J. Chem. Theory Comput. **2006,** *2,* 1203−1219

Journal of Chemical Theory and Computation

Computational Improvements to Quantum Wave Packet ab Initio Molecular Dynamics Using a Potential-Adapted, Time-Dependent Deterministic Sampling Technique

Jacek Jakowski, Isaiah Sumner, and Srinivasan S. Iyengar*

Departments of Chemistry and Physics, Indiana University, 800 E. Kirkwood Avenue, Bloomington, Indiana 47405

Received April 9, 2006

Abstract: In a recent publication, we introduced a computational approach to treat the simultaneous dynamics of electrons and nuclei. The method is based on a synergy between quantum wave packet dynamics and ab initio molecular dynamics. Atom-centered density-matrix propagation or Born−Oppenheimer dynamics can be used to perform ab initio dynamics. In this paper, wave packet dynamics is conducted using a three-dimensional direct product implementation of the distributed approximating functional free-propagator. A fundamental computational difficulty in this approach is that the interaction potential between the two components of the methodology needs to be calculated frequently. Here, we overcome this problem through the use of a time-dependent deterministic sampling measure that predicts, at every step of the dynamics, regions of the potential which are important. The algorithm, when combined with an on-the-fly interpolation scheme, allows us to determine the quantum dynamical interaction potential and gradients at every dynamics step in an extremely efficient manner. Numerical demonstrations of our sampling algorithm are provided through several examples arranged in a cascading level of complexity. Starting from a simple one-dimensional quantum dynamical treatment of the shared proton in $[Cl-H-Cl]^-$ and $[CH_3-H-Cl]^-$ along with simultaneous dynamical treatment of the electrons and classical nuclei, through a complete three-dimensional treatment of the shared proton in $[Cl-H-Cl]^-$ as well as treatment of a hydrogen atom undergoing donor−acceptor transitions in the biological enzyme, soybean lipoxygenase-1 (SLO-1), we benchmark the algorithm thoroughly. Apart from computing various error estimates, we also compare vibrational density of states, inclusive of full quantum effects from the shared proton, using a novel unified velocity−velocity, flux−flux autocorrelation function. In all cases, the potential-adapted, time-dependent sampling procedure is seen to improve the computational scheme tremendously (by orders of magnitude) with minimal loss of accuracy.

Introduction

The time-dependent Schrödinger equation is a starting point for several dynamical methodologies in gas-phase[1] and condensed-phase quantum mechanics.[2] The Born−Oppenheimer approximation, when invoked, allows for separation of electronic and nuclear degrees of freedom and the subsequent propagation of nuclei, quantum-mechanically,[1,3−27] classically,[28−32] or semiclassically[33−48] on fitted electronic surfaces or "on-the-fly"[28−33,40−43,49−51] approximations to the same. For cases where fitted electronic surfaces are used, the required number of quantum chemical calculations to obtain a representation of the surface can be large depending upon the size of the system. It is in this regard that "on-the-fly" approaches to dynamics of nuclei and electrons[28−33,40−43,49−51] have recently become popular, leading to the subfield of ab initio molecular dynamics (AIMD).

In AIMD, an approximation to the electronic wave function is propagated along with the nuclear degrees of freedom to simulate dynamics on the Born−Oppenheimer surface. AIMD approaches can be broadly categorized as

---

* Corresponding author e-mail: iyengar@indiana.edu.

10.1021/ct600131g CCC: $33.50

(a) Born−Oppenheimer molecular dynamics approaches, where the electronic degrees of freedom are represented using self-consistently converged electronic structure calculations (at the level of density functional theory, Hartree−Fock, post-Hartree−Fock or semiempirical methods) or (b) extended Lagrangian approaches,[52,53] where an approximation to the electronic structure is propagated through an adjustment of the relative time-scales of electrons and nuclei.[30,32,50,54,55] The AIMD approaches, when combined with full quantum or semiclassical dynamics schemes, have the potential to treat large problems accurately with the complete machinery of quantum dynamics. Several steps have been taken in this direction both for real and imaginary time quantum propagation.[33,42,56−60]

We have recently developed an approach[59,60] that attempts to overcome some computational bottlenecks in this area. Our approach is quantum-classical[39,61−67] and combines quantum-wave packet dynamics treatment of the time-dependent Schrödinger equation with ab initio molecular dynamics (where the nuclei are treated classical). The latter can be performed using Atom-centered Density-Matrix Propagation (ADMP)[32,55] or Born−Oppenheimer molecular dynamics (BOMD).[28,29,31] The wave packet dynamics is performed through an analytic, banded, Toeplitz representation to the discretized free propagator known as "distributed approximating functionals" (DAF).[27,68−72] [The (i,j)th element of a Toeplitz matrix depends only on $|i - j|$. This property of the free propagator used in the current contribution yields an efficient scheme where only the first (banded) row of the matrix representation of the time-evolution operator needs to be stored.] Some notable features of our wave packet AIMD approach are as follows: (a) The quantum dynamical free-propagation is formally exact and computationally efficient. It includes important quantum-effects such as zero-point effects, tunneling, and over-barrier reflections. (b) The electronic structure is simultaneously treated using hybrid density functionals (for example, B3LYP)[73−75] and derives its strength from established linear scaling electronic structure techniques.[76−78] (The computational complexity of the quantum dynamics formalism scales linearly with grid basis size.[59])

However, there remains an important computational difficulty in our approach.[59] An approximation to the time-dependent interaction potential between the quantum dynamical system and the AIMD system is required on the grid representation of the quantum wave packet, at every instant in time. If ab initio methods are used to represent the electronic degrees of freedom, this presents a major computational task and hinders the direct application of this methodology for higher dimensional quantum dynamical subsystems. We overcome this hurdle here by introducing a potential adapted time-dependent, deterministic sampling measure that allows us to only evaluate the time-dependent interaction potential in important regions. The sampling measure at any given step during the dynamics is cast as a function of the interaction potential, the gradients of the potential with respect to the quantum dynamical particle and wave packet amplitude, all of which are determined at the previous dynamics step.

The paper is organized as follows: For the convenience of the reader, the dynamics formalism is briefly reviewed in section 2. Details can be found in refs 59 and 60. The time-dependent deterministic sampling algorithm is briefly introduced toward the end of section 2 and discussed in more detail in sections 3 and 4. Appendices A−C facilitate this formal discussion. Numerical tests are presented in section 5 and include demonstrations of the computational improvement obtained from the sampling technique for quantum dynamical treatment of a single particle undergoing dynamics along with the surrounding electronic degrees of freedom and classical nuclear degrees of freedom. Examples include treatment of the single quantum particle in one dimension and in full three dimensions, and the systems studied involve hydrogen, hydride or proton-transfer processes in $[\mathrm{Cl{-}H{-}Cl}]^-$, $[\mathrm{CH_3{-}H{-}Cl}]^-$, and the biological enzyme, soybean lipoxygenase-1 (SLO-1). In all cases the computational improvements seen due to the use of time-dependent deterministic sampling are excellent. Our conclusions and outlook toward further improvements are presented in section 6.

II. A Quantum Wave Packet Generalization for ab Initio Molecular Dynamics

In this section we first summarize the governing equations of motion and approximations involved in the quantum wave packet ab initio dynamics formalism. (Further details can be found in refs 59 and 60.) We then proceed to highlight the main computational bottlenecks which form the basis for the rest of this paper.

To efficiently propagate an electron−nuclear system, we start from the time dependent Schrödinger equation (TDSE) and assume that, in large systems, some nuclei may require only classical treatment, but other parts of the system may need to be treated by applying quantum dynamics.[59,60] As a result, we partition the electron−nuclear system into separate, but interacting subsystems:[45−48] Subsystem 1 comprises particles that are treated quantum dynamically. Subsystem 2 contains the nuclei whose description will be given classically, and subsystem 3 comprises the electrons in the system. In the discussion below and the rest of this paper, we represent the position variables for the particles in subsystem 1 as R_{QM}. Similarly, the particles in subsystem 2 are represented using the position variables $\mathbf{R}_C$, and those in subsystem 3 are represented using r_e. We then invoke the time-dependent self-consistent field (TDSCF) mean field separation[45,46,61,79] wherein the full wave function is approximated as a product: $\psi(R_{QM}, \mathbf{R}_C, r_e; t) \equiv \psi_1(R_{QM})\psi_2(\mathbf{R}_C)\psi_3(r_e)\exp[\iota\gamma]$, and $\dot{\gamma}$ (that is the first time-derivative of γ) is proportional to twice the energy of the system. This leads to three separate time-dependent Schrödinger equations, one for each subsystem, and the Hamiltonian for each separate Schrödinger equation is mean-field and depends on the state of the other two subsystems, for example, $H_1 = \langle \psi_2\psi_3|\mathscr{H}|\psi_2\psi_3\rangle$. (See ref 59.)

In the next step we recognize that the particles in subsystem 1 are to be treated using quantum dynamics, and hence we retain the time-dependent Schrödinger equation for subsystem 1, following dynamics under the Hamiltonian

H_1. But, since subsystem 2 comprises nuclei that are not required to be treated within a quantum dynamical formalism (since they are not part of subsystem 1), we enforce the classical limit ($\hbar \rightarrow 0$)[68,64,66,67,80−87] for these particles. In subsystem 3 we choose to enforce the space-time separation of electrons to obtain a stationary state (Born−Oppenheimer) description of electrons. In this fashion we recover a formalism where a portion of the full system is treated using quantum dynamics, another portion of the system is treated classically, while a third portion (the electrons) is described within a stationary state approximation. The description of the electrons can be achieved using Hartree−Fock, DFT, semiempirical, or post-Hartree−Fock formalisms, and if any of the single particle descriptions are used (Hartree−Fock, DFT, or semiempirical), then the electronic subsystem obeys a self-consistent field (SCF) procedure which leads to a converged single particle electronic density matrix, $\mathbf{P}_C$. In this situation, the equations of motion for subsystems 1 and 2 are

$$\imath\hbar\frac{\partial}{\partial t}\chi(R_{QM};t) = \left[-\frac{\hbar^2}{2M_{QM}}\nabla^2_{R_{QM}} + E(\{\mathbf{R}_C, \mathbf{P}_C\}, R_{QM})\right] \times \chi(R_{QM};\, t) \quad (1)$$

and

$$\mathbf{M}\frac{d^2\mathbf{R_C}}{dt^2} = -\left\langle \chi \left| \frac{\partial E(\{\mathbf{R_C},\mathbf{P_C}\}, R_{QM})}{\partial \mathbf{R_C}} \right|_{\mathbf{P}_C} \right| \chi \right\rangle \quad (2)$$

where $\chi(R_{QM};\, t)$ represents the quantum dynamical wave packet, $\mathbf{M}$ denotes the classical nuclear masses, and M_{QM} denotes the mass of the quantum subsystem particle(s). As stated above the equation for $\mathbf{P_C}$ is a self-consistent field equation arising from the single particle methodology of choice. If post-Hartree−Fock treatments are used, then the energy functionals in eqs 1 and 2 are appropriately modified. Subsystem 2 experiences an "averaged" force[88] that depends on the instantaneous wave packet χ.

An alternative description is obtained by employing the recently developed Atom-centered Density Matrix Propagation (ADMP) formalism[32,55,87,89−92] for subsystems 2 and 3. To arrive at this step, we first note that classical dynamics of subsystem 2 and the simultaneous self-consistent treatment of electrons together comprise the now well-known ab initio molecular dynamics (AIMD) paradigm.[28−33,40,49−51,93] It has been noted that under conditions of "adiabatic control",[50,89,90,94] extended-Lagrangian formalisms[52,53] such as ADMP can provide dynamical results in good agreement with Born Oppenheimer dynamics and experiment.[89,95] When ADMP is used to describe the dynamics of the electrons, the equation of motion for subsystem 3 is

$$\underline{\mu}^{1/2}\frac{d^2\mathbf{P_C}}{dt^2}\underline{\mu}^{1/2} = -\left\langle \chi \left| \frac{\partial E(\{\mathbf{R_C},\mathbf{P_C}\}, R_{QM})}{\partial \mathbf{P_C}} \right|_{\mathbf{R}_C} \right| \chi \right\rangle - [\Lambda\mathbf{P_C} + \mathbf{P_C}\Lambda - \Lambda] \quad (3)$$

Here, $\underline{\mu}$ is a fictitious inertia tensor[32,55,89,90] describing the motion of $\mathbf{P_C}$, and Λ is a Lagrangian multiplier matrix used to impose N-representability of $\mathbf{P_C}$. The energy functional, $E(\{\mathbf{R_C},\mathbf{P_C}\}, R_{QM})$, in eqs 3 and 2, is obtained from a density functional treatment (or any other single particle treatment) of the electrons. The equations of motion for subsystem 2 remain the same as in eq 2 apart from the fact that the forces used in ADMP are different from that in Born−Oppenheimer[32,90] through the inclusion of an additional term that depends on the commutator of the single-particle electronic Hamiltonian (or Fock matrix) and $\mathbf{P_C}$. The ADMP subsystem also experiences an "averaged" force similar to subsystem 2. This averaged force and the N-representability constraints propagate $\{\mathbf{R_C}, \mathbf{P_C}\}$ which in turn affects the energy $E(\{\mathbf{R_C}, \mathbf{P_C}\}, R_{QM})$ and determines the dynamics of χ. The system of eqs 1−3 are thus coupled and are solved simultaneously as a single initial value problem.

If Born−Oppenheimer (BO) dynamics is used to represent subsystems 2 and 3, eq 3 is substituted by SCF convergence of $\mathbf{P_C}$. As a result, there arises an important and subtle difference between the ADMP and BO wave packet implementations. In BO, the density matrix, $\mathbf{P_C}$, becomes a function of both $\mathbf{R_C}$ and R_{QM}, but in ADMP, $\mathbf{P_C}$ does not depend on R_{QM}; $\mathbf{P_C}$ only depends on the distribution of the wave packet, $\chi(R_{QM};t)$. Thus in ADMP, the dynamics of $\mathbf{P_C}$ is Ehrenfest-like, while this is not the case when BO is used for the dynamics of $\mathbf{R_C}$, $\mathbf{P_C}$ in conjunction with quantum dynamics. As a result, in ADMP the calculation of $E(\{\mathbf{R_C}, \mathbf{P_C}\}, R_{QM})$ for each additional grid point is simplified, since the terms that depend only on $\mathbf{P_C}$ (the two-electron integrals in the Fock matrix) do not need to be recomputed. This important and subtle difference will be benchmarked in future publications.

The time-evolution of χ, in the coordinate-representation, is approximated using the symmetric split operator[5,96−98]

$$\chi(R^i_{QM};t+\Delta t) = \sum_j \exp\left\{-\frac{\imath V(R^i_{QM})}{2\hbar}t\right\}\tilde{K}(R^i_{QM}, R^j_{QM}) \times \exp\left\{-\frac{\imath V(R^j_{QM})}{2\hbar}t\right\}\chi(R^j_{QM};t) \quad (4)$$

and the free-propagator, $\tilde{K}(R^i_{QM}, R^j_{QM})$, is represented using distributed approximating functionals (DAF):[59,60,68,70]

$$\tilde{K}(R^{\rm i}_{QM}, R^{\rm j}_{QM}) = \frac{1}{\sigma(0)}\exp\left\{-\frac{(R^{\rm i}_{QM} - R^{\rm j}_{QM})^2}{2\sigma(\Delta t_{QM})^2}\right\} \times \sum_{n=0}^{M/2}\left(\frac{\sigma(0)}{\sigma(\Delta t_{QM})}\right)^{2n+1}\left(\frac{-1}{4}\right)^{\rm n}\frac{1}{n!}(2\pi)^{-1/2}H_{2n}\left(\frac{R^{\rm i}_{QM} - R^{j}_{QM}}{\sqrt{2}\sigma(\Delta t_{QM})}\right) \quad (5)$$

In eq 5, H_{2n} are even order Hermite polynomials.[99] The structure $\tilde{K}(R^i_{QM}, R^j_{QM}) \equiv \tilde{K}(|R^i_{QM} - R^j_{QM}|)$, renders a great deal of efficiency to our quantum propagation as discussed in refs 27, 59, 60, and 70. The evolution of $\{\mathbf{R_C}, \mathbf{P_C}\}$ is based on the velocity Verlet integrator[100] as discussed in ref 59.

An important advantage of this dynamical procedure for $\{\chi(R_{QM};t);\mathbf{R_C},\mathbf{P_C}\}$ is that there is no need to prepare the potential energy surface a priori, since the potential in eq 1 is obtained on-the-fly. However, the need to obtain an approximation to the energy and gradients, $\langle\chi|\partial E(\{\mathbf{R_C},\mathbf{P_C}\},$

$R_{QM})/\partial \mathbf{R_C}|_{\mathbf{P_C}}|\chi\rangle$ and $\langle\chi|\partial E(\{\mathbf{R_C},\mathbf{P_C}\},R_{QM})/\partial \mathbf{P_C}|_{\mathbf{R_C}}|\chi\rangle$, at every time-step, also constitutes a computational bottleneck in the current procedure, and the complexity grows with the number of grid points defined as part of the quadrature scheme in eq 4. Hence, the number of grid points where the electronic structure energy and gradients are evaluated needs to be optimized. This should bring the overall scaling of the algorithm down from a dependence on the total number of quantum dynamical grid points to a small fraction where the electronic structure calculations are performed. In section III, we introduce a time-dependent deterministic sampling measure which helps us to predetermine the relevant regions of the potential where the energy and gradients are to be obtained; the values of these parameters in the other regions are obtained through an efficient interpolation scheme. This procedure greatly reduces the computational cost associated with the algorithm, as will be seen in section V.

III. Time-Dependent Deterministic Sampling of the Quantum Potential Surface

To maintain accuracy of the quantum propagation, it is desirable to have the discretized coordinate representation (or grid) cover a large area and assume that the wave packet vanishes outside the grid boundaries. [The boundary conditions used here are the standard Dirichlet type boundary conditions. One could also use Neumann boundary conditions to solve this problem, and this generalization to the methodology is currently in progress.[137]] The density of grid points is then required to be high enough in "important" regions so as to facilitate accurate evaluation of the wave packet. This, however, leads us to a computational bottleneck for an on-the-fly calculation of the electronic energy and gradients. Here, we aim to dynamically determine an optimal and adequate set of grid point locations where the electronic energy and gradients are to be evaluated such that the error in the quantum dynamics is minimal and controllable. We introduce a potential adapted, time-dependent, deterministic sampling measure for this purpose. The formal discussion on the sampling measure in section III and subsequent discussion of the algorithm in section IV are couched in one dimension. The numerical implementation of this methodology in higher dimensions for Cartesian-like grids is discussed in section IV B. Multidimensional benchmarks are provided in sections V B and V C.

Consider a uniformly spaced set of N_Q grid points in the closed interval $[a, b]$. The number N_Q could in principle be large and constitutes the set of points where quantum dynamical propagation is to be performed. Let N_E represent a subset of these N_Q grid points ($N_E \ll N_Q$) where the electronic structure energies and gradients are to be obtained. The location of these N_E points needs to be optimal and *determined dynamically*, so that the propagation error

$$\sigma_{N_E} = \frac{1}{T}\int_0^T \{\|\chi_{ref}(t) - \chi_{N_E}(t)\|_2\}^2 \mathrm{d}t \tag{6}$$

is small for both ADMP and BO wave packet implementations. The quantity $\|A\|_2$ is the L^2 norm[101,102] of vector A, and χ_{ref} denotes the reference wave packet propagated on a potential energy surface determined by electronic structure evaluations on all N_Q uniformly spaced grid points. The function χ_{N_E} denotes the propagated wave packet on a surface obtained using potential and gradient evaluations only on the N_E, in general nonuniformly distributed, grid points. The potential values at the remaining $(N_Q - N_E)$ grid points are to be obtained from a suitable interpolation scheme and used in the quantum propagation of χ_{N_E}. The quantity T in eq 6 is the total time of propagation.

To determine the position of the N_E grid points, we introduce a sampling function, $\omega_0(R_{QM})$, which represents the density of points where the ab initio potential is to be evaluated. While constructing $\omega_0(R_{QM})$, we assume that it depends on the density of the wave packet, $\rho(R_{QM}) \equiv |\chi(R_{QM})|^2$, the potential energy $E(\{\mathbf{R_C},\mathbf{P_C}\}, R_{QM})$, and the magnitude of the potential gradient $[\nabla_{R_{QM}}E]$. We define our sampling function $\omega_0(R_{QM})$ to be inversely proportional to the values of potential energy since, generally, wave packets favor spatial regions with lower potential energy. Similarly, requiring $\omega_0(R_{QM})$ to be directly proportional to the gradients of the potential energy helps to maintain accuracy of integration in areas of the grid where the potential energy changes rapidly.[103] We further assume that $\omega_0(R_{QM})$ is directly proportional to wave packet amplitude $\rho(R_{QM})$. Accordingly our proposed sampling function has the form

$$\omega_0(R_{QM}) \equiv \omega_0(R_{QM};I_V, I_{V'}, I_\chi) = \frac{f_\rho(R_{QM};I_\chi)f_{\mathrm{E}'}(R_{QM};I_{V'})}{f_E(R_{QM};I_V)} \tag{7}$$

where the parameters I_V, $I_{V'}$, and I_χ can take on any numerical value, and the functions f_ρ, $f_{E'}$, and f_E are accordingly defined as

$$f_Y(R_{QM};\, i) = \begin{cases} Y(R_{QM}) & \text{for } i < 0 \\ 1 & \text{for } i = 0 \\ Y(R_{QM}) - Y_{min} + \dfrac{(Y_{max} - Y_{min})}{i} & \text{for } i > 0 \end{cases} \tag{8}$$

Here $Y(R_{QM})$ denotes the wave packet density $\rho(R_{QM})$, potential energy, or gradients. The quantities Y_{max} and Y_{min} denote the maximum and minimum values of Y on the grid. When the index, i, is zero, then f_Y becomes a uniform function independent of Y, and all grid points are considered equally important. When the index is less than zero, then the function $Y(x)$ is used as is in eq 7, and if the index is greater than zero, then $Y(x)$ is shifted as per eq 8. We further assume that ω_0 is L^1 normalized[101] according to

$$\|\omega_0\|_1 = \sum_{i=1}^{N_Q} \omega_0(R^i_{QM})\Delta R_{QM} = \int_a^b |\omega_0(x)|\, dx = 1 \tag{9}$$

where $\{R^i_{QM}\}$ are a set of quantum grid points in the closed interval [a,b], $\Delta R_{QM} = (b - a)/N_Q$ and $\|\cdots\|_1$ defines the L^1-norm. This normalization, in conjunction with the last option in eq 8, is used to regulate the contributions of wave packet density, potential energy, and gradients to the sampling function. This aspect is clear from the fact that for $i > 0$

$$f_{\mathrm{Y}}(R_{QM};\, i) = (Y_{max} - Y_{min})\left[\tilde{Y}(R_{QM}) + \frac{1}{i}\right] \tag{10}$$

where $\tilde{Y}(R_{QM}) = (Y(R_{QM}) - Y_{min})/(Y_{max} - Y_{min})$ and hence

$0 \le \tilde{Y} \le 1$. Thus, for the case of $i > 0$, the normalization condition essentially reduces the sampling function to

$$\omega_0(R_{QM}) \propto \frac{[\tilde{\rho} + 1/I_\chi] \times [\tilde{E}' + 1/I_{V'}]}{\tilde{E} + 1/I_V} \tag{11}$$

where $\tilde{\rho}$, $\tilde{E}'$, and $\tilde{E}$ are each bounded by 1. The contributions from the wave packet, gradients, and energy are thus controlled to allow a flexible scheme to enhance accuracy and efficiency. The performance of eq 7 for various values of I_V, $I_{V'}$, and I_χ and different potentials is presented in section 5. Using the implementations of this approach described in section 4, we choose a set of N_E grid points [$N_E = \alpha N_Q$ and $\alpha \le 1$] where the potential and gradients are evaluated. These data are used in conjunction with an interpolation procedure (discussed in Appendix B) to perform quantum wave packet dynamics on a uniform grid comprising N_Q points.

The interpretation of $\omega_0(R_{QM})$ is such that large values of this quantity indicate regions on the grid where energy and gradient evaluations *should* be conducted accurately using electronic structure methods. When $\omega_0(R_{QM})$ is small, the accuracy of energy and gradients is not critical, and the value of the potential in such regions is obtained though interpolation. Additional interpretations that connect the sampling function in eq 7 to the Wentzel−Kramers−Brillouin (WKB)[80] semiclassical theory and also to Bohmian mechanics[80−87,104−108] are discussed in Appendix A.

IV. Algorithms and Prescriptions for Time-Dependent Deterministic Sampling

A. Conditions on $\omega_0(x)$: The Requirements for a Smooth Transformation. Two important challenges arise while deriving a stable, general, numerical algorithm to carry out grid sampling as per eq 7. Both challenges only arise when the fraction of grid points where electronic structure calculations are to be performed becomes large (i.e., when $\alpha \equiv N_E/N_Q$ is close to 1). While this is not the practically interesting limit, since it is desired that most dynamical calculations are performed for small α, the demand for a stable numerical algorithm requires that we analyze all regions of α, and this analysis is carried out in the current section (section IV A). The computational algorithm of wave packet ab initio dynamics inclusive of sampling is described in section IV B.

For suitably large α the following problems arise. (i) On account of a uniform grid implementation of eq 4 in the current work, the function $\omega_0(R_{QM})$ should tend to a uniform function as $\alpha \to 1$. That is the set of N_E grid points must coincide with the set of N_Q underlying grid points as $\alpha \to 1$. (ii) In general, for any specific grid discretization given by grid spacing ΔR_{QM} and for any normalized function in eq 9, values of $[N_E\omega_0(R_{QM})\Delta R_{QM}]$ could be greater than 1 if α is large. [Note that $N_E\Delta R_{QM} \equiv (b - a)\alpha$.] This would imply that there is a need for more than one potential evaluation in the vicinity of the point R_{QM}, which is clearly impossible. The solutions to both of these challenges are discussed for the remaining portion of this section.

As we see below, the first problem above is trivially solved. The second problem, however, necessitates a modification of the sampling function. Consider a smooth transformation, $\omega_0(R_{QM}) \to \omega(R_{QM})$, that maintains the original functional form of $\omega(R_{QM}) \equiv \omega_0(R_{QM})$ for small values of α (the practically interesting case of $N_E \ll N_Q$) but allows $\omega(R_{QM})$ to smoothly approach a uniform distribution as $\alpha \to 1$:

(a) $\omega(R_{QM})$ tends to a uniform distribution as $N_E/N_Q \to 1$ and

(b) $\omega(R_{QM}) \to \omega_0(R_{QM})$ for $N_Q \gg N_E$.

If $\omega(R_{QM})$ is normalized and bounded according to

$$\|\omega(R_{QM})\|_1 = \sum_{i=1}^{N_Q} \omega(R^i_{QM})\Delta R_{QM} = \int_a^b \omega(x)dx = 1 \tag{12}$$

$$\|N_E\omega(R_{QM})\Delta R_{QM}\|_\infty = N_E \max_i |\omega(R^i_{QM})\Delta R_{QM}| \le 1 \tag{13}$$

it follows from eq 13 that

$$\|\omega(R_{QM})\|_\infty \equiv \max_i |\omega(R^i_{QM})| \le \frac{1}{\alpha(b-a)} \equiv \alpha^* \tag{14}$$

where we have used $\Delta R_{QM} = (b - a)/N_Q$. Using eq 12 we have

$$1 = \int_a^b \omega(x)dx \quad \le \quad (b-a)\|\omega(R_{QM})\|_\infty \tag{15}$$

and thus

$$\|\omega(R_{QM})\|_\infty \quad \ge \quad \frac{1}{b-a} \tag{16}$$

Using eqs 14 and 16 we obtain upper and lower bounds to the L^∞ norm of $\omega(R_{QM})$ as

$$\frac{1}{(b-a)} \quad \le \quad \|\omega(R_{QM})\|_\infty \quad \le \quad \frac{1}{\alpha(b-a)} \tag{17}$$

and hence

$$\lim_{\alpha \to 1} \omega(R^i_{QM}) = (b-a)^{-1} \tag{18}$$

Thus $\omega(R^i_{QM})$ tends to a uniform distribution in the limit $\alpha \to 1$, and condition (a) above is automatically satisfied from the requirements of eqs 12 and 13.

The second requirement, condition (b), is satisfied by introducing a smooth transformation that uses an additive correction function $U(R_{QM})$ to suitably compensate high-density regions (i.e., regions of $\omega_0(R_{QM})$ that are in violation of eq 13) of $\omega_0(R_{QM})$. Since this scheme has the physical effect of equalizing the sampling weights, it is referred to as the “spreading technique” and is the subject of the section below.

1. Smooth Transformation Spreading Technique. We now construct a spreading transformation to satisfy the condition (b) above, i.e., $\omega(R_{QM}) \to \omega_0(R_{QM})$ for $N_Q \gg N_E$. Let us start with a modification of $\omega_0(R_{QM})$ that satisfies eq 13. Toward this we construct two projection operators Ω_0 and Ω_0^c, such that

$$\Omega_0\, f(x) = f(x) \cdot \theta[\omega_0(x) - \alpha^*] \tag{19}$$

where f is an arbitrary function defined in the closed interval

$[a, b]$, the quantity $\theta[\omega_0(x) - \alpha^*]$ is a Heaviside function that is equal to 1 when the argument in parentheses is greater than or equal to zero, and α^* is defined in eq 14. The operator Ω_0^c is the orthogonal complement of Ω_0 and is defined as

$$\Omega_0^c\ f(x) = f(x)\cdot\{1 - \theta[\omega_0(x) - \alpha^*]\} \tag{20}$$

The projection operators Ω_0 and Ω_0^c by definition exhaust the closed interval $[a, b]$, i.e.

$$I = \Omega_0 + \Omega_0^c \tag{21}$$

and I is the identity matrix.

If we assume a grid discretization comprising N_Q evenly spaced grid points in $[a, b]$, it is then our aim to obtain N_E optimally placed points as per the sampling function defined earlier. However, as a consequence of eqs 13 and 14 it follows that the operator Ω_0 represents a projection onto the region in space where $|N_E\omega_0(x)\mathrm{d}x| \equiv \omega_0(x)/\alpha^* \geq 1$. This is precisely the region where potential evaluations need to be performed. However, there is excess sampling density in this region which violates eq 13. This excess sampling density is exactly equal to $||\Omega_0\tilde{\omega}_0||_1$, where $\tilde{\omega}_0 \equiv \{\omega_0(x) - \alpha^*\}$, and needs to be “spread” uniformly in the region represented by the projection operator Ω_0^c. To achieve this, we subtract this excess density from the regions belonging to Ω_0 and add this excess density *uniformly* to the region belonging to Ω_0^c. Consequently an additive correction to ω_0 is introduced

$$\omega_1 = \omega_0 + U(\omega_0) \tag{22}$$

where the additive spreading function $U(\omega_0)$ is defined as

$$U(\omega_0) = -\Omega_0\tilde{\omega}_0 + \frac{||\Omega_0\tilde{\omega}_0||_1}{||\{\Omega_0^c\mathbf{1}\}||_1}\{\Omega_0^c\mathbf{1}\} \tag{23}$$

Again, the first term of the above equation subtracts the excess density from the regions belonging to Ω_0. The second term adds this excess density *uniformly* to the region belonging to Ω_0^c. In the second term, the quantity $\{\Omega_0^c\mathbf{1}\}$ represents the projection of a constant vector ($\mathbf{1(x)} \equiv 1$ for $x \in [a, b]$) onto the region belonging to Ω_0^c.

Equations 22 and 23 summarize the spreading technique. After applying the spreading transformation to $\omega_0(R_{QM})$, the excess sampling density described above is removed. However, in this process a similar excess density may be created in the region belonging to Ω_0^c. Thus, the sampling function ω_1 obtained from eq 22 may suffer the same way as ω_0 in the sense that there probably exists a finite region, $\Omega_1\ f(x)$, on the grid:

$$\Omega_1\ f(x) = f(x)\cdot\theta[\omega_1(x) - \alpha^*] \tag{24}$$

This implies that ω_1 does not satisfy eq 14 either. However, the excess density in the region represented by eq 24 is always smaller in magnitude than $||\Omega\tilde{\omega}_0||_1$ (see Appendix C for details on the convergence properties of the spreading technique) but still *needs to be spread again*. Consequently, the procedure described in eqs 22 and 23 is to be iterated according to

$$\omega_{i+1} = \omega_i + U(\omega_i) \tag{25}$$

where $U(\omega_i)$, as usual, corrects the excess density in ω_i:

$$U(\omega_i) = -\Omega_i\tilde{\omega}_i + \frac{||\Omega_i\tilde{\omega}_i||_1}{||\{\Omega_i^c\mathbf{1}\}||_1}\{\Omega_i^c\mathbf{1}\} \tag{26}$$

Hence, the generalized spreading transformation can be summarized using eqs 25 and 26 as

$$\begin{aligned}\omega_{i+1} &= \left(\Omega_i\alpha^* + \frac{||\Omega_i\tilde{\omega}_i||_1}{||\{\Omega_i^c\mathbf{1}\}||_1}\{\Omega_0^c\mathbf{1}\}\right) + \Omega_i^c\omega_i \\ &= \Omega_i\alpha^* + \Omega_i^c\left(\frac{||\Omega_i\tilde{\omega}_i||_1}{||\{\Omega_i^c\mathbf{1}\}||_1} + \omega_i\right)\end{aligned} \tag{27}$$

The convergence of this iterative technique is analyzed in Appendix C, and the scheme is shown to converge unconditionally.

