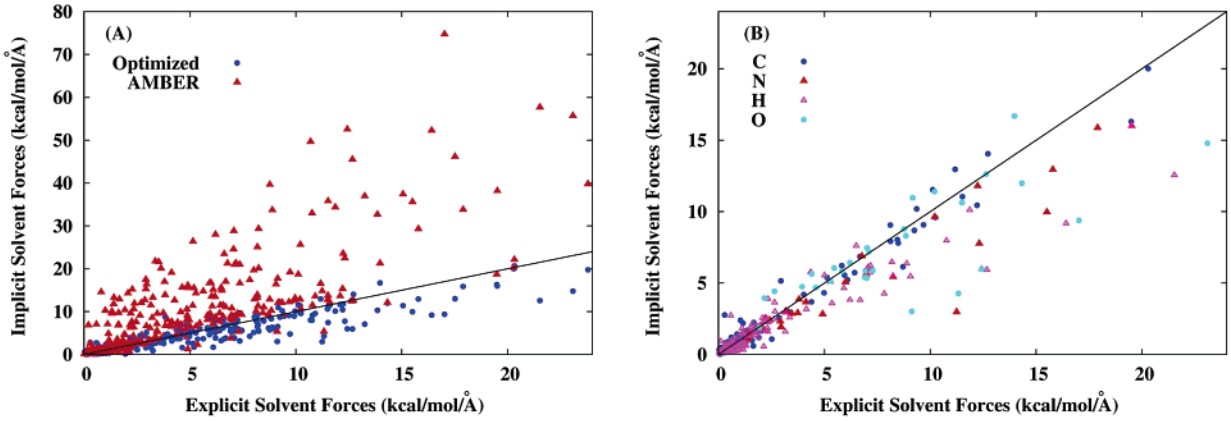


Figure 4. Comparison of explicit solvent and continuum forces. Lines indicate y = x NOT best fit. (A) Continuum forces calculated with AMBER radii in red: correlation coefficient 0.84; slope 1.99; intercept 0.03 kcal mol^{-1} Å^{-1}. Continuum forces calculated with optimized radii in blue: correlation coefficient 0.94; slope 0.80; intercept 0.29 kcal mol^{-1} Å^{-1}. (B) Continuum forces calculated with optimized radii only distinguishing different elements by color.

curacy can be maintained by rescaling abrupt-boundary radii by a single factor that is dependent on the spline window width.[11] This, however, assumes that all radii are equally affected by the spline smoothing. This assumption was tested by optimizing a set of radii for a single spline window of 0.3 Å with the GA. These radii were compared to abrupt boundary radii scaled by a single scaling factor. A simplex optimization showed that the optimal single scaling factor was $R_{spline} = (R_{abrupt} + 0.3)*1.0107$. The independently optimized radii, shown in Table 2, are significantly different from those obtained from the single scaling factor. The AAEs for the independently optimized and single scaled radii were 0.66 and 2.11 kcal/mol, respectively.

Recently, Wagoner et al. suggested comparing continuum and explicit solvent forces instead of molecular solvation energies to test the continuum model.[21] To obtain the explicit solvent electrostatic forces they averaged the total minus nonpolar solvent−solute interactions from a simulation of a 131-residue protein in a fixed conformation. They used the AMBER parm99 vdW radii and demonstrated reasonable correlation between continuum and explicit solvent forces with a correlation coefficient $r = 0.88$. As expected, however, the continuum forces were systematically overestimated by a factor of 2.2, as indicated by the slope from linear regression analysis. To see what effect the optimized radii have on atomic forces, the same procedure was carried out with Trpcage and the C-terminus of G-protein. Explicit solvent forces were reasonably converged after 200 ps of simulation. This was tested by comparing forces from a 500 ps simulation which resulted in a correlation coefficient of 0.9987 and an AAE of 0.10 kcal/mol. Figure 4a shows the improved correlation for atomic forces calculated with the optimized radii, $r = 0.94$, versus the AMBER radii, $r = 0.84$. The AMBER radii still systematically overestimate the continuum forces by a factor of 1.99, while the optimized radii slightly underestimate the forces by a factor of 0.80. Figure 4b shows the optimized atomic forces according to atom type and demonstrates that forces on oxygen atoms are most often underestimated. Comparing specific atomic forces from continuum and explicit solvent models may prove useful in future optimization efforts.

4. Conclusion

Computational biochemistry is a field with multiple levels of models and theories whose computational requirements scale proportionally with accuracy. The success of the lower level, less accurate, more efficient models relies upon their connection to the higher level, more accurate, less efficient models, while the latter are complimented by the former to characterize complex biological systems. As methodology advances, opportunities for improved synergy between levels of theory and models are created. As microscopic, explicit solvent simulations become more accurate and the numerical methods used in continuum models improve, there is an increasing opportunity and need to benchmark continuum models on microscopic simulations.

The main results of this paper are two sets of optimized continuum radii, presented in Table 2, for PB calculations with the AMBER force field using either abrupt or spline smoothed dielectric boundary definitions. Thirty-one radius

groups were defined based on similar chemical environments. Both sets of optimized radii improve quantitative agreement with microscopic simulations when compared to AMBER parm99 vdW, Bondi, and Nina et al. radii as well as the PARSE parameters. The Bondi radii combined with a protein dielectric of 1 and a molecular surface definition were the best alternative choice for AMBER charges. The PARSE parameters, shown to work optimally with a solute dielectric of 2 and a molecular surface definition, were less accurate. The abrupt boundary radii work quite well for protein systems. The smooth boundary radii offer a significant improvement over abrupt boundary radii used on spline smoothed surfaces but still overestimated the solvation energies of four protein-like polypeptides. This overestimation can be compensated for by using a solute dielectric of 1.15 instead of 1.0. The smooth boundary radii greatly improve agreement between continuum and explicit solvent atomic forces.

Future efforts will focus on identifying the source of deviations of the smoothed boundary radii on protein systems as well as small molecule parametrization and coupling these parameters with nonpolar solvation effects. If one hopes to use these radii on systems involving small molecule ligands, it will be important to characterize the appropriate radii for certain small molecules as well. It will also be of great use to test the complementarity of these electrostatic parameters with various nonpolar solvation models by comparing total solvation energies to either explicit solvent simulations or experimental solvation energies. It is expected that the nonpolar and polar contributions to continuum solvation models will be coupled and that the greatest degree of accuracy will require them to be treated as such.

Acknowledgment. The authors thank Dr. Nathan Baker, Dr. Benoit Roux, Dr. Dave Case, and Dr. Michael Feig for valuable discussion and guidance; Dr. Thomas Cheatham III and Dr. Jeffrey Klauda for the AMBER parm99 force field converted to format; and Dr. Robert Konecny and the Center for Theoretical Biological Physics for computing resources. J.M.J.S is supported by the Center for Theoretical Biological Physics. Additional funding for this work comes from National Science Foundation, National Institutes of Health, National Biomedical Computational Resource, Accelrys, Inc., and the Howard Hughes Medical Institute.

Supporting Information Available: Model conformations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Jackson, J. D. *Classical Electrodynamics*; John Wiley and Sons: 1962; Ch. 1, pp 12−14.

(2) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.

(3) Gouy, M. *J. Phys.* **1910**, *9*, 457−468.

(4) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1−20.

(5) Hirata, F.; Redfern, P.; Levy, R. M. *Int. J. Quantum Chem.* **1988**, 179−190.

(6) Roux, B.; Yu, H. A.; Karplus, M. *J. Phys. Chem.* **1990**, *94*, 4683−4688.

(7) Jayaram, B.; Fine, R.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1989**, *93*, 4320−4327.

(8) Sitkoff, D.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(9) Bordner, A. J.; Cavasotto, C. N.; Anagyan, R. A. *J. Phys. Chem.* **2002**, *106*, 11009−11015.

(10) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.

(11) Nina, M.; Im, W.; Roux, B. *Biophys. Chem.* **1999**, *78*, 89−96.

(12) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem.* **2000**, *104*, 6271−6285.

(13) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(14) Deng, Y. Q.; Roux, B. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.

(15) Massova, I.; Kollman, P. A. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.

(16) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67−74.

(17) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413−7422.

(18) Beroza, P.; Case, D. A. *Energet. Biol. Macromol., B* **1998**, *295*, 170−189.

(19) Lu, Q.; Luo, R. *J. Chem. Phys.* **2003**, *119*, 11035−11047.

(20) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2004**.

(21) Wagoner, J. Baker, N. A. *J. Comput. Chem.* **2004**.

(22) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608−640.

(23) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59−75.

(24) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.

(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(26) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244−1253.

(27) Michael, F.; Karanicolas, J.; Brooks, C. L., III. *MMTSB Toolset*; MMTSB NIH Research Resource: The Scripps Research Institute, 2001.

(28) Dunbrack, R. L.; Cohen, F. E. *Protein Sci.* **1997**, *6*, 1661−1681.

(29) Laskowski, R. A.; Hutchinson, E. G.; Michie, A. D.; Wallace, A. C.; Jones, M. L.; Thornton, J. M. *Trends Biochem. Sci.* **1997**, *22*, 488−490.

(30) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

(31) Nina, M.; Simonson, T. *J. Phys. Chem. B* **2002**, *106*, 3696−3705.

(32) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(33) Dong, F.; Vijayakumar, M.; Zhou, H. X. *Biophys. J.* **2003**, *85*, 49−60.

(34) Baker, J. E. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*; Lawrence Erlbaum Associates, Inc.: Mahwah, 1987; pp 14−21.

(35) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.

(36) Bondi, A. *J. Chem. Phys.* **1964**, *64*, 441.

(37) Bashford, D.; Case, D. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.

(38) Roux, B.; Case, D. personal communication.

(39) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nature Struct. Biol.* **2002**, *9*, 425−430.

(40) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657−661.

(41) Tilton, R. F.; Dewan, J. C.; Petsko, G. A. *Biochemistry* **1992**, *31*, 2469−2481.

CT049834O

JCTC Journal of Chemical Theory and Computation

Enzymatic Hydroxylation in *p*-Hydroxybenzoate Hydroxylase: A Case Study for QM/MM Molecular Dynamics

Hans Martin Senn,* Stephan Thiel, and Walter Thiel*

Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany

Received December 21, 2004

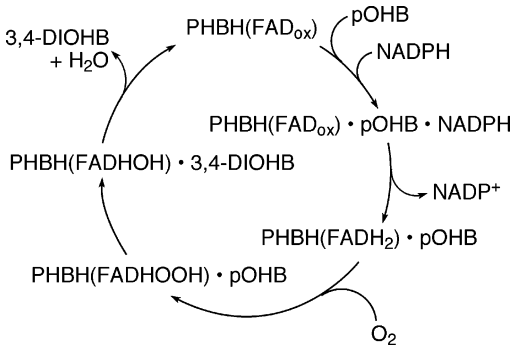
Abstract: We investigate the OH transfer step of the hydroxylation reaction of *p*-hydroxybenzoate in the enzyme *p*-hydroxybenzoate hydroxylase (PHBH) using QM/MM molecular dynamics methods. The QM region (49 atoms) is treated at the AM1 level, while the MM part (ca. 23 000 atoms) is described by the GROMOS force field. Performing pointwise thermodynamic integration from 10 starting structures, we have obtained an average value of the free-energy barrier for this reaction of 101 kJ mol^{-1}. The simulations provide insight into the dynamics of the hydrogen bonding network in the active site along the course of the reaction. In addition, we describe statistical techniques to analyze molecular dynamics data that assess the convergence of averages and yield an error measure. We discuss the effect of different error sources on the free energy.

I. Introduction

The enzyme *p*-hydroxybenzoate hydroxylase[1] (PHBH, EC 1.14.13.2) catalyzes the transformation of *p*-hydroxybenzoate (pOHB) to 3,4-dihydroxybenzoate (3,4-DOHB, also known as protocatechuate). It is active, for example, in soil bacteria where it plays a crucial role in the oxidative degradation of aromatic compounds, such as lignin, 3,4-DOHB being the substrate for subsequent catechol ring-cleavage reactions. PHBH has recently also been proposed as a biocatalyst for the hydroxylation of fluorinated and chlorinated pOHB derivatives.[2] The enzyme contains flavin-adenine dinucleotide (FAD) as a cofactor. The oxygen of the transferred hydroxy group stemming from O_2, PHBH is a flavoprotein classified as a monooxygenase. The second oxygen atom is not utilized for product formation but is ultimately reduced to H_2O.

The catalytic cycle may be divided into a reductive and an oxidative branch, as assessed by the oxidation state of the cofactor (Scheme 1). In the reductive half-reaction, PHBH with oxidized FAD binds the substrate pOHB as well as the two-electron reductant NADPH (reduced nicotinamide-adenine dinucleotide phosphate). The flavin cofactor is reduced by NADPH to $FADH_2$, and $NADP^+$ is released. In the subsequent oxidative half-reaction, $FADH_2$ reacts with molecular oxygen to form a flavin hydroperoxide (FADHOOH) that acts as the active hydroxylation agent. FADHOOH hydroxylates pOHB in the meta position, yielding FADHO (FADHOH after protonation) and a hydroxycyclohexadienone intermediate that tautomerizes rapidly into the aromatic 3,4-DOHB (Scheme 2). To close the cycle, FADHOH loses water, regenerating the oxidized form of FAD, and 3,4-DOHB is liberated.

Scheme 1. Schematic Catalytic Cycle for pOHB Hydroxylation by PHBH

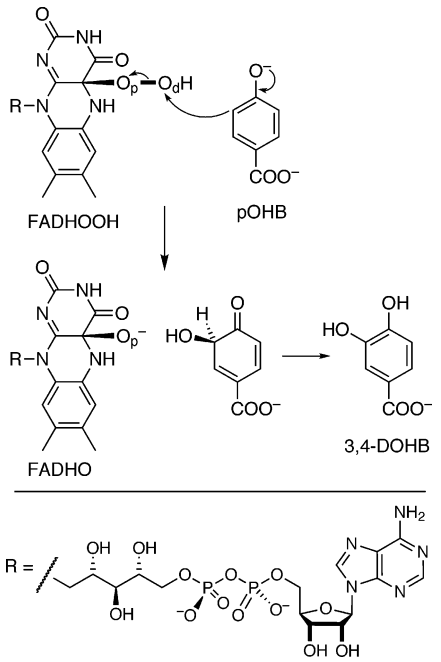


PHBH has been the subject of numerous experimental investigations over the last almost three decades and has become a paradigm for flavoprotein monooxygenases; we

* Corresponding authors e-mail: thiel@mpi-muelheim.mpg.de (W.T.) and senn@mpi-muelheim.mpg.de (H.M.S.).