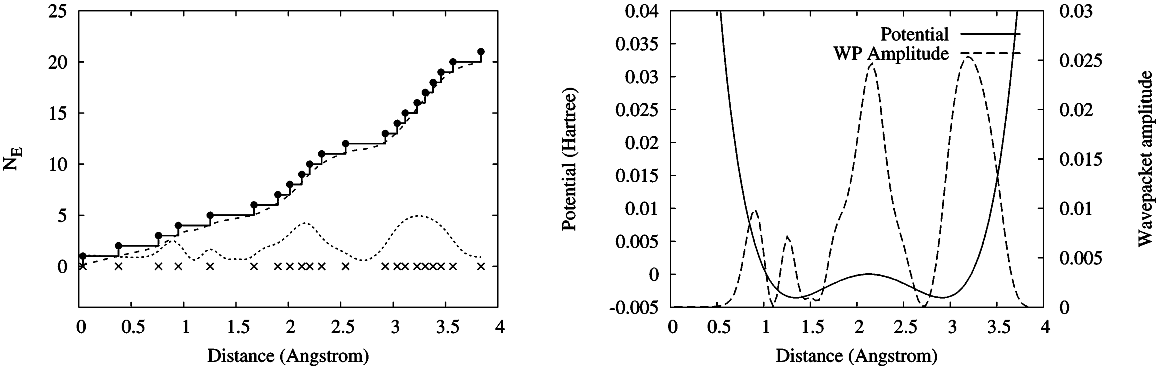
It is further interesting to note that the procedure highlighted above, wherein the family of operators $\{\Omega_i\}$ and $\{\Omega_i^c\}$ modify the sampling function ω_0, has properties reminiscent of multiresolution analysis.[92,109−113] Our scheme, however, differs in that the *multiresolution* is based on spreading the “excess density” $||\Omega_0\tilde{\omega}_0||_1$ as opposed to a translation-dilation process used in standard wavelet theory.

We now comment on the behavior of ω as $\alpha \rightarrow 0$. In this case the upper bound to $||\omega(R_{QM})||_\infty$ in eq 14 goes to infinity, and spreading is not required. Consequently, $\omega(R_{QM}) \rightarrow \omega_0(R_{QM})$ for $N_Q \gg N_E$ (condition (b) in section IV A). However, for values of α greater than zero, the spreading transformation enforces the condition in eq 14 as highlighted above.

B. Computational Implementation of Potential Adapted Time-Dependent Deterministic Sampling for Wave Packet ab Initio Molecular Dynamics. The algorithm to perform mixed quantum wave packet and ab initio dynamics (ADMP or BO) of the $\{\mathbf{R_C}, \mathbf{P_C}, \chi(\mathbf{R_{QM}})\}$ system, based on the sampling scheme, is outlined in this section.

As a preliminary step, a set of N_Q grid points, $\{\mathbf{R_{QM}^i}\}$, is created to represent the discretization of the quantum wave packet. The initial wave packet is defined on this grid. In our studies we have used (a) a real Gaussian function with width chosen such that the wave packet vanishes at the edges of the grid, (b) a thermally sampled linear combination of the eigenstates of the quantum Hamiltonian in eq 1 at $t = 0$, and (c) eigenstates of the quantum Hamiltonian in eq 1 at $t = 0$, as initial wave packets.

During each propagation cycle, the first step is to prepare a potential energy surface and corresponding gradients. The potential is used for quantum propagation, and the gradients are used to construct the AIMD forces (ADMP or BO) for propagation of the classical nuclei and electronic structure. The grid points where the potential and gradients are calculated is controlled by the sampling technique discussed earlier. Toward this, the sampling function is calculated based on the wave packet density, potential energy, and gradients of potential energy obtained from the previous step. [For the first wave packet AIMD step, the potential energy, gradients, and wave packet density are not available. Thus,



(a) (b)

Figure 1. An illustration of the sampling function, $\omega(R_{QM})$, for $N_E = 21$ and $N_Q = 101$. Part (a) shows the sampling function (dotted line), its integral (dashed line), and ceiling function $\lceil \int_{x_0}^{x} \mathrm{d}x'\omega(x') \rceil$. The 21 grid points are set at discontinuities of the ceiling function, and these points are represented in the figure using black dots on the discontinuous ceiling function and "X" at corresponding spots along the abscissa. Part (b) shows the corresponding potential surface and density of wave packet from which the sampling function is calculated.

the initial sampling function is chosen to be uniform.] The condition of eq 13 is then verified, and the iterative spreading scheme described in section IV A1 is applied to ω_0 if required. For small values of α ($\equiv N_E/N_Q$), spreading is generally not needed. Once the sampling function is determined based on the discussion in sections IV A, the grid points where the potential and gradients evaluations are to be conducted are determined based on the dimensionality of the grid as follows.

1. One-dimensional quantum propagation: The potential is sampled at grid points where the ceiling function $\lceil N_E \int_{x_0}^{x} dx'\omega(x') \rceil$ becomes discontinuous and is demonstrated in Figure 1. Here, $\lceil y \rceil$ represents the least integer greater than or equal to y and is called the ceiling of y. Accordingly, a grid map is created that describes the location of the N_E grid points where the potential energy and gradients are to be evaluated. This one-dimensional algorithm is used to obtain acceptable values of I_V, $I_{V'}$, and I_χ, to perform preliminary tests and also in the higher dimensional algorithm described below. The approach here has close connections to the theory of Haar wavelet transforms popular in digital signal processing.[76,92,109−115]

2. Quantum propagation in higher dimensions: There are in turn two options that we consider here. In one case, we construct reduced (or marginal) one-dimensional sampling functions according to

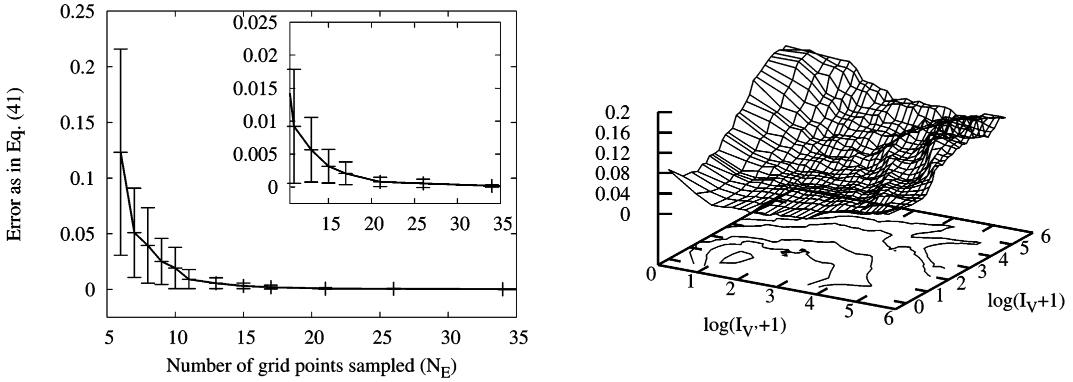
$$\omega_x(x) = \int \mathrm{d}y\mathrm{d}z\, \omega(x, y, z) \tag{28}$$

and similarly for $\omega_y(y)$ and $\omega_z(z)$. The sampling along each direction is then performed using the one-dimensional algorithm discussed above, and the overall grid is constructed as a direct sum. This algorithm is called marginal sampling in further discussion. The simplicity of this approach is appealing, and we find that the results obtained this way are sufficiently good as seen in sections V B and V C. A second algorithm that is more nonisotropic is as follows. The complete set of N_Q grid points are first partitioned into regular rectangular parallelepipeds (or cuboids) such that the integral of ω inside each parallelepiped is roughly a constant. This implies that each parallelepiped must contain the same number of sampling grid points, although their respective sizes may be different. If the number of sampling grid points inside each cuboid is chosen to be a small fraction of N_E, we may perform marginal sampling (eq 28) inside each parallelepiped. Both of these approaches constitute general schemes that we implement in arbitrary dimensions. The second scheme above, again, has strong connections to wavelet theory, and this aspect will be investigated as part of future publications.

Once the grid points are determined as above, the potential and gradients are computed, and the averaged AIMD forces in eqs 2 and 3 are updated. Since the wave packet is defined on the entire grid, while the potential and gradients are evaluated only on N_E grid points, estimation of the gradients on $[N_Q - N_E]$ grid points is required to calculate AIMD forces as per eqs 2 and 3. In this paper we have used a linear interpolation scheme to interpolate all gradients. The evaluated potential on the N_E grid points is used to obtain approximations to the potential energy at the remaining grid points using the Hermite curve interpolation scheme[116,117] described in Appendix B. For this $E(\{\mathbf{R_C},\mathbf{P_C}\}, R_{QM})$ and $\partial E(\{\mathbf{R_C},\mathbf{P_C}\}, R_{QM})/\partial \mathbf{R}_{QM}$ at all N_E grid points are used.

Next, the potential energy on the grid along with the DAF free propagator given by eq 5 are used for causal propagation of wave packet, $\chi(t)$, according to eq 4. The averaged energy gradients $\langle \chi(t)|\partial E(\{\mathbf{R_C},\mathbf{P_C}\},R_{QM})/\partial \mathbf{R_C}|\chi(t)\rangle$ and $\langle \chi(t)|\partial E(\{\mathbf{R_C},\mathbf{P_C}\},R_{QM})/\partial \mathbf{P_C}|\chi(t)\rangle$ are used to propagate $\mathbf{R_C}$ and $\mathbf{P_C}$. (The density matrix, $\mathbf{P_C}$, is obtained through SCF convergence for Born−Oppenheimer dynamics.) The next propagation cycle starts by calculating the sampling function and determining the grid positions for potential evaluation.

The propagation scheme has one additional feature that allows for further reduction in computational cost. As a preliminary step, the distance between all nuclei for all N_E grid points is calculated, and if any of these is found to be smaller than a fixed threshold, then the point is skipped during the potential and gradient evaluation and interpolated later. This aspect is similar to that used in the partial multidimensional grid method of Iordanov et al.,[118] where



(a) (b)

Figure 2. Part (a) shows the average error represented by eq 29 for $N_Q = 101$. Maximum deviations from the average error for individual potentials is displayed using the vertical error bars. As can be seen the error is negligible for all potentials beyond $N_E = 15$. Part (b) shows the behavior of σ_{av} in eq 30 as a function of I_V and I_V'. This part indicates that on average $I_V = 1$ and $I_V' = 3$ are a good choice.

the authors use a fixed energy cutoff based on short-range interaction between particles.

V. Numerical Results

In this section, we first obtain optimal values for sampling parameters I_V, $I_{V'}$, and I_χ described by eq 8. Our general idea toward this issue is to perform a multidimensional search in functional space to determine values of these parameters that would be optimal in an average sense. To simplify the problem, we first consider a series of one-dimensional potential energy functions and use these with the sampling measure to determine I_V and $I_{V'}$. (I_χ is set to zero for this portion of the parametrization.) These are then used to perform a series of quantum wave packet simulations in one dimension, and the accuracy of the dynamics is studied as a function of I_χ values. This helps to determine the value of all three parameters, and these calculations are described in section V A. A physical basis for the choice of I_V, $I_{V'}$, and I_χ is also provided at the end of section V A.

In sections V B and V C we use the parameters thus determined to (a) analyze the error in three-dimensional quantum dynamics for the $[\mathrm{Cl{-}H{-}Cl}]^-$ system and a biological enzyme, soybean lipoxygenase-1 (SLO-1), and (b) to obtain a vibrational density of states inclusive of nuclear quantum effects. The density of states are obtained from a unified velocity−velocity, flux−flux autocorrelation function. We show that our sampling measure is very efficient in determining these spectra since small values of N_E are sufficient to determine the spectral data accurately.

A. Optimal Choice of I_V, I_V', and I_χ. To find optimal values for I_V and $I_{V'}$, 10 different model potentials were considered. The choice of these functions was based on their propensity to appear during molecular dynamics simulations. The potentials used in our tests are as follows:

1. Lennard-Jones 6−12 potential with coefficients $C_6 = C_{12} = 100$. The range of independent variable is [1, 3].
2. Morse potential, with minimum depth $D_e = 10$, located at $x_e = 2$. The exponent was chosen to be $\beta = 0.9$. The range of x is [1, 8].
3. Harmonic potential $^1/_2\, k(x - x_0)^2$, with $x_0 = 4$, $k = 0.002$. The range of x is [0, 8].
4. A realistic one-dimensional potential obtained from the $[\mathrm{Cl{-}H{-}Cl}]^-$ system. The distance between chloride ions were fixed at 4.2 Å, and the position of the hydrogen atom was scanned along the chloride−chloride axis to obtain a total of 101 potential energy points at the B3LYP/6-31G level of theory.
5. A symmetric double-well potential $f_{DW}(x) = (x - 1)^2(x - 3)^2$ for $x \in [0.4, 3.6]$.
6. An asymmetric double-well $f_{ASD}(x) = (x - 2)^2[1 + 3(x - 5)^2]/10$ for $x \in [1, 6]$.
7. A damped double-well potential $f_{\mathrm{DDW}}(x) = [(x - 1)^2(x - 3)^2 - 0.4]\exp(- 1.5x)$. The range of x is [0.6, 8]. The well farther from the origin is relatively weak and may be interpreted as supporting a resonance state from dissociation.
8. A plane wave function $f_{\mathrm{S}}(x) = \sin(x)$ for $x \in [0, {}^5/_2\,\pi]$.
9. A simple Gaussian function $f_{\mathrm{G}}(x) = -6 \exp[-(x - 4)^2]$ for $x \in [0, 8]$.
10. A real coherent state $f_{\mathrm{GM}}(x) = \exp(-x^2/32)\cos(0.7x)$.

The last three functions are not standard potentials. These, however, have characteristics that a wave packet might possess.

For each of the above potentials the sampling function was calculated for a set of N_E values. In each case the values of I_V and $I_{V'}$ were scanned, and the normalized error was calculated according to

$$\sigma = \frac{\| f - f_{app} \|_2}{\| f - f_0 \|_2} \tag{29}$$

where $f_0 = (b - a)^{-1} \int_a^b f(x)\mathrm{d}x$, and $f_{app}(x)$ is the approximation to $f(x)$ obtained from sampling and interpolation. For all calculations $N_Q = 101$ and N_E were scanned in the range 6−51 which corresponds to compression factor α in the range 1/17 through 1/2. To cover even smaller values of α a second set of tests were carried out with $N_Q = 1001$ and the same values N_E as above. This resulted in compression ratios ranging from 1/170 through 1/20.

The behavior of the error is shown in Figure 2(a). In addition, deviations from the average error, due to the individual potentials, are indicated through the use of the vertical bars in the figure. To understand these error values, consider the following interpretation of eq 29. The numerator of eq 29 is the L^2 norm of a difference vector when the potential is normalized as per the denominator. Using simple trigonometry one can see that the hyperangle between the *ket-vectors* f and f_{app} in a N_Q-dimensional linear vector space is approximately $\cos^{-1}[(2 - \sigma^2)/2]$ for small σ. Hence for $\sigma = 0.01$ (which is approximately the case for $N_E = 10$ in Figure 2(a)), the hyperangle between f and f_{app} is 0.57 degrees! Clearly, the error due to deterministic sampling and interpolation is negligible for $N_E > 15$ ($N_Q = 101$) and acceptable for $N_E = 11$. We will see in the next section that even smaller values of N_E can be used with acceptable error in vibrational spectroscopic properties, *and as the dimensionality of the problem increases, the method gets more efficient, exponentially*. This provides the opportunity for a large reduction in computational overhead within the current scheme.

To obtain a single set of values for I_V and $I_{V'}$ that work well on average, we first note that for complex systems the interaction potential is rarely a pure model potential such as those used here. In fact, these could in general possess a mixture of the complexities present in each individual function studied here. Consequently, for an optimal choice of I_V and $I_{V'}$ for realistic potentials we consider the average error

$$\sigma_{av}(I_V, I_{V'}) = \sum_i \frac{\| f^{\,i} - f^{\,i}_{app} \|_2}{\| f^{\,i} - f^{\,i}_{0} \|_2} \tag{30}$$

where the sum runs over all the model potentials. Figure 2(b) shows the normalized average potential error calculated according to eq 30 as a function of natural $\log(I_V + 1)$ and natural $\log(I_{V'} + 1)$ for $N_Q = 101$ and $N_E = 10$. The parameters, I_V and $I_{V'}$, are each scanned in the range 0−256. The minimum in the plot is at approximately $I_V = 1$ and $I_{V'} = 3$. Based on this analysis, we find that $I_V = 1$ and $I_{V'} = 3$ are suitable values that work well for the average case complexity, and, hence, these are the values chosen for the rest of the study in this publication.

To find an optimal I_χ we have conducted a series of one-dimensional wave packet dynamics simulations for the $[Cl-H-Cl]^-$ and $[CH_3-H-Cl]^-$ systems. Each simulation was 100 fs long, and there were a total of 70 simulations which helped us determine an optimal value for I_χ. For the case of the $[Cl-H-Cl]^-$, the shared proton is treated quantum dynamically, and two different sets of simulations have been conducted. In one case, the dynamics of the chloride ions and the electrons are treated using ADMP. In the other case, these are treated using Born−Oppenheimer dynamics. The choice of these systems was based on the following criteria. We needed to test the time-dependent deterministic sampling approach in conjunction with both ADMP and Born−Oppenheimer dynamics approximation. The $[Cl-H-Cl]^-$ system presents a simple test, since it is possible to constrain the Cl atoms at various distances and perform the dynamics; if the Born−Oppenheimer dynamics option were chosen for the chloride ions, the potential energy surface experienced by the wave packet would remain roughly the same. Hence this constitutes the simplest test case. However, if the ADMP dynamics option were chosen for the chloride ions, the potential energy surface would change as the wave packet moves, since the potential is determined in ADMP using the electronic density matrix at the wave packet centroid configuration. Hence the ADMP implementation presents a slightly more challenging test case. For the case of $[CH_3-H-Cl]^-$, again the shared proton is treated using quantum dynamics, while all other atoms are treated using ADMP. We constrain the carbon and chloride atoms to remain at a fixed distance; however, the three remaining hydrogen atoms are completely free to move according to the ADMP forces. Furthermore, since ADMP is used, the potential energy surface experienced by the wave packet changes with time. For all cases, a variety of basis sets was studied, and many values of N_E and I_χ were considered. Our results are summarized in Tables 1−3. In Table 1, the data obtained from Born−Oppenheimer subdynamics of the chloride ion (along with quantum wave packet dynamics of the shared proton) are summarized for two different basis sets, Cl−Cl distances, different N_E, and sampling parameter values. The σ_{N_E} values reported in these tables is described in eq 6, which again has the same physical interpretation as the error in eq 29. Hence, these errors are very small for many different values of I_χ. In Tables 2 and 3, the data obtained from ADMP subdynamics (along with quantum wave packet dynamics of the shared proton) are summarized for $[Cl-H-Cl]^-$ and $[CH_3-H-Cl]^-$, respectively. The reference wave packet for all cases is obtained from a full calculation (without sampling) over $N_Q = 101$ grid points. For all cases, the deterministic sampling parameters for potential and gradient

Table 1: Quantum Wave Packet Propagation Errors for the $[Cl-H-Cl]^-$ System[a]

I_V	I_V'	I_χ	σ_{21}	σ_{15}
		B3LYP/6-31G, *R* = 4.2 Å		
1	3	0	0.0000083	0.0001035
1	**3**	**1**	**0.0000036**	**0.0000720**
1	3	3	0.0000048	0.0000870
1	3	7	0.0000302	0.0003881
1	3	15	0.0003555	0.0069267
1	3	31	0.0079059	0.0491048
1	3	63	0.0331312	0.1068715
1	3	127	0.0581166	0.1515045
1	3	255	0.0732637	0.1826569
1	3	−1	0.0929057	0.2199152
		B3LYP/aug-cc-pVTZ, *R* = 4.2 Å		
1	3	0	0.0000171	0.0002318
1	**3**	**1**	**0.0000066**	**0.0001336**
1	3	−1	0.0066394	0.0492104
		B3LYP/aug-cc-pVTZ, *R* = 3.6 Å		
1	3	0	0.0000955	0.0013409
1	**3**	**1**	**0.0000405**	**0.0007540**
1	3	−1	0.2722858	0.4808945

[a] The shared proton is treated quantum-dynamically, while the chloride ions are treated using Born−Oppenheimer dynamics. Symbols σ_{21} and σ_{15} refer to wave packet errors as determined using eq 6 with $N_E = 21$ and $N_E = 15$, respectively. *R* represents the chloride−chloride distance. The optimal sampling parameters are shown in bold. $N_Q = 101$ is used for all cases.

Table 2: Quantum Wave Packet Propagation Errors for $[Cl-H-Cl]^-$ [a]

I_V	I_V'	I_χ	σ_{21}	σ_{15}
B3LYP/6-31G, $R = 4.2$ Å				
1	3	−1	0.15531148	0.32244568
1	**3**	**1**	**0.00001817**	**0.00015913**
B3LYP/aug-cc-pVTZ, $R = 3.6$ Å				
1	3	−1	0.17838404	0.53823434
1	3	0	0.00163375	0.21019838
1	**3**	**1**	**0.00020302**	**0.04537375**

[a] The shared proton is treated quantum-dynamically, while the rest of the system is treated using ADMP. Symbols σ_{21} and σ_{15} refer to wave packet errors as determined using eq 6 with $N_E = 21$ and $N_E = 15$, respectively. The optimal sampling parameters are shown in bold. $N_Q = 101$ is used for all cases.

Table 3: Quantum Wave Packet Propagation Errors for $[CH_3-H-Cl]^-$ [a]

I_V	I_V'	I_χ	σ_{21}	σ_{15}	σ_{11}
B3LYP/6-311+G(d,p)					
1	3	−1	0.12067507	0.27055252	0.45599940
1	3	0	0.00003690	0.00083307	0.00783050
1	**3**	**1**	**0.00004855**	**0.00032730**	**0.00523513**

[a] The shared proton is treated quantum-dynamically, while the rest of the system is treated using ADMP. Symbols σ_{21}, σ_{15}, and σ_{11} refer to wave packet errors as determined using eq 6 with $N_E = 21$, $N_E = 15$, and $N_E = 11$, respectively. The optimal sampling parameters are shown in bold. $N_Q = 101$ in all cases.

have been set to the previously found optimal values of $I_V = 1$ and $I_{V'} = 3$.

We find that $I_\chi = 1$ is the optimal value for quantum wave packet simulation with $I_V = 1$ and $I_{V'} = 3$. The corresponding entries in Tables 1−3 are shown in bold. To further substantiate the accuracy of the sampling approach, in Figure 3 we present the time-evolution of the L^2-norm of the error in the potential (extrapolated potential based on deterministic sampling minus the reference potential calculated on all grid points) for long simulations (10 ps). As can be seen the error is minimal and well within the acceptable range for DFT electronic structure calculations. Furthermore, $N_E = 11$ represents a sizable reduction in the number of grid points where potential evaluations need to be conducted. We further note that the results for the B3LYP/6-31G, $R = 4.2$ Å calculations on the $[Cl-H-Cl]^-$ system described in Table 1 are rather important. *First, at this geometry, the zero-point energy of the proton (≈2 milli-Hartrees) is comparable to the height of the barrier (≈4 milli-Hartrees) separating the protonated states on each chloride. Furthermore, there are multiple states accessible below the barrier and immediately above the barrier. This provides the opportunity for tunneling, zero point effects, and overbarrier reflection of the wave packet playing an important role in the dynamics.* It is interesting to note that the time-dependent sampling measure performs well even under these highly quantum-mechanical, tunneling oriented conditions.

To analyze the physics behind the choice, $I_\chi = 1$, $I_V = 1$, and $I_{V'} = 3$, using eq 11, let us consider two separate regions of a potential energy surface that are of interest in quantum dynamics. (a) Minimum energy region of the potential surface: The region where the potential energy is low, the wave packet amplitude may be high, and the gradient here is generally small. (b) The region where the potential is high and the gradient is also high represents the region of a potential surface that could mark the advent of tunneling. This region also generally comprises a classical turning point, and the standard semiclassical approximation breaks down[80] in this region. (Also see Appendix A.) Thus, this is an interesting region in quantum dynamical simulations, and, on average, the wave packet amplitude may be expected to be small in such regions. In eq 11 the quantities $\tilde{\rho}$, $\tilde{E}'$, and $\tilde{E}$ are each in the region [0,1] (see discussion before eq 11) and hence for region (a)

$$[\omega_0^{1,3,1}]_{\mathrm{Case(a)}} = constant * \frac{[1 + 1/1] \times [0 + 1/3]}{0 + 1/1} = constant * 2/3 \quad (31)$$

where we have used $I_\chi = 1$, $I_V = 1$, and $I_{V'} = 3$ in eq 11 and also substituted the maximum values of $\tilde{\rho}$ and minimum possible values of $\tilde{E}'$ and $\tilde{E}$ as required by the definition of region (a) above. Now, in region (b), the tunneling and classical turning-point regions

$$[\omega_0^{1,3,1}]_{\mathrm{Case(b)}} = constant * \frac{[0 + 1/1] \times [1 + 1/3]}{1 + 1/1} = constant * 2/3 \quad (32)$$

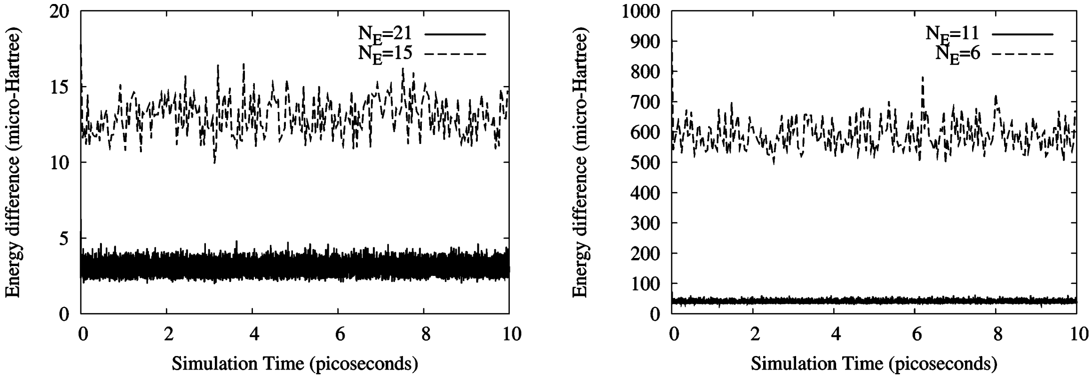
That is, the choice of $I_\chi = 1$, $I_V = 1$, and $I_V' = 3$ provides an equal distribution of grid points in low potential regions as well as tunneling regions.

There can be other possible values of I_χ, I_V, and I_V' that also have the property of producing an equal distribution of grid points in regions (a) and (b). For the special case of $I_\chi = I_V$ it is possible to show that

$$I_V' = (I_\chi + 1)^2 - 1 \quad (33)$$

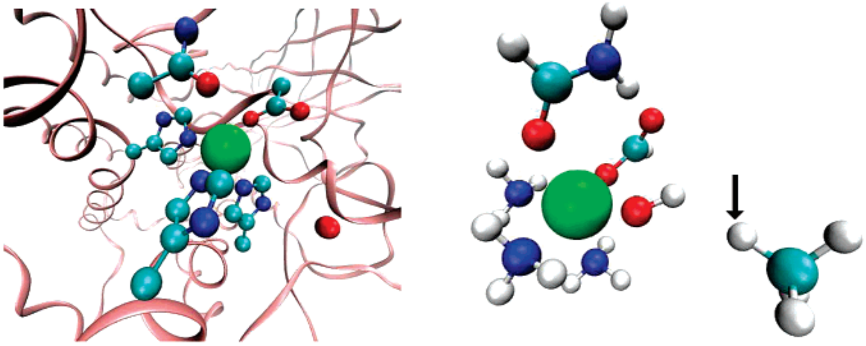
will always provide an equal distribution of points in regions (a) and (b). (We have only considered $I_\chi = I_V$ in this analysis since one would expect the potential and the wave packet amplitude to behave in a mutually complementary fashion.) However, for larger values of I_χ, I_V, and I_V' (larger than 1,3,1) the distribution of points in other classically forbidden regions (regions of high potential and low gradient) becomes smaller, and hence the choice $I_\chi = 1$, $I_V = 1$, and $I_V' = 3$ provides a good physical choice as is seen from the numerical experiments described in this section.

B. Quantum Dynamical Treatment of Hydrogen Transfer in Soybean Lipoxygenase-1 (SLO-1) and Treatment of the Shared Proton in $[Cl-H-Cl]^-$. To study the scaling of errors with dimensionality and system complexity, we have considered (a) the $[Cl-H-Cl]^-$ system again but with full three-dimensional quantum wave packet treatment of the shared proton and (b) soybean lipoxygenase-1 (SLO-1) which has been thought to exhibit hydrogen tunneling in recent literature.[119−125] SLO-1 is a non-heme metalloenzyme that catalyzes the oxidation of unsaturated fatty acids. The active site for the enzyme is shown in Figure 4(a) and the model for the active site used in our study is shown in Figure 4(b). The rate-determining step in the catalytic cycle for SLO-1 involves the abstraction of a hydrogen atom from the fatty acid chain by the octahedral $Fe^{3+}-OH$ complex present in the active site and shows a remarkably large H/D kinetic isotope effect of 81 near room temperature and a weak



(a) (b)

Figure 3. The error in the interpolated potential energy, for the $[Cl-H-Cl]^-$ system, from using time-dependent deterministic sampling during the wave packet dynamics calculations. The shared hydrogen is treated as a one-dimensional quantum wave packet discretized over $N_Q = 101$ regularly spaced grid points. As can be seen the errors in the potential energy for N_E greater than 11 are in the microhartree range.



(a) (b)

Figure 4. Soybean lipoxygenase-1: The active site is shown in part (a), and the model used for the active site in the current study, after geometry optimization, is shown in part (b). The iron atom is represented by the large sphere close to the center of the two figures. The shared hydrogen atom is marked. Note that in our study the three imidazole rings have been terminated with amine groups.

temperature dependence of the reaction rate constant.[120] These observations are thought to reflect the fact that hydrogen transfer occurs extensively via a tunneling mechanism.[120−123]

In this section we consider SLO-1 to study the scaling of errors in the methodology with respect to dimensionality and system complexity and to demonstrate the power of our methodology in attacking large complex problems. We simplify the active site by substituting amine groups for the imidazole rings. Since the purpose of the current section is only to illustrate the scaling of errors in the deterministic sampling methodology, a detailed study of the SLO-1 problem, using wave packet ab initio molecular dynamics along with a more elaborate description of the active site and substrate, to evaluate the extent of nuclear quantum effects will be the subject of a future publication.

To study the evolution of errors, we considered frozen geometries for both $[Cl-H-Cl]^-$ and SLO-1 where a potential energy scan was performed along a three-dimensional Cartesian grid for the shared hydrogen atom. In addition to the electronic energy (at the level of B3LYP/6-31+G**) we also computed gradients of the electronic energy at each grid point for use with the sampling algorithm. (Note: this does not constitute additional effort since the gradients are required to be computed to propagate the classical nuclear degrees of freedom and the electrons for the case of a full wave packet ab initio dynamics calculation.) Once the potential energy surface was obtained, the eigenstates for the quantum nuclei were obtained using the Arnoldi iterative diagonalization procedure.[101,126,127] The Arnoldi scheme is a variant of the Lanczos procedure[101] and involves the repetitive application of the Hamiltonian matrix to an initial vector to form a Krylov basis set.[101] The representation of the Hamiltonian in this new basis set leads to a Hessenberg form or tridiagonal form[101] for the Hamiltonian, which is relatively easy to diagonalize. The action of the Hamiltonian matrix on any vector is calculated easily by taking advantage of the sparse structure of the DAF kinetic energy operator (second derivative of the zero Δt_{QM} limit of eq 5) and the fact that the full three-dimensional kinetic energy operator can be written in direct product form which obviates the need to store the full (million by million) Hamiltonian matrix. For a grid comprising 101 points in each direction (that is, a million points in three dimensions) we only store three 101 × 101 sparse matrices that are used to construct the action of the full Hamiltonian matrix on a vector as required by the Arnoldi scheme. This is also the case for propagation of a wave packet in three dimensions, where again we only

Table 4: Quantum Wave Packet Propagation Errors for [Cl−H−Cl]$^-$ and Soybean Lipoxygenase-1 (SLO-1)[d]

[Cl−H−Cl]$^-$	σ_{101^3}	σ_{11^3}	σ_{15^3}	σ_{21^3}
20th[a]	3.059×10^{-08}	6.660×10^{-04}	1.989×10^{-04}	5.221×10^{-05}
seventh[a]	2.219×10^{-08}	3.999×10^{-03}	5.275×10^{-05}	1.176×10^{-04}
first[a]	1.079×10^{-08}	2.614×10^{-02}	1.308×10^{-03}	3.429×10^{-04}
thermal sampling[b]	1.066×10^{-08}	3.433×10^{-02}	1.491×10^{-04}	2.785×10^{-04}
Gaussian[c]	7.243×10^{-07}	6.037×10^{-03}	1.182×10^{-04}	8.640×10^{-05}

SLO-1	σ_{101^3}	σ_{11^3}	σ_{15^3}	σ_{21^3}
20th[a]	2.812×10^{-03}	7.324×10^{-01}	1.913×10^{-01}	4.421×10^{-01}
first[a]	3.140×10^{-05}	4.500×10^{-01}	8.927×10^{-01}	2.027×10^{-01}
thermal sampling[b]	4.541×10^{-05}	5.558×10^{-01}	7.492×10^{-01}	1.371×10^{-01}
Gaussian[c]	1.593×10^{-03}	6.998×10^{-03}	4.593×10^{-03}	2.356×10^{-03}

[a] Number represents the eigenstates number that is used as initial wave packet. [b] Initial wave packet is constructed as a linear combination of eigenstate with coefficients for the *i*th eigenstate: $\exp[-E_i/kT]$. Temperature = 300 K. [c] Initial wave packet is a Gaussian centered close to the top of the barrier. [d] Symbols σ_{21^3}, σ_{15^3}, and σ_{11^3} refer to wave packet errors as determined using eq 6 with $N_E = 21^3$, $N_E = 15^3$, and $N_E = 11^3$, respectively. $N_Q = 101^3$ in all cases. Total time of propagation for all cases is 500 fs.

store three one-dimensional free propagators, eq 5, to construct the evolution of a three-dimensional wave packet.

The eigenvectors obtained from the Arnoldi process are then used to construct initial wave packets, as described in Table 4, and propagated with and without deterministic sampling, and the propagation errors are compared in Table 4. Here, the wave packet is propagated at each step, and the potential, gradient, and wave packet at N_E grid points are used to compute a sampling function for the next step. This sampling function is then used to obtain a new grid distribution of N_E points (potentially different from the N_E points used in the previous step) which are then used for to propagate the wave packet. The errors in Table 4 indicate that reasonable accuracy in the propagation is obtained using only 11 grid points per dimension.

C. Comparison of Vibrational Properties for Different Values of N_E. As a further test and illustrative calculation, we present the quantum dynamically averaged vibrational density of states from wave packet ab initio dynamics simulations. It is well-known that the Fourier transform of the velocity−velocity autocorrelation represents the vibrational density of states and has been widely used in classical dynamics[128,129] as well as in ab initio molecular dynamics.[95,130] Quantum corrections to such classical correlation functions[131−133] are an important problem in chemical physics.

In our case, the existence of classical and quantum-dynamical nuclei complicates the direct application of the velocity correlation concept. To provide a cogent treatment we exploit the fact that the quantum correspondence to the classical nuclear velocity is given in terms of the probability flux (or probability current)[80]

$$\begin{aligned}\mathscr{J}(x,t) &= \frac{\hbar}{2m\iota}\left[\psi^*(x,t)\nabla\psi(x,t) - \psi(x,t)\nabla\psi^*(x,t)\right] \\ &= \frac{\hbar}{m}\mathscr{I}\left[\psi^*(x,t)\nabla\psi(x,t)\right] \qquad (34)\end{aligned}$$

where $\mathscr{I}[A]$ represents the imaginary portion of the complex number A. Thus to construct the velocity−velocity autocorrelation function, we consider the average flux (or the expectation value of flux) at any given time

$$\mathbf{J}(t) = \langle \mathscr{J} \rangle = \mathscr{R}\left[\left\langle \psi(t) \left| \frac{-\iota\hbar}{m}\nabla \right| \psi(t) \right\rangle\right] \qquad (35)$$

to be used in conjunction with the classical nuclear velocities. The symbol $\mathscr{R}$ represents the real part of the bracketed quantity.

We thus, simultaneously, construct the correlation functions $\langle \mathbf{J}(t)\mathbf{J}(0)\rangle$ and $\langle v(t)v(0)\rangle$ and the full vibrational density of states as the cumulative Fourier transform

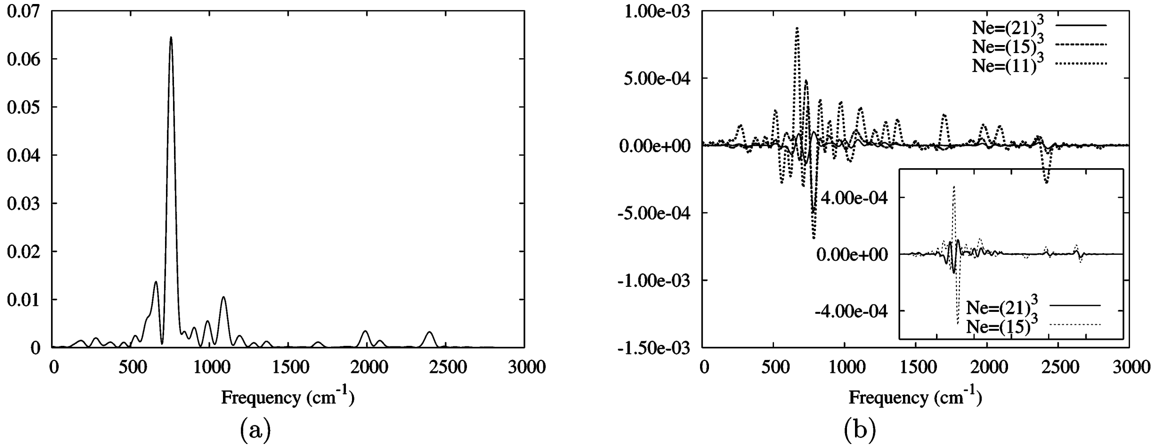
$$C(\omega) = \int_{-\infty}^{+\infty} \exp[-\iota\omega t]\{\langle v(t)v(0)\rangle_{\mathrm{C}} + \langle \mathbf{J}(t)\mathbf{J}(0)\rangle_{\mathrm{Q}}\} \qquad (36)$$

where the symbols $\langle\cdots\rangle_C$ and $\langle\cdots\rangle_Q$ represent the classical and quantum variables ensemble averages.

In Figure 5 we present the vibrational density of states for [Cl−H−Cl]$^-$ from full three-dimensional treatment of the shared hydrogen atom. The results for the uninterpolated study are shown in Figure 5(a), and these are comparable to the experimental results in ref 134. The observed ν_3 band is approximately at 722 cm^{-1}, while the strongest peak in Figure 5(a) is close to 750 cm^{-1}. The discrepancy is explained based on the fact that the Cl−Cl distance used in our simulations are different from those seen in the experiment.[134] In Figure 5(b) we provide the difference spectrum of errors for $N_E = (11)^3$, $(15)^3$, and $(21)^3$ with $N_Q = (101)^3$. It is interesting to note that $N_E = (11)^3$ reproduces the spectrum well in comparison with the full $N_Q = (101)^3$ calculation resulting in a *truly extraordinary compression of the quantum grid and a computational gain of approximately 3 orders of magnitude*. All three-dimensional calculations here used the marginal sampling approach discussed in section IV B.

VI. Concluding Remarks

In this paper, we present a computational scheme to improve the efficiency of our recently developed approach to perform simultaneous dynamics of electrons and nuclei[59] through quantum wave packet ab initio dynamics. A robust and efficient potential adapted time-dependent, deterministic, sampling scheme is derived which improves the efficiency associated with the calculation of the interaction potential between the ab initio system and quantum wave packet. This interaction potential can now be dynamically constructed on a sparse, irregular grid based on deterministic sampling that takes into account the potential energy function at the previous dynamics step, its gradients, and instantaneous density of the wave packet. The sampling criterion has the



(a) (b)

Figure 5. The $[Cl-H-Cl]^-$ vibrational states obtained using eq 36 with the shared hydrogen treated as a three-dimensional wave packet. Part (a) represents the spectrum without time-dependent deterministic sampling for $N_Q = 101$.[3] Part (b) describes the difference spectrum for different values of N_E. All spectra are obtained from 500 fs dynamics data.

attractive feature that it targets regions of the potential that are rapidly varying or lower in energy. In addition, it uses the wave packet density at the previous step to attenuate the distribution of grid points. We have also constructed a cumulative correlation function that includes velocity correlation for the classical degrees of freedom and flux correlation for the quantum dynamical degrees of freedom to obtain the vibrational density of states, inclusive of quantum dynamical effects. The vibrational density of states is used as a tool to check the accuracy of the dynamic sampling algorithm, but the methodology seems general and has promise within the current dynamical framework.

The dynamic sampling function has three adjustable parameters that have been optimized to present an accurate and efficient "on-the-fly" fit of the time-dependent potential. We find that using the sampling function *accurate dynamical and spectral properties can be obtained from only a small fraction of the grid points.* For cases when a single particle is treated quantum mechanically in a bath of surrounding electrons and classical nuclei, the regular grid describing the quantum wave packet can be compressed by several orders of magnitude using the dynamical sampling algorithm and still retain accuracy at a high level. Since the compressed grid is where the electronic energies are determined during the dynamics, this represents an enormous reduction in computational cost. The computational methodology has been demonstrated for both ADMP and Born−Oppenheimer treatment of the classical nuclear and electronic degrees of freedom in conjunction with wave packet propagation.

Acknowledgment. We acknowledge the support of the Camille and Henry Dreyfus New Faculty awards program, the Arnold and Mabel Beckman Young Investigator Award Program, and the Indiana University, Chemistry Department.

Appendix A: A Few Comments on the Physical Interpretation of Eq 7: Connections to WKB Semiclassical Theory and Bohmian Mechanics

Another physical interpretation for the function in eq 7 can be obtained by recognizing its connection to the Wentzel−Kramers−Brillouin (WKB)[80] semiclassical theory. To expound further on this aspect we first introduce the fact that the semiclassical approximation is accurate when

$$\frac{p}{\hbar} \equiv \lambda^{-1} \gg \left(\frac{1}{E - V(x)}\right)\frac{\partial V}{\partial x} \qquad \text{(A1)}$$

i.e. the slowly varying limit of the potential compared to the momentum of the particle and inverse de Broglie wavelength. The sampling function in eq 7 has the property that it is directly proportional to $\partial V/\partial x$ and inversely proportional to the potential. [The modification to the sampling scheme discussed in eq 8 retains this functional dependence.] Hence, $\omega_0(R_{QM})$ is inversely proportional to the WKB length scale λ (or local de Broglie wavelength) in the sense that a greater number of potential evaluations are directed to regions where the potential is rapidly varying and the WKB length scale is smaller. This has the following implications: when the right side of eq A1 is too large, a semiclassical approximation has a greater propensity to failure (unless the momentum of the particle is suitably large). However, for the current methodology, when the right-hand side of eq A1 is large, i.e. when the potential is rapidly varying or when the potential is low, then the sampling function has a larger magnitude which leads to additional potential evaluations in this region. Hence the current approach makes a greater effort to perform accurate quantum dynamics in regions of a potential where a semiclassical approximation may break down. (Also see the discussion at the end of section V A.)

A second important connection may be obtained by understanding the behavior of ω_0 in the vicinity of a wave packet nodal region. This is precisely the region where novel implementations of quantum dynamics based on Bohm's interpretation[80−87,104−108] have trouble on account of the fact that the "quantum potential" $[= (-\hbar^2/2m)\rho^{-1/2}\nabla^2\rho^{1/2}]$ in Bohmian dynamics becomes singular in the vicinity of a node. In our case, the distribution of potential evaluation points is determined not only by the wave packet amplitude but also the potential and its gradients. Hence, if the potential energy is low in the vicinity of a node (as would be the case for nodes enforced by symmetry) or the gradient is high,

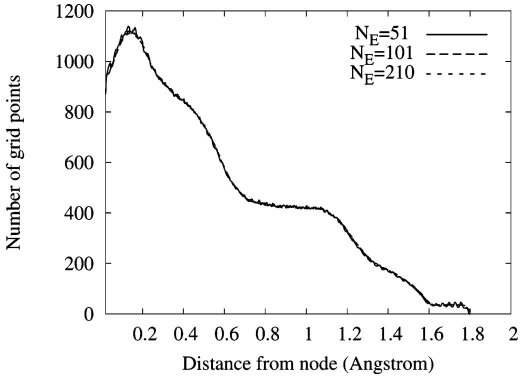


Figure 6. The figure represents the distribution of potential evaluation points obtained from eq 7 in the vicinity of a node. ($N_Q = 901$.) The maximum at 0.15 Å and the substantial probability close to zero indicates that the current formalism maintains an acceptable level of sampling close to nodal points.

then that region is well represented by the current algorithm. Figure 6 shows the number of grid points where potential evaluations are performed as a function of distance from the nearest node. The sampling function in eq 7, along with the algorithm discussed in section IV A 1, is used to determine the location of the N_E grid points. As seen in Figure 6, there exists a substantial probability of finding grid points close to a node when the sampling measure introduced here is used. This was the case over a wide range of N_E. The ADMP potential used for this calculation is a double well potential. The number of nodes is determined dynamically, and the spontaneous appearance or disappearance of nodes is allowed, since the quantum dynamics methodology used here permits wave packet splitting.

Appendix B: Time-Dependent Hermite Curve Interpolation of the Potential Energy Surface

The interpolation scheme used here makes full utilization of the availability of potential and gradients and is described in this section. The scheme discussed here is a special case of cubic spline interpolation[135] and is known as Hermite curve interpolation.[116,117] Consider a function, $f(x)$, defined in an interval $[x_0, x_1]$ with the value of the function and the first derivative at the end points given as $\{f_{x_0}, f_{x_1}, f'_{x_0}, f'_{x_1}\}$. The function can then be approximated using a third-order polynomial in $[x_0, x_1]$ as

$$f(x) \approx f_{app}(x) = f_{x_0}\nu_{x_0}(x) + f_{x_1}\nu_{x_1}(x) + f'_{x_0}\nu'_{x_0}(x) + f'_{x_1}\nu'_{x_1}(x) \tag{B1}$$

where the weights are defined as

$$\nu_{x0}(x) = 2\tilde{x}^3 - 3\tilde{x}^2 + 1 \tag{B2}$$

$$\nu_{x_1}(x) = -2\tilde{x}^3 + 3\tilde{x}^2 \tag{B3}$$

$$\nu'_{x_0}(x) = (\tilde{x}^3 - 2\tilde{x}^2 + \tilde{x})(x_1 - x_0) \tag{B4}$$

$$\nu'_{x_1}(x) = (\tilde{x}^3 - \tilde{x}^2)(x_1 - x_0) \tag{B5}$$

$$\tilde{x} = \frac{x - x_0}{x_1 - x_0} \tag{B6}$$

For our purposes the values f_{x_k} and f'_{x_k} represent the potential energy and the derivative of potential energy with respect to the quantum grid on N_E grid points. The multidimensional generalization is conducted in a similar fashion using the potential and gradient values available at every grid point.

Appendix C: Convergence of the Spreading Transformation

The aim of the spreading transformation is to disperse the excess density $||\Omega_i\tilde{\omega}_i||_1$, as described in section IV A1. This is done through the iterative scheme in eq 27, and it is necessary to evaluate the convergence properties of this scheme. Toward this we define the positive semidefinite convergence measure

$$g_i \equiv ||\Omega_i\tilde{\omega}_i||_1 \geq 0 \tag{C1}$$

When g_i is greater than a numerical threshold, the spreading transformation is not converged at the ith iteration, and the excess density (g_i) needs to be spread onto the region determined by Ω_i^c. As a consequence of eq 14 and the normalization condition in eq 12, it follows that $\Omega_i^c \neq 0$ when $g_i > 0$. Furthermore, since Ω_i^c is a projection operator, $||\{\Omega_i^c\mathbf{1}\}||_1 > 0$. The spreading transformation then leads to new values for ω_{i+1} as per eq 27, Ω_{i+1} and g_{i+1} given by

$$g_{i+1} = ||\Omega_{i+1}\tilde{\omega}_{i+1}||_1 \tag{C2}$$

Using eq 27 we obtain

$$\begin{aligned} g_{i+1} &= \int_a^b \Omega_{i+1}(\omega_{i+1} - \alpha^*)\mathrm{d}x \\ &= \int_a^b \Omega_{i+1}\left\{\Omega_i\alpha^* + \Omega_i^c\left(\frac{||\Omega_i\tilde{\omega}_i||_1}{||\{\Omega_i^c\mathbf{1}\}||_1} + \omega_i\right) - \alpha^*\right\}\mathrm{d}x \\ &= \int_a^b \Omega_{i+1}\Omega_i^c\left\{-\alpha^* + \frac{||\Omega_i\tilde{\omega}_i||_1}{||\{\Omega_i^c\mathbf{1}\}||_1} + \omega_i\right\}\mathrm{d}x \end{aligned} \tag{C3}$$

The operators Ω_{i+1} and Ω_i^c are both projection operators, and Ω_{i+1} is created from spreading g_i onto Ω_i^c. Thus,

$$\Omega_{i+1}\Omega_i^c = (\Omega_{i+1} - \Omega_i) = (\Omega_i^c - \Omega_{i+1}^c) \equiv L_i \tag{C4}$$

Using this in eq C3, we obtain

$$\begin{aligned} g_{i+1} &= \int_a^b (\Omega_{i+1} - \Omega_i)\tilde{\omega}_i dx + g_i\frac{||\{\Omega_i^c\mathbf{1}\}||_1 - ||\{\Omega_{i+1}^c\mathbf{1}\}||_1}{||\{\Omega_i^c\mathbf{1}\}||_1} \\ &= \int_a^b L_i\tilde{\omega}_i dx + g_i\left(1 - \frac{||\{\Omega_{i+1}^c\mathbf{1}\}||_1}{||\{\Omega_i^c\mathbf{1}\}||_1}\right) \end{aligned} \tag{C5}$$

However, since the subspace represented by $(\Omega_{i+1} - \Omega_i) \equiv L_i$ is orthogonal to the subspace represented by Ω_i, this implies $\forall_{x \in L_i}$: $\omega_i(x) < \alpha^*$, and hence $\int_a^b L_i\tilde{\omega}_i dx \leq 0$ where the equality holds only when $\Omega_i = 0$. Hence,

$$g_{i+1} \leq g_i \left(1 - \frac{||\{\Omega_{i+1}^c \mathbf{1}\}||_1}{||\{\Omega_i^c \mathbf{1}\}||_1}\right) \tag{C6}$$

But, by definition $g_{i+1} > 0$, and hence

$$0 \leq g_{i+1} \leq g_i \left(1 - \frac{||\{\Omega_{i+1}^c \mathbf{1}\}||_1}{||\{\Omega_i^c \mathbf{1}\}||_1}\right) \tag{C7}$$

or

$$0 \leq \frac{g_{i+1}}{g_i} \leq \left(1 - \frac{||\{\Omega_{i+1}^c \mathbf{1}\}||_1}{||\{\Omega_i^c \mathbf{1}\}||_1}\right) \tag{C8}$$

Now we consider three different limiting cases for the quantity $||\{\Omega_{i+1}^c \mathbf{1}\}||_1/||\{\Omega_i^c \mathbf{1}\}||_1$ in eq C8 and study the convergence properties of the sequence $\{g_i\}$ for all these cases. It is first important to note that

$$0 \leq \frac{||\{\Omega_{i+1}^c \mathbf{1}\}||_1}{||\{\Omega_i^c \mathbf{1}\}||_1} \leq 1 \tag{C9}$$

Since Ω_{i+1}^c is created after spreading in the *i*th iteration, the range of the subspace Ω_{i+1}^c is smaller than that of Ω_i^c. Hence,

1. $||\{\Omega_{i+1}^c \mathbf{1}\}||_1/||\{\Omega_i^c \mathbf{1}\}||_1 = 1$. It then follows that $0 \leq g_{i+1} \leq 0 g_i$, which means that $g_{i+1} = 0$, that is spreading transformation has converged at the $(i + 1)$th iteration.

2. $||\{\Omega_{i+1}^c \mathbf{1}\}||_1/||\{\Omega_i^c \mathbf{1}\}||_1 = 0$. In this case $||\Omega_{i+1}^c||_1 = 0$. If g_{i+1} is nonzero this implies that there exists an excess density but *no region to "spread" the excess density*. This is in violation of the normalization condition eq 12 and thus $g_{i+1} = 0$ if $||\Omega_{i+1}^c||_1 = 0$ which again implies that the spreading transformation has converged at the $(i + 1)$th iteration.

3. $0 < ||\{\Omega_{i+1}^c \mathbf{1}\}||_1/||\{\Omega_i^c \mathbf{1}\}||_1 < 1$ which leads to $[0 < g_{i+1}/g_i < (1 - ||\{\Omega_{i+1}^c \mathbf{1}\}||_1/||\{\Omega_i^c \mathbf{1}\}||_1)]$, where the upper bound is less that 1. According to de Alembert's ratio test[136] the series comprising $\{g_i\}$ converges when this ratio is less than 1. Hence, by extension the $\{g_i\}$ must converge, is a monotonic decreasing sequence, and is lower bounded by zero.

References

(1) *Dynamics of Molecules and Chemical Reactions;* Wyatt, R. E., Zhang, J. Z. H., Eds.; Marcel Dekker Inc.: New York, New York, 1996.

(2) *Classical and quantum dynamics in condensed phase simulations;* Berne, B. J., Ciccotti, G., Coker, D. F., Eds.; World Scientific: 1997.

(3) Schatz, G. C.; Kupperman, A. *J. Chem. Phys.* **1976**, *65*, 4642.

(4) Delos, J. B. *Rev. Mod. Phys.* **1981**, *53*, 287.

(5) Feit, M. D.; Fleck, J. A. *J. Chem. Phys.* **1982**, *78*, 301.

(6) Kosloff, R. *Annu. Rev. Phys. Chem.* **1994**, *45*, 145.

(7) Leforestier, C.; Bisseling, R. H.; Cerjan, C.; Feit, M. D.; Freisner, R.; Guldberg, A.; Hammerich, A.; Jolicard, D.; Karrlein, W.; Meyer, H. D.; Lipkin, N.; Roncero, O.; Kosloff, R. *J. Comput. Phys.* **1991**, *94*, 59.

(8) DeVries, P. In *Atomic and molecular processes with short intense laser pulses*; Bandrauk, A. D., Ed.; Plenum Press: New York, 1988; Vol. 171 of *NATO ASI Series B.; Physics*, p 481.

(9) Jang, H. W.; Light, J. C. *J. Chem. Phys.* **1995**, *102*, 3262−3268.

(10) Althorpe, S. C.; Clary, D. C. *Annu. Rev. Phys. Chem.* **2003**, *54*, 493−529.

(11) Althorpe, S. C.; Fernandez-Alonso, F.; Bean, B. D.; Ayers, J. D.; Pomerantz, A. E.; Zare, R. N.; Wrede, E. *Nature* **2002**, *416*, 67−70.

(12) Huang, Y.; Iyengar, S. S.; Kouri, D. J.; Hoffman, D. K. *J. Chem. Phys.* **1996**, *105*, 927.

(13) Miller, W. H.; Schwartz, S. D.; Tromp, J. W. *J. Chem. Phys.* **1983**, *79*, 4889.

(14) Makri, N. *Comput. Phys. Comm.* **1991**, *63*, 389−414.

(15) Cao, J.; Voth, G. A. *J. Chem. Phys.* **1994**, *100*, 5106.

(16) Jang, S.; Voth, G. A. *J. Chem. Phys.* **1999**, *111*, 2357.

(17) Feit, M. D.; Fleck, J. A. *J. Chem. Phys.* **1983**, *79*, 301.

(18) Feit, M. D.; Fleck, J. A. *J. Chem. Phys.* **1984**, *80*, 2578.

(19) Kosloff, D.; Kosloff, R. *J. Comput. Phys.* **1983**, *52*, 35.

(20) Kosloff, D.; Kosloff, R. *J. Chem. Phys.* **1983**, *79*, 1823.

(21) Tal-Ezer, H.; Kosloff, R. *J. Chem. Phys.* **1984**, *81*, 3967.

(22) Hartke, B.; Kosloff, R.; Ruhman, S. *Chem. Phys. Lett.* **1986**, *158*, 223.

(23) Iyengar, S. S.; Kouri, D. J.; Hoffman, D. K. *Theor. Chem. Acc.* **2000**, *104*, 471.

(24) Lill, J. V.; Parker, G. A.; Light, J. C. *Chem. Phys. Lett.* **1982**, *89*, 483.

(25) Light, J. C.; Hamilton, I. P.; Lill, J. V. *J. Chem. Phys.* **1985**, *82*, 1400.

(26) Colbert, D. T.; Miller, W. H. *J. Chem. Phys.* **1992**, *96*, 1982−1991.

(27) Huang, Y.; Kouri, D. J.; Arnold, M.; Thomas L. Marchioro, I.; Hoffman, D. K. *Comput. Phys. Comm.* **1994**, *80*, 1.

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

(29) Leforestier, C. *J. Chem. Phys.* **1978**, *68*, 4406.

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

(31) Bolton, K.; Hase, W. L.; Peslherbe, G. H. World Scientific: Singapore, 1998; Chapter: Direct Dynamics of Reactive Systems, p 143.

(32) 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.

(33) Deumens, E.; Diz, A.; Longo, R.; Öhrn, Y. *Rev. Mod. Phys.* **1994**, *66*, 917.

(34) Hack, M. D.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 7917−7926.

(35) Jasper, A. W.; Zhu, C.; Nangia, S.; Truhlar, D. G. *Faraday Discuss.* **2004**, *127*, 1−22.

(36) Miller, W. H. *J. Phys. Chem. A* **2001**, *105*, 2942−2955.

(37) Heller, E. J. *J. Chem. Phys.* **1975**, *62*, 1544−1555.

(38) Fiete, G. A.; Heller, E. J. *Phys. Rev. A* **2003**, *68*, 022112.

(39) Hammes-Schiffer, S.; Tully, J. *J. Chem. Phys.* **1994**, *101*, 4657−4667.

(40) Ben-Nun, M.; Quenneville, J.; Martinez, T. J. *J. Phys. Chem. A* **2000**, *104*, 5161.

(41) Coe, J. D.; Martinez, T. J. *J. Am. Chem. Soc.* **2005**, *127*, 4560.

(42) Martinez, T. J.; Ben-Nun, M.; Ashkenazi, G. *J. Chem. Phys.* **1996**, *104*, 2847.

(43) Martinez, T. J.; Levine, R. D. *J. Chem. Phys.* **1996**, *105*, 6334.

(44) Micha, D. A. *J. Phys. Chem. A* **1999**, *103*, 7562−7574.

(45) Gerber, R. B.; Buch, V.; Ratner, M. A. *J. Chem. Phys.* **1982**, *77*, 3022.

(46) Bisseling, R. H.; Kosloff, R.; Gerber, R. B.; Ratner, M. A.; Gibson, L.; Cerjan, C. *J. Chem. Phys.* **1987**, *87*, 2760−2765.

(47) Makri, N.; Miller, W. H. *J. Chem. Phys.* **1987**, *87*, 5781−5787.

(48) McCoy, A.; Gerber, R. B.; Ratner, M. A. *J. Chem. Phys* **1994**, *101*, 1975.

(49) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. *Rev. Mod. Phys.* **1992**, *64*, 1045.

(50) Marx, D.; Hutter, J. John von Neumann Institute for Computing: Julich, 2000; Vol. 1, Chapter: Ab Initio Molecular Dynamics: Theory and Implementation, pp 301−449.

(51) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514−1527.

(52) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384−2393.

(53) Parrinello, M.; Rahman, A. *Phys. Rev. Lett.* **1980**, *45*, 1196−1199.

(54) Remler, D. K.; Madden, P. A. *Mol. Phys* **1990**, *70*, 921.

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

(56) Pavese, M.; Berard, D. R.; Voth, G. A. *Chem. Phys. Lett.* **1999**, *300*, 93−98.

(57) Tuckerman, M. E.; Marx, D. *Phys. Rev. Lett.* **2001**, *86*, 4946−4949.

(58) Chen, B.; Ivanov, I.; Klein, M. L.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *91*, 215503.

(59) Iyengar, S. S.; Jakowski, J. *J. Chem. Phys.* **2005**, *122*, 114105.

(60) Iyengar, S. S. *Theor. Chem. Acc.* in press. DOI: 10.1007/s00214-005-0010-3. Preprint: http://www.indiana.edu/%7Essiweb/papers/TCA.pdf.

(61) Tully, J. C. *Faraday Discuss.* **1998**, *110*, 407−419.

(62) Kapral, R.; Ciccotti, G. *J. Chem. Phys.* **1999**, *110*, 8919.

(63) Horenko, I.; Salzmann, C.; Schmidt, B.; Schutte, C. *J. Chem. Phys.* **2002**, *117*, 11075−11088.

(64) Donoso, A.; Zheng, Y. J.; Martens, C. C. *J. Chem. Phys.* **2003**, *119*, 5010.

(65) Brooksby, C.; Prezhdo, O. V. *Chem. Phys. Lett.* **2001**, *346*, 463−469.

(66) Prezhdo, O. V.; Brooksby, C. *Phys. Rev. Lett.* **2000**, *86*, 3215−3219.

(67) Gindensperger, E.; Meier, C.; Beswick, J. A. *J. Chem. Phys.* **2000**, *113*, 9369.

(68) Hoffman, D. K.; Nayar, N.; Sharafeddin, O. A.; Kouri, D. J. *J. Phys. Chem.* **1991**, *95*, 8299.

(69) Hoffman, D. K.; Kouri, D. J. *J. Phys. Chem.* **1992**, *96*, 9631.

(70) Kouri, D. J.; Huang, Y.; Hoffman, D. K. *Phys. Rev. Lett.* **1995**, *75*, 49−52.

(71) Marchioro, T. L., II; Arnold, M.; Hoffman, D. K.; Zhu, W.; Huang, Y. H.; Kouri, D. J. *Phys. Rev. E* **1994**, *50*, 2320−2330.

(72) Hoffman, D. K.; Arnold, M.; Kouri, D. J. *J. Phys. Chem.* **1992**, *96*, 6539−6545.

(73) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

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

(75) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.

(76) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.

(77) Scuseria, G. E. *J. Phys. Chem. A* **1999**, *103*, 4782.

(78) White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1994**, *101*, 6593−6605.

(79) Dirac, P. A. M. *The Principles of Quantum Mechanics*, 4th ed.; Oxford University Press: New York, 1958; Vol. 27 of *The International series of monographs on Physics*.

(80) Sakurai, J. J. *Modern Quantum Mechanics;* Addison-Wesley Publishing Company: 1994.

(81) Madelung, E. *Z. Physik* **1926**, *40*, 322−326.

(82) de Broglie, L. *An introduction to the study of wave mechanics;* E. P. Dutton and Company, Inc.: New York, 1930.

(83) Bohm, D. *Quantum Theory;* Prentice Hall Inc.: New York, 1951.

(84) Bohm, D. *Phys. Rev.* **1952**, *85*, 166.

(85) *Bohmian Mechanics: An appraisal;* Cushing, J. T., Fine, A., Goldstein, S., Eds.; Kluwer: Boston, 1996.

(86) Holland, P. R. *The Quantum Theory of Motion;* Cambridge, New York, 1993.

(87) Iyengar, S. S.; Schlegel, H. B.; Voth, G. A. *J. Phys. Chem. A* **2003**, *107*, 7269−7277.

(88) Ehrenfest, P. *Z. Phys.* **1927**, *45*, 455.

(89) 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.

(90) Iyengar, S. S.; Schlegel, H. B.; Voth, G. A.; Millam, J. M.; Scuseria, G. E.; Frisch, M. J. *Isr. J. Chem.* **2002**, *42*, 191−202.

(91) Rega, N.; Iyengar, S. S.; Voth, G. A.; Schlegel, H. B.; Vreven, T.; Frisch, M. J. *J. Phys. Chem. B* **2004**, *108*, 4210−4220.

(92) Iyengar, S. S.; Frisch, M. J. *J. Chem. Phys.* **2004**, *121*, 5061.

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

(94) Tangney, P.; Scandolo, S. *J. Chem. Phys.* **2002**, *116*, 14.

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

(96) Trotter, M. F. *Proc. Am. Math. Soc.* **1959**, *10*, 545.

(97) Nelson, E. *J. Math. Phys.* **1964**, *5*, 332.

(98) Strang, G. *SIAM J. Numer. Anal.* **1968**, *5*, 506−516.

(99) *Handbook of Mathematical Functions;* Abramowitz, M., Stegan, I. A., Eds.; U.S. GPO: Washington, DC, 1964.

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

(101) Golub, G. H.; van Loan, C. F. *Matrix Computations;* The Johns Hopkins University Press: Baltimore, 1996.

(102) Iyengar, S. S.; Kouri, D. J.; Parker, G. A.; Hoffman, D. K. *Theor. Chem. Acc.* **2000**, *103*, 507−517.

(103) Iyengar, S. S. New Computational Approaches to Quantum Dynamics using “Distributed Approximating Functionals, Ph.D. Thesis, University of Houston, 1998.

(104) Lopreore, C. L.; Wyatt, R. E. *Phys. Rev. Lett.* **1999**, *82*, 5190.

(105) Day, B. K.; Askar, A.; Rabitz, H. A. *J. Chem. Phys.* **1998**, *109*, 8770.

(106) Wyatt, R. E.; Kouri, D. J.; Hoffman, D. K. *J. Chem. Phys.* **2000**, *112*, 10730.

(107) Bittner, E. R.; Wyatt, R. E. *J. Chem. Phys.* **2000**, *113*, 8888.

(108) Wyatt, R. E.; Bittner, E. R. *Comput. Sci. Eng.* **2003**, *5*, 22−30.

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

(110) Strang, G. *SIAM Rev.* **1989**, *31*, 613−627.

(111) Daubechies, I. *Ten Lectures in Wavelets;* SIAM: Philadelphia, PA, 1992.

(112) Strang, G.; Nguyen, T. *Wavelets and Filter Banks;* Wellesley-Cambridge Press: 1996.

(113) Strang, G.; Strela, V. *J. Opt. Eng.* **1994**, *33*, 2104−2107.

(114) Arias, T. *Rev. Mod. Phys.* **1999**, *71*, 267−311.

(115) Johnson, B. R.; Modisette, J. P.; Nordlander, P. J.; Kinsey, J. L. *J. Chem. Phys.* **1999**, *110*, 8309−8317.

(116) Bartels, R. H.; Beatty, J. C.; Barsky, B. A. *An Introduction to Splines for use in computer graphics and geometric modeling;* Morgan Kaufman Publishers: Los Altos, CA, 1987.

(117) Carter, S.; Bowman, J. M.; Braams, B. J. *Chem. Phys. Lett.* **2001**, *342*, 636−642.

(118) Iordanov, T.; Billeter, S. R.; Webb, S. P.; Hammes-Schiffer, S. *Chem. Phys. Lett.* **2001**, *27*, 389−397.

(119) Jonsson, T.; Glickman, M. H.; Sun, S. J.; Klinman, J. P. *J. Am. Chem. Soc.* **1996**, *118*, 10319−10320.

(120) Glickman, M. H.; Wiseman, J. S.; Klinman, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 793−794.

(121) Liang, Z.-X.; Klinman, J. *Curr. Opin. Struct. Biol.* **2004**, *14*, 1−8.

(122) Lehnert, N.; Solomon, E. I. *J. Biol. Inorg. Chem.* **2003**, *8*, 294.

(123) Hatcher, E.; Soudackov, A. V.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2004**, *126*, 5763−5775.

(124) Olsson, M. H. M.; Siegbahn, P. E. M.; Warshel, A. *J. Am. Chem. Soc.* **2004**, *126*, 2820−2828.

(125) Olsson, M. H. M.; Siegbahn, P. E. M.; Warshel, A. *J. Biol. Inorg. Chem.* **2004**, *9*, 96−99.

(126) Sorensen, D. C. *SIAM J. Matr. Anal. Apps.* **1992**, *13*, 357−385.

(127) Parlett, B. N.; Saad, Y. *Linear Algebra Applications* **1987**, *88/89*, 575−595.