10.1021/ct049844p CCC: $30.25
Published on Web 04/07/2005

Scheme 2. Hydroxylation Step



refer here to the review by Entsch and van Berkel.[1] As our focus in the present study lies exclusively on the hydroxylation step, we summarize here pertinent experimental results from structural, mechanistic, kinetic, and mutation studies. (1) The hydroxylation reaction follows the aromatic electrophilic substitution mechanism, that is, FADHOOH acts formally as an "OH^+" donor.[3] (2) The substrate reacts in its dianionic (phenolate) form,[4,5] which concurs with (1), the oxido substituent being strongly activating and ortho directing in S_EAr reactions. The pK_a of the phenolic proton in enzyme-bound pOHB is lowered relative to free pOHB.[5,6] (3) A number of specific interactions with active-site residues stabilize the pOHB dianion and contribute to its binding and activation.[5−8] Both the dianionic substrate and the oxidoflavin resulting from OH transfer are stabilized by a positive electrostatic potential in the active site.[9] (4) The different stages of the reaction are coupled with (at least) three distinct conformational states of the flavin, which on one hand makes the active site accessible for substrate binding and interaction of the oxidized FAD with NADPH and on the other hand shields the active hydroperoxy-FAD from reacting with solvent water once it is formed.[10−12] (5) The relative rates of the different reaction steps, and thus the rate-determining step, appear to depend on the external conditions. Under very mildly acidic conditions (pH 6.5) at low temperature (275−279 K), the rate constants for hydroxylation and loss of H_2O from FADHOH are around 47 s^{-1} and 14 s^{-1}, respectively,[4,5,13] that is, hydroxylation is not rate-limiting. The overall turnover rate of the enzyme under these conditions[5] is around 6 s^{-1}. For more basic conditions, the overall turnover rate was determined as 7.5 s^{-1} (pH 8.6, 277 K),[9] 44 s^{-1} (pH 7.9, 298 K),[5] and 55 s^{-1} (pH 8.0, 298 K).[14] Enzyme activity is highest in the pH range 7.5 to 8.5.[15] From temperature-dependent measurements of the overall rate, the activation energy was obtained as 49 kJ mol^{-1} at pH 8.0.[15]

Also from the theoretical side the hydroxylation step of the PHBH catalytic cycle has found considerable attention and has become a kind of prototype system for computational studies on enzyme reactions. An early study found a good correlation of experimental turnover rates of fluoro-substituted pOHB derivatives with the energies of the highest occupied molecular orbital (HOMO) calculated for the free substrates in the gas phase at the semiempirical AM1 level.[16] This observation supported the notion that the hydroxylation is rate-determining under the conditions of the measurements[14] (pH 8.0, 298 K). It also provided evidence for the dianionic state of the active substrate as well as for the S_EAr mechanism. A series of contributions by Ridder et al.[17−20] realized a considerable methodical step forward by using a combined molecular-mechanical/quantum-mechanical (QM/MM) approach. Applying AM1 to the substrate and the isoalloxazine part of FADHOOH and the CHARMM force field to the remaining part of the cofactor as well as for the whole protein including crystal water, they calculated reaction profiles along a predefined reaction coordinate describing the attack of the hydroperoxy-"OH^+" on the meta position of the substrate, optimizing at each point along the reaction coordinate the atoms within a given distance of the active region. The activation barriers thus obtained correlated again well with the experimental turnover rates for fluoro-substituted pOHB derivatives[17] and with experimental rate constants of the hydroxylation step[3] for a series of modified flavins and substrates,[19] corroborating the S_EAr mechanism and the dianionic state of the substrate. These conclusions were strengthened by analyzing in more detail[18] the structural and electronic changes and interactions with specific residues along the reaction path. The overall effect of the surrounding protein on the energetics of the reaction was found to be relatively small and the dominant contributions to stem from the QM part. The role of the enzyme in this case thus appears to consist mainly in activating the substrate by deprotonation, preparing the "OH^+" donor FADHOOH, and keeping the two reactants in a position favorable for reacting. The latest contribution[20] improved upon the computational level, employing HF/6-31G(d) instead of AM1 for the QM part in the reaction-profile calculations. In addition, single-point calculations were done on the isolated QM part in the gas phase with B3LYP/6-311+G(d, p) and LMP2/6-31+G(d). The results largely confirmed the conclusions obtained before with regard to the mechanism and the role of the protein.

The barrier calculated for the hydroxylation of pOHB is 73 kJ mol^{-1} with AM1/CHARMM.[17] HF/CHARMM provides a considerably higher value (about 125 kJ mol^{-1}), while the single-point calculations on HF/CHARMM geometries yield 51 and 47 kJ mol^{-1} with B3LYP and LMP2, respectively.[20] They provide a value of about −230 kJ mol^{-1} for the reaction energy. The comparison of different methods against gas-phase calculations up to MP2/6-31++G(d, p)//HF/6-31++G(d, p) on model reactions for the PHBH hydroxylation showed that HF drastically overestimates the barrier height, whereas B3LYP underestimates it; AM1 also overestimates, by about the same amount as B3LYP is too low.[21] Structurally, AM1 is fairly accurate, except for the length of the peroxide O−O bond, which is underestimated.[18,21] Together with the finding that AM1/CHARMM reproduces also the H-bonding patterns obtained with HF/

CHARMM,[20] AM1 can be considered sufficiently accurate and reliable for the reaction under study, at significantly lower computational cost than DFT or correlated ab initio methods.

A slightly different methodological approach was taken by Billeter et al.,[21] who combined AM1 with the GROMOS force field to optimize the geometry of reactant and product states and the transition state for the entire enzymatic system. In addition, they ran molecular dynamics (MD) simulations at the same level of theory, performing thermodynamic integration (TI) along the $C_{meta}-O_d$ distance as the reaction coordinate. They obtained 89 kJ mol^{-1} for the (static) potential energy barrier and 49 kJ mol^{-1} for the free-energy barrier. The corresponding reaction energies were -286 kJ mol^{-1} and -211 kJ mol^{-1}, respectively.

The purpose of the present contribution is 2-fold. First, we have investigated the hydroxylation step in PHBH with more extensive MD simulations and thermodynamic integration calculations. While retaining AM1/GROMOS, we have enlarged the model by including a shell of water molecules around the protein. To account better for changes in the protein environment, we have performed a series of simulations starting from several snapshots derived from a classical MD run. We have also employed a more accurate reaction coordinate. Second, we have used these calculations as a case study for clarifying a number of methodological and procedural issues that arise in the context of QM/MM MD simulations. We have focused on well-defined criteria for monitoring and establishing equilibration and stationary averages as well as on a careful identification and quantification of errors.

II. Methods

A. System Definition and Preparation. The setup of the PHBH system was based on the one described in ref 21. The system was prepared by means of fully classical MD simulations using GROMOS96.[22] The GROMOS87[23] force field parameters (set 37C4 with SPC water) were used, augmented by aliphatic hydrogen atoms lacking van der Waals and electrostatic interactions (which are absorbed into the atom the H is bound to). This setup facilitates QM/MM coupling as it allows one to define and change the QM region without modifying the MM topology. The enzyme (394 amino acids), including 219 crystallographic water molecules, FADHOOH, and dianionic pOHB (7004 atoms in total), was placed in a cubic box ($a = 9$ nm) of water molecules. The whole system was subjected to a series of minimizations and MD runs (periodic boundary conditions, *NVT*, 300 K) with gradually decreasing harmonic positional restraints acting on the nonwater atoms. Finally, the restraints were released except for the cofactor and the substrate (102 atoms), which were kept restrained with $k = 1.046 \times 10^4$ kJ mol^{-1} nm^{-2}. With this setup, an MD run of 200 ps was performed, from which snapshots were taken every 40 ps.

In each snapshot, a solvation shell was defined by the water molecules located between 0.29 and 1.1 nm from any protein atom. Water molecules further away were discarded; those closer to, or inside, the protein were left free. The water molecules in the shell were kept fixed, forming a rigid enclosing wall around the protein, preventing the free water from escaping into the vacuum and effectively providing a kind of boundary potential to the system. Details about the composition of the snapshots are summarized in Table 1.

Table 1. Composition of the Snapshots

	no. of atoms				
snapshot (ps)	protein	FADHOOH + pOHB	free water	fixed water	total
40	6245	102	2445	13 980	22 772
80	6245	102	2556	13 863	22 766
120	6245	102	2625	14 016	22 988
160	6245	102	2619	14 001	22 967
200	6245	102	2598	14 136	23 081

Each snapshot was then used as starting structure for locating stationary points (minima and transition state) of the hydroxylation reaction at the AM1/GROMOS level by means of a protocol involving a linear-scaling microiterative optimization algorithm working in hybrid delocalized coordinates[24] as implemented in ChemShell.[25] The QM region included the substrate and the isoalloxazine part of FADHOOH (49 atoms); full details on these calculations will be reported elsewhere.[26] The stationary points thus obtained served as initial structures for the QM/MM MD simulations described here.

B. QM/MM MD Setup. All calculations reported were done at the AM1[27]/GROMOS level within the modular program package ChemShell,[25,28] which provided the QM/MM coupling engine and the MD driver. The energy and gradient evaluations for the QM and MM part were performed by the MNDO99[29] and GROMOS96[22] codes, respectively, interfaced to ChemShell. The QM region contained the dianionic substrate and the isoalloxazine part of FADHOOH up to the first methylene unit of the ribityl side chain (see Scheme 2) and was terminated by a H link atom (49 QM atoms in total). The QM density was electrostatically embedded into the field of the fixed MM point charges by including them into the QM Hamiltonian. The charge-shift scheme[25] was applied at the QM/MM boundary; no electrostatic QM/MM cutoff was employed. The MM electrostatic interactions were evaluated in GROMOS96 for all atom pairs within a distance of 1.4 nm and approximated by a generalized Poisson−Boltzmann reaction field[30] beyond. The parameters for the reaction field were $\epsilon_1 = 1.0$ and $\epsilon_2 = 54.0$ (i.e., the permittivity of SPC water[31]) for the relative permittivities of the inner region and the continuum, respectively; the ionic strength was zero (i.e., inverse Debye screening length $\kappa = 0$).

The MD simulations were performed under *NVT* (canonical) conditions at $T = 300$ K. The temperature was controlled by a Berendsen thermostat[32] (coupling time $\tau_c = 0.1$ ps) during the heating phase. For equilibration and production, a Nosé−Hoover chain (NHC) thermostat[33−36] was implemented into the ChemShell dynamics module, together with a reversible noniterative leapfrog-type integrator[37] (thermostat chain length: 4; characteristic time of the thermostat coupling to the physical system $\tau_{NHC} = 0.02$ ps, corresponding to a thermostat wavenumber $\tilde{\nu}_{NHC} = 375$ cm^{-1}). All hydrogen

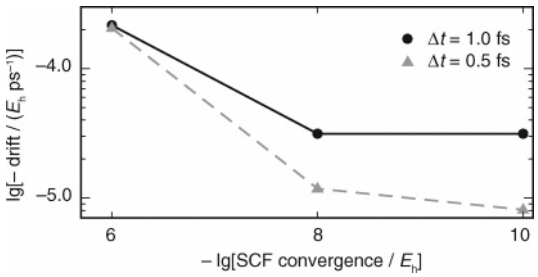


Figure 1. Drift of the total energy in AM1 Born−Oppenheimer *NVE* simulations as a function of the SCF convergence criterion and the time step.

atoms were assigned the mass of deuterium, $m_H = 2.014$ u; the free water molecules were kept rigid using SHAKE constraints.[38,39] The time step was $\Delta t = 1$ fs.

Each starting structure (optimized reactant and transition states of snapshots 40, 80, 120, 160, and 200 ps and the product state of snapshot 40 ps) was equilibrated as follows: Starting with random velocities corresponding to a Boltzmann distribution at 300 K, the system was run for 10 ps with the Berendsen thermostat, followed by 25 ps of NHC dynamics. For simulations starting from a transition state (TS), the reaction coordinate was kept fixed at its value in the TS using a SHAKE constraint for a difference of distances. The equilibration process was assessed by monitoring (1) the temperatures of different subsystems (protein, free water, FADHOOH + pOHB) and (2) the value of the (unconstrained) reaction coordinate (RC) for simulations of a reactant or product state or the constraint force acting on the RC for simulations of a TS, respectively. For the system to be considered equilibrated, these quantities were required to be stationary according to the criteria and the procedure described below (section I.D). The average subsystem temperatures were in addition required to coincide and be equal to 300 K within the confidence interval. Typically, it took around 25 ps for the temperatures to become equilibrated. For most systems, 35 ps of total equilibration time were sufficient for all criteria to be met. If not, the simulation was extended until this was the case.

An important issue in QM/MM MD simulations (or, more generally, in any type of Born−Oppenheimer MD) is the balance between the time step, which controls the accuracy of the integration of the equations of motion, and the SCF convergence, which determines the accuracy of the QM forces. Using FADHOOH (88 atoms, $m_H = 2$ u) as a test system, we ran Born−Oppenheimer *NVE* (microcanonical) simulations with AM1 to evaluate the influence of the time step and the SCF convergence criterion on the long-term stability of the simulations as measured by the drift in the total energy. As shown in Figure 1, either of the two parameters can be the limiting factor. If the SCF convergence is not strict enough, decreasing the time step does not improve stability. On the other hand, tightening the convergence criterion is only able to reduce the drift down to a certain level, at which the time step becomes limiting. Since the computation time is not significantly increased by a stricter SCF convergence criterion, whereas depending linearly on the time step, we chose as the optimum combination a time step of 1 fs and an SCF convergence of 10^{-8} E_h. Furthermore, we modified ChemShell to avoid any loss of accuracy in energy, gradient, and coordinate data when they are exchanged between the different modules (QM code, MM code, QM/MM coupling engine, MD driver), which proved to be crucial for the stability of long simulations.

C. Thermodynamic Integration. The free energy of activation (and in one case also the reaction free energy) was obtained from thermodynamic integration along the difference-of-distances reaction coordinate RC = $d(C_{meta}-O_d) - d(O_p-O_d)$, which captures both the formation of the bond between C_{meta} of pOHB and the distal oxygen of the hydroperoxy group and the cleavage of the peroxide bond, thus providing an accurate description of the full reactive process from reactants to products. Note that according to this definition, the reaction runs from positive values of the RC at the reactants to negative values at the products, that is, from "right to left" on a RC scale.

The formalism of thermodynamic integration[40] can be summarized by the relations

$$\Delta A_{a\to b} = \int_a^b \mathrm{d}\xi \frac{\partial A}{\partial \xi} = \int_a^b \mathrm{d}\xi \left\langle \frac{\partial V}{\partial \xi} \right\rangle_\xi \approx \int_a^b \mathrm{d}\xi \langle F_c \rangle_\xi \qquad (1)$$

where $A = A(\xi)$ is the Helmholtz free energy parametrized by ξ, V is the potential energy, and F_c is the force of constraint; $\langle\cdot\rangle$ designates averaging over the canonical ensemble. Identifying ξ with the reaction coordinate, the free-energy difference $\Delta A_{a\to b}$ between two states a and b, characterized by ξ_a and ξ_b, is evaluated as the potential of the mean force of constraint,[41−43] that is, the average force acting on the constraint that keeps the reaction coordinate at a certain value. In the last equality in eq 1, we made the approximation that the metric-tensor correction[41,44−47] is negligible (see Appendix A.1 for explicit expressions of the constraint force and the metric-tensor correction).

To evaluate the integral of the average constraint force over the reaction coordinate, it is discretized. One performs simulations at a series of fixed values of ξ, calculates $\langle F_c \rangle$ at each point, and numerically integrates the average constraint force over the reaction coordinate.[48]

We used an integration step along the RC of 0.1 Å. At each point, the system was sampled for at least 5 ps, until the force of constraint was stationary according to the procedure described in section I.D below. After 1 ps, the system was stepped to the next point. The end point of the integration on the reactant side was determined from the average value of the RC in the reactant (typically 1.7 Å), obtained from the converged equilibration run. Correspondingly, the end point on the product side was at RC = −1.6 Å. For the cases where we were interested only in the forward barrier, we integrated from the reactant up to RC = 0.0 Å, the TS being typically at RC ≈ 0.4 Å. For all snapshots, we integrated forward (from the equilibrated reactant to the TS) and backward (from the equilibrated TS to the reactant), obtaining two independent values for the forward free energy of activation, $\Delta^{\ddagger}A_{fw}$. A typical force curve and its integral are shown in Figure 2. For the 40-ps snapshot, we also integrated further from the TS to the product, which yielded the reaction free energy for the forward reaction, $\Delta_r A_{fw}$.

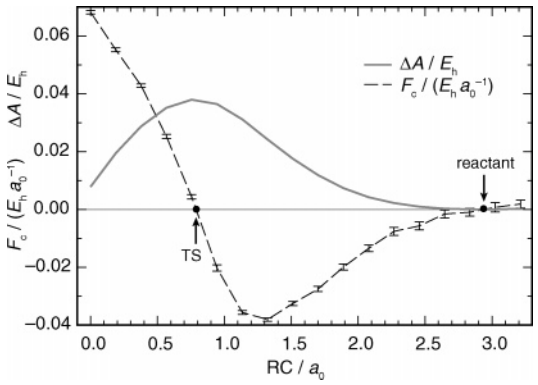


Figure 2. Average constraint force and its integral (i.e., the free-energy difference) along the RC. 40-ps snapshot, sampling from reactant to TS.

Details on the numerical integration and the error analysis are provided in Appendix A.1.

D. Analysis of MD Data. When calculating the average of any dynamical quantity from an MD simulation, e.g., for monitoring the equilibration process or obtaining a "measurement" such as the force of constraint, one needs a procedure to calculate the average and its variance from the MD data and criteria to assess if the average has become stationary. The former issue is somewhat complicated by the fact that MD data is inherently correlated, which means that the usual statistical formulas, e.g., to calculate the variance of the mean, are not applicable. Two main approaches to deal with this problem are conceivable. One can either (1) treat the correlations explicitly by extracting correlation information from the data and accounting for them in the calculation of the variance or (2) transform the data such as to remove the correlations and apply the standard formulas for statistically independent data. Both strategies have been used in the literature and their relative merits discussed.[49−54] The weakness of the first approach is that obtaining reliable estimates for the correlation coefficients[51] is not straightforward (see, e.g., the succinct treatment in ref 53) as different estimators involving different approximations may be chosen depending on the nature and the properties of the correlations present in the data. We have therefore chosen a decorrelation approach, which appears to be more robust. In particular, we have adopted the procedure suggested by Schiferl and Wallace,[50] which we summarize here.

The main idea of this treatment is to remove correlations by coarse-graining and then to subject the decorrelated data to statistical tests checking for stationarity and lack of correlation. The coarse-grained data set is generated simply by collecting successive original data points $\{X_i, i = 1, 2, ..., N\}$ into n nonoverlapping segments or bins of width m. (N is adjusted to make m an integer divider by deleting extraneous points from the beginning of the series.) We denote the bin means and variances by $\bar{x}_k$ and s_k^2 ($k = 1, 2, ..., n$). The $\bar{x}_k$ now constitute the coarse-grained data set, which is free of correlations, provided an appropriate choice of m, as discussed below.

The procedure can be outlined as follows:

(1) Plot the X_i and determine visually when the data series is apparently constant (no obvious drift or change in bandwidth). Discard the data before that point. Also determine the oscillation period, e.g., by counting the number of maxima over a certain interval at the end of the series.

(2) Coarse-grain the data, using the oscillation period as a trial value for m.

(3) Apply statistical tests for (a) lack of trend in the $\bar{x}_k$ (Mann−Kendall test for trend); (b) lack of trend in the s_k^2 (Mann−Kendall test for trend); (c) normal distribution of the $\bar{x}_k$ (Shapiro−Wilk W test for normality); and (d) lack of correlation in the $\bar{x}_k$ (von Neumann test for serial correlation). Details on these tests are provided in Appendix A.2.

(4) If either of the trend tests fails, drop more points from the start of the series. If they both succeed, but the normality and/or the correlation test fails, increase m.

(5) If so many points need to be eliminated or m increased so much that n falls below 24, extend the simulation to generate more data. (This minimum value for n secures the resolving power of the statistical tests.)

(6) If all tests pass, calculate mean, variance, and confidence interval for the series:

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n}\bar{x}_k \tag{2}$$

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(\bar{x}_k - \bar{X})^2 \tag{3}$$

$$\langle X\rangle = \bar{X} \pm t_{n-1}^{(\alpha/2)}\frac{S}{\sqrt{n}} \tag{4}$$

$t_{n-1}^{(\alpha/2)}$ is the quantile of Student's t distribution with $n - 1$ degrees of freedom at the chosen significance level α. We have worked with $\alpha = 0.025$ (i.e., a confidence level of 95%) throughout.

This procedure affords a well-defined and well-justified protocol to assess whether a given quantity has converged to a stationary value and at the same time yields this average value and a measure for its statistical error. The importance of applying well-defined statistical criteria has recently been highlighted also by Karplus and co-workers,[55] who elaborated on the idea of reverse cumulative averaging.[56]

Regarding the sampling times required to achieve stationary averages, we note that after the relatively long initial equilibration phase (tens of picoseconds), the sampling time at each value of the RC could be kept rather short (a few picoseconds). We did not observe any systematic dependence of the sampling time on the RC, the snapshot, or the starting point. The complete data from the thermodynamic integration simulations including the number of steps accepted in the final average, bin width, and average constraint force are provided as Supporting Information.

Results for PHBH and Discussion

A. Energetics. The free energies of activation for the hydroxylation of pOHB in PHBH calculated for the different snapshots (40, 80, 120, 160, and 200 ps) and starting points (reactant R and transition state TS) are collected in Table 2. The values lie within a range of 95 to 108 kJ mol^{-1} with an average of 101 kJ mol^{-1}, which is thus our best estimate for

Table 2. Free Energies of Activation, Values of the RC at the End-Points of the Thermodynamic Integration, and Values of the Components of the RC at the TS for the Hydroxylation of pOHB in PHBH

snapshot (ps) (start[b])	$\Delta^{\ddagger}A_{fw}$ (kJ mol^{-1})[c]	RC (Å)		$\langle d\rangle_{TS}$[a] (Å)	
		R	TS	O_p-O_d	C_m-O_d
40 (R)	101 ± 2	1.56	0.42	1.66	2.06
40 (TS)	101 ± 2	1.55	0.41	1.63	2.05
80 (R)	108 ± 2	1.62	0.42	1.66	2.03
80 (TS)	100 ± 2	1.53	0.42	1.64	2.04
120 (R)	106 ± 2	1.48	0.42	1.65	2.05
120 (TS)	105 ± 2	1.54	0.41	1.64	2.04
160 (R)	98 ± 2	1.77	0.44	1.65	2.05
160 (TS)	95 ± 2	1.65	0.45	1.68	2.08
200 (R)	102 ± 2	1.51	0.42	1.65	2.05
200 (TS)	97 ± 2	1.45	0.42	1.65	2.05

[a] Average distance at RC = 0.42 Å for 40 ps (TS), at RC = 0.4 Å otherwise. The error according to eq 4 is 0.01 Å, except for $\langle d(C_m-O_d)\rangle_{TS}$ in 80 ps (R), where it is 0.05 Å. [b] R: Integration from reactant to TS; TS: integration from TS to reactant. [c] Activation free energy of the forward reaction; error calculated according to eq 14.

$\Delta^{\ddagger}A_{fw}$. This value is significantly higher than the 49 kJ mol^{-1} obtained previously[21] also with AM1/GROMOS. However, those calculations differ from the ones reported here in several aspects, such as the absence of a solvation shell, a different reaction coordinate, and positional restraints on the atoms outside a region of about 7 Å around the active site, which precludes a direct comparison. With respect to corresponding experimental data, there is the value for the overall activation energy[15] of 49 kJ mol^{-1} (derived from an Arrhenius plot covering a temperature range of 21 K) that could serve as a reference, provided that hydroxylation is rate-limiting under the experimental conditions (pH 8.0).[16,17] Accepting this assumption, one can also convert[57] overall rate constants obtained under the same conditions[5,14] into a free energy of activation, affording about 64 kJ mol^{-1}. Our calculations thus appear to overestimate the barrier.

This discrepancy is partly accounted for by the known deficiency of AM1 to yield barriers that are too high. For instance, AM1 overestimates the barrier for the hydroxylation of pOHB by H_2O_2 by 45% compared to MP2.[21] Another factor that may contribute is the uncertainty about the protonation state of the enzyme. There are five histidine residues within 15 Å of the active site that were all taken as neutral but can carry an additional proton depending on pH and local pK_a. A mutation study that replaced a negatively charged glutamate some 10 Å away from pOHB by a neutral glutamine (Glu49Gln[9]), thus enhancing the positive electrostatic field in the active site, found that this seemingly subtle change leads to a 4-fold increase of the hydroxylation rate. Protonating one or more of the histidines around the active site is hence likely to decrease the barrier. However, we have not investigated this possibility in the present work.

The barriers obtained for both starting points of the same snapshot agree within 5 kJ mol^{-1} in all cases except the 80-ps snapshot, where the difference is 8 kJ mol^{-1}. This variability as well as the spread between different snapshots is a consequence of incomplete sampling. In the ideal case of exhaustive sampling, we would expect to obtain the same value irrespective of the history of the system. For real sampling conditions, however, the system retains some memory of the previous treatment (that is, the classical MD simulations, the QM/MM optimizations, and the previous TI steps) that is not averaged out fully. The observed variability thus reflects differences in the environment that affect the height of the free-energy barrier.

Table 3. Convergence Behavior of the Activation Free Energy with Increasing Sampling Time for the 80-ps Snapshot

sampling length[a]	$\Delta^{\ddagger}A_{fw}$ (kJ mol^{-1})	
	R[b]	TS[b]
first 5 ps	108.0 ± 1.9	99.8 ± 2.1
first 10 ps	105.6 ± 1.6	100.8 ± 1.5
first 20 ps	105.6 ± 1.0	101.5 ± 1.1
first 30 ps	104.6 ± 1.0	102.2 ± 1.1
first 40 ps	103.8 ± 0.9	101.1 ± 1.1
all data	103.6 ± 0.8	100.6 ± 0.7

[a] Amount of raw data considered. Following the analysis procedure, a variable number of steps is discarded from the beginning of the series at each integration point. [b] R: Integration from reactant to TS; TS: integration from TS to reactant.

To investigate the influence of sampling time, we chose the 80-ps snapshot, where the results from the two starting points differed most. We sampled for an additional 40 ps at each value of the RC, thus accumulating at least 45 ps per integration point. We evaluated the free energy by including successively more data into the analysis. The results of this convergence study are shown in Table 3. While the value for snapshot 80 ps (TS) remains essentially constant upon extending the sampling time, the activation free energy calculated from snapshot 80 ps (R) decreases by about 4 kJ mol^{-1}. This brings it close to the average of the other snapshots, and the difference between the results from the two starting points is reduced from 8 to 3 kJ mol^{-1}. We conclude that the observed variability in the results is indeed due to incomplete sampling and can be reduced by increasing the sampling time. Furthermore, the approach to use several starting points provides a relatively cost-efficient way to validate the results and identify "outliers"; bringing these to convergence can, however, require considerable additional effort.

From the full thermodynamic integration of snapshot 40 ps from the TS to both the reactant and the product state, we obtained the free energy of reaction as $\Delta_r A_{fw} = -212$ kJ mol^{-1}, which is in perfect in agreement with the result of the previous AM1/GROMOS study.[21] Combining the reaction energy with the barrier for the forward reaction, we obtain for the activation free energy of the backward reaction $\Delta^{\ddagger}A_{bw} = 313$ kJ mol^{-1}. The value of the RC at the product was −1.51 Å.

B. Structural Features. With regard to structural features, we focus on interactions of pOHB with active site residues and their changes along the RC. There is quite a detailed understanding from experimental studies of how the dianionic substrate is accommodated in the active site (see Figure 3). The phenolate oxygen is stabilized by a hydrogen bond to Tyr 201, from which a H-bond chain extends via Tyr 385 and several water molecules to His 72.[5] The carboxy group forms a salt bridge with Arg 214[6] and accepts hydrogen

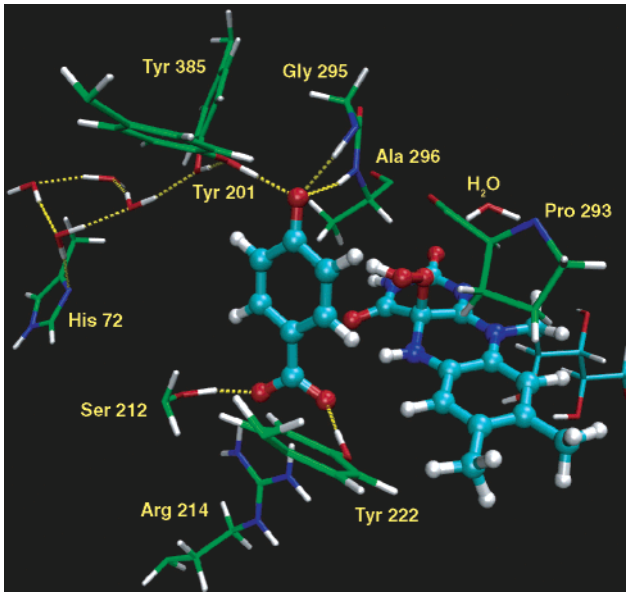


Figure 3. Active site of PHBH with FADHOOH, bound pOHB (both cyan), and selected residues (green). QM region in ball-and-sticks representation, MM part as sticks. Unless necessary, only the side chains of amino acids are shown.

bonds from Ser 212[8] and Tyr 222.[7] In addition, we find hydrogen bonds from the backbone amide NH of Gly 295 and Ala 296 to the pOHB phenolate. Special attention has been given to the role of Pro 293. Previous calculations[18,20] found that the FADHOOH hydroperoxy group forms a hydrogen bond to the backbone amide carbonyl oxygen of Pro 293 as the reaction proceeds toward the transition state. This stabilization was identified as the only significant specific interaction of the protein with the transition state.[20] Moreover, as the peroxide O−O bond breaks, the negative charge developing on the proximal oxygen is stabilized by a hydrogen bond from a nearby water molecule.[18]

We have followed the hydrogen bonding pattern in the active site along the reaction coordinate. Figure 4 shows, as a representative example, pertinent distances for snapshot 40 ps (R). The features we discuss here, however, apply to all 10 simulations, if not mentioned otherwise. pOHB is kept in place by several permanent interactions that remain stable over the course of the reaction. The carboxylate is stabilized through hydrogen bonds from the hydroxy groups of Tyr 222 and Ser 212 as well as by the salt bridge to the guanidinium moiety of Arg 214 (Figure 4e). The phenolate oxygen accepts stable hydrogen bonds from the hydroxy group of Tyr 201 and the amide NH of Gly 295 (Figure 4c). The hydrogen bond from the relatively acidic phenol-OH is somewhat shorter than the one from the amide NH. Until before the transition state, there is also a hydrogen bond from

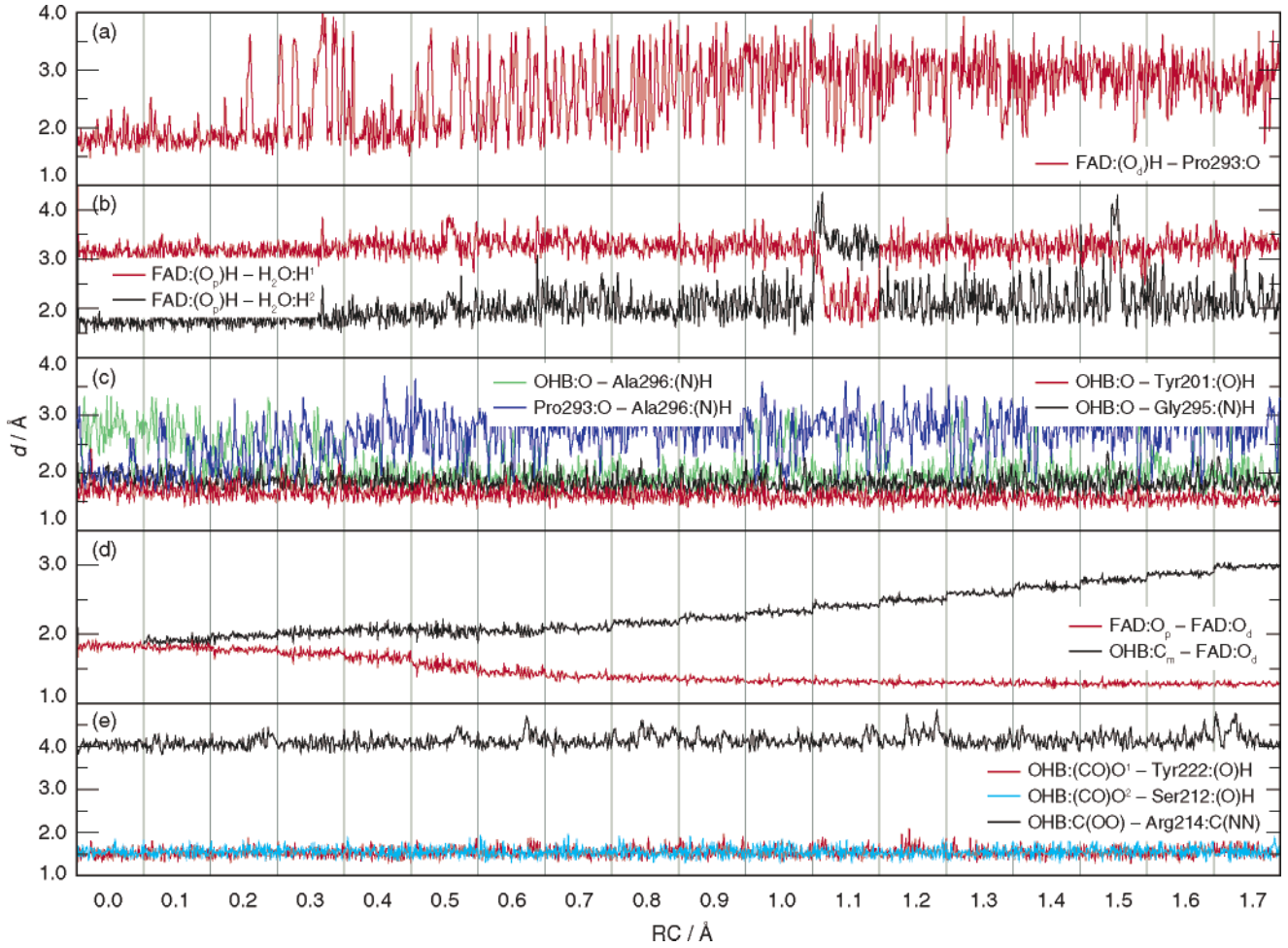


Figure 4. Selected distances along the RC for snapshot 40 ps (R). The reaction proceeds from right to left; the TS is at 0.4 Å. Each vertical panel shows the last 4 ps of sampling at the given value of the RC.