(128) Rahman, A. *Phys. Rev.* **1964**, *136*, A405.

(129) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9080−9089.

(130) Iyengar, S. S. *J. Chem. Phys.* **2005**, *123*, 084310.

(131) Berens, P. H.; White, S. R.; Wilson, K. R. *J. Chem. Phys.* **1981**, *75*, 515.

(132) Bader, J. S.; Berne, B. J. *J. Chem. Phys.* **1994**, *100*, 8359.

(133) Lawrence, C. P.; Nakayama, A.; Makri, N.; Skinner, J. L. *J. Chem. Phys.* **2004**, *120*, 6621.

(134) Kawaguchi, K. *J. Chem. Phys.* **1988**, *88*, 4186−4189.

(135) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C;* Cambridge University Press: New York, 1992.

(136) Arfken, G. *Mathematical Methods for Physicists;* Academic Press: New York, 1985.

(137) Li, X.; Iyengar, S. S. Manuscript in preparation.

CT600131G

JCTC Journal of Chemical Theory and Computation

Comparative Static and Dynamic Study of a Prototype S_N2 Reaction

Laurent Joubert,† Michele Pavone,‡ Vincenzo Barone,‡ and Carlo Adamo*,†

Laboratoire d'Electrochimie et de Chimie Analytique, UMR CNRS 7575,
Ecole Nationale Supérieure de Chimie de Paris, 11, rue Pierre et Marie Curie,
75231 Paris Cedex 05, France, and Laboratorio di Struttura e Dinamica Molecolare,
Dipartimento di Chimica, Complesso Universitario Monte Sant'Angelo, Via Cintia,
I-80126 Napoli, Italy

Received January 10, 2006

Abstract: Ab initio molecular-dynamic simulations, using density functional theory (DFT) and the recent atom-centered density-matrix propagation method (ADMP), were used to study the bond formation process in a prototypical S_N2 reaction, namely the Walden inversion. Using the real space partition schemes of both electronic density and electron localization function gradient fields, we analyzed different quantum chemical topology (QCT) properties along the ADMP trajectory. In particular, atomic charges derived from the Bader's atoms-in-molecules (AIM) theory were used to analyze intra- and intermolecular charge transfers between atoms, while the electronic population of the forming bonding basin obtained from the electron localization function (ELF) gradient field was employed to describe the bond formation process. These results were compared to the corresponding QCT properties issuing from a static approach based on the intrinsic reaction path (IRP). Although similar features are found for both static and dynamic approaches, the dynamic QCT analysis provides some explanation of the differences observed during the formation of the ion−molecule complex. In particular, it suggests a stronger electron exchange leading to an effective maximization of both covalent and noncovalent interactions.

1. Introduction

The interpretation of the numerical results issuing from quantum chemical calculations in terms of classical chemical concepts is not a straightforward task. For instance, electron transfer is strongly related to the definition of atoms and bonds in a molecular system. Beyond the basic picture of Lewis[1] based on intuitive concepts, at least two different models have been proposed with the aim of obtaining quantitative information on the nature of chemical bond. On one hand, classical localization procedures of the one-electron density matrix and/or overlap matrix lead to different definitions of atomic charges (e.g. Mulliken or Löwdin ones[2,3]) or to hybrid orbitals (see for instance refs 4−6). On the other hand, quantum chemical topological (QCT) analyses based on the electronic density, i.e., the Bader's atoms in molecules (AIM) partition scheme[7,8] or on derived functions, such as the electron localization function (ELF),[9−11] have convincingly shown their reliability and robustness. In particular, these methods are well suited to characterize chemical bonds[12−15] or to get insights into the reactivity of molecular systems.[7,8,16−20]

Nowadays, a wide choice of tools is available to chemists for the theoretical analysis of different classes of chemical reactions. These tools are even more powerful when coupled with efficient algorithms for the exploration of potential energy surfaces (PES), such as those based on the intrinsic reaction coordinate (IRC) concept.[21] Calculation of topological or electronic variables along the reaction path provides a clear-cut picture of the reaction mechanism and for rationalizing the structural/energetic properties of the stationary points on the PES (reactants, transition states, and products).[22,23] It must be pointed out, however, that all these approaches have been developed in the framework of a time-

---

* Corresponding author e-mail: carlo-adamo@enscp.fr.

† Ecole Nationale Supérieure de Chimie de Paris.

‡ Complesso Universitario Monte Sant'Angelo.

10.1021/ct0600159 CCC: $33.50

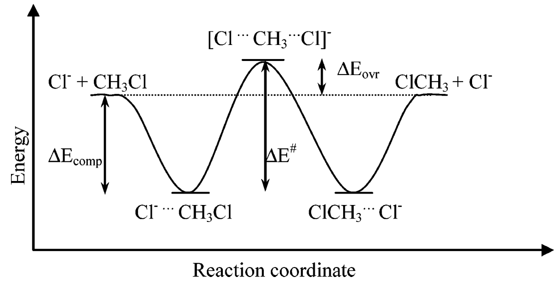


Figure 1. Sketch of energy profile for the S_N2 reaction under study.

independent (i.e. static) view. Only recently Gross and coworkers have formulated a suitable but rather complex approach introducing a time dependent ELF.[24]

At the same time, methods rooted on the so-called ab initio molecular dynamics, both in Born−Oppenheimer and extended Lagrangian versions, are today well assessed and have been widely used for reactivity studies.[25−27] In particular, the so-called Car−Parrinello approach,[28] using an extended Lagrangian scheme, is based on the evaluation of energies and gradients at the density functional theory level (DFT) and on the contemporary propagation of electronic and nuclear degrees of freedom.[29] The time evolution of any property is then easily computed a posteriori from suitable statistical averages of single point evaluations on snapshots extracted from the trajectory.[29]

An attempt to explore the difference between static and dynamic ab initio approaches has been performed,[30] but it has been focused mainly on the thermochemistry, since the choice of unique and meaningful parameters for comparison is not trivial.

In the present article, we report a study concerning the complex charge-transfer processes driving the bond breaking and formation processes in a prototype reaction. The static approach is based on the standard IRC model, while the dynamic one rests on an extended Lagrangian approach employing atom-centered (Gaussian) functions, the so-called atom-centered density matrix propagation (ADMP) model.[31] Our aim is to elucidate the differences, and the common aspects, between both approaches in the analysis of bond breaking/formation. To this end, we have chosen topological quantities as probe molecular descriptors. The so-called Walden inversion, the S_N2 reaction sketched in Figure 1, has been chosen as a model, since it is a simple and well-studied reaction, from both a static and dynamical point of view.[32−52] In the present study we limit our analysis to the gas-phase reaction, and we do not consider solvent effects which are known to strongly modify the potential energy surface (PES).[35]

2. Computational Details

All the calculations were carried out with the Gaussian 03 program,[53] using the hybrid PBE0 functional[36,54] and the 6-31+G(d,p) basis set.[55] Starting from the transition state, we calculated reaction pathways by both static and dynamic approaches. On one hand, we followed the intrinsic reaction path (IRP) between the transition state and the stable ion−molecule complex.[21] On the other hand, the dynamic simulation was performed at 298 K in the canonical ensemble, for a total simulation time of 1.5 ps. The starting point (transition state) was previously optimized with the same basis set and functional used in the simulation. The velocity-Verlet algorithm[56] was employed for the integration of the equations of motions using a time step of 0.25 fs. The fictitious mass of the electron was set to 0.20 amu with a scaling both for core and valence electrons as described in ref 31. The velocities of the nuclei were scaled each five time steps to ensure a constant temperature within a $\Delta T = 5$ K tolerance. The stability of the simulations was monitored by checking at each step the idempotency of the density matrix (within a 10^{-12} threshold) and the so-called adiabaticity index (within a 10^{-4} threshold, see ref 31 for more details). In practice, we restricted our trajectory investigation to the time frame of interest, analyzing the path only during the formation of the stable ion−molecule complex. Along this truncated trajectory (shorter than 100 fs in time scale), we extracted a number of snapshots, spaced by 2 fs, for the subsequent structural and electronic analyses of the reaction.

The charge-transfer processes were investigated using a QCT approach, namely the AIM partition scheme of the electronic density.[7,8] According to this model, a topological atom is defined as a region of the real space consisting of a bundle of electron density gradient paths attracted by a nucleus. This partition allows evaluating atomic properties, defined as volume integrals over nonoverlapping atomic basins. In particular, the electron population associated with an atom is simply the volume integral of the electronic charge density over the basin. Other useful properties can be derived such as the delocalization index, providing a quantitative measure of the sharing of electrons between two atomic basins.[13]

Another partition of the electronic density can be achieved by considering the electron localization function ($\eta(\mathbf{r})$, ELF)[9] which can be viewed, to a good approximation, as a compact form of the spin-pair composition function, relying on a chemical approach to describe the nature of the electron cloud in terms of parallel and antiparallel spin-pair concentrations.[57] In practice, portions of the real space characterized by a high probability to find antiparallel spin-pair concentrations correspond to values of $\eta(\mathbf{r})$ close to 1. In contrast, other portions where the probability to find parallel spin-pair concentrations is low are characterized by $\eta(\mathbf{r})$ values close to 0. During the 1990s, Silvi and co-workers proposed a gradient field topological analysis of this bounded function,[10,11] providing a quantum chemical reformulation of the empirical Lewis chemical bond model. The topology of this gradient field is more complex than the one of the electronic density, presenting nonatomic basins such as bonding basins, shared between two or more neighboring atoms (see ref 11 for details on the nature of these basins). The electronic population of these basins can be viewed as a quantitative measure of bond strengths in terms of covalence, i.e., the sharing of strongly localized, antiparallel spin pairs. As already mentioned in the Introduction, this “bonding” electronic population will be used in the following to describe the formation of the carbon−chlorine covalent bond, as a complement to AIM calculations. All QCT calculations (AIM

Table 1. Computed Main Geometry Parameters (Å) for the Equilibrium Ion−Dipole Complex and the Transition State

	complex		transition state
	d(C−Cl$_1$)	d(C−Cl$_2$)	d(C−Cl)
MP2/TZ3P+R+(2f,d)[a]	3.27	1.81	2.32
CCSD(T)/TZ+2Pf+dif[b]	3.15	1.82	2.30
BP/TZ+2P[b]	3.10	1.84	2.34
OLYP/TZ+2P[c]	3.37	1.83	2.36
mPW1PW/6-31+G(d,p)[d]	3.01	1.83	2.33
PBE0/6-31+G(d,p)	3.16	1.82	2.33
PBE0/6-311++G(3df,2dp)	3.14	1.81	2.31

[a] Reference 37. [b] Reference 45. [c] Reference 46. [d] Reference 35.

Table 2. Complexation Energies, Overall Energies, and Barrier Heights (in kcal/mol) for the Studied S$_N$2 Reaction

	ΔE_{comp}	ΔE_{ovr}	$\Delta E^{\#}$
MP2[a]/TZ3P+R+(2f,d); ZPE corrected	−10.5	3.5	14.0
MP4[b]TZ+2P; ZPE corrected	−10.6	1.8	12.4
G3;[c] ZPE corrected	−11.2	1.8	13.0
CBS-QB3(+);[c] ZPE corrected	−10.7	2.4	13.1
best ab initio results[d]	−11.3	1.6	12.9
BP/TZ+2P;[b] ZPE corrected	−10.3	−5.7	4.6
OLYP/TZ+2P;[e] ZPE corrected	−9.0	−0.1	8.9
B3LYP/6-31G(d);[f] ZPE corrected	−9.5	−0.9	8.7
mPW1PW/6-31+G(d,p);[g] ZPE corrected	−9.8	0.7	10.5
PBE0/6-31+G(d,p); ZPE corrected	−10.1	0.5	10.6
experiment[h]	−12 ± 2	3/1 ± 1	13 ± 2

[a] Reference 37. [b] Reference 38. [c] Reference 39. [d] Reference 45. [e] Reference 46. [g] Reference 35. [h] Reference 51.

and ELF) have been performed with the Topmod package[58] and local scripts.

3. Results

Before discussing in some detail the results of our analysis, we want to add some general remarks on the topological approaches chosen. Recently, Chevreau and co-workers[59] have combined AIM calculations with dynamic simulations to explore charge-spin transfers in model molecules introducing what they refer to as time-dependent topological analyses. In contrast, we prefer to follow the proposition of Popelier[8] and use the more specific definition of quantum chemical topology (QCT). The main point is that QCT embraces both AIM and ELF topological analyses and any approaches that adopt the topological idea.[60] Therefore, we will refer in the following to dynamic quantum chemical topology (DQCT) for QCT analyses performed in the framework of a dynamic treatment of the reactivity, since time is not an explicit variable neither for the property nor for the electron density. Similarly, a QCT study realized along an intrinsic reaction path, i.e., a static approach of the reactivity, will be referred to as a static quantum chemical topology (SQCT) analysis.

3.1. Tuning the DFT Approach. The prototypical S$_N$2 reaction studied in the present work (see Figure 1) is characterized in the gas phase by a double-well potential energy surface with two equivalent local minima corresponding to the formation of a pre- and a postreaction ion−molecule complex (Cl$^-$•••CH$_3$Cl) and a transition state (TS) of D_{3h} symmetry ([Cl•••CH$_3$•••Cl]$^-$). Due to the overall symmetry of the reaction, we will limit our study to a reaction path from the transition state to one of the equivalent minima.

We have chosen the PBE0 functional, which provides reliable thermodynamic data for the S$_N$2 reaction under study, when coupled to extended basis sets (e.g. 6-311+(3df,3pd)). Since a large number of calculations are needed in order to evaluate the variations of all the topological quantities along the trajectories, use of extended basis sets becomes prohibitive. For this reason, we performed some preliminary tests to assess the validity of the chosen functional with the medium size 6-31+G(d,p) basis set. We first examined the geometrical features of the key structures, i.e., the geometries of the ion−molecule complex and of the transition state. Structural and energetic results are summarized in Tables 1 and 2, respectively. From a structural point of view, our data are very close to MP2 results[38] for the transition state structure, with a deviation of about 0.01 Å for the carbon−chlorine distance. The same trends are observed for the C−Cl$_2$ bonding distance in the ion−molecule complex. In contrast, a larger error (0.20 Å) is observed for the C•••Cl$_1$ distance, this effect being related to basis set incompleteness. In fact, when the larger 6-311+G(2d,p) basis set is considered this distance increases significantly (3.23 Å), toward the MP2 value.

The most significant thermodynamic quantities are the complexation energy of the ion−molecule species (ΔE_{comp}), the activation energy, i.e. the relative energy of the D_{3h} saddle point with respect to the ion−molecule complex ($\Delta E^{\#}$), and the overall barrier (ΔE_{ovr}), defined as the difference between ΔE_{comp} and $\Delta E^{\#}$. These data are reported in Table 2 and compared with selected previous theoretical and experimental results. While the computation of initial closed-shell reagents does not involve particular difficulties, the determination of the charged transition state energy [Cl•••CH$_3$•••Cl]$^-$ by DFT approaches is more involved. As a matter of fact, most of the standard functionals, and in particular those resting on the GGA, fail in determining the energy barriers. This effect has been related to the self-interaction error both in the exchange and in the correlation parts, which implies a too large delocalization of the electron density with the consequent overstabilization of the transition state (for a recent discussion about this point see ref 61). In contrast, PBE0 calculations provide accurate results, close to the best post-HF methods and well in the range of the experimental estimates. Of course, we cannot exclude that this result is due to a fortuitous error compensation between the GGA and HF contributions.

3.2. Static and Dynamic Topological Analyses. The energy variations along the IRP and the ADMP trajectory are reported in Figures 2 and 3, respectively. We recall that the IRP corresponds to a unique and constrained minimum energy path and does not contain any dynamic information, whereas the ADMP profile is computed along a trajectory issuing from a dynamic simulation. Therefore any straight (i.e. point-to-point) comparison can be misleading, but we think that global differences remain meaningful.

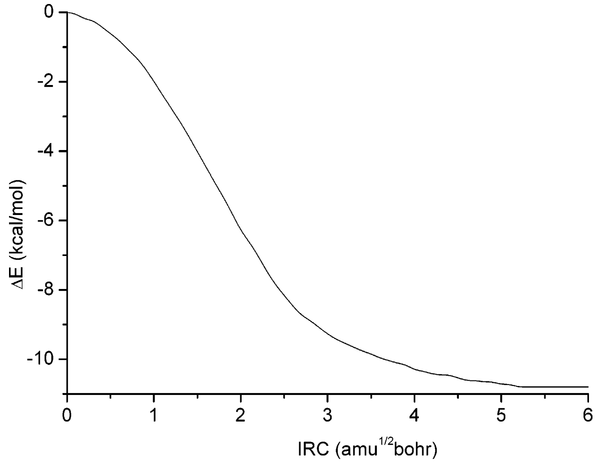


Figure 2. Plot of the potential energy profile along the IRP from the transition state to the ion−molecule complex.

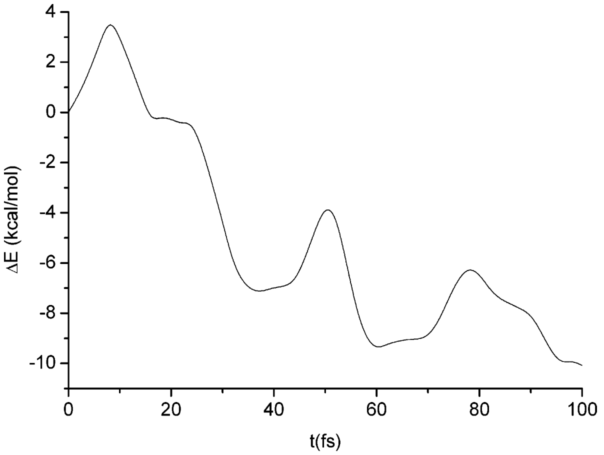


Figure 3. Plot of the potential energy profile along the ADMP trajectory from the transition state to the ion−molecule complex.

Table 3. Computed AIM Basin Populations (e^-) for the Transition State and the Ion−Molecule Complex

atom	transition state	ion−molecule complex	atom	transition state	ion−molecule complex
Cl1	17.70	17.95	H3	0.83	0.86
H1	0.83	0.84	C	6.11	6.10
H2	0.83	0.88	Cl2	17.70	17.38

The minimum energy constraint determines a unique IRP, on which the energy continuously decreases from the transition state to the minimum. In contrast, the energy globally decreases along the ADMP trajectory, but it appears clearly that the reaction does not follow this minimum energy path. Actually, the minimum is rapidly reached after 96 fs, but the curve exhibits some energy risings corresponding to local destabilizations of the whole system.

To rationalize the differences observed between static and dynamic approaches, we investigated the intra- and intermolecular charge transfers along the different reaction paths. As a first step, we analyzed the AIM atomic populations for the two states of interest, i.e., the transition state and ion−molecule complex. The results are collected in Table 3. At the transition state, each chlorine atom bears a negative charge of −0.70 $|e^-|$. This electron excess corresponds to the sharing of the unit charge of the complex increased by a supplementary electron transfer of approximately 0.20 e^- from hydrogen atoms to each of the chlorine atom. Along the IRP, the formation of the ion−dipole molecule corresponds to a global charge transfer of 0.25 e^- from the incoming chloromethane moiety to the leaving chlorine atom. This charge transfer is slightly increased (by 0.01 e^-) when considering the dynamic process. This global molecule−ion interaction is accompanied by an intramolecular charge transfer in the forming chloromethane molecule. This phenomenon is evidenced by a slight increase of the electron population on the methyl moiety (0.05 and 0.06 e^- at the end of the static and dynamic process, respectively).

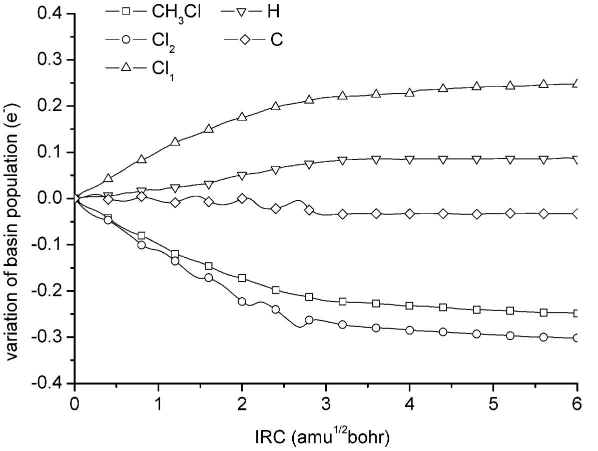


Figure 4. Variations of AIM populations along the IRP. The "superbasin" corresponding to the forming chloromethane molecule (5 basins) is represented by the solid curve.

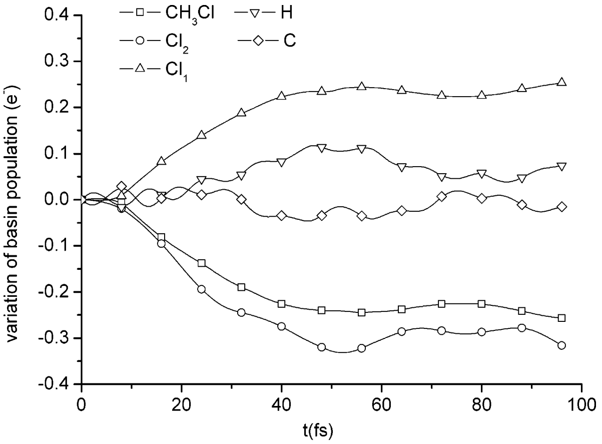


Figure 5. Variations of AIM populations along the ADMP trajectory. The "superbasin" corresponding to the forming chloromethane molecule (5 basins) is represented by the solid curve.

A deeper insight into the charge-transfers processes can be gained by examining the variations of the atomic electronic populations along the IRP and ADMP trajectory. Figures 4 and 5 show the relative variations of selected AIM basin populations with respect to the initial populations in the transition state configuration for the static and dynamic approach, respectively. We first examine the global mechanism, i.e., the molecule-to-ion charge transfer, evidenced by two essential curves. The first one corresponds to the monotonically increasing atomic population of the leaving chlorine atom Cl1, tending to the population of an isolated chloride anion. In parallel, the second curve, grouping the atomic populations of the whole CH_3Cl molecule, shows the corresponding and decreasing electronic population of the forming molecule. When the ion−dipole complex is formed, the global electronic transfer between the two moieties corresponds to the aforementioned net charge of 0.25 $|e^-|$. Three additional curves are presented on the same graphs, corresponding to the variations of the atomic populations of (i) the chlorine atom Cl1 of the forming chloromethane molecule, (ii) the three hydrogen atoms taken together, and (iii) the carbon atom. If we examine these variations, substantial differences are observed between static and dynamic approaches.

A careful examination of the atomic population variations along the static IRP (Figure 4) allows us to propose a three-step charge-transfer mechanism. In the first part of the path, with a length of about 1.0 $amu^{1/2}$ Bohr, the electron excess on the Cl2 atom is transferred almost totally to the other chlorine through the atomic basins of the hydrogen atoms. This particular point is evidenced by examining the variations of the atomic populations of the two chlorine atoms that almost compensate each other. In other words, the electronic flux coming from the Cl2 basin and entering the three hydrogen basins corresponds to the exiting flux that penetrates the Cl1 basin. Besides this main electron transfer, a second one is observed between the carbon and the three hydrogen atoms. In fact, along the path, the carbon−hydrogen bond lengths systematically increase when the methyl group adopts a C_{3v} geometry in the forming molecule and points toward the leaving chlorine atom. This phenomenon is enhanced by an increasing favorable electrostatic interaction with this chlorine atom. Therefore, the lengthening of these bonds explains the local charge transfer observed from carbon to hydrogen atoms. Finally, we note substantial fluctuations in the population variations of the linked carbon and chlorine atoms, which, once again, compensate each other. In more detail, a small fraction of the electron flux between chlorine atoms penetrates the carbon basin, resulting in a small increase of the corresponding population that tends to be slightly in excess, whereas, at the same time, the formation of the carbon−chlorine bond counteracts this flux by inducing a small charge transfer from the carbon to the chlorine atom. In the second step of the charge-transfer mechanism, i.e., up to 3.0 $amu^{1/2}$ Bohr, the electron loss of the Cl2 atom becomes stronger than the electron gain of the leaving Cl1 atom. Therefore, the electron flux that enters the hydrogen basins becomes larger than the exiting flux with a consequent increase of the electron population in the hydrogen basins. The electron charge transfer from carbon to hydrogen atoms increases as well, though to a lower extent, thus reinforcing the electronic population of these basins. Furthermore, we note that the fluctuations observed along the carbon and chlorine (Cl2) curves are amplified when approaching the equilibrium structure of the chloromethane. The last step of the charge-transfer mechanism corresponds approximately to the second half of the IRP. Negligible structural changes are observed in the chloromethane molecule, resulting in equilibrium for carbon and hydrogen basin populations. The Cl1 atom continues to move away from the molecule leading to a strong decrease of the chlorine-to-chlorine charge transfer, the population of the Cl1 atom approaching that of a chloride anion.

The DQCT analysis of atomic population variations along the ADMP trajectory (see Figure 5) suggests a charge-transfer mechanism quite different from that issuing from the SQCT calculations. We can decompose it in five steps. We first note that the system is destabilized during the first 8 fs of the simulation (see Figure 2), certainly due to an excess of initial kinetic energy. During that period, population variations are negligible, except for the periodic fluctuations of the carbon and hydrogen populations corresponding to the C−H stretching mode. These fluctuations are present along the whole trajectory. The second dynamic step, occurring between 8 and 16 fs, is very similar to the first part of the IRC path, where an electron excess on the Cl2 atom is transferred almost totally to the other chlorine through the atomic basins of the hydrogen atoms. The third step corresponds roughly to time steps between 16 and 30 fs and can be compared to the second step of the IRC charge-transfer mechanism. In fact, the curve corresponding to the population variations of the Cl2 chlorine atom strongly deviates from the one corresponding to the population variations of the whole chloromethane molecule. Meanwhile, both electronic populations of carbon and hydrogen atoms substantially increase, indicating an intramolecular charge transfer from the Cl2 basin to the carbon and hydrogen basins. In contrast with the static approach, no fluctuations are observed for the carbon and Cl2 chlorine curves but for the periodic variations on the carbon curve corresponding to the C−H stretching vibration mode.

At this point, we want to remark that all the results discussed above are preserved when different initial conditions (kinetic energy and temperature) are considered, since the excess energy mainly affects the population of the CH vibrational states. This is not surprising since the $Cl-CH_3$-Cl starting ion-complex exhibits a rather poor energy transfer between the inter- and intramolecular vibrational modes.[42,62]

The third step corresponds to time frames between 30 and 50 fs. We observe that the electronic population of the hydrogen basins strongly increases due to a substantial electron transfer from the carbon basin. Moreover, the corresponding population variations almost compensate to each other. All these subtle variations are related to structural changes during the dynamics. On Figure 6, we isolated the population variations of the carbon and hydrogen basins together with the variations of C−H distances. During the time frame of interest, we note a substantial lengthening of these distances. This phenomenon was also observed along the static IRP but with smaller amplitude. Here, the C−H stretching mode increases the corresponding bond lengths, strengthening the electrostatic attraction between the leaving chlorine atom and hydrogen atoms and favoring the electron flux between them. This noncovalent effect should favor the stabilization of the system, but we know that the complex is strongly destabilized within this time frame, as evidenced

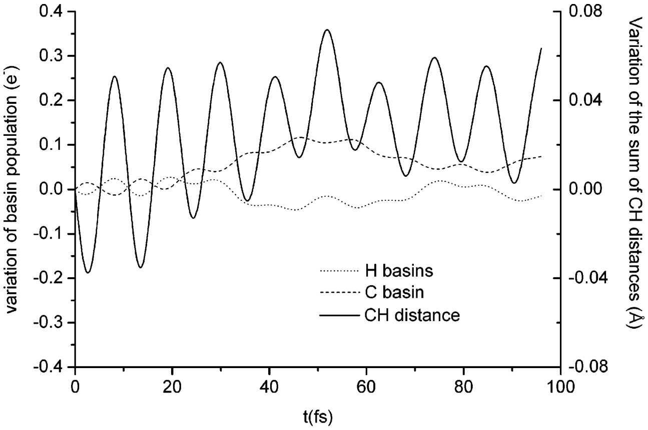


Figure 6. Comparison between the variations of selected basin populations (C and H) and the sum of C−H distances (solid line) along a dynamic pathway from the transition state to the ion−molecule complex.

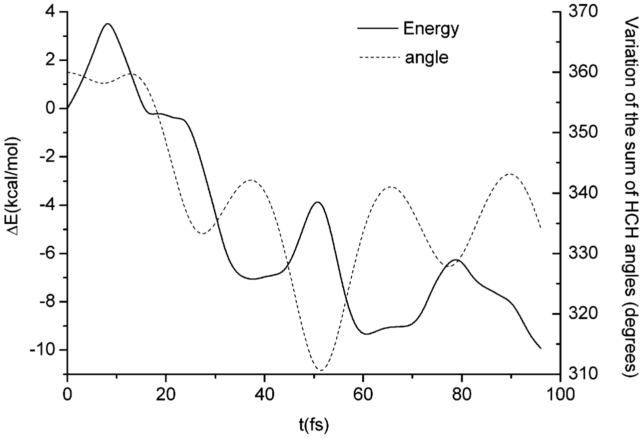


Figure 7. Comparison between the potential energy variations (in bold) and H−C−H valence angles (dash) along a dynamic pathway from the transition state to the ion−molecule complex.

by a sudden potential energy rising in Figure 3. This surprising behavior can be better rationalized by examining the structural evolution of the system. Figure 7 shows the variations of the potential energy together with the evolution of the H−C−H valence angles, i.e., we plotted the variations of the sum of these three angles along the trajectory. In the time interval of interest, we can see that the sum of angles reaches a minimum value (about 310°). In other words, the strong electrostatic attraction between the Cl1 chlorine atom and hydrogen atoms induces not only an increase of the C−H bond lengths but also a substantial narrowing of the H−C−H valence angles. The latter effect may explain at least partly the destabilization of the system by increasing electrostatic and Pauli repulsions between the hydrogen atoms. A deeper insight on this unexpected and sudden potential energy rising can be obtained from a topological description of the forming C−Cl2 bond. Even if AIM is, in our opinion, adequate for such a description,[14] Cioslowski and co-workers[63−65] and Feinberg and Ruedenberg[66] pointed out some problems inherent to the definition of a bond path in AIM. In this paper, we follow an alternative route offered by the DQCT analysis of the ELF gradient field. We choose here to focus on the forming bonding basin between the carbon and the incoming chlorine atom (Cl2). The variations of both the C−Cl2 bond distance and the electronic population of the corresponding bonding basin are represented in Figure 8. In the time interval corresponding to the second peak of energy increase, a continuous decrease of the C−Cl2 distance is observed, down to a global minimum value of 1.682 Å at t = 50 fs. This bond distance is about 0.15 Å shorter than the equilibrium value. Meanwhile, the electron population of the bonding basin dramatically increases up to reach a final value of 1.45 e$^-$, emphasizing a strong covalent character which is quite unexpected for such highly polarized interaction. In short, the electron sharing between carbon and chlorine is maximized during the dynamic simulation, leading to a very short interatomic distance and to the consequent destabilization of the whole system by a strong increase of the Pauli repulsion between atoms. It is noteworthy that a SQCT analysis (see Figure 9) does not exhibit such patterns, but, rather, points out a continuous decrease of the C−Cl2 bonding distance, corresponding to a normal increase of the bonding electronic population reaching an equilibrium value of 1.18e$^-$. Finally, the two last steps of the dynamic charge-

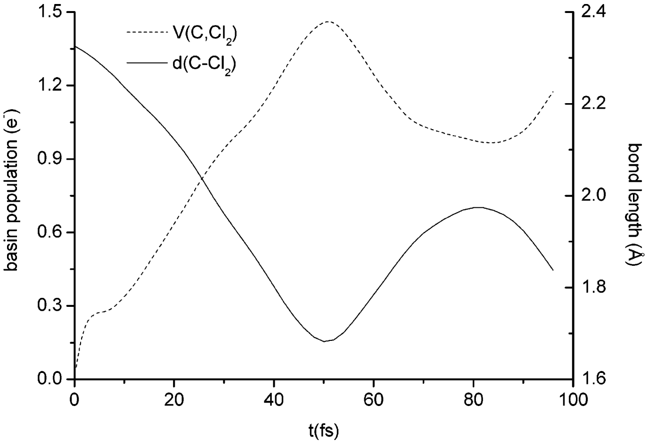


Figure 8. Left *Y* axis: variation of the C−Cl2 distance (in bold) along the ADMP trajectory. Right *Y* axis: variations of the forming bonding basin (C,Cl2) electronic population (dashed).

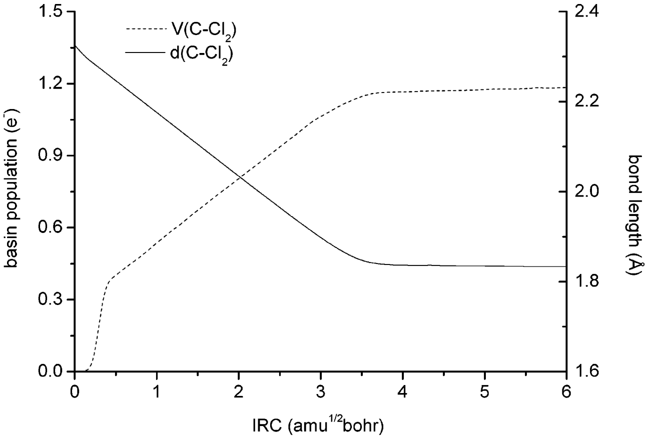


Figure 9. Left *Y* axis: variation of the C−Cl2 distance (in bold) along the IRP. Right *Y* axis: variations of the forming bonding basin (C,Cl2) electronic population (dashed).

transfer mechanism correspond to the second half of the trajectory. First, between 50 and 78 fs, the system, strongly destabilized, tends to reduce quickly its potential energy. Thus, we note a sudden lengthening of the C−Cl2 bond and an opening of the H−C−H valence angles (see Figure 7) coupled with a shortening of the C−H bond lengths (see Figure 6). This last effect induces a strong lowering of the charge density transferred from the carbon to the hydrogen atoms. Moreover, it is important to point out an inverted electron flux from the ion to the chloromethane molecule, which tries to equilibrate the destabilized system. Finally, the last part of the dynamic trajectory, beyond 78 fs, corresponds to the effective formation of a stable ion−molecule complex and to the end of the charge transfer between the molecule and the chlorine atom stabilized by the completion of its external valence shell.