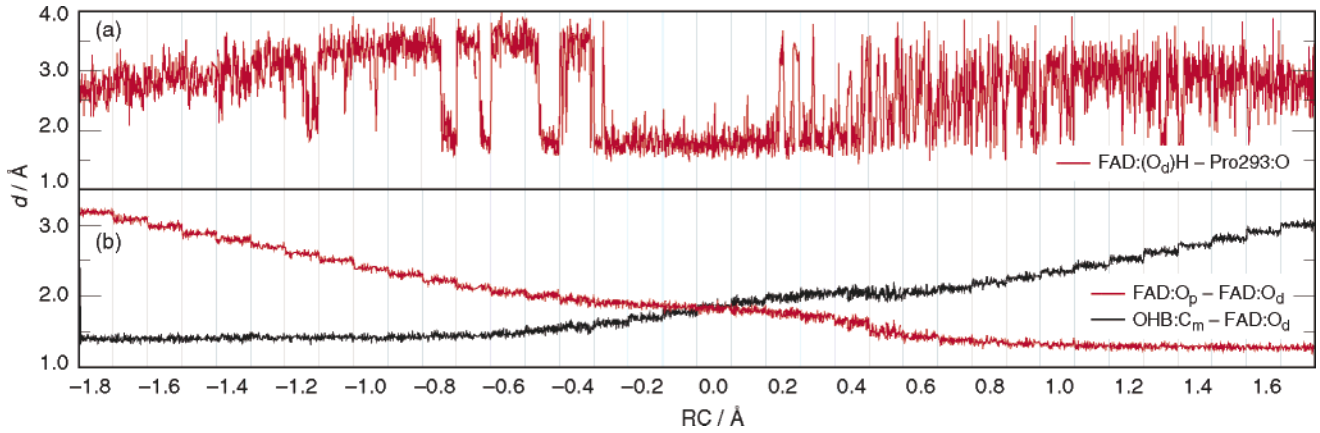


Figure 5. Selected distances along the RC for snapshot 40 ps (TS). The complete reaction from reactants (on the right) to products (on the left) is covered; the TS is at 0.4 Å. Each vertical panel shows the last 4 ps of sampling at the given value of the RC.

the amide NH of Ala 296 to the pOHB phenolate. This interaction is lost around the TS, when the phenolate is transformed into a carbonyl moiety, which is less basic and hence a weaker hydrogen-bond acceptor. Likewise, the remaining two hydrogen bonds from Tyr 201 and Gly 295 lengthen by about 0.2 Å as the system proceeds from reactants to the TS.

On the other hand, the oxido oxygen of FADHO resulting from breaking the O−O bond in FADHOOH is a much better hydrogen-bond acceptor than the hydroperoxy group. This is reflected in the development of a strong hydrogen bond from a water molecule (Figure 4b). Initially, this water molecule is only weakly associated with the proximal hydroperoxy oxygen. The average distance is around 2.3 Å, with large fluctuations, and it occasionally rotates (and binds through the other hydrogen atom, e.g., at RC = 1.1 Å) or drifts away (RC = 1.5 Å). Toward the TS, the interaction becomes more stable, the distance fluctuations subside, and a strong, stable hydrogen bond is established from the TS onward.

The proposed capability of Pro 293 of "shepherding" the transferring OH moiety by a hydrogen bond to its amide oxygen is of particular interest. As seen in Figure 4a, the frequency of hydrogen bonding events from the transferred OH to the Pro 293 amide oxygen increases as the reaction proceeds toward the TS. However, only at the TS (RC = 0.4 Å) a stable hydrogen bond is established for the first time. It is broken and reformed several times after the TS, before becoming permanent toward RC = 0.0 Å. From the complete thermodynamic integration carried out for snapshot 40 ps (TS) (see Figure 5a), we find the reverse pattern as the system moves away from the TS toward the products, the hydrogen bond between OH and Pro 293 breaking more and more frequently and finally staying broken. Unlike the other hydrogen bonds discussed above, the OH−Pro 293 interaction varies widely among different snapshots. In some cases, a stable interaction exists at the TS; in other cases, the bond is only established later along the RC or not at all.

Another factor of interest in this context is the hydrogen bond from the Ala 296 amide NH to the amide oxygen of Pro 293 as it competes for Pro 293 against the transferred OH group. This interaction is complementary to the bond from Ala 296 to the pOHB phenolate oxygen, that is, one of the bonds is present at any given time (see Figure 4c). In most simulations, the Ala 296−pOHB interaction is almost stable until the TS, with a few switching events to Ala 296−Pro 293, and is permanently replaced by Ala 296−Pro 293 somewhere after the TS. However, the exact location of the switch from Ala 296−pOHB to Ala 296−Pro293 varies for different snapshots or does not happen at all.

To obtain a more quantitative measure for the presence or absence of these two hydrogen bonds, we determined for each point along the RC a bonding index (or bonding probability), that is, the fraction of the time the bond is present over the last 4 ps. The criterion for a bond is a distance of 2.5 Å or less. We find the following general trends for the bond indices along the RC: (i) The OH−Pro 293 bond index is around 0.1 from the reactants up until RC = 1.0 Å, when it starts rising. (ii) At the TS (RC = 0.4 Å), the OH−Pro 293 bond index scatters widely, between 0.44 and 0.93; at RC = 0.0 Å, the bond index is greater than 0.8 for most snapshots. (iii) The Ala 296−Pro 293 bond index remains around 0.1 up until the TS, where it increases sharply to 0.9 or larger at RC = 0.0 Å in all cases but one.

To elucidate specifically the effect on the transition state and the barrier, we averaged the bond indices over the TS and the two adjacent points (i.e., over RC = 0.5, 0.4, and 0.3 Å). The results of this analysis are shown in Table 4. The average bond index for OH−Pro 293 shows some correlation with the activation free energy, e.g., snapshots 160 ps (TS) and 200 ps (TS) have a low barrier and a high bond index, which would be expected from the argument that this hydrogen bond stabilizes the TS. However, it fails in other cases, where the bond index is low, but the barrier is not, e.g., 160 ps (R) or 80 ps (TS). The corresponding inverse correlation of the barrier height with the Ala 296−Pro 293 bond index is similarly inconclusive. The most consistent trend is found for the difference between the two bond indices, but there are exceptions here as well. For instance, 40 ps (TS) should have a particularly low activation free energy and 80 ps (TS) a high one, which is not the case. A remarkable detail in Table 4 is the observation that only in the two 80-ps snapshots both hydrogen bonds are equally

Table 4. Bond Indices for the Hydrogen Bonds to the Pro 293 Amide Oxygen from the FADHOOH OH and the Ala 296 Amide NH, Averaged over the TS Region

snapshot (ps) (start[b])	average bond index[a]		
	$O_dH-Pro293$	Ala296−Pro293	difference
40 (R)	0.68	0.38	0.30
40 (TS)	0.72	0.20	0.52
80 (R)	0.49	0.55	−0.06
80 (TS)	0.48	0.54	−0.06
120 (R)	0.56	0.32	0.24
120 (TS)	0.68	0.42	0.25
160 (R)	0.51	0.12	0.38
160 (TS)	0.75	0.38	0.37
200 (R)	0.66	0.41	0.25
200 (TS)	0.78	0.49	0.29

[a] The bond index was determined for each point of the RC as described in the text. Tabulated are the values averaged over RC = 0.3, 0.4, 0.5 Å. [b] R: Integration from reactant to TS; TS: integration from TS to reactant.

probable around the TS, while the bond to the transferred OH is otherwise clearly more abundant.

In conclusion, we can identify two distinct groups of hydrogen bonds in the active site. There is a set of permanent hydrogen bonds that stabilize the substrate and keep it in place. Another set of hydrogen bonds rearranges as the reaction proceeds, adapting to the changing charge distribution in the active site. In particular, we find a rearrangement of the hydrogen bonding network around the amide oxygen of Pro 293 in the TS region that is very likely to stabilize specifically the transferred OH group and thus the transition state. However, it has not been possible to single out this interaction as the determining factor controlling the barrier height.

Finally, we consider the course of the two components of the reaction coordinate, the hydroperoxy O_p-O_d and pOHB C_m-O_d distances (Figure 4d). Until RC = 1.0 Å, the O−O bond is almost unaffected, the RC acting to bring the substrate closer to the cofactor. The O−O bond then starts lengthening, slowly until RC = 0.7 Å, more rapidly between RC = 0.7 and 0.3 Å. During this latter period, the C_m-O_d distance remains almost constant. This is the phase where the O−O bond is broken. From the full simulation on snapshot 40 ps (TS) (Figure 5b) we find that from RC = 0.2 to −0.6 Å, the C−O distance smoothly shortens, that is, the C−O bond is forming, while the O−O distance increases, that is, 3,4-DOHB drifts away from the cofactor. Only at about RC = −0.7 Å, the C−O bond is fully formed and remains constant. From RC = 0.3 Å onward, we can also observe the change in hybridization of the attacked C_m atom from sp^2 to sp^3, as reflected in the dihedral angle $C_o-C_m-C_p-H_m$.

IV. Conclusions

We have investigated the OH transfer step of the hydroxylation reaction of *p*-hydroxybenzoate in the enzyme *p*-hydroxybenzoate hydroxylase (PHBH). Using thermodynamic integration with AM1/GROMOS QM/MM molecular dynamics simulations, we have calculated the activation free energy of the reaction. We have used 10 different starting structures, obtained from classical MD simulations and QM/MM optimizations of stationary points. Our model included a shell of solvating water molecules around the protein, and the full system except the outer water shell was left unconstrained. From these calculations, we have obtained a free-energy barrier of $\Delta^{\ddagger}A_{fw} = 101$ kJ mol^{-1} as our best estimate. As AM1 is known to overestimate reaction barriers, this value is likely to be too high. For the reaction free energy, we have obtained $\Delta_r A_{fw} = -212$ kJ mol^{-1}. The simulations have also provided detailed insight into the changes in the hydrogen bonding network in the active site and the cleaving and forming of the covalent bonds along the course of the reaction. With respect to a stabilization of the transition state by hydrogen bonds, we have confirmed that Pro 293 interacts with the transferring OH in the transition state region. However, it has not been possible to pinpoint this interaction, or any other, as the controlling factor for the barrier height.

Moreover, we have described a procedure involving coarse-graining of the data and statistical tests to ensure that converged averages are obtained from MD simulations. The procedure also yields a measure for the statistical error and hence allowed us to assess the magnitude of different sources of error. For the simulation conditions used, the error in the final free-energy difference is dominated by the error due to incomplete sampling, as seen by the differences in the results for different starting points. The pure statistical error, on the other hand, is mainly caused by the variance of the average constraint force, whereas the error from the numerical integration procedure is negligible.

QM/MM calculations treating the QM region at a higher level of theory are currently underway in our laboratory, which will provide a better assessment of the performance of AM1.

Acknowledgment. We thank Dr. S. Billeter, IBM Zürich Research Laboratory, Rüschlikon, Switzerland, for helpful discussions and the implementation of the difference-of-distances constraint in ChemShell.

Appendix

A.1. Calculation of the Free Energy and Error Analysis. The difference-of-distances constraint involving the three atoms *A*, *B*, and *C* has the holonomic form

$$\sigma(\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C) = |\mathbf{R}_{BA}| - |\mathbf{R}_{CB}| - \xi = 0 \qquad (5)$$

where $|\mathbf{R}_{IJ}| = |\mathbf{R}_J - \mathbf{R}_I|$ and ξ is the value of the constraint. The iterative calculation of the constraint force with the SHAKE algorithm[38,58] requires the derivatives with respect to the atomic positions involved

$$\nabla_A \sigma = \hat{\mathbf{R}}_{BA}$$

$$\nabla_B \sigma = -(\hat{\mathbf{R}}_{BA} + \hat{\mathbf{R}}_{CB})$$

$$\nabla_C \sigma = \hat{\mathbf{R}}_{CB} \qquad (6)$$

where $\hat{\mathbf{R}}$ designates unit vectors. The constraint force F_c is identified with the converged Lagrangean multiplier γ that guarantees that the constraint is exactly satisfied in the next

time step. Generalizing to L constraints, the positions are propagated as

$$\mathbf{R}_I(+) = \bar{\mathbf{R}}_I(+) - \frac{\Delta^2}{M_I}\sum_{k=1}^{L}\gamma_k \nabla_I \sigma_k|_0 \quad (7)$$

where $\bar{\mathbf{R}}_I(+)$ is the new position propagated *without* constraints, M_I is the mass of atom I, and Δ is the time step.

The complete expression for the free-energy difference including the metric-tensor correction reads[47]

$$\Delta A_{a\to b} = \int_a^b \mathrm{d}\xi \langle F_c \rangle_\xi - k_\mathrm{B}T \ln \frac{\langle |\mathbf{Z}|^{-1/2} \rangle_{\xi=b}}{\langle |\mathbf{Z}|^{-1/2} \rangle_{\xi=a}} \quad (8)$$

where $|\mathbf{Z}|$ is the determinant of the matrix

$$Z_{mn} = \sum_J M_J^{-1} \nabla_J \xi_m \nabla_J \xi_n \quad (9)$$

(also referred to as the Fixman determinant[59]), where the sum runs over all atoms involved in any constraint. For the case of only one constraint, $|\mathbf{Z}|$ simplifies to

$$z = \sum_J M_J^{-1} (\nabla_J \xi)^2 \quad (10)$$

Evaluating explicitly for the difference-of-distances constraint, eq 5, yields

$$\begin{aligned} z &= M_A^{-1}\hat{\mathbf{R}}_{BA}^2 + M_B^{-1}(-\hat{\mathbf{R}}_{BA} - \hat{\mathbf{R}}_{CB})^2 + M_C^{-1}\hat{\mathbf{R}}_{CB}^2 \\ &= M_A^{-1} + 2M_B^{-1}[1 + \cos(\hat{\mathbf{R}}_{BA}, \hat{\mathbf{R}}_{CB})] + M_C^{-1} \end{aligned} \quad (11)$$

Apart from constant mass terms that will cancel out in eq 8, the correction thus depends only on the angle between the two bonds involved in the constraint or, rather, on the change in this angle between the end points of the integration. In the present application, the angle $\angle(\mathbf{R}_{OO}, \mathbf{R}_{CO})$ varies between about 140° and 170° for reactant and TS, respectively, which amounts to a correction of the order of -1 kJ mol^{-1}.

Provided the average constraint force at a series of points ξ_i along the RC, $\langle F_c \rangle_i$, polynomials of third degree were used to interpolate between successive points with Mathematica.[60] The resulting analytical representation was used to determine the value of the RC at the stationary points (where $\langle F_c \rangle = 0$) and to obtain the free-energy difference between these points by integration.