4. Some Comments on the Static vs Dynamic Descriptions

Our results show how both DQCT and SQCT approaches suggest that the driving force for the formation of the ion−molecule complex are the local charge-transfer processes. These processes lead to an increase and maximization of the number of electron pairs exchanged between atoms during the bond formation process. This picture is consistent with previous analyses, based on different theoretical models. Among others, we want to recall that similar even if much more qualitative conclusions were drawn several years ago by Shaik on the basis of arguments resting on valence bond (VB) calculations.[67] In the same philosophy, but more recently, it is has been suggested that, within a VB approach, the overlap between the active orbitals of the incoming Cl^- and CH_3Cl moieties rules the overall reactive process.[62] Moreover, the maximization of electron pairs sharing between atoms strongly reminds one of the “maximum localization hybrid orbitals” overlap principle suggested more than 30 years ago by Del Re.[68] This criterion has been used as a driving force in a chemical reactivity study based on semiempirical qualitative approaches.[69] Analogous results were also reached by Toro-Labbé using the principle of maximum hardness.[70]

Beyond the static approach of reactivity, we observe that this force is so enhanced in the dynamics simulation that it allows an overgrowing electronic population in the bonding basin, as evidenced by the QCT of the ELF function. This results in a very short C−Cl2 distance and in high energy. The subsequent increase of the Pauli repulsion is permitted to eventually reach the final equilibrium state, through a large amplitude C−Cl2 motion. Before this reactive state, the reactants are in a kind of “harmonic reversible” regime where each part still preserves its own electronic identity. What is new in our analysis is that the combination of time dependent ab initio calculations combined to different quantum chemical topology approaches well underlines this charge-transfer phenomenon and its amplitude as a driving force, revisiting and assessing, in a modern and accurate way, working hypotheses carried out in more approximate schemes.

5. Conclusion

In the present work, we have analyzed in some detail the bond formation mechanism for a prototype S_N2 reaction using static and dynamic approaches, both based on DFT and localized (Gaussian) basis sets. Combined with different QCT partitions of the molecular system in real space, we performed static and dynamic QCT analyses to examine intra- and intermolecular electronic charge transfers along the static reaction profile and the dynamic trajectory. Although our results exhibit similar global features in the static and dynamic approaches, only a dynamic QCT approach reveals the crucial role of the electron charge transfers during the formation of the ion−dipole complex, leading to spontaneous maximizations of covalent and noncovalent interactions.

Acknowledgment. The authors thank Prof. Zuccarello (Catania, Italy) for providing some reprints.

References

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

(2) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(3) Löwdin, P. O. *J. Chem. Phys.* **1950**, *18*, 365.

(4) Del Re, G. *Theor. Chim. Acta* **1963**, *1*, 188.

(5) Pipek, J. P.; Mezey, G. *J. Chem. Phys.* **1989**, *90*, 4916.

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

(7) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, England, 1990.

(8) Popelier, P. L. A. *Atoms in Molecules: An Introduction*; Pearson Education: London, England, 2000.

(9) Becke, A. D.; Edgecombe, K. E. *J. Chem. Phys.* **1990**, *92*, 5397.

(10) Silvi, B.; Savin, A. *Nature* **1994**, *371*, 683.

(11) Savin, A.; Silvi, B.; Colonna, F. *Can. J. Chem.* **1996**, *74*, 1088.

(12) Bader, R. F. W.; Johnson, S.; Tang, T. H.; Popelier, P. L. A. *J. Phys. Chem.* **1996**, *100*, 15398

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

(14) Cortés-Guzmán, F.; Bader, R. F. W. *Coord. Chem. Rev.* **2005**, *249*, 633.

(15) Merino, G.; Vela, A.; Heine, T. *Chem. Rev.* **2005**, *105*, 3812.

(16) Bader, R. F. W.; MacDougall, P. J. *J. Am. Chem. Soc.* **1985**, *107*, 6788.

(17) Chaudry, U. A.; Popelier, P. L. A. *J. Phys. Chem. A* **2003**, *107*, 4578.

(18) Berski, S.; Andrés, J.; Silvi, B.; Domingo, L. R. *J. Phys. Chem. A* **2003**, *107*, 6014.

(19) Michelini, M. D. C.; Russo, N.; Alikhani, M. E. *J. Comput. Chem.* **2005**, *26*, 1284.

(20) Santos, J. C.; Andrés, J.; Aizman, A.; Fuentealba, P.; Polo, V. *J. Phys. Chem. A* **2005**, *109*, 3687.

(21) Fukui, K. *Acc. Chem. Res.* **1981**, *14*, 363. Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154.

(22) Solà, M.; Toro-Labbé, A. *J. Phys. Chem. A* **1999**, *103*, 8847.

(23) Polo, V.; Andrés, J. *J. Comput. Chem.* **2005**, *26*, 1427.

(24) Burns, T.; Marques, M. A. L.; Gross, E. K. U. *Phys. Rev. A* **2005**, *71*, 10501.

(25) Ziegler, T.; Autschbach, J. *Chem. Rev.* **2005**, *105*, 2695.

(26) Gleich, D.; Hutter, J. *Chem. Eur. J.* **2004**, *10*, 2435.

(27) Tateyama, Y.; Blumberger, J.; Sprik, M.; Tavernelli, I. *J. Chem. Phys.* **2005**, *122*, 234505.

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

(29) Marx, D.; Hutter, J. In *Modern Methods and Algorithms of Quantum Chemistry Proceedings*; Grotendorst, J., Ed.; Forschungszentrum Jülich: NIC Series, 2000; Vol. 3, p 339.

(30) Ammal, S. C.; Yamataka, H.; Aida, M.; Dupuis, M. *Science* **2003**, *299*, 1555.

(31) Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frish, M. J. *J. Chem. Phys.* **2001**, *114*, 9758. Iyengar, S. S.; Shlegel, H. B.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frish, M. J. *J. Chem. Phys.* **2001**, *115*, 10291. Shlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frish, M. J. *J. Chem. Phys.* **2002**, *117*, 8694.

(32) Ho, W. P.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 10726.

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

(34) Ensing, B.; Meijer, E. J.; Blöchl, P. E.; Baerends, P. E. *J. Phys. Chem. A* **2001**, *105*, 3300.

(35) Cossi, M.; Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *297*, 1.

(36) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

(37) Wladkowski, B. D.; Lim, K. F.; Allen, W. D.; Brauman, J. I. *J. Am. Chem. Soc.* **1992**, *114*, 9136 and references therein.

(38) Deng, L.; Branchadell, V.; Ziegler, T. *J. Am. Chem. Soc.* **1994**, *116*, 10645.

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

(40) Yang, S.-Y.; Fleurat-Lessard, P.; Hristov, I.; Ziegler, T. *J. Phys. Chem A* **2004**, *108*, 9461.

(41) Streitwieser, A.; Choy, G. S.; Abu-Hasanayn, F. J. *J. Am. Chem. Soc.* **1997**, *110*, 5013.

(42) Sun, L.; Song, K., Hase, W. L. *Science* **2002**, *296*, 875.

(43) Schmatz, S.; Botschwina, P.; Hauschildt, J.; Schinke, R. *J. Chem. Phys.* **2001**, *114*, 5233.

(44) Botschwina, P. *Theor. Chem. Acc.* **1998**, *99*, 426.

(45) Gonzales, J. M.; Allen, W. D.; Schaefer, H. F. *J. Phys. Chem. A* **2005**, *109*, 10613.

(46) Bento, A. P.; Solà, M.; Bickelhaupt, F. M. *J. Comput. Chem.* **2005**, *26*, 1497.

(47) Lee, I.; Kim, C. K.; Soh, C. K.; Li, H. G.; Lee, H. W. *J. Phys. Chem. A* **2002**, *106*, 1081.

(48) Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974.

(49) Gritsenko, O. V.; Ensing, B.; Schipper, P. R. T.; Baerends, E. J. *J. Phys. Chem. A* **2000**, *104*, 8558.

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

(51) Barlow, S. E.; Van Doren, J. M.; Bierbaum, V. M. *J. Am. Chem. Soc.* **1988**, *110*, 7240.

(52) Larson, J. W.; McMahon, T. B. *J. Am. Chem. Soc.* **1985**, *107*, 766. Li, C.; Ross, P.; Szulejko, J. E.; McMahon, T. B *J. Am. Chem. Soc.* **1996**, *118*, 9360. Wladkowski, B. D.; Brauman, J. I. *J. Phys. Chem.* **1993**, *97*, 13158.

(53) 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, S.; 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.; Gonzelez, C.; Pople, J. A. *Gaussian 03, Revision B.05*; Gaussian, Inc.: Wallingford, CT, 2004.

(54) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *109*, 911

(55) Francl, M. M.; Petro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M.-H.; DeFree, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654.

(56) Swope, W. C.; Andersen, H. C.; Berend, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637.

(57) Silvi, B. *J. Phys. Chem. A* **2003**, *107*, 3081.

(58) Noury, S.; Krokidis, X.; Fuster, F.; Silvi, B. Topmod package, Université Pierre et Marie Curie, 1997.

(59) Soler, P.; Bergès, J.; Fuster, F.; Chevreau, H. *Chem. Phys. Lett.* **2005**, *411*, 117. Chevreau, H. *J. Chem. Phys.* **2005**, *122*, 244316.

(60) Popelier, P. L. A.; Aicken, F. M. *Chem. Phys. Chem.* **2003**, *4*, 824.

(61) Toulouse, J.; Savin, A.; Adamo, C. *J. Chem. Phys*. **2002**, *117*, 10465.

(62) Blavins, J. J.; Cooper, D. L.; Karadakov, P. B. *J. Phys. Chem. A* **2004**, *914*, 108.

(63) Cioslowski, J.; Karkowski, J. *Quantum-Mechanical Theory of Atoms in Molecules: A Relativistic Formulation in Fundamentals of Molecular Similarity*; Kluwer Academic: 2001; p 101

(64) Cioslowski, J.; Mixon, S. T. *Can. J. Chem.* **1992**, *70*, 443.

(65) Cioslowski, J.; Mixon, S. T. *J. Am. Chem. Soc.* **1992**, *114*, 4382.

(66) Feinberg, M. J.; Ruedenberg, K. *J. Chem. Phys*. **1971**, *54*, 1495

(67) Shaik, S. S. *J. Am. Chem. Soc.* **1981**, *103*, 3692

(68) Rastelli, A.; Bozzoli, A. S.; Del Re, G. *J. Chem. Soc., Perkin Trans 2* **1972**, 1571

(69) Zuccarello, F.; Del Re, G. *J. Comput. Chem.* **1987**, *6*, 816.

(70) Toro-Labbé, A. *J. Phys. Chem.* **1999**, *103*, 4398.

CT0600159

JCTC Journal of Chemical Theory and Computation

What NMR Relaxation Can Tell Us about the Internal Motion of an RNA Hairpin: A Molecular Dynamics Simulation Study

Alessandra Villa* and Gerhard Stock

Institute for Physical and Theoretical Chemistry, J. W. Goethe University, Max-von-Laue-Strasse 7, 60438 Frankfurt am Main, Germany

Received May 5, 2006

Abstract: Classical molecular dynamics simulations of a 14-mer UUCG RNA hairpin are performed to study its conformational dynamics and corresponding NMR relaxation parameters. The direct calculation of the relaxation rates from the trajectory yields good agreement with experiment, indicating the validity of the theoretical model. Various ways to provide a link between theory and experiment are considered, including the "model-free approach" of Lipari and Szabo and Gaussian axial fluctuation model of Brüschweilwer. It is studied if the underlying assumptions of these approaches are satisfied in the case of a flexible RNA hairpin. Being consistent with the analysis of the NMR experiments, Lipari−Szabo fits of the first 100 or 1000 ps of the internal correlation functions lead to a nice agreement between calculated and experimental order parameters and internal correlation times. Finally, the relation between NMR order parameters and the underlying internal motion of the RNA hairpin is discussed in detail. A principal component analysis reveals that the principal motions of the molecule account only partially for the measured NMR order parameters, because the latter are insensitive to internal dynamics occurring on a nanosecond time scale due to molecular tumbling.

1. Introduction

Conformational dynamics may play a key role in the function of biomolecules such as proteins, DNA, and RNA. The flexible parts of a protein or ribonucleic acid, such as loop regions, are often involved in mediating specific interactions, for example, between protein and RNA during a binding process.[1−3] In the case of RNA, the flexibility of the loop may directly affect both the specificity and the affinity of the binding.[4] To account for the function of RNA systems, a site-specific dynamic description is therefore an important complement to static structural information.[5,6]

Internal motions in biomolecules occur on a wide range of time scales, from femtoseconds to seconds. Molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) spin relaxation measurements are valuable tools to gain access to fast (i.e., subnanosecond) internal motions. MD simulations[7] provide directly information at the atomistic level on inter- and intramolecular motions using an empirical force field as molecular model. NMR relaxation measurements[8] yield the dipolar correlation function, from which dynamical quantities such as generalized order parameters S^2 and effective correlation times τ_e can be extracted.[9−12] In particular, the combination of NMR and MD investigations has been shown to provide a comprehensive description of fast conformational dynamics of proteins.[13−20] On one hand, we may use experimental NMR results as benchmark data to study the accuracy of the MD description; on the other hand, we may employ the MD trajectory to provide a microscopic interpretation of the NMR experiments. While the description of structure and dynamics of proteins is well established, RNA systems have been comparatively little studied using NMR relaxation[21−25] or MD simulation. (For general reviews on RNA simulations see refs 26−33.)

---

* Corresponding author phone: +49-69-79829711; fax: +49-69-79829709; e-mail: villa@theochem.uni-frankfurt.de.

10.1021/ct600160z CCC: $33.50

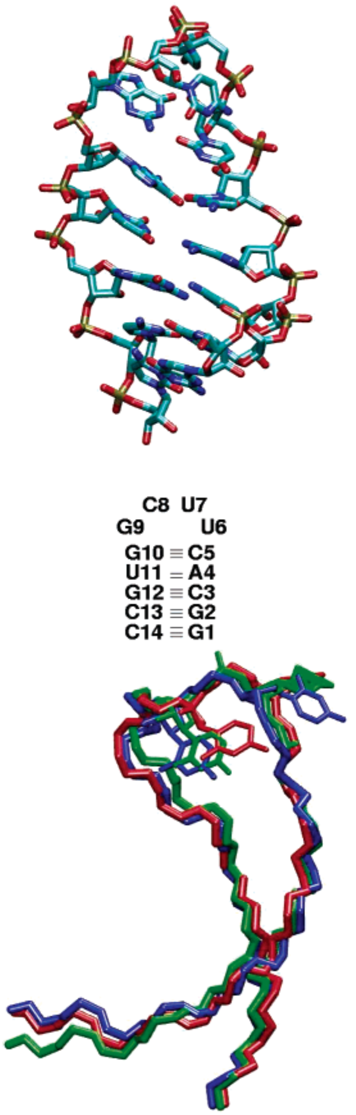


Figure 1. The 14-mer RNA loop. Top: representative MD snapshot at 50 ns. Center: secondary structure including base-pair hydrogen bonds and residue numbering. Bottom: backbone atoms and residues U7 and C8, describing the main hairpin motion along the first three principal components of the trajectory. Shown are snapshots at 10 ns (blue), 30 ns (green), and 50 ns (red). The pictures were performed using the graphical package VMD.[59]

In this work we present an MD study of the fast dynamics of the RNA hairpin ggcacUUCGgugcc (Figure 1). This hairpin belongs to one of the most stable tetraloop families[34] and is therefore a popular model system for theoretical[35−38] and experimental[39−41] investigations. The present study has also been motivated by recent work on UUCG loops by Duchardt and Schwalbe,[24] who performed detailed NMR ^{13}C relaxation measurements for the carbon atoms $C_{1'}$ in the ribose moiety and for the carbon atoms C_6 and C_8 in the pyrimidine and purine residues. Previously, ^{15}N relaxation measurements have been used by Akke et al.[21] to characterize the base dynamics of the UUCG loop.

First, the MD results are compared to the structures of available NMR[39,41] and crystallographic[40] studies on similar UUCG loops. Second, the MD trajectory is used to directly calculate the NMR relaxation rates,[18] in order to avoid most of the assumptions usually employed in the experimental analysis. Then we consider various ways to provide a link between theory and experiment, including the so-called model-free approach developed by Lipari and Szabo[9] and the Gaussian axial fluctuation model.[42] In particular, we study if the underlying assumptions of these approaches (such as the separation of overall and internal motion and the use of monoexponential Lipari−Szabo fits of the internal correlation function) are satisfied in the case of an RNA hairpin. Finally, the relation between NMR order parameters and the underlying internal motion of the RNA hairpin is discussed in detail.

2. Theory and Methods

2.1. Molecular Dynamics Simulations. The MD simulations were performed using the GROMACS suite of programs (version 3.2).[43,44] The AMBER force field (parm98)[45,46] was employed to describe the 14-mer UUCG RNA loop. The loop was placed in a rhombic dodecahedron box (edge length approximately 5 nm), which was subsequently filled with 2713 TIP3P water molecules.[47] To neutralize the system, 13 sodium ions were placed randomly in the simulation box.

A twin range cutoff was used for the Lennard-Jones interactions, that is, interactions between atoms within 1.0 nm were evaluated every step, while interactions between atoms within 1.4 nm were evaluated every 5 steps. The particle mesh Ewald method[48] was employed to treat Coulomb interactions, using a switching distance of 1.0 nm, a grid of 0.12 nm, and a beta value of 3.1 nm^{-1}. Constant pressure p and temperature T were maintained by weakly coupling the system to an external bath at 1 bar and 298 K, using the Berendsen barostat and thermostat, respectively.[49] The RNA, the ions, and the solvent were independently coupled to the temperature bath with a coupling time of 0.1 ps. The pressure coupling time was 0.5 ps, and the isothermal compressibility was $4.5 \cdot 10^{-5}$ bar^{-1}. The bond distances and the bond angle of the solvent water were constrained using the SETTLE algorithm.[50] All other bond distances were constrained using the LINCS algorithm.[51] A leapfrog integrator with an integration time step of 2 fs was used.

To obtain the starting structure of the UUCG hairpin, the loop was modeled based on the crystallographic structure of Ennifar et al.,[40] while the stem structure was built using tools of the AMBER6 software.[52] Following 20 ns of equilibration, the system was simulated for 50 ns. Analysis of the trajectories was performed with tools from the GROMACS package and with modified versions of them. To define the presence of a hydrogen bond, an acceptor−donor distance smaller than 0.35 nm was requested.

2.2. NMR Relaxation Parameters. We have focused on the relaxation of the ^{13}C nuclear spin through dipolar interaction with the attached ^{1}H. According to the relaxation theory of Bloch, Wangness, and Redfield[8], the spin−lattice (R_1), the spin−spin (R_2) relaxation rates, and the nuclear Overhauser enhancement (NOE) are given by

$$R_1 = \frac{d}{4}[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_C + \omega_H)] + cJ(\omega_C) \quad (1)$$

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

$$\mathrm{NOE} = 1 + \frac{d}{4R_1}\frac{\gamma_C}{\gamma_H}[6J(\omega_C + \omega_H) - J(\omega_H - \omega_C)] \quad (3)$$

where $d = (\mu_0^2/4\pi^2)(\hbar^2\gamma_C^2\gamma_H^2/r_{CH}^6)$, $c = \omega_C^2\Delta\sigma_C^2/3$, and $J(\omega)$ represents the spectra density defined in eq 5. Here, μ_0 is the vacuum permeability, $\hbar$ is Planck's constant, and γ_X is the gyromagnetic ratio of nucleus X. r_{CH} is the intermolecular distance between the two nuclei; for the bond $C_{1'}-H_{1'}$ a distance of 0.109 nm and for C_6-H_6 and C_8-H_8 a distance of 0.108 nm was used. $\Delta\sigma_C$ is the ^{13}C chemical shift anisotropy; for $C_{1'}$ a value of 45 ppm, for C_6 a value of −179 ppm, and for C_8 a value of −134 ppm were used. A value of 600.13 MHz was used for ω_H and 150.90 MHz for ω_C. All these values have been chosen in line with the NMR relaxation experiments of Duchardt and Schwalbe.[24]

2.3. Correlation Functions. The NMR relaxation due to the dipole−dipole interaction between two nuclei (i.e., carbon and hydrogen) can be described by the correlation function[8]

$$C(t) = \langle P_2(\vec{\mu}(0)\cdot\vec{\mu}(t))\rangle \quad (4)$$

where $\vec{\mu}$ is a unit vector pointing along the C−H bond, $P_2(x) = (1/2)(3x^2 - 1)$ is the second Legendre polynomial, and $\langle..\rangle$ denotes an equilibrium average. The spectral density

$$J(\omega) = 2\int_0^\infty C(t)\cos(\omega t)\,\mathrm{d}t \quad (5)$$

which determines the relaxation parameters in eqs 1−3, is given by the Fourier transform of the correlation function.

Assuming that overall and internal motions of the molecule are independent, the total correlation function $C(t)$ can be factorized in the correlation functions for overall motion, $C_O(t)$, and for internal motion, $C_I(t)$, respectively:

$$C(t) = C_O(t)C_I(t) \quad (6)$$

The total correlation functions were calculated for the $C_{1'}-H_{1'}$, C_6-H_6, and C_8-H_8 dipoles of all residues according to eq 4. To obtain the internal correlation functions, each conformation was translated and rotated to give the best fit to a reference structure. Since no large conformational arrangement took place during the 50 ns simulation, the molecule-fixed frame is unambiguously defined by this approach. Subsequently, the correlation functions for overall motion were calculated using eq 6. Assuming that the overall motion of the molecule is isotropic, this correlation function is given by

$$C_O(t) = \frac{1}{5}e^{-t/\tau_c} \quad (7)$$

where the rotational correlation time τ_c is proportional to the inverse of the rotation diffusion constant.

In the model-free approach of Lipari−Szabo,[9] the internal and overall motions are assumed to be independent, and the internal correlation function is given by the following relation

$$C_I(t) = S^2 + (1 - S^2)e^{-t/\tau_e} \quad (8)$$

where S^2 is the order parameter and τ_e is the effective (or internal) correlation time τ_e for the C−H dipole. Insertion of eq 8 in eqs 5−7 yields the spectral density

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_c}{1 + \tau_c^2\omega^2} + \frac{(1 - S^2)\tau}{1 + \tau^2\omega^2}\right) \quad (9)$$

with $\tau^{-1} = \tau_c^{-1} + \tau_e^{-1}$.

2.4. Order Parameters. We have employed three different approaches to calculate the order parameters.

Lipari−Szabo Fit. Employing the Lipari−Szabo form of the internal correlation function (eq 8), S^2 was fitted using the first 100 ps or the first 1 ns of the MD internal correlation.

Equilibrium Average. Using the general property of correlation functions that $\lim_{t\to\infty}\langle A(0)B(t)\rangle = \langle A\rangle\langle B\rangle$, the order parameter can be determined by[8]

$$S_{\mathrm{eq}}^2 = \lim_{t\to\infty} C_I(t) = \frac{4\pi}{5}\sum_{m=-2}^{m=2} |\langle Y_{2m}(\theta, \varphi)\rangle|^2 \quad (10)$$

where Y_{2m} is the spherical harmonic function of rank 2, $\theta(t)$ and $\varphi(t)$ are the polar angles defining the orientation of the dipole C−H at each snapshot of the trajectory, and $\langle\cdots\rangle$ denotes the average over all snapshots. This corresponds to a Lipari−Szabo fit using the full time range of the internal correlation function. It should be stressed that eq 10 avoids the cumbersome calculation of time-dependent correlation functions. In particular, this allows us to use highly efficient Monte Carlo schemes (e.g., like the popular replica exchange MD[53]) to calculate the equilibrium average in eq 10.

GAF Model. Assuming that the nucleobase flexibility monitored by the order parameters of C_6/C_8 is exclusively caused by base motions along the glycosidic torsional angle χ, the order parameter can be related to motions around the $C_{1'}-N_1$ or $C_{1'}-N_9$ bonds. Assuming furthermore a Gaussian distribution for the dihedral angle χ, the Gaussian axial fluctuation (GAF) model[42] leads to the following expression for the order parameter:

$$S_{\mathrm{gaf}}^2 = 1 - 3\sin^2\chi\left\{\cos^2\chi(1 - e^{-\sigma_\chi^2}) + \frac{1}{4}\sin^2\chi(1 - e^{-4\sigma_\chi^2})\right\} \quad (11)$$

Here the dihedral angle χ is defined by $O_{4'}-C_{1'}-N_1-C_2$ in the pyrimidine and by $O_{4'}-C_{1'}-N_9-C_4$ in the purine, and σ_χ is its standard deviation.

2.5. Principal Component Analysis. Principal component analysis is an efficient method to represent the motion of a 3*N*-dimensional system in terms of a few "principal" components.[54−56] The basic idea is that the correlations of the motions are represented by the covariance matrix

$$\sigma_{ij} = \langle(q_i - \langle q_i\rangle)(q_j - \langle q_j\rangle)\rangle \quad (12)$$

where $q_1, ..., q_{3N}$ are the mass-weighted Cartesian coordinates of the solute molecule, and $\langle\cdots\rangle$ denotes the average over all sampled conformations. By diagonalizing σ, we obtain $3N$ eigenvectors v_n and eigenvalues λ_n, which are rank-ordered in descending order, i.e., λ_1 is the largest eigenvalue. We may expand the MD trajectory $q(t) = \{q_i(t)\}$ on the basis of the eigenvectors υ_n according to

$$x^{(n)}(t) = \sum_{i=1}^{n} [\upsilon_i \cdot q(t)]\upsilon_i \qquad (13)$$

While for $n = 3N$ the expansion becomes exact [$x^{(3N)}(t) = q(t)$], for small n (in practice, $n = 1-5$) $x^{(n)}(t)$ approximates the motion of the system in terms of a few principal components representing the "essential" dynamics of the system.[56] The projection of the MD trajectory on the first n eigenvectors is used in the calculation of the order parameters.

3. Results and Discussion

3.1. Structural Features of the UUCG Loop. The 14-mer tetraloop was simulated for 50 ns in explicit water at 298 K. The root-mean-square deviation for all atom coordinates, after fitting to a reference structure, had an average value of 0.19 (±0.03) nm. As an illustration, Figure 1 shows a representative snapshot of the hairpin at the end of the trajectory. The residues forming the stem are all involved in Watson−Crick base-pairs and stacking interactions. The UUCG loop is mainly stabilized by hydrogen bonds between the residues U6 and G9. The residue U7 is looped out, and the residue C8 is unpaired. All bases are in an anti conformation, except for G9 which shows a syn conformation and for U7 which is in an equilibrium between syn and anti conformations.

The structures sampled by the MD simulation are in overall agreement with previous investigations on similar UUCG loops, including MD studies[35−38] as well as NMR[39,41] and crystallographic[40] experiments. In particular, experiments and simulation largely agree on the interactions that stabilize the secondary structure of the hairpin. The loop residues U6 and G9 are involved in base−base and base−sugar hydrogen bonds, and there is a hydrogen bond between the C8 base and the U7 phosphate oxygen as observed in the experimental structures. Moreover, a weak interaction is observed between the 2′-OH group of U7 and the base G9. In the crystallographic structure, the U7 sugar oxygen is hydrogen bonded to the G9 base oxygen, while this is not the case in the NMR structures. The analysis of the calculated backbone dihedral angles of the loop residues U6−G9 indicates the presence of at least two conformational states (see section 3.4). Two clusters of structures are also obtained by the NMR refinement, while only one structure is reported in the X-ray study. One of the two MD conformers shows a better agreement with the crystal structure and with one of the NMR clusters as shown in Tables 4 and 5 in the Supporting Information. In the simulation, dihedral angle transitions are observed for the residues U6−U7, while the two groups of NMR structures differ in residues C8−G9.

3.2. NMR Relaxation Parameters. The internal and total correlation functions of all $C_{1'}-H_{1'}$ sugar bonds and C_6-H_6 and C_8-H_8 base bonds have been calculated according to eq 4, using the 50 ns trajectory with and without subtracting the overall motion, respectively (see Methods). The total correlation functions decay on average within a nanosecond (see Figure 2), while the internal correlation functions generally show a decay on a time scale of 10 ps (see Figure 3 for some representative examples). Monoexponential fits of the internal correlation function are seen to be appropriate in most cases, except for the loop residues U6−G9. In the latter cases, the internal correlation functions exhibit a multiexponential decay on pico- and nanosecond time scales.

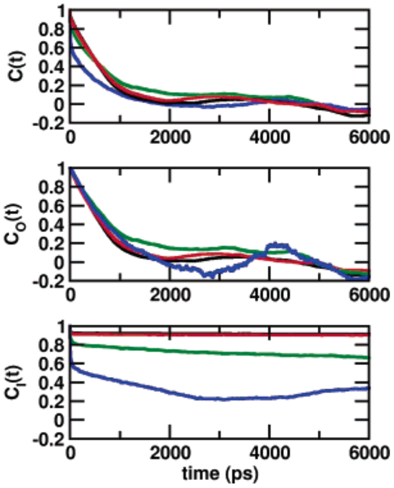


Figure 2. Total (top), overall-motion (middle), and internal (bottom) correlation functions of the dipoles $C_{1'}-H_{1'}$ (black and green line) and C_6-H_6 (red and blue line) for the residues C3 and U7, respectively.

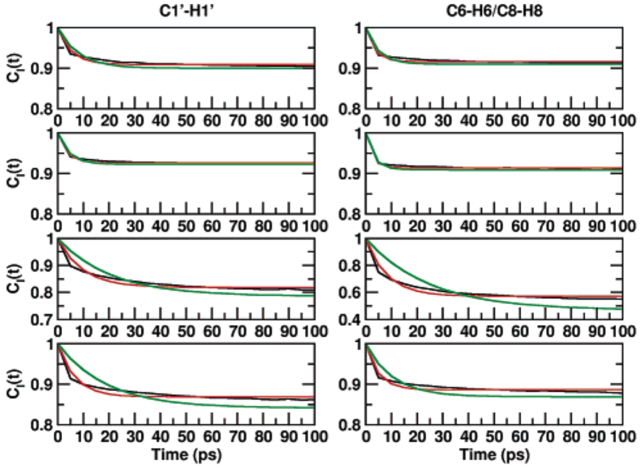


Figure 3. Internal correlation functions (black lines) of the dipoles $C_{1'}-H_{1'}$ (left side) and C_6-H_6 and C_8-H_8 (right side) for the residues G2, C3, U7, and G9 (from top to bottom). Exponential fits using the first 100 ps and the first 1 ns of the correlation function are shown in red and green, respectively.

Let us first study the validity of the assumption that overall and internal motions are separable, which leads to the factorization of the total correlation function $C(t)$ into components $C_O(t)$ and $C_I(t)$ describing overall and internal motion, respectively. As a representative example, Figure 2 shows the three correlation functions for the sugar and base dipole motions of the relatively rigid stem residue C3 and the most flexible loop residue U7. The overall and internal correlation functions of the stem residue exhibit an obvious separation of time scales (1 ns vs 10 ps) and are therefore clearly separable. In the case of the loop residue U7, on the other hand, both correlation functions decay on a nanosecond time scale, and one may expect a coupling of overall and internal motions. However, for the relatively short times (0.1−1 ns) that are relevant in the analysis of the experimental NMR data, the internal correlation function $C_I(t)$ may be approximated by a 24 ps decay time (see Figure 3) and $C_O(t)$ decays just as the other overall-motion correlation functions. That is, for short times the factorization approximation is not expected to change the results of the calculation of NMR data.

The correlation times τ_c were obtained by fitting the overall-motion correlation function of each dipole to the monoexponential function e^{-t/τ_c}. The fitted τ_c has an average value of 0.7 ns, which is clearly shorter than values obtained by using a hydrodynamics model[24] for the whole hairpin (2.35 ns for $C_{1'}$ and 2.17 ns for C_6/C_8). The main reason for this deviation appears to be the different viscosity of the solvent in experiment and simulation. The diffusion constant for TIP3P water is around 5.56−5.70 10^{-5} cm^2 s^{-1},[57,58] that is, about two times larger than the corresponding experimental value. To quantitatively calculate NMR observables, we therefore cannot directly take the MD correlation functions. Instead, we assume that the correlation function can be factorized (eq 6) and use the correlation time τ_c obtained from the NMR analysis.[24]

Using the experimental correlation time τ_c and the calculated internal correlation functions, we have employed eqs 1−7 to compute the relaxation parameters R_1, R_2, and NOE for all investigated dipoles. As shown by black circles in Figure 4, the calculated values are in good agreement with the experimental results.[24] The relative error $\sum_i|x_i^{\mathrm{MD}} - x_i^{\mathrm{exp}}|/\sum_i x_i^{\mathrm{exp}}$ is 0.04 for R_1, 0.14 for R_2, and 0.03 for NOE, respectively. Employing an axially symmetric diffusion model of overall rotation (with $\tau_c = 2.44$ and 2.18 ns and a diffusion anisotropy of 1.32 and 1.37 for $C_{1'}$ and C_6/C_8, respectively),[21,24] the relative errors are almost identical with respect to the isotropic diffusion model (0.04 for R_1, 0.13 for R_2, and 0.03 for NOE). The relatively large discrepancy obtained for the spin−spin relaxation rate R_2 may be related to the fact that in numerous cases the experimental R_2 values have been corrected for conformational exchange contributions during the experimental analysis.[24] To assess the validity of the Lipari−Szabo approach, the relaxation constants have also been obtained by fitting the first 100 and 1000 ps of C_I(t) to the Lipari−Szabo spectral density (red squares in Figure 4). Here, the relative errors for the 100 ps and 1 ns fit are 0.05 and 0.06 for R_1, 0.12 and 0.13 for R_2, and 0.02 and 0.03 for NOE, respectively. The relative errors of both fits are virtually identical and do not differ from the errors obtained in the direct MD evaluation of $C_I(t)$.