To quantify the error in the free energy, we consider two contributions: (1) The error of the numerical integration procedure and (2) the error due to the uncertainty in the $\langle F_c \rangle_i$. To assess the magnitude of these two error sources, we perform the analysis for integration according to Simpson's rule[61]

$$I = \int_a^b \mathrm{d}x\, f(x) \approx \frac{h}{3}(f_1 + 4\sum_{\substack{i=2 \\ i\ \mathrm{even}}}^{N-1} f_i + 2\sum_{\substack{i=3 \\ i\ \mathrm{odd}}}^{N-2} f_i + f_N) \quad (12)$$

where $f_i = f(x_i)$ are the values of the function $f(x)$ at the (equispaced) grid points x_i, h is the grid spacing, and N is the total number of points; N must be odd. Simpson's rule approximates the function by second-degree polynomials defined by three successive points. The error of Simpson's rule is given by

$$|e_I| \le \frac{(b-a)h^4}{180}|f^{(4)}(x^*)| = \frac{(N-1)h^5}{180}|f^{(4)}(x^*)| \quad (13)$$

where $|f^{(4)}(x^*)|$ is an upper limit for the absolute value of the fourth derivative of $f(x)$ on $[a, b]$. For values as they have typically occurred in the present application (integration interval 2.28 a_0, $h = 0.19$ a_0, $|f^{(4)}(x^*)| \approx 3.5$ $E_\mathrm{h}a_0^{-5}$), we estimate $|e_I| \le 0.15$ kJ mol^{-1}. ($|f^{(4)}(x^*)|$ was obtained numerically from typical force curves using Mathematica.)

To calculate the error contribution of the uncertainty in the f_i, given by their variances s_i^2, to the value of the integral, we apply Gaussian error propagation to Simpson's rule, obtaining

$$s_I^2 = \sum_i \left(\frac{\partial I}{\partial f_i}\right)^2 s_i^2 = \left(\frac{h}{3}\right)^2 (s_1^2 + 16\sum_{\substack{i=2 \\ i\ \mathrm{even}}}^{N-1} s_i^2 + 4\sum_{\substack{i=3 \\ i\ \mathrm{odd}}}^{N-2} s_i^2 + s_N^2) \quad (14)$$

With $h = 0.19$ a_0, $N = 13$, and assuming a constant, typical value of $s_i = 10^{-3}$ $E_\mathrm{h}a_0^{-1}$ for the uncertainty of the constraint force, we estimate $s_I = 1.8$ kJ mol^{-1}.

We therefore conclude from this analysis that the error of the integral, that is, of the free-energy difference, is dominated by the error due to the uncertainty in the constraint force for the simulation conditions and parameters used in the present work. Accordingly, the reported error bars for ΔA have been calculated from eq 14.

A.2. Statistical Tests for Establishing Converged Averages. The first test to be applied to the coarse-grained data is the Mann−Kendall test for trend[62,63] because it is a nonparametric (i.e., distribution-free) test, unlike alternative trend tests which assume a normal distribution, and because the absence of a trend is a prerequisite for the following tests. The Mann−Kendall test was implemented as in ref 50.

The Shapiro−Wilk W statistic[64] is a very powerful and universal measure for departure from normality, testing only for shape, but not for a specific mean and variance of the normal distribution. We have used a modern implementation of the Shapiro−Wilk test[65−67] that is valid also for larger sample sizes. Code for the algorithm[65] and the required auxiliary routines[68−70] was obtained from StatLib.[71]

To check for correlation in the trend-free, normal data, we have used the von Neumann test for serial correlation[72] as described in ref 50.

Supporting Information Available: Full data from the thermodynamic integration simulations for each sampling run, including number of steps sampled, information on the averaging procedure, and average force of constraint. This information is available free of charge via the Internet at http://pubs.acs.org/.

References

(1) Entsch, B.; van Berkel, W. J. H. *FASEB J.* **1995**, *9*, 476−483.

(2) Jadan, A. P.; Moonen, M. J. H.; Boeren, S.; Golovleva, L. A.; Rietjens, I. M. C. M.; van Berkel, W. J. H. *Adv. Synth. Catal.* **2004**, *346*, 367−375.

(3) Ortiz-Maldonado, M.; Ballou, D. P.; Massey, V. *Biochemistry* **1999**, *38*, 8124−8137.

(4) Entsch, B.; Ballou, D. P.; Massey, V. *J. Biol. Chem.* **1976**, *251*, 2550−2563.

(5) Entsch, B.; Palfey, B. A.; Ballou, D. P.; Massey, V. *J. Biol. Chem.* **1991**, *266*, 17341−17349.

(6) Clarkson, J.; Palfey, B. A.; Carey, P. R. *Biochemistry* **1997**, *36*, 12560−12566.

(7) van der Bolt, F. J. T.; Vervoort, J.; van Berkel, W. J. H. *Eur. J. Biochem.* **1996**, *237*, 592−600.

(8) Moran, G. R.; Entsch, B.; Palfey, B. A.; Ballou, D. P. *Biochemistry* **1999**, *38*, 6292−6299.

(9) Ortiz-Maldonado, M.; Cole, L. J.; Dumas, S. M.; Entsch, B.; Ballou, D. P. *Biochemistry* **2004**, *43*, 1569−1579.

(10) Palfey, B. A.; Moran, G. R.; Entsch, B.; Ballou, D. P.; Massey, V. *Biochemistry* **1999**, *38*, 1153−1158.

(11) Palfey, B. A.; Basu, R.; Frederick, K. K.; Entsch, B.; Ballou, D. P. *Biochemistry* **2002**, *41*, 8438−8446.

(12) Ortiz-Maldonado, M.; Entsch, B.; Ballou, D. P. *Biochemistry* **2003**, *42*, 11234−11242.

(13) Entsch, B.; Ballou, D. P. *Biochim. Biophys. Acta* **1989**, *999*, 313−322.

(14) Husain, M.; Entsch, B.; Ballou, D. P.; Massey, V.; Chapman, P. J. *J. Biol. Chem.* **1980**, *255*, 4189−4197.

(15) van Berkel, W. J. H.; Müller, F. *Eur. J. Biochem.* **1989**, *179*, 307−314.

(16) Vervoort, J.; Rietjens, I. M. C. M.; van Berkel, W. J. H.; Veeger, C. *Eur. J. Biochem.* **1992**, *206*, 479−484.

(17) Ridder, L.; Mulholland, A. J.; Vervoort, J.; Rietjens, I. M. C. M. *J. Am. Chem. Soc.* **1998**, *120*, 7641−7642.

(18) Ridder, L.; Mulholland, A. J.; Rietjens, I. M. C. M.; Vervoort, J. *J. Mol. Graphics Modell.* **1999**, *17*, 163−175.

(19) Ridder, L.; Palfey, B. A.; Vervoort, J.; Rietjens, I. M. C. M. *FEBS Lett.* **2000**, *478*, 197−201.

(20) Ridder, L.; Harvey, J. N.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Phys. Chem. B* **2003**, *107*, 2118−2126.

(21) Billeter, S. R.; Hanser, C. F. W.; Mordasini, T. Z.; Scholten, M.; Thiel, W.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2001**, *3*, 688−695.

(22) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf and BIOMOS b.v.: Zürich, Switzerland, and Groningen, The Netherlands, 1996.

(23) van Gunsteren, W. F.; Berendsen, H. J. C. *GROMOS87 Program and Manual*; Laboratory of Chemical Physics, University of Groningen: Groningen, The Netherlands, 1987.

(24) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177−2186.

(25) 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. R.; 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.

(26) Thiel, S.; Senn, H. M.; Thiel, W. Manuscript in preparation.

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

(28) *ChemShell*, V. 3.0a3, 2004.

(29) Thiel, W. *MNDO99*, V. 6.1; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2003.

(30) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451−5469.

(31) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 3169−3174.

(32) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

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

(34) Nosé, S. *Mol. Phys.* **1984**, *52*, 255−268.

(35) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.

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

(37) Jang, S.; Voth, G. A. *J. Chem. Phys.* **1997**, *107*, 9514−9526.

(38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(39) Kutteh, R.; Straatsma, T. P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1998; Vol. 12; pp 75−136.

(40) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.

(41) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472−477.

(42) Ciccotti, G.; Ferrario, M.; Hynes, J. T.; Kapral, R. *Chem. Phys.* **1989**, *129*, 241−251.

(43) Paci, E.; Ciccotti, G.; Ferrario, M.; Kapral, R. *Chem. Phys. Lett.* **1991**, *176*, 581−587.

(44) Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737−7744.

(45) den Otter, W. K.; Briels, W. J. *J. Chem. Phys.* **1998**, *109*, 4139−4146.

(46) Schlitter, J.; Klähn, M. *J. Chem. Phys.* **2003**, *118*, 2057−2060.

(47) Schlitter, J.; Klähn, M. *Mol. Phys.* **2003**, *101*, 3439−3443.

(48) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175−1188.

(49) Jorgensen, W. L. *Chem. Phys. Lett.* **1982**, *92*, 405−410.

(50) Schiferl, S. K.; Wallace, D. C. *J. Chem. Phys.* **1985**, *83*, 5203−5209.

(51) Straatsma, T. P.; Berendsen, H. J. C.; Stam, A. J. *Mol. Phys.* **1986**, *57*, 89−95.

(52) Bishop, M.; Frinks, S. *J. Chem. Phys.* **1987**, *87*, 3675−3676.

(53) Flyvbjerg, H.; Petersen, H. G. *J. Chem. Phys.* **1989**, *91*, 461−466.

(54) Morales, J. J.; Nuevo, M. J.; Rull, L. F. *J. Comput. Phys.* **1990**, *89*, 432−438.

(55) Yang, W.; Bitetti-Putzer, R.; Karplus, M. *J. Chem. Phys.* **2004**, *120*, 2618−2628.

(56) Pearlman, D. A. *J. Comput. Chem.* **1994**, *15*, 105−123.

(57) $\Delta^{\ddagger}G = RT \ln[k_B T/(hk_{obs})]$, where k_{obs} is the rate constant, T is the temperature, R is the universal gas constant, k_B is Boltzmann's constant, and h is Planck's constant.

(58) Ryckaert, J.-P. *Mol. Phys.* **1985**, *55*, 549−556.

(59) Fixman, M. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3050−3053.

(60) *Mathematica*, V. 5.0; Wolfram Research: Champaign, IL, 2003.

(61) Weisstein, E. W. *Simpson's Rule*; 2004; http://mathworld.wolfram.com/SimpsonsRule.html.

(62) Mann, H. B. *Econometrica* **1945**, *13*, 245−259.

(63) Kendall, M. G.; Gibbons, J. D. *Rank correlation methods*; Arnold: London, 1990.

(64) Shapiro, S. S.; Wilk, M. B. *Biometrika* **1965**, *52*, 591−611.

(65) Royston, P. *Appl. Statist.* **1995**, *44*, 547−551.

(66) Royston, P. *Statistician* **1993**, *42*, 37−43.

(67) Royston, P. *Statist. Comput.* **1992**, *2*, 117−119.

(68) Royston, J. P. *Appl. Statist.* **1982**, *31*, 176−180.

(69) Hill, I. D. *Appl. Statist.* **1973**, *22*, 424−427.

(70) Wichura, M. J. *Appl. Statist.* **1988**, *37*, 477−484.

(71) Vlachos, P. *StatLib − Applied Statistics algorithms*; Department of Statistics, Carnegie Mellon University: Pittsburgh, PA, 2004; http://lib.stat.cmu.edu/apstat/.

(72) von Neumann, J. *Ann. Math. Stat.* **1941**, *12*, 367−395.

CT049844P

JCTC Journal of Chemical Theory and Computation

Determination of Intermolecular Hydrogen Bonded Conformers of α-Aryloxypropanoic Acids Using Density Functional Theory Predictions of Vibrational Absorption and Vibrational Circular Dichroism Spectra

Jiangtao He and Prasad L. Polavarapu*

Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235

Received October 29, 2004

Abstract: The density functional theoretical predictions of vibrational absorption (VA) and vibrational circular dichroism (VCD) spectra for monomeric chiral α-aryloxypropanoic acids are found to be in serious disagreement with corresponding experimental spectra. This disagreement made it difficult to establish the predominant conformations and absolute configuration of chiral α-aryloxypropanoic acids. Since carboxylic acids tend to form dimers through intermolecular hydrogen bonding in solution, such hydrogen bonding should be either eliminated in the experimental measurements or modeled properly in the theoretical calculations for achieving agreement between theoretical predictions and experimental observations. Esterification eliminates hydrogen bonding in solution, but this approach does not provide information about the conformations of the acids. As a consequence, the option to properly model the dimers of α-aryloxypropanoic acids and to evaluate the dimer conformations in reproducing the experimental spectra becomes important. Here we report a method for obtaining the dimer conformations of α-aryloxypropanoic acids using density functional theory and evaluate their reliability by comparing the predicted and experimental VA and VCD spectra. The population weighted predicted VA and VCD spectra of dimers matched well with the corresponding experimental spectra in solution, thereby indicating the predominant dimer conformers and absolute configuration of α-aryloxypropanoic acids.

Introduction

α-Aryloxypropanoic acids are an important class of chiral chemicals used as herbicides. These herbicides include 2-(4-chloro-2-methylphenoxy)propanoic acid and 2-(2,4-dichlorophenoxy)propanoic acid, which are marketed under brand names Mecoprop and Dichloroprop, respectively. Each of these two compounds can exist in two enantiomeric forms. Only their (+)-enantiomers are herbicidally active, while the (−)-enantiomers are devoid of herbicidal activity but are powerful antiauxins.[1−3] These compounds may undergo racemization by inversion of the absolute configuration at the asymmetrically substituted C-atom.[4] It was also found that microbial degradation of racemic herbicides preferentially eliminates one enantiomer over the other.[5] To understand such biological interactions at the molecular level it is necessary to know the absolute configuration and predominant conformations of chiral herbicides.

Two other α-aryloxypropanoic acids of importance are 2-(2-chlorophenoxy)propanoic acid and 2-(3-chlorophenoxy)-propanoic acid, which are closely related to Mecoprop and Dichloroprop. To eliminate the influence of hydrogen bonding effects in solution, methyl esters of propanoic acids, namely methyl 2-(4-chloro-2-methylphenoxy)propanoate, methyl 2-(2,4-dichlorophenoxy)propanoate, methyl 2-(2-chlorophenoxy)propanoate, and methyl 2-(3-chlorphenoxy)-propanoate, have been synthesized[6,7] and subjected to experimental and theoretical investigations. From these investigations the absolute configurations of esters, and those of parent acids there from, have been determined. Although these prior studies helped in determining the absolute configurations and provided insight into the conformations of esters, they did not provide any information on the

* Corresponding author phone: (615)322-2836; fax: (615)322-4936; e-mail: Prasad.L.Polavarapu@vanderbilt.edu.

10.1021/ct049905j CCC: $30.25
Published on Web 03/25/2005

predominant conformations of parent acids. The objective of this manuscript is to determine the predominant conformations of α-aryloxypropanoic acids.

Hydrogen bonding plays a very important role in the structures of molecules ranging from water[8] to biological molecules such as DNA[9] and proteins.[10] This is also the case for small carboxylic acids such as formic acid, acetic acid, and propanoic acid. The experimental and theoretical studies show that these carboxylic acids tend to form dimers through intermolecular hydrogen bonding.[11] There are also several efforts in understanding hydrogen bonding using different theoretical methods, including ab initio and density functional theory (DFT).[12]

Vibrational circular dichroism (VCD) is an important chiroptical technique for studying the absolute configuration and conformations of chiral molecules.[13−16] The combination of experimental VCD spectra and corresponding density functional theory predictions provides a powerful method for determining the absolute configuration and predominant conformations of chiral molecules.[16] But, most theoretical VCD spectral calculations to date have dealt with monomer molecules. The calculations for dimer molecules with flexible structure are less common because such studies will require much more computational effort. A few attempts have been made to investigate dimers, where monomers have only a limited number of possible starting conformation.[17] For monomers with several possible starting conformations, we are not aware of any attempts trying to address intermolecular hydrogen bonded dimers. In the present study, we address this problem and suggest an approach to deal with monomers possessing several possible starting conformations.

In this manuscript we report investigations for four α-aryloxypropanoic acids, which are 2-(2-chlorophenoxy)propanoic acid **1**, 2-(3-chlorophenoxy)propanoic acid **2**, 2-(4-chloro-2-methylphenoxy)propanoic acid **3**, and 2-(2,4-dichlorophenoxy)propanoic acid **4**. We have optimized the dimer conformations of these propanoic acids and calculated their VA and VCD spectra. The good agreement obtained between the population weighted predicted spectra for dimer conformations and the experimental spectra in solution further validates the predominant conformations of four acids investigated.

Although the biological or agricultural relevance of compounds **1**−**4** relates to their use in aqueous environments, infrared absorption or VCD spectroscopy is not a convenient technique for studying aqueous solutions due to strong absorption of infrared radiation by water. Furthermore, in aqueous environments intermolecular hydrogen bonding between solute and solvent molecules becomes as important as intermolecular hydrogen bonding between solute molecules. These competing hydrogen-bonding effects make it difficult to obtain reliable theoretical predictions. For this reason, the present experimental investigations are carried out in $CDCl_3$ solvent, which is not expected to interfere with intermolecular hydrogen bonding between solute molecules. As a result, it is sufficient to carry out the theoretical predictions in a vacuum without explicit inclusion of solvent influence.

Experimental Methods

Racemic **1** and **2** were obtained from TCI America and Aldrich Chemical Co., respectively. The enantiomers of **1** and **2** were separated[7] in-house at 25 °C on a 4.6 × 250 mm chiral column (Chiralpak AD-H column from Daicel Chemical Industries, Ltd.) using a mobile phase of 3% 2-propanol in hexane, with 0.03% trifluoroacetic acid, run at 1 mL/min with UV detection at 254 nm. The experimental specific rotations of the first elutants of **1** and **2**, respectively, are $[\alpha]_D^{20} = -26.6$ (c=1.23, CH_3COCH_3) and $[\alpha]_D^{20} = -49.7$ (c=0.58, CH_3COCH_3). The experimental specific rotations of the second elutants of **1** and **2**, respectively, are $[\alpha]_D^{20} = +28.5$ (c=0.97, CH_3COCH_3) and $[\alpha]_D^{20} = +53.5$ (c=0.79, CH_3COCH_3). Analytical grade samples of (+)-enantiomers of **3** ($[\alpha]_D^{20} = +16.3$ (c=2.31, $CHCl_3$)) and **4** ($[\alpha]_D^{20} = +19.0$ (c=2.29, $CHCl_3$)) were obtained[6] as a gift from A. H. Marks Co. (U.K.).

The infrared and VCD spectra were recorded on a commercial Fourier transform VCD spectrometer, Chiralir (Bomem-BioTools, Canada). A second PEM was added to improve the signal/noise ratio and reduce the artifacts.[18] The VCD spectra were recorded with 3-h data collection time at 4 cm^{-1} resolution. Spectra were measured in $CDCl_3$ solvent. The concentrations are 0.150 M for **1** (path length 180 μm); 0.150 M for **2** (path length 180 μm); 0.158 M for **3** (path length 120 μm); and 0.155 M for **4** (path length 70 μm). The sample was held in a variable path length cell with BaF_2 windows. In both the absorption and VCD spectra presented, the corresponding solvent spectra obtained under identical conditions were subtracted out.

Theoretical Calculations

The geometry optimization, vibrational frequencies, absorption intensities, and VCD intensities were calculated by Gaussian 03 program.[19] The calculations used the density functional theory with B3LYP functional and 6-31G* basis set. The procedure for calculating the VCD intensities using DFT theory is as implemented in Gaussian 03 program.[19] The theoretical absorption and VCD spectra were simulated with Lorentzian band shapes and 5 cm^{-1} bandwidths. Because the predicted band positions are higher than the experimental band positions, the calculated B3LYP/6-31G* frequencies were scaled with 0.9613, as recommended by Wong.[20] The use of larger basis sets can overcome the need for arbitrary scaling of frequencies, but such larger basis sets also would require more computational resources than are available to us. One approach to lower the time demands involved in geometry optimization calculations is to do the conformational search using Monte Carlo methods using empirical force fields. However, for consistency all calculations here were carried out at B3LYP/6-31G* level of theory.

Results and Discussion

The results will be divided into two sections, one on monomer conformations and the other on dimer conformations. The results on monomeric acid conformations have been reported[6,7] before, but first we briefly summarize those results for clarity and completeness.

Monomer Conformations. The conformational flexibility in these acids arises from the rotation around four single

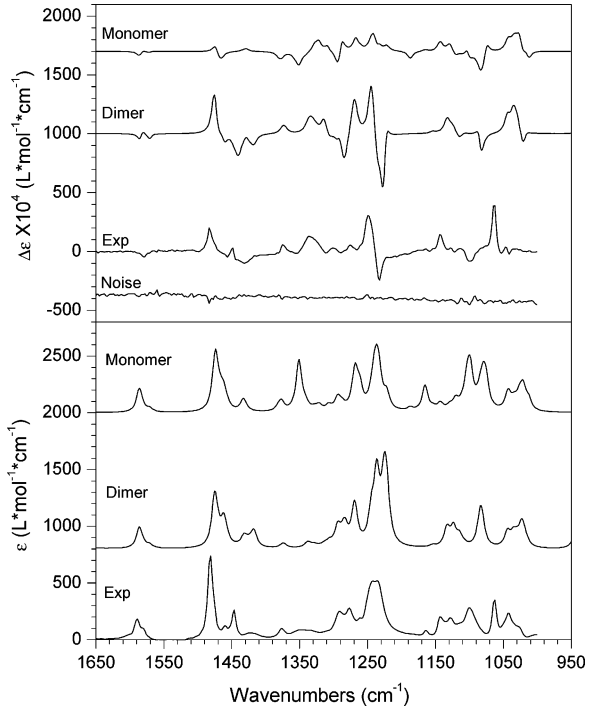
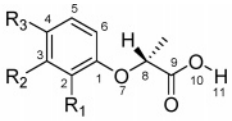


Figure 1. Comparison between experimental and B3LYP/6-31G* predicted vibrational absorption (bottom panel) and VCD (top panel) spectra of 2-(2-chlorophenoxy)propanoic acid. The predicted frequencies are scaled by 0.9613. The experimental measurements were made for (+)-enantiomer and calculations were done for (*R*)-configuration. The trace labeled "noise" represents the reproducibility level in the experimental VCD spectrum.

Chart 1. Structure and Atomic Numbering of α-Aryloxypropanoic Acids



1, R_1=Cl, R_2=H, R_3=H; **2**, R_1=H, R_2=Cl, R_3=H;

3, R_1=CH_3, R_2=H, R_3= Cl; **4**, R_1=Cl, R_2=H, R_3= Cl;

bonds: C^1O^7, O^7C^8, C^8C^9, and C^9O^{10} (see Chart 1). The initial dihedral angles D1, D2, D3, and D4, respectively, representing the segments (see Chart 1) $C^2C^1O^7C^8$, $C^1O^7C^8C^9$, $O^7C^8C^9O^{10}$, and $C^8C^9O^{10}H^{11}$ are varied in 120° increments. This resulted in $3^4 = 81$ possible conformations for each of the four acids. All these conformations have been subjected to geometry optimizations. *(A) (R)-2-(2-Chlorophenoxy)-propanoic acid*, (*R*)-**1**: For this acid, four lowest energy stable conformers, labeled Con1, Con2, Con3, and Con4, with fractional populations of 0.28, 0.27, 0.24, and 0.21 are found. The four dihedral angles, D1, D2, D3, and D4, for these conformers are as follows: 177, 71, 35, and 178 for Con1; −164, 146, −4, and 1 for Con2; 177, 73, −159, and −176 for Con3; 81, 80, −169, and −178 for Con4. The population weighted predicted vibrational absorption and VCD spectra are compared to the experimental spectra in Figure 1. From this figure it is evident that the predicted spectra of monomeric (*R*)-1 do not compare well[7] with the experimental spectra. *(B) (R)-2-(3-Chlorophenoxy)propanoic acid*, (*R*)-**2**: The four lowest energy conformers, labeled Con1, Con2, Con3, and Con4, with fractional populations of 0.25, 0.25, 0.23, and 0.20 are identified. The four dihedral angles, D1, D2, D3, and D4, for these conformers are as follows: 177, 71, 36, and 179 for Con1; 0.4, 72, −158, and −177 for Con2; 178, 73, −158, and −177 for Con3; and −2, 70, 33, and 179 for Con4. The population weighted predicted vibrational absorption and VCD spectra are compared to the experimental spectra in Figure 2. From this figure it is evident that the predicted spectra of monomeric (*R*)-**2** do not compare well with the experimental spectra for this acid either. *(C) (R)-2-(4-Chloro-2-methylphenoxy)-propanoic acid*, (*R*)-**3**: Two lowest energy conformers, labeled Con1 and Con2 with fractional populations of 0.46 and 0.43, are identified. The four dihedral angles, D1, D2, D3, and D4 for these conformers are as follows: 178, 70, 35, and 179 for Con1 and 180, 72, −156, and −177 for Con2. The population weighted predicted vibrational absorption and VCD spectra are compared to the experimental spectra in Figure 3. From this figure it is evident that the predicted spectra for monomeric (*R*)-**3** do not compare well with the experimental spectra. *(D) (R)-2-(2,4-Dichlorophenoxy)propanoic acid*, (*R*)-**4**: Three lowest energy conformers, labeled Con1, Con2 and Con3 with fractional populations of 0.31, 0.28, and 0.26, are identified. The four dihedral angles, D1, D2, D3, and D4, for these conformers are as follows: 177, 70, 36, and 178 for Con1; 178, 72, −160,

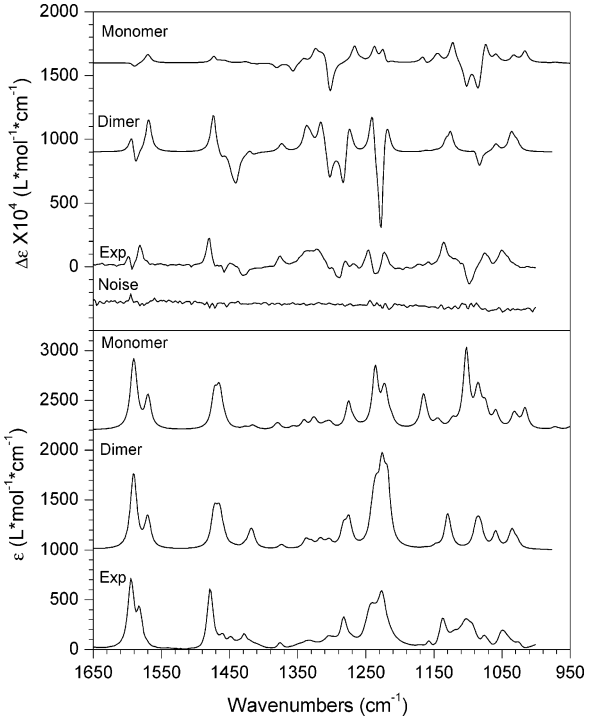


Figure 2. Comparison between experimental and B3LYP/6-31G* predicted vibrational absorption (bottom panel) and VCD (top panel) spectra of 2-(3-chlorophenoxy)propanoic acid. The predicted frequencies are scaled by 0.9613. The experimental measurements were made for (+)-enantiomer and calculations were done for (*R*)-configuration. The trace labeled "noise" represents the reproducibility level in the experimental VCD spectrum.

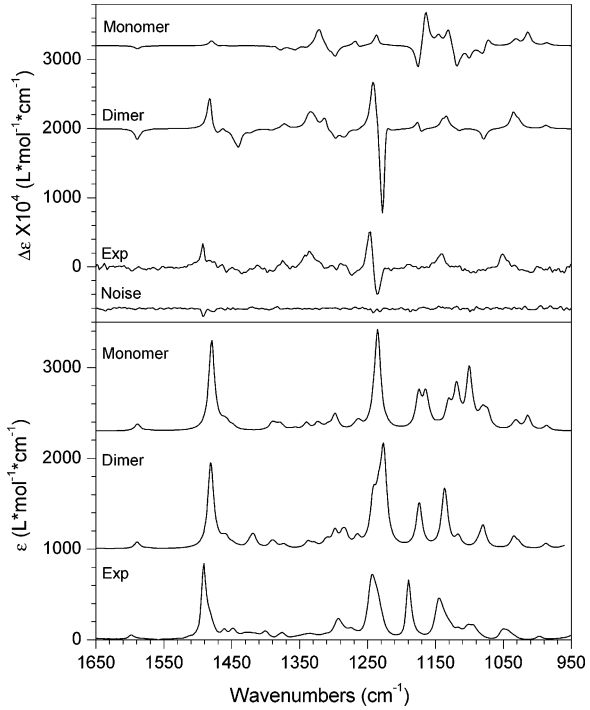


Figure 3. Comparison between experimental and B3LYP/6-31G* predicted vibrational absorption (bottom panel) and VCD (top panel) spectra of 2-(4-chloro-2-methylphenoxy)-propanoic acid. The predicted frequencies are scaled by 0.9613. The experimental measurements were made for (+)-enantiomer and calculations were done for (*R*)-configuration. The trace labeled "noise" represents the reproducibility level in the experimental VCD spectrum.

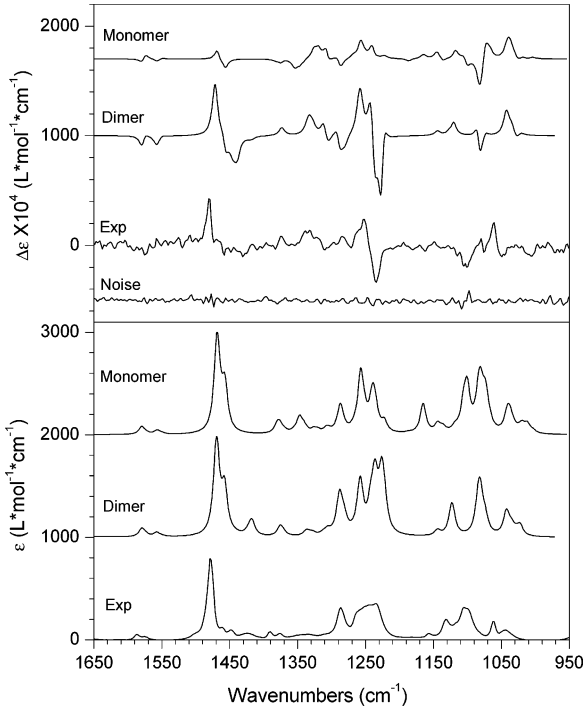
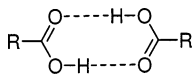


Figure 4. Comparison between experimental and B3LYP/6-31G* predicted vibrational absorption (bottom panel) and VCD (top panel) spectra of 2-(2,4-dichlorophenoxy)propanoic acid. The predicted frequencies are scaled by 0.9613. The experimental measurements were made for (+)-enantiomer and calculations were done for (*R*)-configuration. The trace labeled "noise" represents the reproducibility level in the experimental VCD spectrum.

and −177 for Con2; and 82, 80, −171, and −179 for Con3. The population weighted predicted vibrational absorption and VCD spectra are compared to the experimental spectra in Figure 4. From this figure it is evident that the predicted spectra for monomeric (*R*)-**4** do not compare well with the experimental spectra.

In summary the calculations for monomeric propanoic acids failed to provide a reasonable agreement[6,7] between the predicted and experimental VA and VCD spectra. However, after methyl esterification of the acids, the experimental VA and VCD spectra obtained for esters were correctly reproduced[6,7] by the B3LYP/6-31G* predictions of these spectra for the esters. This observation indicated that the mismatch between experimental and predicted spectra for acids is not due to the theoretical level used for the calculations but is likely to be associated with intermolecular hydrogen bonding prevalent for carboxylic acids in solution. The presence of intermolecular hydrogen bonding in solution for the four acids considered is further confirmed[6,7] by the concentration dependent absorption spectra. Although methyl esterification[6,7] removed the hydrogen bonding influence in the experimental spectra and provided a good match between experimental and predicted spectra for esters, this approach does not answer the question about the predominant conformations of acids in solution, which is critical to explain the biological activity of acids. Thus we found it necessary to pursue calculations on dimer conformations of acids.