The qualitative agreement between the directly calculated MD data and the NMR relaxation parameters indicates that the force field and the simulation time scale used in this study are appropriate to describe the relaxation of the C−H bonds monitored in the NMR experiment. Thus, the MD simulation may be employed to reveal the dynamic information included in the experimental NMR data. Furthermore, we may use the MD data to compare and validate various methods to calculate the order parameters of the RNA hairpin.

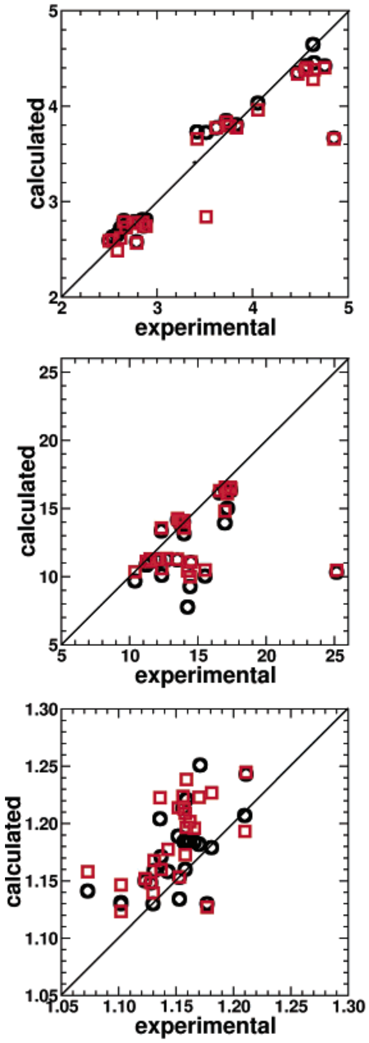


Figure 4. Experimental[24] vs calculated values of spin−lattice (top) and spin−spin (middle) relaxation rates (s^{-1}) and NOE (bottom) for all C−H dipoles. The black circles and red squares correspond to calculations using the spectral densities obtained directly from the correlation function and from a 100 ps Lipari−Szabo fit, respectively.

3.3. Order Parameters. Experimental and calculated order parameters for the $C_{1'}-H_{1'}$, C_6-H_6, and C_8-H_8 dipoles of all residues are compared in Figure 5 and in Tables 1 and 2. In the experimental study,[24] two models have been used to analyze the relaxation data, assuming either isotropic or axially symmetric diffusion. Both models are seen to give quite similar results for the order parameters and show that the loop residues, in particular U7, exhibit enhanced conformational fluctuations. The experimental effective correlation times τ_e listed in Tables 1 and 2 are mostly below 10 ps, with the exception of the residues C5, U7, and C14.

As detailed in the methods section, three different approaches have been used to calculate NMR order parameter from the MD trajectory. In the first approach, the order parameters S^2 and the internal correlation times τ_e were obtained by fitting the first 100 ps and the first 1 ns of the MD internal correlation to eq 8. Both Lipari−Szabo fits yield a good agreement with the experimental order parameters. In Figure 5 the results for the 100 ps fit are reported together with the experimental values. Tables 1 and 2 show that the

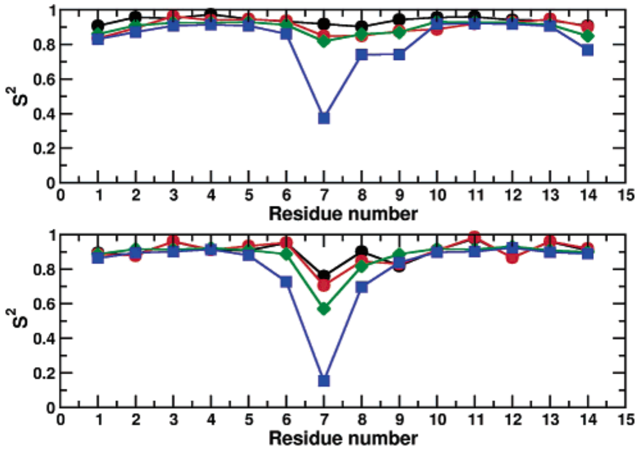


Figure 5. Order parameter S^2 for $C_{1'}$ (top) and C_6/C_8 (bottom) as a function of the residue number. The experimental values[24] are shown in black (using the isotropic diffusion model) and in red (using the axially symmetric diffusion model). Calculated values are shown in green (100 ps fit) and in blue (using eq 10).

Table 1: Internal Correlation Times τ_e and Order Parameters of $C_{1'}$ for the 14 Residues of the Hairpin[a]

	MD (100 ps fit)		MD (1 ns fit)		MD (eq 10)	experiment	
	τ_e (ps)	S^2	τ_e (ps)	S^2	S_{eq}^2	τ_e (ps)	S^2
G1	6.3	0.859	9.1	0.847	0.830	<10	0.835
G2	5.2	0.909	8.1	0.900	0.871	<10	0.896
C3	3.6	0.926	4.4	0.922	0.908	<10	0.963
A4	2.7	0.925	3.2	0.923	0.914	<10	0.939
C5	3.2	0.928	4.4	0.924	0.908	<10	0.946
U6	5.4	0.913	14.6	0.897	0.861	<10	0.936
U7	9.7	0.818	20.6	0.786	0.374	15.74	0.848
C8	9.9	0.858	23.8	0.829	0.741	<10	0.850
G9	6.6	0.869	19.1	0.841	0.742	<10	0.877
G10	3.2	0.931	4.3	0.926	0.917	<10	0.888
U11	2.5	0.927	2.9	0.925	0.919	<10	0.920
G12	3.1	0.926	3.6	0.924	0.918	<10	0.928
C13	3.8	0.914	4.7	0.910	0.906	<10	0.946
C14	8.3	0.849	52.9	0.782	0.767	412.14	0.902

[a] Reported values are derived from the MD simulation (by 100 ps fitting, 1 ns fitting, and by using eq 10) and from the NMR experiment[24] (isotropic model).

100 ps fit is more appropriate to be compared to the experimental data, since it better reproduces the experimental effective correlation times.

In the second approach, we have used the equilibrium average in eq 10 to calculate the order parameters. This corresponds to a Lipari−Szabo fit using the full time range of the internal correlation function. Figure 5 demonstrates that the resulting order parameters only agree for the relatively rigid stem residues but not for the flexible loop residues. The reason for this discrepancy is that, by using the entire internal correlation function, the order parameters contain information also on internal motions occurring on a nanosecond time scale. For example, the loop conformational rearrangement and the anti−syn transitions of the U7 base represent internal motions on a time scale longer than 5 ns. The experimental analysis is limited by the decay of the overall-motion correlation function due to molecular tumbling (≈2 ns). As a consequence, possibly existing internal motions on a nanosecond time scale are not reflected in the experimental data. Although S_{eq}^2 correctly reflects the fluctuations of the system, it may therefore not be suited for the comparison to NMR order parameter.

Table 2: Internal Correlation Times τ_e and Order Parameters of C_6/C_8 for the 14 Residues of the Hairpin[a]

	MD (100 ps fit)		MD (1 ns fit)		MD (eq 10)	experiment		GAF
	τ_e (ps)	S^2	τ_e (ps)	S^2	S_{eq}^2	τ_e (ps)	S^2	S_{gaf}^2
G1	4.4	0.886	6.4	0.877	0.864	<10	0.886	0.998
G2	3.5	0.916	4.9	0.910	0.894	<10	0.878	0.995
C3	2.5	0.914	3.2	0.909	0.900	<10	0.961	0.993
A4	2.3	0.921	2.8	0.918	0.913	<10	0.912	0.987
C5	2.6	0.909	3.5	0.904	0.880	121.59	0.933	0.977
U6	4.1	0.887	19.1	0.858	0.727	<10	0.953	0.928
U7	8.3	0.571	24.3	0.469	0.153	13.38	0.706	
C8	7.3	0.816	18.7	0.780	0.696	<10	0.845	0.938
G9	4.9	0.886	10.5	0.869	0.838	<10	0.830	0.770
G10	2.4	0.917	3.3	0.911	0.897	<10	0.909	0.991
U11	2.3	0.914	2.7	0.911	0.901	<10	0.984	0.986
G12	1.9	0.931	2.2	0.929	0.923	<10	0.866	0.998
C13	2.9	0.909	3.7	0.904	0.898	<10	0.960	0.993
C14	3.6	0.899	4.6	0.894	0.889	38.08	0.919	0.959

[a] Reported values are derived from the MD simulation (by 100 ps fitting, 1 ns fitting, and by using eq 10), from the NMR experiment[24] (isotropic model), and from the GAF model.

Finally, we have applied the GAF model (eq 11), which assumes that the nucleobase flexibility monitored by the order parameters of C_6/C_8 is exclusively caused by base motions along the glycosidic torsional angle χ. To this end, we have calculated the distribution function of χ, which exhibits a single peak for all residues except U7. In the latter case, the base adopts both anti and syn conformations during the simulation, and the GAF model is not applicable (see Methods). Table 5 in the Supporting Information lists the mean and the variance of χ for all residues as calculated from the MD trajectory. All results are found to be in good agreement with the experimental data.[39,24] The order parameters obtained from the GAF model are reported in Table 2. Except for the residue G9, the values for S_{gaf}^2 are significantly larger (from 0.928 to 0.998) than the ones obtained from experiment and the Lipari−Szabo fit. The failure of the GAF model to correctly reproduce the order parameters of the UUCG hairpin clearly demonstrates that the motion of the base C−H dipole is not only caused by fluctuations of the base but is also due to the flexibility of the sugar ring and the backbone. For the UUCG hairpin under consideration, the GAF model on average accounts for about 20% of the fluctuations contributing to the order parameter.

Experimentally, the order parameters depend on the reorientation of the dipole and on the reorientation of the chemical shift anisotropy tensor. The latter contribution may affect the data on the aromatic carbon more than the dipole relaxation and is not easy to discriminate in practice. To support our analysis above, we have also calculated the order parameters and the internal correlation times of the N−H dipoles, by fitting the first 100 ps of the MD internal correlation function to eq 8. In Table 3, the calculated values

Table 3: Internal Correlation Times τ_e and Order Parameters of N_1/N_3 for the Guanine and Uracil Bases of the Hairpin[a]

	MD (100 ps fit)		experiment[21]		experiment[24]	
	τ_e (ps)	S^2	τ_e (ps)	S^2	τ_e (ps)	S^2
G1	3.7	0.890		0.74		
G2	2.5	0.907		0.787		0.938
U6	3.5	0.851	9.0	0.773		
U7	7.2	0.544				
G9	3.2	0.885	4.0	0.807		0.955
G10	2.0	0.917		0.776		0.907
U11	1.7	0.902		0.760		0.905
G12	1.7	0.918		0.780		0.905

[a] Reported values are derived from the MD simulation (by 100 ps fitting) and from the NMR experiment[21,24] (axially symmetric diffusion model except for G1).

Table 4: Results of $1 - S^2$ Obtained for the Complete Trajectory (all) and by Including the First Three (PCA3) and the First Ten (PCA10) Principal Components, Respectively[a]

	$1 - S^2$ (100 ps fit)			$1 - S_{eq}{}^2$ (eq 10)		
sugar	all	PCA3	PCA10	all	PCA3	PCA10
U6	0.087	0.006	0.024	0.139	0.022	0.065
U7	0.182	0.053	0.089	0.626	0.511	0.578
C8	0.142	0.043	0.073	0.259	0.144	0.200
G9	0.131	0.033	0.056	0.258	0.140	0.186
base	all	PCA3	PCA10	all	PCA3	PCA10
U6	0.113	0.018	0.047	0.273	0.063	0.191
U7	0.429	0.082	0.232	0.847	0.195	0.680
C8	0.184	0.035	0.070	0.304	0.121	0.192
G9	0.114	0.013	0.035	0.162	0.033	0.074

[a] The order parameter are calculated either from a 100 ps Lipari–Szabo fit or from an equilibrium average (eq 10).

for $N_1{-}H_1$ and $N_3{-}H_3$ dipoles of guanine and uracil bases are reported together with the available experimental data.[21,24] The N–H values fully agree with the values calculated for C–H dipole of the same base. Hence it may be expected that they characterize the identical base dynamics. Moreover, the calculated S^2 for N–H dipoles show a similar trend as in the experimental data, but they quantitatively agree only with one of the two available experimental sets.

3.4. Internal Dynamics of the UUCG Loop. In practice, the above studied GAF model is not used to calculate the order parameter but to rationalize the internal motions described by the experimental order parameter. Having validated our theoretical model in section 3.2, we are now in a position to use the 50 ns all-atom trajectory for this purpose. In what follows, we first characterize the main motions of the RNA hairpin. Then we analyze to what extent these motions are reflected in the calculated order parameter.

Principal component analysis (PCA) represents a standard method to identify the "principal" motions of a molecular system.[54−56] The approach represents the motion in terms of an orthogonal basis, the principal components (PCs), which are ordered according to their content of root-mean-square fluctuations (see Methods). For the 50 ns simulation of the UUCG loop, the first three PCs already contain 60% of the overall fluctuations of the hairpin, and to cover 80, 90, and 95% of the fluctuations, only 10, 25, and 50 out of 1442 PCs are required, respectively. In this sense, the first few PCs represent the main motions of the system. As an illustration, Figure 1 shows the motion along the first three PCs, which mainly consists of a conformational rearrangement involving the loop region. The presence of (at least) two conformational states is also confirmed by the analysis of the backbone dihedral angles of the loop residues U6–G9, see Table 4 in the Supporting Information. To assess the influence of this conformational rearrangement on the NMR order parameters, we have recalculated S^2 for the first and second half of the trajectory. Both 100 and 1000 ps Lipari–Szabo fits yield virtually identical results for S^2. The highest discrepancy is shown by the U7 base in the 1 ns fit (0.43 vs 0.50). This finding indicates that a simulation time of 50 ns is enough to investigate the C–H relaxation of the RNA loop and that its backbone rearrangement at 30 ns does not affect the order parameter values.

Let us now study to what extent the principal motions of the system account for the order parameters. To this end, we have expanded the MD trajectory in its first *n* PCs (see Methods) and calculated S^2 from this approximated trajectory, using a 100 ps Lipari–Szabo fit. Figure 6 in the Supporting Information shows the resulting order parameters as a function of the number of included PCs. As may be expected, the first PCs in general make the largest contribution to the decay of S^2. Compared to the rapid convergence of the overall fluctuations, however, the order parameters converge relatively slowly to their value obtained for the complete trajectory. Focusing on the loop residues U6–G9, Table 4 compares the results of $1 - S^2$ obtained for the complete trajectory (all), for the first three PCs (PCA3), and for the first ten PCs (PCA10). On average, the first three and 10 PCs yield about 20 and 40% of the value of $1 - S^2$ for the complete trajectory, respectively. Recalling that the first three and ten PCs contain 60 and 80% of the overall fluctuations, respectively, the order parameter apparently accounts only partially for the principal motions of the system. This is because the motion along the first few PCs may (i) be only weakly correlated with the orientation of the C–H dipoles and (ii) contains slow motion which is not seen by the Lipari–Szabo fit.

The latter issue can be studied by recalculating $1 - S_{eq}{}^2$ via an equilibrium average that covers all time scales of the trajectory. As shown in Table 4, in this case the first three and ten PCs on average yield about 40 and 70% of the value of $1 - S_{eq}{}^2$ for the complete trajectory, respectively, that is, quite similar to the values obtained for the overall fluctuations. In particular, we find that the $S_{eq}{}^2$ for the sugars U7, C8, and G9 to a large extent (80 and 55%) are caused by the motion along the first three PCs shown in Figure 1. Hence, if all time scales of the trajectory are taken into account, $S_{eq}{}^2$ are well described by the principal motions of the system.

4. Conclusions

A 50 ns MD simulation has been used to study the fast dynamics of the 14-mer UUCG hairpin. The comparison of the simulation with the NMR relaxation data allows us to achieve several interesting results. First, the force field and the 50 ns simulation time used in this study appear appropriate to describe the fast relaxation of the C−H bonds monitored in the NMR experiment. Although we find slow conformation rearrangements of the RNA loop which may not be sampled appropriately, these motions do not influence the calculated NMR parameters. Internal motions on a time scale longer than 1 ns have been shown not to affect the calculated NMR relaxation rates and, as a consequence, will not be reflected in the order parameters.

Second, the overall and internal motions of the hairpin are found to be virtually independent, and the factorization approximation of the correlation function holds. Even in the case of the flexible U7, the approximation is not expected to change the calculated NMR data, since only the dynamics at short times ($\leq$1 ns) affect the relaxation parameters. Moreover, the internal correlation function can be replaced by a monoexponential function, fitted on its first 100 or 1000 ps, with no change of the relaxation parameters. The latter two observations demonstrate the validity of the Lipari−Szabo approach for the RNA hairpin.

Different ways to calculate the NMR order parameter have been compared. The best approach was found to be Lipari−Szabo fits of the first 100 ps (or 1 ns) of the MD internal correlation functions. The resulting order parameters S^2 and internal correlation times τ_e nicely agree with the experimental values. The calculation of order parameters via an equilibrium average was shown to deteriorate for residues undergoing slow internal dynamics, since the latter cannot be seen in NMR relaxation experiment limited by molecular tumbling. Although the equilibrium-average calculation of order parameters correctly reflects the fluctuations of the system, it may therefore not be suited for the comparison to experimental results. Finally the GAF model only yielded about 20% of the correct value for $1 - S^2$. This indicates that the motion of the base C−H dipoles not only maybe caused by fluctuations of the base but also are due to the flexibility of the sugar ring and the backbone, in particular for the loop residue C8.

The principal component analysis of the 50 ns trajectory has confirmed that a conformational rearrangement involving the loop region represents the main motion of the system. This principal motion, however, accounts only partially for the measured NMR order parameters S^2, because the latter are not sensitive to internal dynamics on a nanosecond time scale. Calculating the order parameter via an equilibrium average that covers all time scales of the trajectory, we obtain a direct correspondence between calculated S_{eq}^2 and principal motions. In particular, we find that the order parameters S_{eq}^2 for the sugars of residues U7, C8, and G9 are to a large extent caused by the motion along the first three principal components shown in Figure 1.

Acknowledgment. The authors wish to express their thanks to Elke Duchardt and Harald Schwalbe for numerous inspiring and helpful discussions and for sharing their results prior to publication. They also thank Jessica Koplin and Yuguang Mu for their contributions in the early stages of this work. This work has been supported by the Frankfurt Center for Scientific Computing, the Fonds der Chemischen Industrie, and the Deutsche Forschungsgemeinschaft (SFB 579 “RNA-ligand interactions”).

Supporting Information Available: Tables presenting the calculated and experimental mean values of the main dihedral angles of the 14-mer RNA hairpin and a figure reporting the order parameters for $C_{1'}$ and C_6/C_8 as a function of the number of included principal components. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Wand, A. J. *Nature Struct. Biol.* **2001**, *8*, 926−931.

(2) Leulliot, N.; Varani, G. *Biochemistry* **2001**, *40*, 7947−7956.

(3) Al-Hashimi, H. M. *ChemBioChem* **2005**, *6*, 1506−1519.

(4) Persson, T.; Hartmann, R. K.; Eckstein, F. *ChemBioChem* **2002**, *3*, 1066−1071.

(5) Koplin, J.; Mu, Y.; Richter, C.; Schwalbe, H.; Stock, G. *Structure* **2005**, *13*, 1255−1267.

(6) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. *Science* **2006**, *311*, 653−656.

(7) Frenkel, D.; Smit, B. *Understanding molecular simulations: from algorithms to applications*; Academic Press: London, 1996.

(8) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Oxford University Press: New York, 2004.

(9) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.

(10) Brüschweiler, R.; Case, D. A. *Prog. Nucl. Magn. Reson. Spectrosc* **1994**, *26*, 27−58.

(11) Fischer, M. W. F.; Majumdar, A.; Zuiderweg, E. R. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *33*, 207−272.

(12) Korzhnev, D. M.; Billeter, M.; Arseniev, A. S.; Orekhov, V. Y. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 197−266.

(13) Levy, R. M.; Karlpus, M.; Wolynes, P. G. *J. Am. Chem. Soc.* **1981**, *103*, 5998−6011.

(14) Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289−2302.

(15) Palmer, A. G., III; Case, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 9059−9067.

(16) Chatfield, D. C.; Szabo, A.; Brooks, B. R. *J. Am. Chem. Soc.* **1998**, *120*, 5301−5311.

(17) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 7305−7313.

(18) Peter, C.; Daura, X.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *20*, 297−310.

(19) Case, D. A. *Acc. Chem. Res.* **2002**, *35*, 325−331.

(20) Lange, O. F.; Grubmüller, H.; de Groot, B. L. *Angew. Chem.* **2005**, *44*, 3394−3399.

(21) Akke, M.; Fiala, R.; Jiang, F.; Patel, D.; Palmer, A. G., III. *RNA* **1997**, *3*, 702−709.

(22) Vallurupalli, P.; Kay, L. E. *J. Am. Chem. Soc.* **2005**, *127*, 6893−6901.

(23) Showalter, S. A.; Baker, N. A.; Tang, C.; Hall, K. B. *J. Biomol. NMR* **2005**, *32*, 179−193.

(24) Duchardt, E.; Schwalbe, H. *J. Biomol. NMR* **2005**, *32*, 295−308.

(25) Chiarparin, E.; Rüdisser, S.; Bodenhausen, G. *ChemPhysChem* **2001**, *2*, 41−45.

(26) Cheatham, T., III; Kollman, P. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435−471.

(27) Zacharias, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 311−317.

(28) Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **2000**, *8*, 227−236.

(29) Auffinger, P.; Westhof, E. *Biopolymers* **2001**, *56*, 266−274.

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

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

(32) Norberg, J.; Nilsson, L. *Acc. Chem. Res* **2002**, *35*, 465−472.

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

(34) Woese, C.; Winker, S.; Gutell, R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8467−8471.

(35) Williams, J.; Hall, K. *Biophys. J.* **1999**, *76*, 3192−3205.

(36) Williams, J.; Hall, K. *J. Mol. Biol.* **2000**, *297*, 1045−1061.

(37) Miller, J.; Kollman, P. *J. Mol. Biol.* **1997**, *270*, 436−450.

(38) Nina, M.; Simonson, T. *J. Phys. Chem. B* **2002**, *106*, 3696−3705.

(39) Allain, F.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333−353.

(40) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. *J. Mol. Biol.* **2000**, *304*, 35−42.

(41) Duchardt, E.; Richter, C.; Ohlenschläger, O.; Görlach, M.; Wöhnert, J.; Schwalbe, H. *J. Am. Chem. Soc.* **2004**, *126*, 1962−1970.

(42) Brüschweiler, R.; Wright, P. *J. Am. Chem. Soc.* **1994**, *116*, 8426−8427.

(43) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Comm.* **1995**, *91*, 43−56.

(44) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.

(45) Cheatham, T.; Cieplak, P.; Kollman, P. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−861.

(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

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

(48) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(49) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(50) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.

(51) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(52) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber 6*; University of California: San Francisco, CA, 1999.

(53) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(54) Ichiye, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 205−217.

(55) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696−2699.

(56) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−425.

(57) Mark, P.; Nilsson, L. *J. Comput. Chem.* **2002**, *23*, 1211−1219.

(58) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(59) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1990**, *14*, 33−38.

CT600160Z

JCTC Journal of Chemical Theory and Computation

Multiconfiguration Molecular Mechanics Based on Combined Quantum Mechanical and Molecular Mechanical Calculations

Hai Lin,*[,†,‡] Yan Zhao,[†] Oksana Tishchenko,[†] and Donald G. Truhlar*[,†]

Chemistry Department and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431, and Chemistry Department, University of Colorado at Denver and Health Science Center, Denver, Colorado 80217-3364

Received May 15, 2006

Abstract: The multiconfiguration molecular mechanics (MCMM) method is a general algorithm for generating potential energy surfaces for chemical reactions by fitting high-level electronic structure data with the help of molecular mechanical (MM) potentials. It was previously developed as an extension of standard MM to reactive systems by inclusion of multidimensional resonance interactions between MM configurations corresponding to specific valence bonding patterns, with the resonance matrix element obtained from quantum mechanical (QM) electronic structure calculations. In particular, the resonance matrix element is obtained by multidimensional interpolation employing a finite number of geometries at which electronic-structure calculations of the energy, gradient, and Hessian are carried out. In this paper, we present a strategy for combining MCMM with hybrid quantum mechanical molecular mechanical (QM/MM) methods. In the new scheme, electronic-structure information for obtaining the resonance integral is obtained by means of hybrid QM/MM calculations instead of fully QM calculations. As such, the new strategy can be applied to the studies of very large reactive systems. The new MCMM scheme is tested for two hydrogen-transfer reactions. Very encouraging convergence is obtained for rate constants including tunneling, suggesting that the new MCMM method, called QM/MM-MCMM, is a very general, stable, and efficient procedure for generating potential energy surfaces for large reactive systems. The results are found to converge well with respect to the number of Hessians. The results are also compared to calculations in which the resonance integral data are obtained by pure QM, and this illustrates the sensitivity of reaction rate calculations to the treatment of the QM-MM border. For the smaller of the two systems, comparison is also made to direct dynamics calculations in which the potential energies are computed quantum mechanically on the fly.

1. Introduction

An accurate potential energy surface is an essential element in carrying out reliable dynamics simulations of a chemical reaction. Direct dynamics,[1−12] where the potential energy surface is calculated on the fly by electronic structure theory, is limited to relatively small systems or low levels of electronic structure theory due to the high computational costs of reliable electronic structure calculations. Multiconfiguration molecular mechanics (MCMM),[13−17] which is an extension of conventional molecular mechanics (MM)[18−47] so that it can be used for chemical reactions, is an alternative way to generate reactive potential energy surfaces with significantly reduced computational effort.

---

* Corresponding author fax: (612)624-9390; e-mail: truhlar@umn.edu (D.G.T.); hai.lin@cudenver.edu (H.L.).

[†] University of Minnesota.

[‡] University of Colorado at Denver and Heath Science Center.

10.1021/ct600171u CCC: \$33.50

In MCMM,[13−17] as in earlier semiempirical[48−55] and empirical[56−62] valence bond formulations, the actual representation for a system at a given geometry is obtained by mixing multiple (usually two: reactant and product) MM configurations; and as in empirical valence bond theory,[56−62] one constructs an electronically nonadiabatic (i.e., diabatic) Hamiltonian matrix **V** whose diagonal elements (V_{11} and V_{22}) are given by MM. The MM configuration of reactant (or product) is the bonding pattern for the system in the reactant (or product) state, and the corresponding energy is interpreted as a diagonal element of a valence bond configuration interaction Hamiltonian matrix. These valence bond states are also sometimes called diabatic electronic states. For example, in an atom transfer reaction $AB + C \rightarrow A + BC$, the electronic wave function for any geometry can be considered to be composed of two interacting valence-bond or diabatic configurations (A−B, C) and (A, B−C), where "−" denotes a bond in MM and paired set of bonding orbitals in valence bond theory. Configuration interaction between the reactant and product configurations of different bonding patterns leads to the ground-state potential energy surface.

The resonance integral (V_{12}, which equals V_{21}) is the most critical element in MCMM, and it is at the heart of the MCMM methodology. In our implementation of MCMM,[13,14] the resonance integral is obtained by Shepard interpolation of quadratic expansions around a set of points where electronic structure data are available. The Born−Oppenheimer potential energy surface is approximated by the lowest eigenvalue of the matrix **V**, and the resonance integral reproduces a quadratic expansion[57] of the QM data in the vicinity of each electronic structure data point. This kind of nondiagonal representation of the Hamiltonian has been widely used in a variety of contexts for modeling reactive systems,[4,48−66] but in work prior to MCMM, the resonance integral was usually a simple functional form (even a constant) fit to one or more data and potential energy surface feature (e.g., barrier height) rather than, as in MCMM, a systematically improvable function fit to reproduce data at as many geometries as required to achieve convergence.

The MCMM approach not only is similar to the empirical valence-bond[56−62] (EVB) treatment but also contains critical differences.[13−15] (1) As mentioned in the previous paragraph, the resonance between valence-bond configurations is calibrated to the electronic-structure calculations instead of to often-limited kinetic data from experiments or accurate data only at the saddle point. (2) A very general scheme (the Shepard interpolation[67,68] scheme) is used to generate a semiglobal potential energy surface. In work reported so far and in the present article, the semiglobal surface is designed to be accurate in the kinetically important region for reaction-path rate calculations and large-curvature tunneling calculations (called the reaction swath,[7,69] i.e., the region near the reaction path including the region associated with small-curvature tunneling[7,70−74] and also the larger region on the concave side of the reaction valley that is critical for large-curvature tunneling[7,69,74−79]). In future work, MCMM will be extended to cover the entire region of the potential energy surface required to run classical trajectories.

The MCMM method can be considered to be a general and efficient fitting scheme for creating potential energy surfaces for reactive systems. A merit of the MCMM method is that it does not require the human judgment traditionally associated with the "art" of fitting multidimensional functions. The MCMM fitting process is unique and automatic, except for the choice of geometries to include, the choice of MM functions, and the choice of QM level. In principle the results converge to a numerically accurate interpolation of the potential energy surface for any reasonable scheme of adding data, although in practice one strategy of adding data points may converge the results faster than another strategy does. Moreover, MCMM does not require a uniform grid for adding data points; this flexibility of using a nonuniform distribution of data points allows one to obtain a fairly high accuracy from a reasonably small amount of electronic structure data by putting a dense set of data points in an interesting region, while a sparse set of data points is used in other regions and one can always improve the accuracy by including more data points whenever needed. Once constructed, the MCMM surface is inexpensive to calculate, with a cost roughly the same as conventional MM. Another notable merit of the MCMM method is that the constructed potential energy surface is full dimensional, since both the diagonal and off-diagonal elements in the diabatic Hamiltonian are full dimensional. The full-dimensional nature of the potential energy surface is critical for realistic modeling of chemical reactions.

An alternative approach is to restrict the functional form of the resonance matrix element and to obtain the parameters in the assumed functional form by global least-squares fitting.[59] This is a promising approach, but it is not systematically improvable like the MCMM algorithm based on Shepard interpolation. An alternative to straight Shepard interpolation that could be employed in MCMM is to combine it with interpolant moving least squares, as proposed by Ishida and Schatz.[80] In this method, one first performs local least-squares fitting with restricted functional forms (e.g., polynomials) for the energies from electronic-structure calculations at a set of given geometries, and one then computes the gradients and Hessians at these geometries by employing the locally fitted functions. Finally, one carries out Shepard interpolation based on these electronic-structure energies and approximate gradients and Hessians. This is a very promising method, because it maintains the merit of Shepard interpolation and avoids the electronic-structure calculations of gradients and Hessians, which are more expensive than for energies. However, we note that, as a tradeoff, this method requires a large amount of energetic data in order to ensure the approximated derivatives are accurate representations of the true (electronic-structure) derivatives. Whether one uses straight Shepard interpolation or an intermediate interpolant least squares step, the key feature that reduces the data requirements in MCMM is that one interpolates the relatively smooth V_{12} and combines it with powerful preexisting MM functions, whereas other work employing Shepard interpolation[67,68,81] interpolates the more structured ground-state potential energy surface and does not take advantage of MM for the dependence of the potential on spectator degrees of freedom.

The usefulness of the MCMM scheme has been established for rate constant calculations as demonstrated by tests,[13,14,16,17,82] against a diverse set of reactions involving hydrogen transfer using variational transition state theory with multidimensional tunneling[7,73,74,79,83−86] (VTST/MT). It should be noted that the reactions used for tests are challenging ones, because of significant variational and tunneling effects. It is encouraging that a scheme involving predetermined locations for the points was able to predict rates in good agreement (~10%) with direct dynamics for all reactions studied with only a dozen (or less) electronic-structure Hessians[13] or partial electronic-structure Hessians.[16]

As a fitting scheme for creating potential energy surfaces for reactive systems, the success of the MCMM method relies on the electronic-structure data at the Shepard input points. Electronic-structure calculations are feasible for small- and medium-size systems comprising tens to hundreds of atoms, depending on the level of theory involved. The partial electronic-structure Hessian scheme,[16] which uses electronic-structure calculated Hessian elements for active atoms and interpolated Hessian elements for spectator atoms for non-stationary Shepard points, reduces the computational cost by up to an order of magnitude, making it feasible to apply MCMM dynamics to large organic reactions. However, for even larger reactive systems such as enzymes, electronic-structure calculations are too expensive to be applied to the whole system, even just at the saddle point, and an alternative solution must be found.

Combined quantum mechanical and molecular mechanical (QM/MM) methods[87−145] are very promising as a practical way to extend electronic-structure calculations to large reactive systems where a reaction takes place in a localized region. (Extensions to cases where the localized region can change are also available.[95,110,132,145]) A QM/MM method treats this localized region, e.g., the active site and its neighbors in an enzyme (called the primary subsystem, PS), by a QM method and includes the influence of the surroundings (e.g., the protein and/or solvent environment, called the secondary subsystem, or SS) at the MM level. The QM/MM energy for the entire system (ES) can be formally defined by

$$E(\mathrm{QM/MM;ES}) = E(\mathrm{QM;PS}) + E(\mathrm{MM;SS}) + E(\mathrm{QM/MM;PS|SS}) \quad (1)$$

i.e., as a summation of the energy of the PS, the energy for the SS, and the interaction energy between them.

The PS is also called the QM subsystem, and the SS is often called the MM subsystem. The coupling between PS and SS can be treated either by mechanical embedding (ME) schemes,[92,103] where the electrostatic interaction is evaluated classically at the MM level, or by electrostatic embedding schemes[103,107] that describe the electrostatic interaction between the PS and SS as one-electron operators that enter the QM Hamiltonian. The interactions other than electrostatic, e.g., stretching, bending, torsion, and van der Waals interactions, are computed at the MM level in both mechanical embedding and electrostatic embedding schemes. Including the interactions between the active center and its environment allows a more realistic description of the system, in comparison with isolated QM calculations on subsystems, which are often called model systems. QM calculations on such model systems will be called cluster QM (CQM). Although the electronic structure for the surroundings does not change during the reaction (that is why we can use MM to describe it), the presence of the surroundings may affect the electronic structure of the active center (that is why we need to take it into account).