Chart 2. Carboxylic Acid Dimer



Dimer Conformations. It is known that carboxylic acids tend to form dimers via intermolecular hydrogen bonding between two monomer units, as shown in Chart 2. The problem in constructing dimer conformations for the four acids under consideration is that these acids have too many possible starting conformations for monomers, making the calculations for dimers a tedious task. To construct a dimer conformation, the dihedral angle, D4, of $C^8C^9O^{10}H^{11}$ (see Chart 1) can be fixed at ∼180 degree, but there are three other dihedral angles, D1, D2, and D3, which can be varied. If we take 60, 180, and −60 as possible starting angles for each of these three dihedral angles, then one has to consider 27 starting structures for each monomer unit in the dimer. Combining the two monomer units, the dimer formed will then have a total of (27+1)*27/2 = 378 possible starting conformations for each of the four acids considered.

However, upon looking at the predicted lowest energy conformers of monomeric acids and their methyl esters,[6,7] we found that the dihedral angles D1, D2, and D3 of lowest energy conformers of monomeric acids and their esters are similar. Based on this observation, we developed the following working hypothesis for constructing dimer conformations. Since the dihedral angles D1, D2, and D3 are not influenced significantly by esterification, they probably

Table 1. Optimized Geometries, Gibbs Energies, and Populations of Conformers of (*R*)-2-(2-Chlorophenoxy)propanoic Acid Dimer[a]

	monomer 1				monomer 2					
conformer	D1	D2	D3	D4	D1	D2	D3	D4	energy (kJ/mol)	population
Con3-Con4	178	72	−159	−176	79	81	−169	−177	0	0.3366
Con3-Con3	177	72	−154	−176	177	72	−154	−176	1.696	0.1698
Con1-Con3	176	71	35	178	178	72	−155	−176	1.859	0.1590
Con1-Con4	177	71	31	178	79	81	−169	−178	2.308	0.1326
Con1-Con1	177	72	32	178	177	72	32	178	2.977	0.1013
Con4-Con4	79	80	−167	−177	79	80	−167	−177	2.990	0.1007

[a] The labels Con1, Con2, Con3, and Con4 are for the optimized conformations of monomeric acid. The dihedral angle definitions for D1, D2, D3, and D4 are respectively, $C^2C^1O^7C^8$, $C^1O^7C^8C^9$, $O^7C^8C^9O^{10}$, and $C^8C^9O^{10}H^{11}$. See Chart 1 for atom numbering. The four optimized dihedral angles, D1−D4, for monomeric acid conformers are as follows: 177, 71, 35, and 178 for Con1; −164, 146, −4, and 1 for Con2; 177, 73, −159, and −176 for Con3; and 81, 80, −169, and −178 for Con4.

would not be influenced by hydrogen bonding as well. Then the lowest energy conformers of monomeric acids can be used to represent the monomer units in a dimer. This hypothesis, along with the requirement that the $C^8C^9O^{10}H^{11}$ dihedral angle, D4, should be ∼180° for intermolecular hydrogen bonding, will greatly reduce the number of possible starting dimer conformations. With these guidelines, we have investigated the possible dimer conformations using the lowest energy monomeric acid (or their ester) conformations as the starting point. The notation used for dimers is as follows: if a dimer is formed from combining conformer 1 (labeled Con1) of one monomer unit and conformer 3 (Con3) of a second monomer unit, then the dimer is labeled as Con1-Con3. The reliability of final dimer conformations is then evaluated by comparing the predicted VA and VCD spectra for these conformers with the corresponding experimental spectra in solution.

*(A) (R)-2-(2-Chlorophenoxy)propanoic Acid, (R)-**1***. For (*R*)-**1**, we constructed the dimer conformations using the optimized structures of conformers 1, 3, and 4 of the monomeric acid[7] (labeled as Con1, Con3, and Con4). The structures of these three conformations are similar to those of conformers 1, 2, and 3 of methyl ester.[7] The conformer 2 of monomeric acid was not used because the dihedral angle D4 for this conformer (1.2°) does not permit dimer formation. The length of hydrogen bond O—H- - - - -O is taken from the literature[21] to be 2.70 Å. The three starting monomer conformations generated six starting dimer conformations, and all of these conformers were submitted to geometry optimization. The final optimized structures are summarized in Table 1. In the optimized structures, the geometries of individual monomer units of dimer are very close to those of starting monomeric acids. The optimized hydrogen bond length, O—H- - - - -O, is around 2.69 Å. The optimized structure of lowest energy dimer conformer, Con3-Con4, is shown in Figure 5 A.

To evaluate if these dimer conformers are in fact the ones that represent the spectra in solution, we calculated the vibrational frequencies and intensities for the six stable dimer conformers listed in Table 1. From the vibrational frequency calculation, we noted that all six structures represent energy minima, as there were no imaginary frequencies. The relative Gibbs free energies and populations generated there from are also listed in Table 1. The population weighted VA and VCD spectra for (*R*)-**1** are compared to the corresponding experimental spectra in Figure 1.

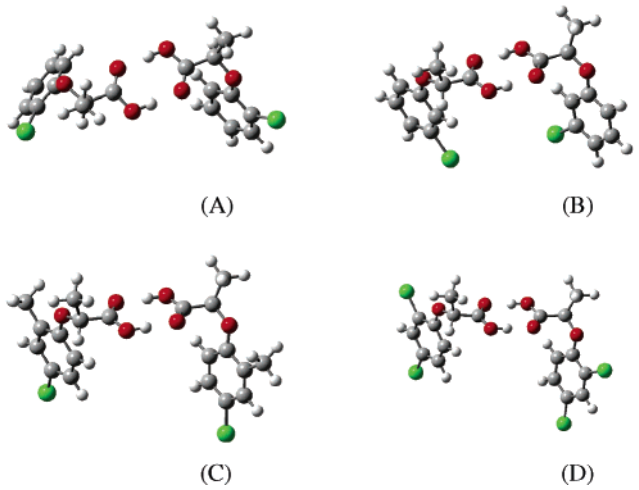


Figure 5. The structures of lowest energy conformers of dimeric acids: (A) Con3-Con4 of (*R*)-2-(2-chlorophenoxy)-propanoic acid; (B) Con2-Con2 of (*R*)-2-(3-chlorophenoxy)-propanoic acid; (C) Con2-Con2 of (*R*)-2-(4-chloro-2-methylphenoxy)propanoic acid; and (D) Con2-Con2 of (*R*)-2-(2,4-dichlorophenoxy)propanoic acid.

In the VA spectra of Figure 1, the predicted dimer spectra can be seen to be in much better agreement with experimental spectrum, while predicted monomeric acid spectrum is not. The region from ∼1650 cm^{-1} to 1000 cm^{-1} has a good agreement between experimental and predicted dimer spectra. The predicted monomer VA spectrum on the other hand shows two extra peaks around 1340 and 1160 cm^{-1}. In the VCD spectra of Figure 1, the predicted dimer spectrum has a much better agreement, compared to the spectrum calculated for monomeric acid, with the experimental spectrum in the ∼1600−1000 cm^{-1} region. One discrepancy is in the region near 1275 cm^{-1} where predicted spectrum shows an extra negative−positive couplet. Nevertheless, the predicted spectrum of dimer satisfactorily reproduces the positive−negative couplet seen around 1240 cm^{-1} in the experimental spectrum, while the predicted monomer spectrum failed to reproduce this feature. From the calculated vibrational displacements, the vibration that generated this positive−negative couplet is found to involve the hydrogen atoms in the carboxylic groups of dimer. This explains why the predicted spectrum for the monomer acid cannot reproduce this couplet.

Since a reasonable agreement, in both VA and VCD spectra, is seen between the predicted dimer spectra and

Table 2. Optimized Geometries, Gibbs Energies, and Populations of Conformers of (*R*)-2-(3-Chlorophenoxy)propanoic Acid Dimer[a]

	monomer 1				monomer 2					
conformer	D1	D2	D3	D4	D1	D2	D3	D4	energy (kJ/mol)	population
Con2-Con2	0.1	73	−158	−176	0.1	72	−158	−176	0	0.1898
Con3-Con3	178	72	−154	−176	179	72	−154	−176	0.181	0.1764
Con2-Con3	0.1	72	−160	−177	179	71	−157	−177	0.328	0.1662
Con3-Con4	178	72	−156	−176	−1	71	32	179	1.000	0.1267
Con1-Con3	177	71	35	178	178	72	−156	−177	1.082	0.1227
Con1-Con2	177	71	33	178	0.1	72	−157	−177	1.166	0.1186
Con2-Con4	−0.4	72	−157	−177	−2	71	33	178	1.599	0.0996

[a] The labels Con1, Con2, Con3, and Con4 are for the optimized conformations of monomeric acid. The dihedral angle definitions for D1, D2, D3, and D4 are, respectively, $C^2C^1O^7C^8$, $C^1O^7C^8C^9$, $O^7C^8C^9O^{10}$, and $C^8C^9O^{10}H^{11}$. See Chart 1 for atom numbering. The four optimized dihedral angles, D1−D4, for monomeric acid conformers are as follows: 177, 71, 36, and 179 for Con1; 0.4, 72, −158, and −177 for Con2; 178, 73, −158, and −177 for Con3; and −2, 70, 33, and 179 for Con4.

Table 3. Optimized Geometries, Gibbs Energies, and Populations of Conformers of (*R*)-2-(4-Chloro-2-methylphenoxy)propanoic Acid Dimer[a]

	monomer 1				monomer 2					
conformer	D1	D2	D3	D4	D1	D2	D3	D4	energy (kJ/mol)	population
Con2-Con2	180	71	−155	−176	180	71	−155	−176	0	0.5465
Con1-Con2	180	70	33	178	180	71	−156	−176	1.452	0.3042
Con1-Con1	179	71	31	178	179	71	31	178	3.216	0.1493

[a] The labels Con1 and Con2 are for the optimized conformations of monomeric acid. The dihedral angle definitions for D1, D2, D3, and D4 are, respectively, $C^2C^1O^7C^8$, $C^1O^7C^8C^9$, $O^7C^8C^9O^{10}$, and $C^8C^9O^{10}H^{11}$. See Chart 1 for atom numbering. The four optimized dihedral angles, D1−D4 for monomeric acid conformers are as follows: 178, 70, 35, and 179 for Con1 and 180, 72, −156, and −177 for Con2.

experimental spectra in solution, the dimer conformations used in calculation can be considered to represent the conformations in the solution. Furthermore, since experimental VCD spectra were obtained for (+)-enantiomer and calculations were done for the (*R*)-configuration, the absolute configuration can be assigned as (+)-(*R*), as was concluded from prior investigations[7] on methyl ester of this acid.

(B) (R)-2-(3-Chlorophenoxy)propanoic Acid, (R)-2. Location of chlorine atom at position 3 of phenyl group in (*R*)-**2** makes the energy barrier less when the dihedral angle $C^2C^1O^7C^8$ is varied. As a result there is one additional conformation here than in the case of (*R*)-**1**. Four stable conformers 1, 2, 3, and 4 of monomeric acid are similar[7] to conformers 1, 2, 4, and 3 of the corresponding methyl ester. These conformers (labeled as Con1, Con2, Con3, and Con4) are used to construct dimer conformers, which lead to 10 starting dimer structures. All of these 10 conformers are submitted to geometry optimization. The seven most stable conformers are listed in Table 2. The remaining three conformers with higher energy are not listed because their populations are negligible for the present purposes. Again, the monomer components in dimer have the geometries similar to those optimized for monomeric acids. The structure of the lowest energy dimer conformer, Con2-Con2, is shown in Figure 5B. The vibrational spectra of (*R*)-**2** dimers are calculated at the optimized geometries. The populations of dimer conformers deduced from Gibbs energies are also listed in Table 2. The VA and VCD spectra are compared to the experimental spectra in Figure 2. In the VA spectra of Figure 2, we see an excellent agreement again between experimental and predicted dimer spectra but not between experimental and predicted monomer spectra. The relative intensities of predicted dimer VCD bands in the 1360 cm^{-1} to 1200 cm^{-1} region are somewhat larger than the experimental values. When positive and negative VCD bands appear next to each other, separation between their band positions will influence their relative intensities due to mutual cancellation. Thus unless band positions are predicted accurately, which is difficult at B3LYP/6-31G* level, the relative intensities of overlapping VCD bands cannot be reproduced well. The overall agreement between experimental and predicted dimer VCD spectra is much better than that between the experimental and predicted monomer VCD spectra. In particular, the signs of VCD bands in predicted dimer and experimental spectra agree very well in the region from ∼1650 cm^{-1} to 1000 cm^{-1}. Thus the dimer conformations used can be considered to represent the conformations in solution. From the comparison between predicted dimer and experimental spectra, the absolute configuration of **2** can be assigned as (+)-(*R*), as concluded from investigations on corresponding methyl esters.[7]

(C) (R)-2-(4-Chloro-2-methylphenoxy)propanoic Acid, (*R*)-**3.** For this acid, the conformers 1 and 2 (labeled as Con1 and Con2) in both acid and its methyl ester[6] account for about 90% of the population. Therefore, only conformers 1 and 2 are chosen to construct dimer structures. This will generate three starting dimer structures. All of these three conformers are submitted for geometry optimization. Their optimized structures are listed in Table 3. Here also the monomer components in dimer have the geometries similar to those optimized for monomeric acids. The structure of the lowest energy dimer conformer, Con2-Con2, is shown in Figure 5C.

The VA and VCD spectra are calculated for these three conformers at their optimized geometries. The population weighted predicted spectra for dimeric acids are shown in Figure 3. In the VA spectra of Figure 3, we notice an excellent agreement again between experimental and predicted dimerspectra. There are some minor differences in the relative absorption intensities among predicted dimer and experimental spectra, but the relative intensities cannot be expected to match well at B3LYP/6-31G* level used here. In the VCD spectra of Figure 3, the major positive and

Table 4. Optimized Geometries, Gibbs Energies, and Populations of Conformers of (*R*)-2-(2,4-Dichlorophenoxy)propanoic Acid Dimer[a]

	monomer 1				monomer 2					
conformer	D1	D2	D3	D4	D1	D2	D3	D4	energy (kJ/mol)	population
Con2-Con2	178	72	−155	−176	178	72	−155	−176	0	0.2340
Con1-Con2	177	71	31	178	178	72	−158	−176	0.415	0.1979
Con2-Con3	178	73	−162	−176	79	81	−171	−178	0.848	0.1662
Con3-Con3	80	81	−171	−178	80	81	−171	−178	0.898	0.1628
Con1-Con3	177	71	34	178	80	80	−168	−178	1.286	0.1392
Con1-Con1	177	71	32	178	177	71	32	178	2.108	0.0999

[a] The labels Con1, Con2, and Con3 are for the optimized conformations of monomeric acid. The dihedral angle definitions for D1, D2, D3, and D4 are, respectively, $C^2C^1O^7C^8$, $C^1O^7C^8C^9$, $O^7C^8C^9O^{10}$, and $C^8C^9O^{10}H^{11}$. See Chart 1 for atom numbering. The four optimized dihedral angles, D1−D4, for monomeric acid conformers are as follows: 177, 70, 36, and 178 for Con1; 178, 72, −160, and −177 for Con2; and 82, 80, −171, and −179 for Con3.

negative VCD bands agree well between experimental and predicted dimer spectra. The experimental positive−negative VCD couplet around 1240 cm^{-1}, which does not show up in the predicted spectrum of monomeric acid is reproduced very nicely in predicted dimer spectrum. Thus the dimer conformations used can be considered to represent the conformations in solution. Since the signs of VCD bands in predicted dimer and experimental spectra agree very well in Figure 3, the absolute configuration of **3** can be assigned as (+)-(*R*), as deduced from investigations on corresponding methyl ester.[6]

(D) (R)-2-(2,4-Dichlorophenoxy)propanoic Acid, (*R*)-**4**. For (*R*)-**4**, three conformers (labeled Con1, Con2, Con3) account for most of the population of monomer acid conformers. Thus these three conformers are used to construct the dimer structures, resulting in six starting dimer conformations. The optimized dimer structures are listed in Table 4. Here also the monomer components in dimer have the geometries similar to those optimized for monomeric acids. The structure of the most stable conformer, Con2-Con2, is shown in Figure 5D.

The predicted VA and VCD spectra for dimeric acid are shown in Figure 4. The agreement between experimental and predicted dimer spectra is quite good, while that between experimental and predicted monomer spectra is rather poor. Some minor discrepancies however should be noted. In the 1250 cm^{-1} region, the experimental absorption spectrum shows a broad band, while the predicted dimer absorption spectrum shows three resolved bands. In the region from 1060 to 1150 cm^{-1}, the predicted dimer spectrum shows two well-separated peaks, while experimental spectrum shows two partially separated peaks. These discrepancies are attributed to the errors in predicted band positions at the B3LYP/631G* level. Nevertheless, the fact the experimental spectra are reasonably well reproduced by the predicted dimer spectra and not by the predicted monomer spectra suggests that the dimer conformations used here represent those in solution. Moreover, since the signs of all major VCD bands in predicted dimer and experimental spectra agree very well, the absolute configuration of **4** can be assigned as (+)-(*R*), agreeing with the conclusion arrived at from the investigations on corresponding methyl ester.[6]

The vibrational absorption spectra in the carbonyl region are shown in Figure 6 for all four acids considered. Concentrations used for the experimental absorption spectra in Figure 6 are the same as those for the spectra in Figures 1−4. The experimental VCD spectra in this region are not shown because VCD bands in this region are very weak. The predicted dimer absorption spectra have better agreement with experimental absorption spectra compared to the predicted monomer spectra. The strong peak around 1730 cm^{-1} in the experimental spectra is reproduced well in the predicted dimer spectra. The high frequency shoulder at 1774 cm^{-1} in the experimental spectra is assigned[7] to the carbonyl vibration of monomer. Although some amount of monomer conformers is present in solution, the absorption and VCD in the 1650−950 cm^{-1} region are mostly determined by the dimer conformations as can be inferred from Figures 1−4 and from concentration dependent studies (vide infra). If the goals were to compare the experimental and calculated spectra quantitatively, then one would have to simulate the predicted spectra with a certain amount of monomer conformers. Alternately, it would be possible to determine the percent populations of monomers and dimers using a regression analysis of the experimental and predicted intensities,[22] if the predicted absorption and VCD intensities are quantitatively accurate. But it was pointed out[22] that B3LYP predictions of absorption and VCD intensities are not sufficiently accurate for quantitative comparisons, even when large basis sets are used for small molecules. For this reason

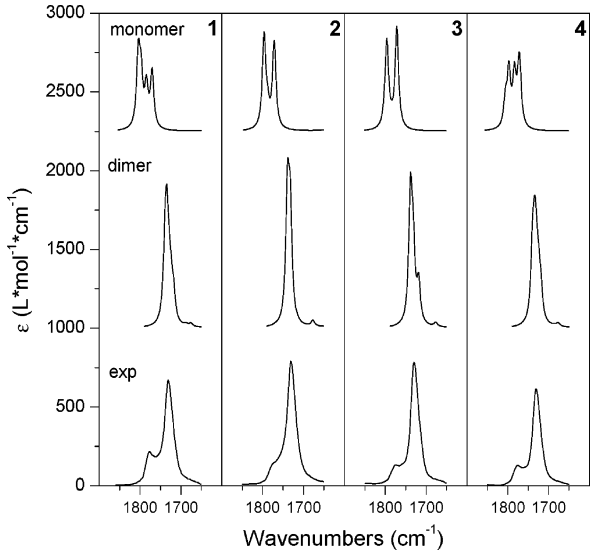


Figure 6. The experimental (bottom trace) and predicted (top two traces) vibrational absorption spectra for acids **1**−**4** in the carbonyl region. The predicted frequencies are scaled by 0.9613.

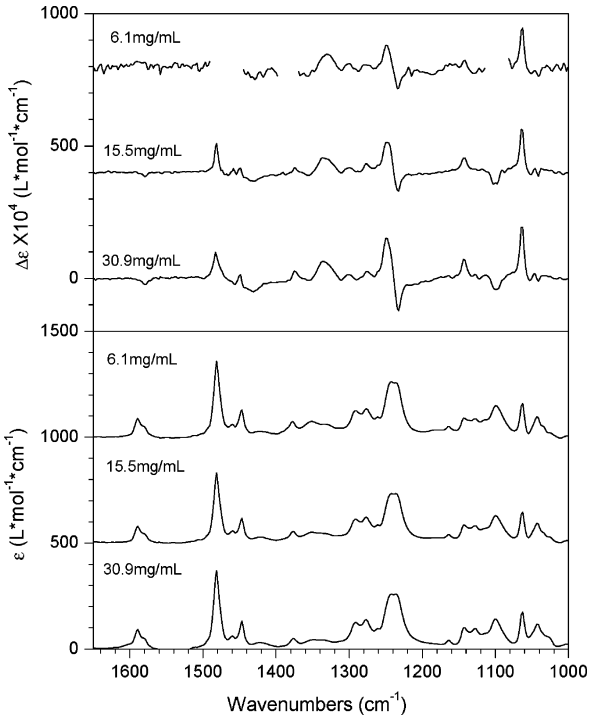


Figure 7. Concentration dependent vibrational absorbance and VCD spectra of (+)-2-(2-chlorophenoxy)propanoic acid in $CDCl_3$. The path length is 180 μm for 30.9 mg/mL, 400 μm for 15.5 mg/mL, and 900 μm for 6.1 mg/mL. In the topmost VCD spectrum (at 6.1 mg/mL) the regions around 1470, 1360, and 1080 cm^{-1} have been removed because of noise resulting from solvent absorption interference.

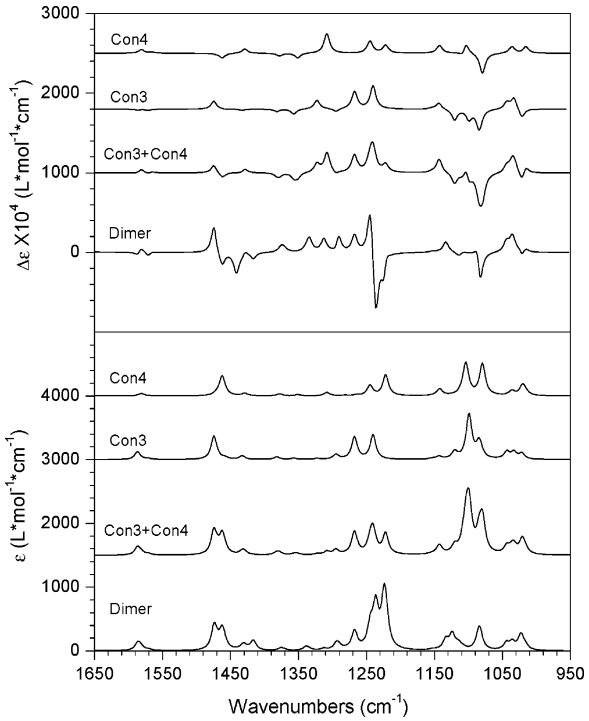


Figure 8. Evaluation of spectra additivity of B3LYP/6-31G* predicted spectra for 2-(2-chlorophenoxy)propanoic acid: (a) vibrational absorption (bottom panel) and (b) VCD (top panel) Con3 and Con4 are the monomer spectra predicted for these two conformations. Con3+Con4 represents the numerical sum of the spectra predicted for monomer conformers 3 and 4, while dimer spectra represent those predicted for the dimer Con3-Con4. The predicted frequencies are scaled by 0.9613.

we did not attempt to obtain the simulated spectra with an admixture of monomer and dimer contributions.

Concentration Dependence. It is known that studies on dilute solutions are appropriate to avoid the influence of dimers and to increase the population of monomers in solution. This strategy, although perfectly valid, is not particularly useful for VCD spectra because, extremely dilute concentrations are needed for realizing predominant monomer conformations and at such dilute concentrations solvent absorption interference prevents VCD measurements with sufficient signal-to-noise. To illustrate this point concentration dependent absorption and VCD spectra for (+)-2-(2-chlorophenoxy)propanoic acid in the 1650−950 cm^{-1} region are shown in Figure 7. As the concentration is lowered, the VCD signal magnitudes are seen to decrease, but the onset of solvent interference becomes an issue long before monomer contributions begin to appear in the region studied. As can be seen in Figure 7, the regions 1491−1445, 1398−1369, and 1115−1082 cm^{-1} had to be blocked out at a concentration of 6.1 mg/mL, due to excessive noise resulting in these regions from the solvent absorption interference. Thus it is not possible to conduct VCD studies on solutions that are continuously diluted to the point where monomers dominate.

Spectral Additivity. When dealing with large molecular systems, it would be tempting to undertake the calculations on different fragments of the molecule of interest and add the results to represent the property of full molecule. Such an approach would be valid if the interactions between different fragments do not influence the property of interest. In the case of vibrational properties, however, the very nature of normal coordinates imposes cross talk between fragments of a molecule. So it is of interest to know if the vibrational spectra of fragments can be additive to represent the vibrational spectrum of full molecules. Since dimers are viewed as a combination of two monomers, it would be useful to verify the additivity of the spectra of monomers. In Figure 8, the predicted vibrational spectra for monomer conformers 3 and 4, their numerical sum and of dimer derived from these monomers are compared for 2-(2-chlorophenoxy)propanoic acid. In the 1150−950 cm^{-1} region, the numerically added VCD spectrum compares reasonably well with the predicted VCD spectrum of dimer. But the numerically added absorption spectrum in this region does not match with the predicted absorption spectrum of dimer. In the 1400−1150 cm^{-1} region, both absorption and VCD spectra predicted for dimer are significantly different from the corresponding numerically added spectra for monomers contained therein. The vibrational modes involving intermolecular hydrogen bonding appear in the 1400−1150 cm^{-1} region, as noted for other hydrogen-bonded systems[23] such as butanol. Thus the concept of additivity for vibrational absorption and VCD spectra of α-aryl propanoic acids does not work in the spectral regions where vibrational modes involving intermolecular hydrogen bonds occur.

Conclusions

The predicted VA and VCD spectra for monomer conformations of α-aryloxypropanoic acids did not reproduce the corresponding experimental spectra. This is believed to be due to dimer formation for acids in solution. Noting that the predicted structures of monomeric acids are very similar to those of corresponding methyl esters, we hypothesized that the lowest energy structures of monomeric acids can be used to construct the starting structures for dimeric acids. The starting geometries of dimer structures were optimized at B3LYP/6-31G* level. In the optimized dimer structures, the individual monomer components are found to have structures similar to the optimized conformers of monomeric acids. To verify the reliability of optimized dimer structures, VA and VCD spectra have been calculated at the optimized geometries. The population weighted predicted VA and VCD spectra for dimeric acids are found to be in good agreement with the corresponding spectra in $CDCl_3$ solution.

Based on the present work, we can suggest a general approach for interpreting the vibrational spectra of carboxylic acids in solution. This approach involves first finding the lowest energy conformers of a monomeric acid and its methyl ester. If the lowest energy conformers of acid and its ester are similar, then one can use these lowest energy optimized structures for generating dimer conformers. Vibrational spectra calculated for the optimized dimer conformers can then be compared to the experimental spectra to evaluate the validity of these conformers in solution. In this manner, one can determine the predominant conformations of dimeric acids as well as their absolute configuration using density functional predictions of VA and VCD spectra.

Acknowledgment. Grants from NSF (CHE0092922) and Vanderbilt University are gratefully acknowledged. We thank Jeremy Wardman of A.H. Marks Co., for providing the analytical grade samples of Mecoprop and Dichloropop. Contract grant sponsor: NSF; contract grant number: CHE0092922.

Supporting Information Available: Optimized coordinates, vibrational frequencies, and intensities for lowest energy conformers are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Buser, H. R.; Muller, M. D. *Environ. Sci. Technol.* **1997**, *31*, 1960−1967.

(2) Sheldon, R. A. *Chirotechnology: Industrial Synthesis of Optically Active Compounds*; Marcel Dekker: New York, 1993; pp 62−64.

(3) Loos, M. A. *Herbicides: Chemistry, Degradation and Mode of Action*; Marcel Dekker: New York, 1975; Vol. 1, pp 1−128.

(4) Testa, B.; Trager, W. F. *Chirality* **1990**, *2*, 129−133.

(5) Zipper, C.; Suter, M. J. F.; Haderlein, S. B.; Gruhl, M.; Kohler, H. P. E. *Environ. Sci. Technol.* **1998**, *32*, 2070−2076.

(6) He, J.; Wang, F.; Polavarapu, P. L. *Chirality* **2005**, *17*, S1−S8.

(7) He, J.; Polavarapu, P. L. *Spectrochim Acta* **2005**, in press.

(8) Jeffrey, G. A. *An introduction to hydrogen bond*; Oxford University Press: New York, 1997.

(9) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737−738.

(10) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205−211.

(11) Scheiner, S. *Hydrogen bond, a theoretical perspective*; Oxford University Press: New York, 1997; pp 99−101.

(12) Hadzi, D. *Theoretical treatments of hydrogen bonding*; John Wiley & Sons: 1997; pp 13−74.

(13) Barron, L. D. *Molecular Light Scattering and Optical Activity*, 2nd ed.; Cambridge University Press: 2004; pp 331−342.

(14) Polavarapu, P. L. *Vibrational spectra: principles and applications with emphasis on optical activity*; Elsevier: New York, 1998; pp 143−182.

(15) Nafie, L. A. *Annu. Rev. Phys. Chem.* **1997**, *48*, 357−386.

(16) (a) Polavarapu, P. L.; He, J. *Anal. Chem.* **2004**, *76*, 61A−67A. (b) Freedman, T. B.; Cao, X.; Dukor, R. K.; Nafie, L. A. *Chirality* **2003**, *15*, 743−758. (c) Devlin, F. J.; Stephens, P. J.; Osterle, C.; Wiberg, K. B.; Cheeseman, J. R.; Frisch, M. J. *J. Org. Chem.* **2002**, *67*, 8090−8096. (d) Andrushchenko, V.; van de Sande, J. H.; Wieser, H. *Biopolymers* **2003**, *72*, 374−390. (e) Keiderling, T. A.; Xu, Q. *Adv. Protein Chem.* **2002**, *62*, 111−161.

(17) (a) Wang, F.; Polavarapu, P. L.; Drabowicz, J.; Kielbasinski, P.; Potrzebowski, M. J.; Mikolajczyk, M.; Wieczorek, M. W.; Majzner, W. W.; Lazewska, I. *J. Phys. Chem.* **2004**, *108*, 2072−2079. (b) Cappelli, C.; Corni, S.; Mennucci, B.; Cammi, R.; Tomasi, J. *J. Phys. Chem. A* **2002**, *106*, 12331−12339.

(18) Nafie, L. A. *Appl. Spectrosc.* **2000**, *54*, 1634−1645.

(19) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. C., J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. B., J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. M., M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. N., M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. H., M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. K., X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. J., R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. C., C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. V., P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. D., A. D. Daniels, M. C. Strain, O. Farkas, D. K. M., A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. O., Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. S., G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. M., D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. N., M. Challacombe, P. M. W. Gill, B. Johnson, W. C., M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision B.01; Gaussian, Inc.: Pittsburgh, PA, 2003.

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

(21) Almenningen, A.; Bastiansen, O.; Motzfeldt, T. *Acta Chem. Scand.* **1969**, *23*, 2848−2864.

(22) He, J.; Petrovic, A. G.; Polavarapu, P. L. *J. Phys. Chem.* **2004**, *108*, 1671−1680. He, J.; Petrovic, A. G.; Polavarapu, P. L. *J. Phys. Chem B* **2004**, *108*, 20451−20457.

(23) Wang, F.; Polavarapu, P. L. *J. Phys. Chem. A* **2000**, *104*, 10683−10687.

CT049905J