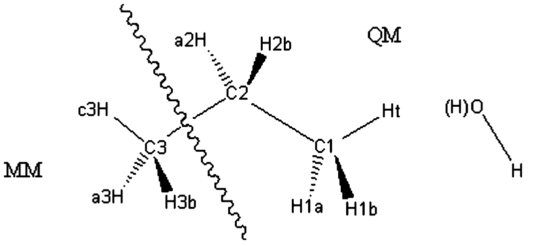


Figure 1. Reaction of OH with C_3H_8 to produce H_2O and CH_2-CH_2CH_3. The QM/MM boundary cuts through the C2−C3 bond.

It therefore appears attractive to replace full QM calculations by QM/MM calculations in generating input data points for MCMM. This scheme can be called QM/MM-MCMM, and it will be the central topic of the present contribution. In comparison, the conventional MCMM scheme that is based on a full QM calculations can be called QM-MCMM or simply MCMM. Section II presents the QM/MM-MCMM scheme in detail.

The QM/MM and MCMM methods both represent ways to add QM elements to MM. The QM/MM method incorporates QM contributions based on size extension, while MCMM (or other valence-bond theories) builds in QM contributions vertically through interconfigurational resonance. Thus, the QM/MM-MCMM is a method that includes the QM contributions both laterally and vertically.

Another way to combine QM and MM both laterally and vertically can be called MCMM/MM or (QM-MCMM)/MM, where MCMM (also called QM-MCMM) is inserted into the QM part of the QM/MM methodology rather than, as in the present paper, inserting QM/MM into MCMM. (In other words, the MCMM surface describes only the primary system in MCMM/MM but describes the entire system in QM/MM-MCMM.) In MCMM/MM, one constructs an MCMM surface for the primary system based on QM/MM calculations for the entire system or based on CQM calculations on a model system, whichever is appropriate or practical, and uses this MCMM calculation to replace full QM calculations in subsequent QM/MM calculations. The MCMM/MM method may be useful in some circumstances where the interaction between the PS and SS can be well described at the MM level.

Our new QM/MM-MCMM method will be demonstrated by calculations on two H-transfer reactions. The first reaction, R1, which is illustrated in Figure 1, is the OH radical reacting with propane at the primary carbon; this reaction has been studied previously with the QM-MCMM scheme.[16] The second reaction, R2, is the OH radical reacting with camphor, as shown in Figure 2. In particular, we will compute the

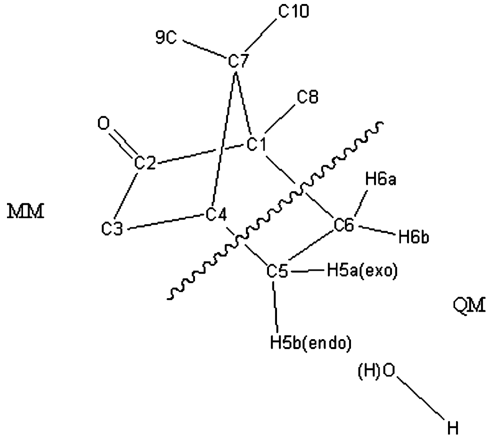


Figure 2. Reaction of OH with camphor. The QM/MM boundary cuts through the C4−C5 bond the C1−C6 bond. The hydrogen atoms are shown explicitly only in the QM region.

abstraction of the *exo*-hydrogen (H5a) at the C5 position. We note that the reaction of OH with camphor is important for atmospheric chemistry.[146−148] However, the reader should be aware that the reaction occurs along several pathways and does not occur with the greatest probability at the C5 carbon position.[147] One expects the H abstraction from a tertiary C−H bond at the C4 carbon position to have a larger rate than the abstraction from secondary and primary C−H bonds at the other carbon positions. Since the emphasis of the present study is to demonstrate the new QM/MM-MCMM methodology instead of to compare with experiment for the overall rate of reaction of OH with camphor (or to explain atmospheric chemistry), we selected the reaction at the C5 position because it is a better test for QM/MM.

The dynamics algorithm, calculations, and results will be presented in sections III, IV, and V, respectively, followed by a discussion of the results in section VI and concluding remarks in section VII.

II. The MCMM Algorithm Based on QM/MM Calculations

As explained in the Introduction, the QM/MM-MCMM method involves replacing full QM calculations by QM/MM calculations in generating input data points for MCMM. The QM-MCMM method was described in detail in ref 13. The QM/MM methodology used here and its implementation were also published previously.[143] Here, we only summarize briefly those aspects of MCMM and QM/MM related to the computations in this work, and we refer to the original references[13,143] for details.

II.A. QM-MCMM. MCMM approximates the Born−Oppenheimer potential energy at a geometry defined in internal coordinates **q** as the lowest eigenvalue of a 2 × 2 diabatic electronic Hamiltonian matrix $\mathbf{V}(\mathbf{q})$

$$\mathbf{V}(\mathbf{q}) = \begin{pmatrix} V_{11}(\mathbf{q}) & V_{12}(\mathbf{q}) \\ V_{12}(\mathbf{q}) & V_{22}(\mathbf{q}) \end{pmatrix} \tag{2}$$

where the V_{11} and V_{22} elements, as mentioned in the Introduction, are classical MM potential functions that describe the reactant and product valence bond configurations, and the V_{12} matrix element is the resonance integral. In our implementation, the reactant (or product) configuration is the reactant well (or product well) valence bond configuration (for example, the reactant van der Waals complex of OH with propane or the product van der Waals complex of propyl radical with water). The lowest eigenvalue of eq 2 is

$$V(\mathbf{q}) = \frac{1}{2}\{(V_{11}(\mathbf{q}) + V_{22}(\mathbf{q})) - [(V_{11}(\mathbf{q}) - V_{22}(\mathbf{q}))^2 + 4V_{12}(\mathbf{q})^2]^{1/2}\} \tag{3}$$

The potential energy surface $V(\mathbf{q})$, and its first and second derivatives, which are required for the dynamical calculations, are obtained by analytic differentiation of eq 3 after we know $V_{11}(\mathbf{q})$, $V_{22}(\mathbf{q})$, and $V_{12}(\mathbf{q})$. The terms associated with $V_{11}(\mathbf{q})$ and $V_{22}(\mathbf{q})$ are readily available from MM calculations, and the central problem is how to obtain the resonance integral $V_{12}(\mathbf{q})$ and its derivatives.

The construction of $V_{12}(\mathbf{q})$ is carried out in two steps. First, at a number of selected geometries $\mathbf{q}^{(k)}$, with $k = 1, 2, ..., M$, one computes the energies $V(\mathbf{q}^{(k)})$, gradients $\mathbf{g}^{(k)}$, and Hessian matrices $\mathbf{f}^{(k)}$ by electronic-structure calculations, and one evaluates $V_{12}(\mathbf{q};k)$ and its derivatives by using eq 2 inversely thereafter, i.e., evaluates $V_{12}(\mathbf{q};k)$ from $V_{11}(\mathbf{q};k)$, $V_{22}(\mathbf{q};k)$, and $V(\mathbf{q};k)$.[57] The set of M data points $\mathbf{q}^{(k)}$ are called Shepard points. Second, based on this set of M data points of $V_{12}(\mathbf{q};k)$, V_{12} at a desired geometry **q** can be evaluated by Shepard interpolation[67,68,81] as a linear combination of the quadratic expansions around these Shepard points

$$V_{12}^{S}(\mathbf{q}) = \sum_{k=1}^{M} W_k(\mathbf{q}) V_{12}^{\mathrm{mod}}(\mathbf{q};k) \tag{4}$$

where $W_k(\mathbf{q})$ are normalized weights, and $V_{12}^{\mathrm{mod}}(\mathbf{q};k)$ is a modified quadratic function, as defined in ref 13. Once the interpolation is constructed, $V_{12}(\mathbf{q})$ and its derivatives are available analytically at any desired geometry, and one can calculate $V(\mathbf{q})$ according to eq 4.

Because the MCMM representation of the potential surface is based on both electronic structure calculations and MM energies, and because the latter have different zeros of energy at reactant and product configurations, a unique energy scale must be defined across the potential energy surface. We use the same procedure here as in previous work.[166] In particular, we define the zero of energy to be the electronic structure theory energy at the MM equilibrium geometry corresponding to the reactant valence bond configuration (where the two reactants are infinitely far apart in the case of a bimolecular reaction), and we adjust the MM energies for V_{11} and V_{22} as follows

$$V_{11} = V_{11}^{\mathrm{MM}} - V_{\mathrm{o}}(\mathrm{I}) \tag{5}$$

$$V_{22} = V_{22}^{\mathrm{MM}} - V_{\mathrm{o}}(\mathrm{II}) + \Delta E \tag{6}$$

where $V_{\mathrm{o}}(\mathrm{I})$ and $V_{\mathrm{o}}(\mathrm{II})$ are MM energies at the MM equilibrium geometries corresponding to valence bond configurations I (where the two reactants are infinitely far apart in the case of a bimolecular reaction) and II (where the two products are infinitely far apart in the case of two products), and ΔE is the electronic-structure theory energy of the product relative to the reactant.

As one can see from the above procedure, the accuracy of an MCMM surface depends on the accuracy of the Shepard data points obtained by QM or QM/MM calculations and used as input. In principle the results converge to a numerically accurate interpolation of the potential energy surface with the inclusion of more Shepard points. In practical applications, one wishes to limit the computational cost by using only a small number of data points. We have proposed and tested a general scheme[14] where only one to ten electronic-structure points, including the saddle point and 0−9 nonstationary points, are needed to generate a potential energy that produces reasonably accurate rate constants in the VTST/MT calculations.[14,82] We also demonstrated that, when several Hessians were required in the original scheme, the efficiency of this scheme could be significantly improved up to a factor of 11 by employing partial electronic-structure Hessians.[16] We emphasize that this scheme might not be the best scheme for a specific reaction,[17] but it was found to be applicable to a diverse set of reactions, and it provides a start from which future refinement can be performed.

II.B. QM/MM. As mentioned in the Introduction, the QM/MM method can be viewed as a generalized electronic-structure method for large reactive systems where reaction takes place in a localized region (the primary system, PS). The PS is described at the QM level, and its environment (the secondary subsystem, SS) is treated at the MM level. The electrostatic interaction between PS and SS is either calculated classically at the MM level in the mechanical embedding schemes[92,103,144] or modeled as one-electron operators that enter the QM Hamiltonian in the electrostatic embedding schemes.[103,107,144] The interactions other than electrostatic interaction between the PS and the SS are computed at the MM level in both the mechanical embedding and electrostatic embedding schemes.

A question arises in QM/MM calculations: how should one select MM parameters for the PS atoms in calculation of the coupling between PS and SS atoms? These MM parameters are not necessarily the same for the PS atoms in the reactant and product because the atom types are changed for some atoms, e.g., a carbon atom may change from C═O type to C−O−H type. This will particularly be a concern if the boundary between the PS and the SS is very close to the active site. An extensive discussion has been given in our recent review,[144] and here we note that there is no unambiguous answer. Our suggestion is to use one set of MM parameters and to examine whether the errors introduced by using one set of parameters exceed the errors produced by other approximations that are introduced by the QM/MM framework. Although our treatment is not a perfect solution, it is very practical, and it appears to be reasonable.[143] Note that this problem occurs for QM/MM but not for MCMM. (In the present study we used the product parameters; therefore, we need parameters for a carbon radical atom type. See the Supporting Information for details.)

Another important issue in QM/MM is how to treat the dangling bonds of the PS when a QM/MM boundary cuts through covalent bonds. Various schemes have been developed to handle such a situation. Those schemes differ from each other mainly in two aspects: (i) how to treat the degrees of freedom of the nuclei of the SS atoms that are directly bond to the PS atoms and (ii) how to treat the charge distribution at the boundary. The charge distribution at the boundary is simply ignored in the mechanical embedding scheme, and its inclusion in the electrostatic embedding schemes is the major source of differences among these schemes. Our treatment[143] of the first aspect, following others,[88,89] is to saturate the dangling bond at the PS by a normal hydrogen atom (the so-called hydrogen-link atom, HL). The coordinates of HL are determined as a linear combination of the coordinates of the PS and SS boundary atoms between which the covalent bond is cut. More precisely, we place the HL atom on the Q1−M1 bond, where the Q1 and M1 denote the PS and SS boundary atoms, respectively. We scale the Q1−HL distance R(Q1-HL) with respect to the Q1−M1 distance R(Q1−M1) by a scaling factor C_{HL}, in the same way as proposed by Morokuma and co-workers[99,120]

$$R(\mathrm{Q1-HL}) = C_{\mathrm{HL}} R(\mathrm{Q1-M1}) \tag{7}$$

$$C_{\mathrm{HL}} = R_0(\mathrm{Q1-H})/R_0(\mathrm{Q1-M1}) \tag{8}$$

where R_0(Q1−H) and R_0(Q1−M1) are the MM bond distance parameters for the Q1−H and Q1−M1 stretches, respectively, in the employed MM force field. Such a treatment eliminates the extra degrees of freedom due to the artificially introduced link atoms, making dynamics calculations meaningful.

It is important to retain a qualitatively correct charge distribution near the QM/MM boundary, especially when there are MM atoms carrying substantial partial charges close to the boundary, as demonstrated in a recent publication.[143] Such a situation motivates one to use an electrostatic embedding scheme, which allows the electronic structure of the PS to respond to the presence of the charge distribution of the SS. Here we use our recently developed redistributed charge and dipole (RCD) scheme[143] as the electrostatic embedding scheme. The RCD scheme for electrostatic embedding involves distributing the M1 charge evenly onto midpoints of the M1−M2 bonds, where M2 is an SS atom that directly bonds to the M1 atom, and further modifies both the redistributed charges and the charges on the M2 atoms in order to preserve the M1−M2 bond dipoles. For example, suppose that there are n M2 atoms bonded to the M1 atom and that the charges on the M1 and M2 atoms are q_{M1} and $q_{\mathrm{M2},k}$ ($k = 1, 2, ..., n$), respectively. The initially redistributed charge q_0 is determined by

$$q_0 = q_{\mathrm{M1}/n} \tag{9}$$

and the final redistributed charge q_0^{RCD} is given by

$$q_0^{\mathrm{RCD}} = 2q_0 \tag{10}$$

and the modified M2 charges $q_{\mathrm{M2},k}^{\mathrm{RCD}}$ are

$$q_{\mathrm{M2},k}^{\mathrm{RCD}} = q_{\mathrm{M2},k} - q_0 \tag{11}$$

The QM calculations are carried out in the presence of the redistributed charges and the charges on the SS atoms.

The QM/MM gradient and Hessians are calculated by use of the chain rule, as described in refs 99 and 120.

II.C. QM/MM-MCMM. QM/MM-MCMM differs from QM-MCMM in the use of QM/MM calculations instead of QM calculations to provide energies, gradients, and Hessians at the Shepard points. A central concern in such a replacement of QM calculations by QM/MM calculations is how accurately the QM/MM description approximates the QM one or, since the QM one is itself approximate, how accurate QM/MM is in comparison to experiment. Attention should be paid to various issues. The first is how much the geometries and energies will change, e.g., at the stationary points of the potential energy surface. If used with care, QM/MM calculations may produce reasonably accurate geometric and energetic data in comparison with full QM calculations. Due to its intrinsic limitations, e.g., the prohibition of charge transfer between PS and SS, the QM/MM method is not designed to give the highest possible quantitative accuracy. It is probably more useful to stress the qualitative conclusions, and, when quantitative comparisons are made, one should focus on relative energies where the errors may cancel to some extent.

The second issue is the Hessian, which determines the vibrational frequencies and their associated eigenvectors. The QM and QM/MM Hessians usually show substantial differences, due to the different QM and MM frequencies as well as the additional approximation at the boundary. This is especially a concern for modes that involve simultaneous motions of both PS and SS atoms. This makes precise correlation of QM modes with QM/MM modes ambiguous. Visualization of the vibrations helps to some extent, but its use is limited if the system is big. Thus, a one-to-one comparison between the QM, MM, and QM/MM vibrational modes is only approximate.

The vibrational frequencies are important for determination of the zero-point vibrational energies and the vibrational partition functions. In many cases, the deviations of the QM/MM frequencies from the QM frequencies are similar for the reactant state, at the saddle point, at the product, and along the reaction path that connects these three geometries. Since it is often the relative zero-point energy rather than the absolute zero-point energy that matters, the systematic underestimation or overestimation of the frequencies in QM/MM computations can lead to at least partial cancellation of errors. The error cancellation is, however, less complete for the low-frequency modes, for which the harmonic partition functions are very sensitive to the change of frequencies. One should take this into consideration when assessing the performance of any rate calculation employing the harmonic approximation.

The eigenvector of the imaginary-frequency mode at the saddle point, which is often highly localized and is easily characterized, is used in determination of the reaction path. Usually, this imaginary-frequency mode involves motions of the PS atoms at or close to the reaction center, and its eigenvector is often very similar in the QM and QM/MM calculations.

The eigenvectors of the generalized normal mode vibrational modes along the reaction path are also important in determination of the coupling between the motion of reaction coordinate and the spectator coordinates. Such coupling plays a critical role in certain types of calculations, e.g., tunneling calculations. One needs considerably more experience testing QM/MM calculations to assess their performance for such detailed dynamical issues.

III. Dynamics

The dynamics calculations in this work were carried out in the framework of VTST/MT,[73,74,79,83−86] whose validity has been well-established.[149−151] The conventional transition state theory (TST) rate constants were determined by

$$k^{\mathrm{TST}} = (\sigma/\beta h)(Q^{\ddagger}/\Phi^{\mathrm{R}}) \exp(-\beta V^{\ddagger}) \tag{12}$$

where σ is the symmetry factor that represents the reaction path multiplicity, i.e., the number of equivalent reaction paths from reactant to product, and σ is 1 for the reactions in this work, $\beta = k_{\mathrm{B}}T$, k_{B} is Boltzmann's constant, T is the temperature, h is Planck's constant, $Q^{\ddagger}$ is the generalized partition functions for the system at the saddle point, Φ^{R} is the partition function per unit volume of the reactant, and $V^{\ddagger}$ is the classical barrier height. The generalized partition function is a product of partition functions for electronic, vibrational, and rotational degrees of freedom as well as, for bimolecular reactions, the relative translational partition function per unit volume:

$$Q(s,T) = Q_{\mathrm{el}}(s,T)Q_{\mathrm{vib}}(s,T)Q_{\mathrm{rot}}(s,T) \tag{13}$$

$$\Phi^{\mathrm{R}}(T) = \Phi^{\mathrm{R}}{}_{\mathrm{el}}(T)\Phi^{\mathrm{R}}{}_{\mathrm{vib}}(T)\Phi^{\mathrm{R}}{}_{\mathrm{rot}}(T)\Phi^{\mathrm{R}}{}_{\mathrm{rel}}(T) \tag{14}$$

Here s is the signed distance from the saddle point along the minimum energy path (MEP) in mass-scaled coordinates,[152] where $s = 0$ corresponds to the saddle point.

The canonical variational transition state (CVT) rate constants were computed by

$$k^{\mathrm{CVT}} = (\sigma/\beta h)(Q^{\mathrm{CVT}}/\Phi^{\mathrm{R}})\exp[-\beta V_{\mathrm{MEP}}(s^*)] \tag{15}$$

where Q^{CVT} is the generalized total partition function at the CVT bottleneck location s^* on the reaction path, $V_{\mathrm{MEP}}(s^*)$ is the potential energy on $V_{\mathrm{MEP}}(s)$ at s^*, and $V_{\mathrm{MEP}}(s)$ is the potential energy along the MEP. The CVT bottleneck location, s^*, is the place on the reaction path where the generalized free energy of activation is a maximum.[84]

We considered two kinds of tunneling calculations, both of which involve the vibrationally adiabatic ground-state potential energy curve, which is expressed as

$$V_{\mathrm{a}}{}^{\mathrm{G}}(s) = V_{\mathrm{MEP}}(s) + \sum_{m=1}^{3N-7} \epsilon_m^{\mathrm{G}}(s) \tag{16}$$

where N is the total number of atoms in the system, m denotes a generalized normal-mode vibration transverse to the reaction coordinate, and $\epsilon_m{}^{G}(s)$ is the zero-point energy of these generalized normal-mode vibrations at s. The maximum of $V_{\mathrm{a}}{}^{\mathrm{G}}(s)$ is called $V_{\mathrm{a}}{}^{\mathrm{AG}}$. The first kind of tunneling calculations is the small-curvature tunneling (SCT) approximation,[73,74] which calculates the transmission coefficients semiclassically including the effects of reaction-path curvature,[70,72,83,84,152−155] which enters the calculations through an effective mass[73,74] for the reaction-path motion. The second type of tunneling contribution is the large-curvature tunneling (LCT)[74,79,86] approximation, which assumes that tunneling occurs along the most direct path between the reactant and product valleys in the potential energy surface without necessarily assuming vibrational adiabaticity relative to the reaction-coordinate motion. The LCT calculations include tunneling into vibrationally excited diabatic vibrational states. In direct dynamics calculations, excited states are included to the extent that they contribute; however, in MCMM we include up to the vibrational quantum number of the highest excited state determined by a prespecified protocol. In this protocol,[16] the highest vibrational quantum number is increased successively until the tunneling contribution is converged within 1% or until the contribution from the newly added state reaches a local minimum (with respect to vibrational quantum number), whichever happens first. Such a prespecified protocol for the determination of how many final states to include in the LCT calculation has been shown[16] to be helpful in avoiding artificially large tunneling rates in MCMM dynamics due to the possibly inaccurate potential energy surface far away from the reaction swath where Shepard points are missing. In analyzing the results it is often useful to also consider the LCT(0) approximation, which includes tunneling only into the ground vibrationally diabatic state. The final best tunneling approximation is microcanonically optimized multidimensional tunneling (μOMT), which is obtained by accepting the larger of the SCT and LCT results at each tunneling energy.[86] All LCT and LCT(0) calculations in this paper were carried out with the version 4[79] algorithm.

A special note is made for the calculations of vibrational partition functions of the low-frequency modes. An accurate calculation of thermodynamic functions for these low-frequency modes is, generally speaking, very challenging. Such low-frequency modes are often associated with large-amplitude motions, and the routinely used harmonic approximation is suspicious for the evaluation of vibrational partition functions for these modes. For modes of very low frequencies (<10 cm^{-1}), the use of harmonic approximation can be qualitatively incorrect and lead to large errors. However, an accurate and effective approach to include the anharmonicity is yet to be developed. Therefore, we make a compromise in this work by use of the harmonic approximation for all vibrational modes with a cutoff frequency in the thermodynamic calculations, i.e., whenever the frequency of a mode is smaller than the prespecified (cutoff) frequency, the cutoff frequency will be used instead in calculations of vibrational partition functions. This procedure of using a cutoff frequency has been adopted[16] in our previous studies. Although one cannot expect very high accuracy from such a crude treatment, it is hoped that this procedure may provide a more realistic estimation of the vibrational partition functions than simply using the calculated harmonic frequency.

IV. Calculations

IV.A. QM Calculations. The first reaction (R1) was studied recently (as reaction R6 in ref 16) by direct dynamics and by QM-MCMM dynamics employing QM calculations at the MPW1K/6-31+G(d,p) level in generation of a potential energy surface. Briefly, the modified Perdew-Wang 1-parameter model for kinetics (MPW1K)[156,157] is a hybrid Hartree−Fock density functional theory (HF-DFT) model, whose parameters had been optimized against a selected database that consists of 20 forward barrier heights (BHs), 20 reverse BHs, and 20 energies of reaction. The one-parameter hybrid Fock-Kohn−Sham operator can be written as follows

$$F = F^{\mathrm{H}} + (X/100)F^{\mathrm{HFE}} + [1 - (X/100)](F^{\mathrm{SE}} + F^{\mathrm{GCE}}) + F^{\mathrm{C}} \tag{17}$$

where F^{H} is the Hartree operator, F^{HFE} is the Hartree−Fock exchange operator, X is the percentage of HF exchange, F^{SE} is the Dirac−Slater local density functional for exchange, F^{GCE} is the gradient corrections for the exchange functional, and F^{C} is the total correlation functional including both local and gradient-corrected parts. The MPW1K model uses the Adamo and Barone's modified Perdew-Wang 1991 exchange functional (mPW)[158] for F^{GCE} and Perdew and Wang's 1991 correlation functional (PW91)[159] for F^{C}, and set $X = 42.8$. The 6-31+G(d,p)[160] basis set is also denoted as the DIDZ basis set in this paper.

For the second reaction (R2), we used the newly developed MPWB1K[161] density function and again the 6-31+G(d,p) basis set. The MPWB1K model is very similar to the MPW1K model, but it uses the Becke95 functional[162] for F^{C}, and X was adjusted to 44. The calculations were carried out by use of the *Gaussian03* package.[163]

IV.B. MM Calculations. The MM calculations were performed by use of a locally modified TINKER program.[164] For reaction R1, the MM3 force field[20−22] was selected for both the QM/MM and MCMM calculations, and the same force field parameters as used in the previous studies[14,16] were employed, so that the results are comparable with the previous ones.

For reaction R2, the CHARMM27 force field[37] was selected for the QM/MM calculations, and the MM3 force field (modified as discussed below) was used for the MCMM calculations. This illustrates that the QM/MM calculation and the MM calculation that are combined by the MCMM scheme need not be related to one another in any special way.

Note that for the reaction R2, CHARMM27 is the same as CHARMM22. The CHARMM27 force field used for the QM/MM calculations needs to be augmented by a set of parameters for atom types (e.g., carbon radical) that are not defined in the standard CHARMM27 force field. These augmented MM parameters were set to the parameters for similar atom types, with adjustment of the point charges to maintain the neutrality of the camphor and camphor radical species. The point charges for the OH radical are obtained in a different manner, i.e., they are derived using the Merz−Singh−Kollmann[88,165] electrostatic potential (ESP) fitting procedure from electronic-structure calculations for the OH radical at the MPWB1K/DIDZ level. We assume that the final results are not especially sensitive to these parameter choices, and we are satisfied if the selected MM parameters provide a qualitatively corrected zero-order description of the system.

A detailed description of the CHARMM parameters used here is given in Table S1 of the Supporting Information. These MM parameters are validated by comparisons between MM- and QM-optimized geometries of the reactant and product. Reasonably good agreement between the MM and QM results is achieved, as demonstrated in Table S2 in the Supporting Information, indicating that the augmented MM parameters are appropriate.

As mentioned above, the MM3 force field used for reaction R2 was modified. In particular, we use the MM3[20−22] parametrization with one exception. In MM3 the van der Waals interaction is given by the Buckingham Exp-6 potential

$$V_{\mathrm{Exp-6}}(r) = \epsilon\left[Ae^{-Br/r_o} - C\left(\frac{r_o}{r}\right)^6\right] \qquad (18)$$

where r_o is the sum of the van der Waals radii, and $\in$ is the energy parameter for the interaction between two atoms. The van der Waals term in the MC-TINKER program[166] that we have used in previous work for MCMM calculations is written as a linear combination of eq 18 and an r^{-12} term that represents repulsion

$$E_{\mathrm{vdw}}(r) = \epsilon\left[Ae^{Br/r_o} - C\left(\frac{r_o}{r}\right)^6 + D\alpha\left(\frac{r_o}{r}\right)^{12}\right] \qquad (19)$$

where α is defined as

$$\alpha = \frac{V_{\mathrm{Exp-6}}(r)}{\epsilon((r_o/r))^{12}}\bigg|_{r=(1/2)r_o} \qquad (20)$$

which yields

$$\alpha = [Ae^{-1/2} - C(2.0)^6]/2.0^{12} \qquad (21)$$

The values for the unitless parameters A, B, and C are the same as in the original MM3 formulation,[20] viz. 184 000.0, 12.0, and 2.25, respectively. In the original version of MC-TINKER, D was arbitrarily set to 0.2. However, a better value is usually in the range 0.005−0.02. We found that the optimum value for D for the camphor reactions is 0.01. The optimum value is found by running a few MCMM(0) calculations with values of D in the range 0.005−0.2 and selecting the one that gives the most realistic contour map. For reaction R1, we used the original MC-TINKER value of 0.2.

IV.C. QM/MM Calculations. The QM/MM calculations were performed by use of the QMMM program,[167] which calls a QM package and an MM package to perform the required single-level calculations and then combines the single-level calculations into multilevel results. In this work, the QM package *Gaussian03*[163] and the MM program TINKER[164] are called. Geometry optimizations were carried out using the optimizer[168] in *Gaussian03* through the *external* option available in *Gausian03*.

The partitioning of the systems into PS (QM) and SS (MM) regions is shown in Figure 1 for reaction R1 and in Figure 2 for reaction R2. The QM/MM boundary cuts through one C−C bond in R1 and two C−C bonds in R2. The PSs capped by the hydrogen-link atoms in both reactions are the same, i.e., OH + C_2H_6. The mechanical embedding scheme was selected for R1, and the RCD electrostatic embedding scheme described in section II.B was chosen for R2. The MPW1K/6-31+G(d,p) and MPWB1K/6-31+G(d,p) levels were selected for R1 and R2, respectively, being the same as those in pure QM calculations. Also, the same MM parameters were employed as those used in the pure MM calculations, i.e., MM3 for R1 and CHARMM27 for R2. One motivation to use different QM methods, different MM force fields, and different embedding schemes for these two reactions is to demonstrate the generality of the QM/MM-MCMM algorithm as well as the capacity of the programs in handling a diversity of QM, MM, and QM/MM methods. Note that the QM/MM calculations do not require special/extra parametrization for the given QM level of theory and/or for the given MM force field.

IV.D. Direct Dynamics Calculations. The direct dynamics calculations for reaction R1 have been reported previously by use of the GAUSSRATE program,[169] and we reran the calculations by using a different cutoff frequency in evaluation of the thermodynamic functions. The current cutoff frequency is set to 50 cm^{-1}, and it is larger than the previous cutoff frequency (10 cm^{-1}),[16] which we think is too conservative. Due to the very high computational cost, we did not carry out direct dynamics calculations for reaction R2.

IV.E. QM-MCMM Dynamics Calculations. The QM-MCMM dynamics calculations for reaction R1 have been reported previously[16] employing a cutoff frequency of 10 cm^{-1} in evaluation of the thermodynamic functions. In the current study, we reran the calculations by using a different cutoff frequency of 50 cm^{-1}. In calculating statistical mechanical quantities such as partition functions and free energies, any frequency below the cutoff is increased to the cutoff value as a way to account, very approximately, for anharmonic effects on low-frequency vibrations. (Often, for H transfer reactions, there is one low frequency near the saddle point, although when one extends the reaction path to large $|s|$, there are others because some frequencies transform to orbital-rotational motions.)

The QM-MCMM dynamics calculations for reaction R2 were performed following the same well-established procedure[14] as that used in the calculations for reaction R1. Briefly, one begins dynamics calculations with an MCMM potential energy surface that is Shepard-interpolated on the basis of three stationary points (the reactant well, the product well, and the saddle point), and one successively adds up to 10 nonstationary points (seven in the valley of MEP and three in the concave-side region that is important for LCT computations) to improve the accuracy of the MCMM potential energy surface. Dynamics calculations based on the 3 stationary and n ($n = 0, 1, ..., 10$) nonstationary points are called MCMM−n dynamics calculations, and the MCMM-10 results are the final result that are reported in this work.

More details about how to locate the nonstationary points have been given in ref 14 and summarized in Table S1 (Supporting Information) in ref 16, and they are not repeated here.

The Page−McIver[170] method was chosen to follow the MEP in mass-scaled coordinates, which were scaled to a reduced mass μ of 1 amu. A step size of 0.001 a_0 was used for the gradient, and a new Hessian was calculated every 0.01 a_0 along the MEP. The reaction path was calculated out to $s = -3.0$ a_0 on the reactant side and to $s = 2.0$ $a0$ on the product side. We included the electronically excited $^2\Pi_{1/2}$ state of OH with an excitation energy of 140 cm^{-1} in calculating the reactant partition functions. The generalized normal-mode analyses were carried out using redundant internal coordinates; in particular, for reaction R1, we use the same redundant internal coordinates as used in our previously study,[16] and for reaction R2, we use the set of internal coordinates listed in Table S3 in the Supporting Information. The harmonic approximation was employed for all modes. A cutoff frequency of 50 cm^{-1} was selected for computing vibrational partition functions (that is, all frequencies below 50 cm^{-1} were replaced by the value of 50 cm^{-1} to simulate anharmonicity in low-frequency modes).

The QM-MCMM dynamics calculations were carried out by use of the MC-TINKERATE program.[171]

IV.F. QM/MM-MCMM Dynamics Calculations. The QM/MM-MCMM dynamics calculations for both reactions were carried out following the same procedure as the one used for the QM-MCMM dynamics calculations (see Section IV.E). The same setting was also used for following the MEP and for computing thermodynamic functions. The QM/MM-MCMM dynamics calculations were carried out by use of the MC-TINKERATE program.[171]

V. Results

The energetic and saddle-point geometric data are collected in Table 1. The QM data for reaction R1 have been reported previously, and they are tabulated here for comparison. A statistical summary of the comparison between QM and QM/MM vibrational frequencies is listed in Table 2, while the frequencies are given in Table S4 for reaction R1 and Table S5 for reaction R2 in the Supporting Information. In Table 3, we compare the locations of the first 7 nonstationary points in the QM-MCMM and QM/MM-MCMM dynamics calculations for reaction R1, the locations being indicated by the value of s, the signed distance from the saddle point along the MEP. The reaction-path curvature, κ, is illustrated as a function of s along the MEP for reaction R1 in Figure 3. The convergence of the rate constants with respect to the number of nonstationary Hessians is shown in Tables 4 and 5. After this all MCMM results shown in the article (in subsequent tables) are the final MCMM-10 ones.

Table 1. Reaction Energies, Barrier Heights, Saddle Point Geometries, and the Vibrational Imaginary Frequency at the Saddle Point for the Reactions OH + Propane (R1) and OH + Camphor (R2)[a]

	QM	QM/MM	CPS
		R1[b]	
ΔE	−12.6	−13.2	−13.1
$V^{\ddagger}$	4.9	4.7	3.9
$R_{\text{O(H)−Ht}}$	1.337	1.343	1.343
$R_{\text{C1−Ht}}$	1.195	1.191	1.192
$\theta_{\text{C1−Ht−O(H)}}$	177.2	178.2	177.7
$\omega^{\ddagger}$	1004i	958i	969i
		R2[c]	
ΔE	−16.3	−14.0	−13.1
$V^{\ddagger d}$	1.8	2.9	3.9
$R_{\text{O(H)−H5a}}{}^{d}$	1.392	1.337	1.347
$R_{\text{C5−H5a}}{}^{d}$	1.173	1.198	1.190
$\theta_{\text{C5−H5a−O(H)}}{}^{d}$	172.6	167.8	178.1
$\omega^{\ddagger d}$	538i	966i	861i

[a] Energy in kcal/mol, distance in Å, angles in degree, and frequency in cm^{-1}. The capped primary system (CPS) is OH + C_2H_6 for both reactions. [b] Reaction takes place at the C1 position. The MPW1K/DIDZ level of theory is used in QM calculations, and the MM3 force field is used for MM calculations. See Figure 1 for the QM/MM boundary setup. The mechanical embedding scheme was used in QM/MM computations. [c] The MPWB1K/DIDZ level of theory in QM calculations and CHARMM27 force field for MM calculations. See Table S1 in the Supporting Information for augmented MM parameters. See Figure 2 for the QM/MM boundary setup. The electrostatic embedding (EE) scheme redistributed charge and dipole (RCD) was used in QM/MM computations. [d] Abstraction of *exo* H5a.

The rate constants for reactions R1 and R2 are presented in Tables 6 and 7, respectively; Table 8 shows the percentage error for the rate constants of R1 calculated by the QM-MCMM dynamics and by the QM/MM-MCMM dynamics with respect to those calculated by direct dynamics. Listed in Table 9 are the percentage errors for the rate constants of R2 calculated by QM/MM-MCMM dynamics with respect to QM-MCMM. In general, a percentage deviation is calculated as

$$\text{PD} = \frac{y^{\text{Cal}} - y^{\text{Std}}}{y^{\text{Std}}} \times 100\% \tag{22}$$

where y^{Cal} is the quantity to be examined (e.g., rate constants calculated by MCMM dynamics), and y^{Std} is the standard result (e.g., rate constants calculated by direct dynamics) that the calculation tries to reproduce.

For an assessment of the overall performance of the calculations, we calculated the mean signed deviation (MSD) and mean unsigned deviation (MUD) for a set of quantities as follows

$$\text{MSD} = \frac{1}{N}\sum_{i=1}^{N}(y_i^{\text{Cal}} - y_i^{\text{Std}}) \tag{23}$$

$$\text{MUD} = \frac{1}{N}\sum_{i=1}^{N}|y_i^{\text{Cal}} - y_i^{\text{Std}}| \tag{24}$$

where y_i^{Cal} is the quantity to be examined, y_i^{Std} is the standard result that the calculation tries to reproduce, and N is the number of quantities for which the comparisons are made.

Table 2. Statistic Summary of the Deviations between the QM and the QM/MM Vibrational Frequencies for the Reactions OH + Propane (R1) and OH + Camphor (R2)[a]

	reactant	product	saddle point	
		R1[b]		
num. of modes				
total	27	24		32
$\|\Delta\nu\| > 100\ \mathrm{cm^{-1}}$	3	3		3
$\|\Delta\nu\| < 10\ \mathrm{cm^{-1}}$	11	5		14
MSD[c] [cm^{-1}]	−32	−35		−26
MUD[d] [cm^{-1}]	39	50		34
$\|\Delta\nu\|_{\max}$ [cm^{-1}]	244	247		249
		R2[e]		
num. of modes				
total	75	72	80[f]	
$\|\Delta\nu\| > 100\ \mathrm{cm^{-1}}$	13	13	13[d]	13[e]
$\|\Delta\nu\| < 10\ \mathrm{cm^{-1}}$	11	9	8[d]	14[e]
MSD[c] [cm^{-1}]	−61	−62	−62[d]	−61[e]
MUD[d] [cm^{-1}]	67	72	71[d]	71[e]
$\|\Delta\nu\|_{\max}$ [cm^{-1}]	332	374	352[d]	360[e]

[a] Frequency in cm^{-1}. The deviation is calculated as $\Delta\nu = (\nu_{\mathrm{QM/MM}} - \nu_{\mathrm{QM}})$, where the modes are arranged in order of decreasing magnitude. The OH and H_2O species are always treated at the pure QM level, and they are not included in the comparisons. The imaginary-frequency mode has been shown in Table 1, and it is excluded from the comparison in this table. [b] Reaction takes place at the C1 position. The MPW1K/DIDZ level of theory is used in QM calculations, and the MM3 force field is used for MM calculations. See Figure 1 for the QM/MM boundary setup. The mechanical embedding (ME) scheme was used in QM/MM computations. [c] Defined by eq 21. [d] Defined by eq 22. [e] The MPWB1K/DIDZ level of theory in QM calculations and CHARMM27 force field for MM calculations. See Table S1 in the Supporting Information for augmented MM parameters. See Figure 2 for the QM/MM boundary setup. The electrostatic embedding (EE) scheme redistributed charge and dipole (RCD) was used in QM/MM computations. [f] Abstraction of *exo* H5a.

The mean signed percentage deviation (MSPD) and mean unsigned percentage deviation (MUPD) are defined in a manner similar to the MSD and MUD

$$\mathrm{MSPD} = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{y_i^{\mathrm{Cal}} - y_i^{\mathrm{Std}}}{y_i^{\mathrm{Std}}}\right) \times 100\% \qquad (25)$$

$$\mathrm{MUPD} = \left(\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i^{\mathrm{Cal}} - y_i^{\mathrm{Std}}}{y_i^{\mathrm{Std}}}\right|\right) \times 100\% \qquad (26)$$

Obviously, the MSPD and MUPD do not give an even-handed representation of the cases in which the quantities are underestimated (those cases are limited to a percentage error of 100%); however, they are still very instructive.

VI. Discussion

As can be seen Table 1, the QM/MM computations reproduce the QM barrier heights and reaction energies reasonably well, and the QM/MM results outperform the capped primary system (CPS) studies at the same QM level considerably. For reaction R1, the QM/MM barrier height is only 0.2 kcal/mol lower than the QM barrier height. The good agreement in energetics leads to excellent agreement of the MEPs calculated by the QM/MM-MCMM and QM-MCMM schemes, as illustrated in Figure 3(A). For reaction R2, the difference between the QM/MM and QM barrier heights is larger (about 1 kcal/mol) than that for reaction R1.

Table 3. Comparison of the Locations of the Shepard Points in QM-MCMM and in QM/MM-MCMM Dynamics Calculations for the Reaction OH + Propane (R1)[a]

	QM-MCMM		QM/MM-MCMM	
*n*th point	*s*	*E*	*s*	*E*
0[b]	0.00	4.91	0.00	4.72
1	−0.51	4.26	−0.52	4.07
2	0.19	4.09	0.19	3.93
3	−1.22	3.08	−1.21	2.95
4	0.29	2.48	0.29	2.36
5	−2.11	1.85	−2.08	1.77
6	−0.45	4.29	−0.45	4.13
7	0.53	−3.49	0.53	−3.72
8H	−2.54	N/a	−2.51	N/a
8L	0.51	N/a	0.51	N/a
8	N/a	7.75	N/a	7.54
9H	−2.79	N/a	−2.77	N/a
9L	0.96	N/a	0.97	N/a
9	N/a	6.15	N/a	6.07
10H	−2.79	N/a	−2.77	N/a
10L	0.96	N/a	0.97	N/a
10	N/a	2.56	N/a	2.17

[a] Only Shepard points for which QM or QM/MM calculations are needed are listed, thus excluding the reactant well and the product well. Energy in kcal/mol, and *s* in bohr. The scheme that was presented in ref 14 and summarized in Table S1 in ref 16 is adopted for determination of the locations of the Shepard points. See the cited references for details. The *n*th Shepard point is determined on the basis of the MCMM-($n-1$) potential energy surface, and, therefore, the *s* value listed for the *n*th point is corresponding to the MEP in the MCMM-($n-1$) potential energy surface. Reaction take place at the C1 position. The MPW1K/DIDZ level of theory in QM calculations and MM3 force field for MM calculations. See Figure 1 for the QM/MM boundary setup. The mechanical embedding scheme was used in QM/MM computations. [b] The saddle point.

The QM geometries for the saddle points are also reproduced well by the QM/MM computations. The bond distances for the breaking/forming bonds generally agree within 0.02 Å between the QM and QM/MM results. The only exception is for reaction R2, where the distance between the O atom in the OH radical and the H atom being transferred has been shortened by 0.04 Å for the *exo* H5a atom in QM/MM computations.

Turning to the vibrational frequencies, one finds in Table 2 that there are usually several tens of wavenumbers deviations on average between the QM/MM and QM calculations, as indicated by the mean unsigned deviation (MUD). However, it is the mean signed deviation (MSD) that is more important in determination of the zero point energy, because the difference in the errors of the transition state and reactant zero point energies has a large effect on dynamics calculations if vibrational frequencies are not scaled in the dynamics step. The MSD is −32 cm^{-1} for the reactant, −26 cm^{-1} for the saddle point, and −35 cm^{-1} for the product in reaction R1. For reaction R2, the MSDs are close to −60 cm^{-1}, with a variation as small as 1 cm^{-1}. As noted previously in section II.C, the same trend of over- or

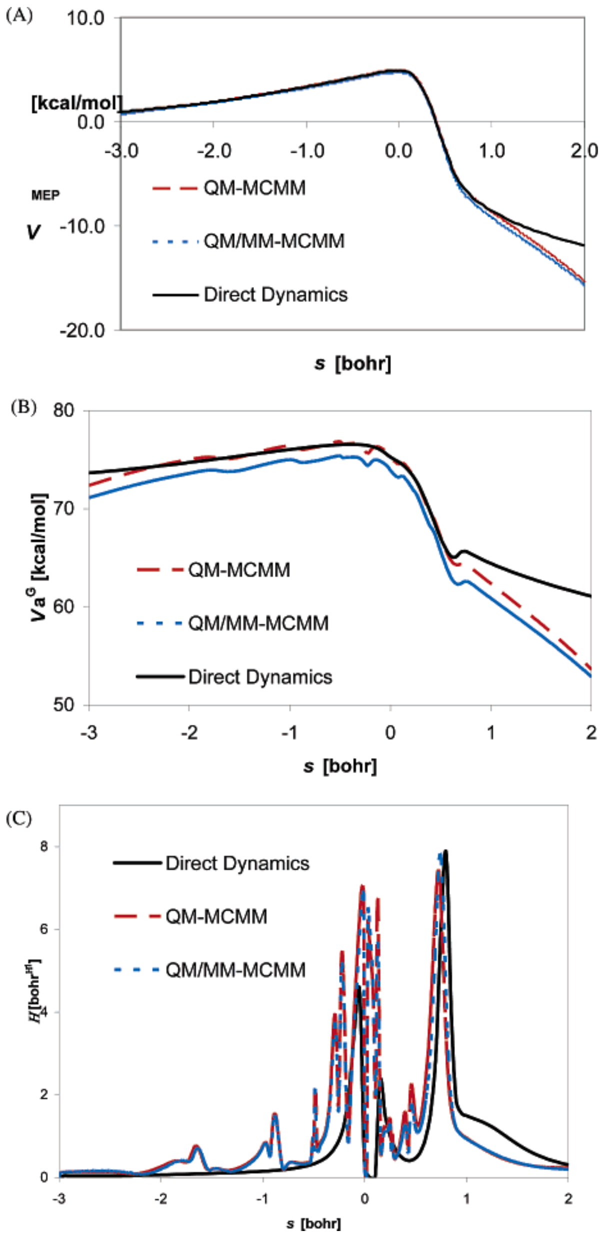


Figure 3. (A) The classical potential energy curve, V_{MEP}, (B) the vibrational adiabatic ground-state potential energy curve, V_a^G, and (C) the reaction-path curvature reaction, κ, as functions of reaction coordinate *s* for the reaction OH with propane (R1).

Table 4. Convergence of TST, CVT, CVT/SCT, and CVT/μOMT with Respect to the Number of Nonstationary Hessians for QM/MM-MCMM-*X* ($X = 2, 4, 6, 8, 10$) at $T = 300$ and 600 K for the R1 Reaction OH + Propane[a]

T (K)	TST	CVT	CVT/SCT	CVT/μOMT
		MCMM-0		
300	0.0	108.8	1346.3	1334.4
600	0.0	40.0	153.2	151.2
		MCMM-2		
300	0.0	−34.3	36.8	35.7
600	0.0	−25.0	−0.8	−1.6
		MCMM-4		
300	0.0	−33.3	−26.4	−27.0
600	0.0	−24.2	−12.1	−12.8
		MCMM-6		
300	0.0	0.7	0.8	0.0
600	0.0	2.4	2.4	1.6
		MCMM-8		
300	0.0	0.0	0.0	−0.8
600	0.0	0.8	0.8	0.0
		MCMM-10		
300	0.0	0.0	0.0	0.0
600	0.0	0.0	0.0	0.0

[a] Convergence is indicated by $(k_{QM/MM-MCMM-X} - k_{QM/MM-MCMM-10})/k_{QM/MM-MCMM-10} \times 100$.

Table 5. Convergence of TST, CVT, CVT/SCT, and CVT/μOMT with Respect to the Number of Nonstationary Hessians for QM/MM-MCMM-*X* ($X = 2, 4, 6, 8, 10$) at $T = 300$ and 600 K for the R2 Reaction OH + Camphor[a]

T (K)	TST	CVT	CVT/SCT	CVT/μOMT
		MCMM-0		
300	0.0	26.5	26.8	26.8
600	0.0	−8.8	−9.8	−9.8
		MCMM-2		
300	0.0	3.2	−19.5	−19.5
600	0.0	2.9	−3.3	−3.3
		MCMM-4		
300	0.0	4.5	11.0	11.0
600	0.0	3.6	5.9	5.9
		MCMM-6		
300	0.0	0.0	−1.3	−1.3
600	0.0	0.0	0.0	0.0
		MCMM-8		
300	0.0	0.0	1.1	1.4
600	0.0	0.0	0.7	0.7
		MCMM-10		
300	0.0	0.0	0.0	0.0
600	0.0	0.0	0.0	0.0

[a] Convergence is indicated by $(k_{QM/MM-MCMM-X} - k_{QM/MM-MCMM-10})/k_{QM/MM-MCMM-10} \times 100$.

underestimation of a vibrational frequency along the reaction path leads to significant cancellation of error in evaluating the relative free energy profile and vibrational partition functions. This rationalizes why QM/MM-MCMM gives reasonable rate constants compared with QM-MCMM. As a result, the vibrationally adiabatic ground-state potential energy curve, V_a^G, looks very similar in the QM/MM-MCMM and the QM-MCMM dynamics calculations, as demonstrated in Figure 3(B) for reaction R1.

The coupling between the motion along the reaction coordinate and the spectator coordinates is very similar in the QM/MM-MCMM and QM-MCMM dynamics calculations. Taking reaction R1 as an example, the reaction-path curvature, which is an indication of such coupling, shows very small differences between QM/MM-MCMM and QM-MCMM dynamics calculations, as depicted in Figure 3(C). The underlying mechanism is that usually the coupling is significant only for those atoms that are close to the atoms directly involved in bond breaking and bond forming, and often those atoms are included in the PS, which is treated at the QM level. Thus, QM/MM-MCMM dynamics essentially

Table 6. Rate Constants Calculated by the Direct Dynamics, by the QM-MCMM Dynamics, and by the QM/MM-MCMM Dynamics for the Reaction OH + Propane (Reaction R1)[a]

T (K)	TST	CVT	CVT/SCT	CVT/LCT(0)	CVT/LCT	CVT/μOMT
			Direct Dynamics			
200	6.55E−16	2.89E−17	5.59E−17	4.64E−17	4.77E−17	5.59E−17
300	1.12E−14	1.46E−15	1.92E−15	1.77E−15	1.80E−15	1.92E−15
400	5.43E−14	1.19E−14	1.37E−14	1.31E−14	1.32E−14	1.37E−14
500	1.58E−13	4.60E−14	4.97E−14	4.84E−14	4.86E−14	4.97E−14
600	3.53E−13	1.22E−13	1.28E−13	1.25E−13	1.26E−13	1.28E−13
1000	2.73E−12	1.23E−12	1.24E−12	1.23E−12	1.23E−12	1.24E−12
			QM-MCMM			
200	6.55E−16	1.54E−17	5.25E−17	4.12E−17	4.35E−17	5.34E−17
300	1.12E−14	9.82E−16	1.74E−15	1.56E−15	1.61E−15	1.76E−15
400	5.43E−14	8.99E−15	1.25E−14	1.17E−14	1.20E−14	1.26E−14
500	1.58E−13	3.74E−14	4.64E−14	4.45E−14	4.51E−14	4.67E−14
600	3.53E−13	1.04E−13	1.21E−13	1.18E−13	1.19E−13	1.22E−13
1000	2.73E−12	1.11E−12	1.04E−12	1.02E−12	1.03E−12	1.04E−12
			QM/MM-MCMM			
200	9.71E−15	2.82E−17	8.15E−17	6.49E−17	6.78E−17	8.26E−17
300	1.46E−13	1.48E−15	2.42E−15	2.18E−15	2.23E−15	2.44E−15
400	6.60E−13	1.24E−14	1.63E−14	1.54E−14	1.57E−14	1.64E−14
500	1.84E−12	4.82E−14	4.94E−14	4.76E−14	4.80E−14	4.96E−14
600	3.99E−12	1.25E−13	1.24E−13	1.21E−13	1.22E−13	1.25E−13
1000	2.90E−11	1.22E−12	1.17E−12	1.16E−12	1.16E−12	1.17E−12

[a] Reaction takes place at the C1 position. The MPW1K/DIDZ level of theory in QM calculations and MM3 force field for MM calculations. See Figure 1 for the QM/MM boundary setup. The mechanical embedding (ME) scheme was used in QM/MM computations. The MCMM-10 results are listed for MCMM dynamics calculations. The rate constant including tunneling is then given by $k^{\text{CVT/MT}} = \kappa^{\text{MT}} k^{\text{CVT}}$, where κ^{MT} is the transmission coefficient, and MT is SCT, LCT(0), LCT, or μOMT. The definitions of these abbreviations are given in footnote *a* of Table 8.

Table 7. Rate Constants Calculated by the QM-MCMM Dynamics and by the QM/MM-MCMM Dynamics for the Reaction OH + Camphor (Reaction R2)[a,b]

T (K)	TST	CVT	CVT/SCT	CVT/LCG0	CVT/LCG	CVT/μOMT
			QM-MCMM			
200	1.29E−13	3.92E−15	9.28E−15	8.26E−15	8.64E−15	9.30E−15
300	2.22E−13	2.52E−14	3.81E−14	3.59E−14	3.68E−14	3.81E−14
400	3.39E−13	7.42E−14	9.41E−14	9.08E−14	9.22E−14	9.42E−14
500	4.89E−13	1.57E−13	1.83E−13	1.79E−13	1.81E−13	1.83E−13
600	6.82E−13	2.78E−13	3.10E−13	3.05E−13	3.07E−13	3.10E−13
1000	2.01E−12	1.08E−12	7.65E−13	7.60E−13	7.62E−13	7.65E−13
			QM/MM-MCMM			
200	3.53E−14	1.95E−16	5.38E−16	4.35E−16	1.14E−15	1.16E−15
300	1.17E−13	4.00E−15	6.30E−15	5.71E−15	1.01E−14	1.02E−14
400	2.52E−13	2.08E−14	2.69E−14	2.54E−14	3.66E−14	3.67E−14
500	4.51E−13	6.19E−14	7.27E−14	7.01E−14	9.00E−14	9.03E−14
600	7.27E−13	1.37E−13	1.53E−13	1.50E−13	1.79E−13	1.79E−13
1000	2.93E−12	9.47E−13	9.85E−13	9.76E−13	1.05E−12	1.05E−12

[a] The MPWB1K/DIDZ level of theory in QM calculations and CHARMM27 force field for MM calculations. See Table S1 in the Supporting Information for augmented MM parameters. See Figure 2 for the QM/MM boundary setup. The electrostatic embedding (EE) scheme redistributed charge and dipole (RCD) was used in QM/MM computations. The MCMM-10 results are listed. The rate constant including tunneling is then given by $k^{\text{CVT/MT}} = \kappa^{\text{MT}} k^{\text{CVT}}$, where κ^{MT} is the transmission coefficient, and MT is SCT, LCT(0), LCT, or μOMT, which are defined in footnote *a* of Table 8. [b] Abstraction of *exo* H5a.

retains the feature of reaction-path curvature of the QM-MCMM dynamics. This also implies that, under certain circumstances, one could use effective reduced-dimensional resonance matrix element, V_{12}^{eff}, to replace the full-dimensional resonance matrix element, V_{12}, in the MCMM expression eq 2, without losing much accuracy in MCMM dynamics calculations.

Tables 4 and 5 are the main results of the present paper. They show that the rate constants converge well with respect to the number of Hessians at nonstationary points. In particular, all MCMM-6 and MCMM-8 calculations in the two tables agree with the MCMM-10 calculations within 2%. The noticeably better convergence for reaction R2 as compared to R1 comes primarily from a more realistic description of the van der Waals energy for R2 in the repulsion region (see section IV.B). In another article, we will show that by using appropriate MM functions, more accurate MCMM surfaces can be generated with a fewer number of Hessians.[82]

If QM/MM provides good agreement with QM for barrier heights, vibrations, energies of reaction, and barrier heights,

Table 8. Percentage Deviations of Reaction Rates in QM-MCMM and QM/MM-MCMM Dynamics (with Respect to Direct Dynamics) for the Reaction of OH + Propane (R1)[a]

	T (K)	TST	CVT	CVT/ SCT	CVT/ LCT(0)	CVT/ LCT	CVT/ μOMT
			QM-MCMM				
	200	0	−47	−6	−11	−9	−4
	300	0	−33	−9	−12	−11	−8
	400	0	−24	−9	−11	−9	−8
	500	0	−19	−7	−8	−7	−6
	600	0	−15	−5	−6	−6	−5
	1000	0	−10	−16	−17	−16	−16
MSPE[b]		0	−25	−9	−11	−10	−8
MUPE[c]		0	25	9	11	10	8
			QM/MM-MCMM				
	200	1382	−2	46	40	42	48
	300	1204	1	26	23	24	27
	400	1115	4	19	18	19	20
	500	1065	5	−1	−2	−1	0
	600	1030	2	−3	−3	−3	−2
	1000	962	−1	−6	−6	−6	−6
MSPE[b]		1126	2	14	12	12	14
MUPE[b]		1126	3	17	15	16	17

[a] Reaction takes place at the C1 position. The MPW1K/DIDZ level of theory in QM calculations and MM3 force field for MM calculations. See Figure 1 for the QM/MM boundary setup. The mechanical embedding scheme was used in QM/MM computations. The MCMM-10 results are listed for MCMM dynamics calculations. The rate constant including tunneling is then given by $k^{\mathrm{CVT/MT}} = \kappa^{\mathrm{MT}} k^{\mathrm{CVT}}$, where κ^{MT} is the transmission coefficient, and MT is SCT, LCT(0), LCT, or μOMT. The transmission coefficients are defined as follows: SCT, small-curvature tunneling; LCT(0), large-curvature tunneling only considering unexcited diabatic tunneling states; LCT, full large-curvature tunneling; μOMT, microcanonically optimized multidimensional tunneling. [b] MeanSigned Percentage Deviation (MSPD) averaged over six temperatures. [c] MeanUnsigned Percentage Deviation (MUPD) averaged over six temperatures.

Table 9. Percentage Deviations of Reaction Rates in QM/MM-MCMM Dynamics (with Respect to QM-MCMM Dynamics) for the Reaction of OH + Camphor (Reaction R2)[a,b]

	T (K)	TST	CVT	CVT/ SCT	CVT/ LCT(0)	CVT/ LCT	CVT/ μOMT
	200	−73	−95	−94	−95	−87	−88
	300	−47	−84	−83	−84	−73	−73
	400	−26	−72	−71	−72	−60	−61
	500	−8	−61	−60	−61	−50	−51
	600	7	−51	−51	−51	−42	−42
	1000	46	−12	29	28	38	37
MSPE[c]		−17	−62	−55	−56	−46	−46
MUPE[d]		34	62	65	65	58	59

[a] The MPWB1K/DIDZ level of theory in QM calculations and CHARMM27 force field for MM calculations. See Table S1 in the Supporting Information for augmented MM parameters. See Figure 2 for the QM/MM boundary setup. The electrostatic embedding (EE) scheme redistributed charge and dipole (RCD) was used in QM/MM computations. The MCMM-10 results are listed. The rate constant including tunneling is then given by $k^{\mathrm{CVT/MT}} = \kappa^{\mathrm{MT}} k^{\mathrm{CVT}}$, where κ^{MT} is the transmission coefficient, and MT is SCT, LCT(0), LCT, or μOMT. The definitions of the transmission coefficients are given in footnote *a* of Table 8. [b] Abstraction of the *exo* H5a. [c] MeanSigned Percentage Deviation (MSPD) averaged over six temperatures. [d] MeanUnsigned Percentage Deviation (MUPD) averaged over six temperatures.

QM/MM-MCMM dynamics should yield rate constants comparable with QM-MCMM dynamics calculations. The R1 reaction in the present study is a nice example that confirms this expectation: the QM and QM/MM barrier heights differ slightly (by 0.2 kcal/mol), and the QM and QM/MM rate constants agree with each other generally within 25%, as can be seen from Tables 4 and 6. Here, we should point out that it is more meaningful to compare the rate constants with tunneling contributions, e.g., CVT/SCT rate constants, than the rate constants determined by just one point at the reaction path, e.g., the CVT rate constants, because the nontunneling calculations are more vulnerable to the numerical noise in the computations; the calculation of the tunneling contributions involves integration over the reaction path, and this reduces the numerical noises significantly (since performing an integral is like performing an average). The barrier heights for the pathway of reaction R2 that we studied show larger discrepancies (about 1 kcal/mol) in QM and QM/MM computations, in comparison with the R1 case. The bigger errors in the R2 barrier height are not surprising, since the QM/MM boundary in R2 is closer to the C atom that is directly involved in bond-breaking, and one therefore expects more pronounced boundary effects due to the mismatch between the QM and MM energy approximations. Generally speaking, a 1 kcal/mol difference from QM calculations is quite reasonable for QM/MM calculations. Unfortunately, though, a difference as small as 1 kcal/mol in energy can produce a large difference in rates, the effect being more pronounced at low temperatures, as demonstrated in Tables 5 and 7. Reactions R1 and R2 serve as examples at two extremes to illustrate how the accuracy of QM/MM energetic information affects the accuracy of QM/MM-MCMM dynamics.

VII. Concluding Remarks

In this paper, we introduced a new MCMM algorithm, which is called QM/MM-MCMM, for studies of very large reactive systems. The QM/MM-MCMM method uses QM/MM calculations to generate the resonance matrix element V_{12} of MCMM calculations, and it is tested for two hydrogen-transfer reactions by comparing QM/MM-MCMM calculations to MCMM calculations in which the resonance matrix element is generated by full QM calculations, which is called QM-MCMM, and also, for one reaction, by comparing to full QM direct dynamics calculations where the potential energies are computed quantum mechanically on the fly. Very encouraging results are obtained for rate constants including tunneling contributions, when the QM/MMM method adequately reproduces the barrier height of the QM calculations, thereby validating the QM/MM-MCMM method as a very general and efficient procedure for generating potential energy surfaces for large reactive systems. As full QM calculations are usually not possible for very large systems such as proteins or large nanoparticle catalysts, QM/MM-MCMM is an alternative practical way to QM-MCM in study of these reactive systems. However, as illustrated by the cautionary example of reaction R2, one will not get as reliable results if the QM/MM calculations introduce an error in the barrier height.

The resonance matrix element V_{12} plays the central role in the MCMM method and in other methods based on valence bond theory. An advantage of our MCMM algorithm, as discussed in the Introduction, is that the full dimensionality of the resonance matrix element V_{12} is included. Does one really need a full-dimensional resonance integral in all cases? For a large system, a full-dimensional resonance integral might cause technical problems in handling (computing and storing) very large Hessian matrices. Could one replace V_{12} by a reduced-dimensional effective resonance integral in the MCMM expression eq 2 without losing much accuracy in the MCMM dynamics calculations? If yes, how can we construct the effective reduced-dimensional resonance integral?

It seems that, in many cases, significant coupling between the motion along the reaction coordinate and other motions is present only for atoms that are within a small distance from the active center. That is, dynamical coupling seems rather localized near to the bond-breaking and bond-forming atoms. This justifies the use of a reduced-dimensional resonance integral, provided that the dynamical coupling is adequately described. Please note that even though the reduced-dimensional resonance integral is used, the MCMM potential is still full dimensional. In other words, one approximates the full-dimensional MCMM potential based on the full-dimensional resonance integral by a full-dimensional MCMM based on a reduced-dimensional resonance integral.

Now we consider the second question: how to construct the effective reduced-dimensional resonance integral? The most simplified possible reduced-dimensional effective resonance integral is a constant. Although for some reactions, the resonance matrix element is almost constant along the MEP, we found that it must sometimes vary appreciably off the MEP in order to give reasonable frequencies.[13] Therefore, the assumption that the resonance matrix element is constant in the reaction swath is probably oversimplified for many cases.

By fitting the resonance integral to a simple analytic function, like an exponential function or Gaussian, one increases the flexibility of a reduced-dimensionality scheme, and by selecting the parameters carefully, one might be able to reproduce some desired properties such as the experimental barrier height or theoretically calculated vibrational frequencies at the saddle point. The problem of this kind of treatment is that it is difficult to find an analytical function that is generally applicable for different reactions and systematically improvable by including more and more kinetic and electronic-structure information.

In comparison with simple analytic functions, Shepard interpolation is a general and systematically improvable algorithm for constructing the resonance surface. Therefore a way to construct a reduced-dimensional effective resonance integral is by Shepard interpolation based on QM or QM/MM calculations in which the resonance integral explicitly depends on only a subset of the atoms, e.g., the atoms in the model system (the atoms in the active center and possibly some nearby spectator atoms). We will focus our discussion on this point in the remaining text.

The most direct treatment for reducing the dimensionality of the resonance integral is to perform cluster QM calculations for a model system. As an example, we note that Sierka and Sauer[123] explored a idea of combining a method they call QM-Pot with the empirical valence-bond theory in studying proton-transfer reactions in zeolite catalysts; QM-Pot,[101] which stands for the "combined quantum mechanics-interatomic potential functions approach", is a mechanical-embedding QM/MM algorithm in the language we use, but they prefer to reserve "molecular mechanics" for traditional organic and biological force fields. In ref 123, the authors started by doing QM calculations on a model system, then parametrized these with an EVB potential function, and proceeded to generate an initial guess (geometry and Hessian) for minima and saddle points for the entire system based on the parametrized EVB potential for the model system and the interatomic potential between the model system and its surroundings; finally they optimized those stationary points (the minima and the saddle point) employing the QM-Pot approach. Their way of combining the EVB for a model system and the interatomic potentials between the model system and its environment can be regarded as Cluster-QM-MCMM, or CQM-MCMM, because the resonance matrix element V_{12} is derived from a cluster model for the active site, which differs from our QM/MM-MCMM treatment where V_{12} is derived from QM/MM calculations for the entire system.

Cluster QM calculations are much less expensive, and they can produce good results in many cases if used with care. However, a key problem of the cluster QM calculations is whether the model system is a faithful model for the active center of the entire system as the electronic structure changes during the reaction. Although the dynamical coupling seems rather localized near to the bond-breaking and bond-forming atoms, the environment (e.g., the solvent) may perturb the electronic structure of the atoms in the active center by imposing geometric constraints and/or by polarization. Such effects can be critical in some cases, e.g. when the model system contains transition metals and has more than one energetically low-lying electronic state.[172] A model system in the gas phase might not be a wise choice in such a situation, and QM/MM calculations, in particular electrostatic-embedding QM/MM calculations, are recommended, if full QM calculations for the entire system are not feasible. If a model system does not provide a good representation of the electronic structure of the reaction center, we doubt that one can get a correct potential surface for the entire system by improving only the diagonal terms V_{11} and V_{22}, with the resonance integral based on that inappropriate model system. In particular, we question whether it is appropriate to use a resonance integral calculated in the gas phase for condensed-phase systems. A detailed and systematic examination of this issue would be worthwhile.

Then how should one get the effective reduced-dimensional resonance integral from full-dimensional QM or QM/MM calculations? One possible way is to treat explicitly the degrees of freedom of the atoms in the model system while considering the environment as a bath. One might optimize the geometry of the environment for a given geometry of the active center, i.e., force the environment to follow the reaction in the active site adiabatically, and evaluate the resonance integral accordingly. A more elaborate scheme would be to perform an ensemble average and use the averaged resonance integral as the effective resonance integral. After doing that, one simply retains in the resonance integral the degrees of freedom of the atoms in the active site. These schemes are surely more complicated than the scheme based on cluster QM calculations, but they have the reward that the resonance integral is obtained, instead of by a gas-phase model, by a more realistic model that it is probably more accurate.

In the Introduction we compared the present QM/MM-MCMM approach to another possible approach, called MCMM/MM. Whereas the former corresponds to putting QM/MM into MCMM, the latter corresponds to putting MCMM into QM/MM. We pointed out that MCMM/MM would not contain the effects of the SS in the quantum mechanical description of the PS, that is, it would correspond to mechanical embedding, whereas the present approach allows one to use, as is done in the present applications, either a mechanical embedding QM/MM method or an electrostatic embedding QM/MM method, which provides a more realistic description. (We demonstrated both mechanical and electrostatic embedding schemes in this study.) In addition to this question of the interaction of the subsystems, there are also computational considerations involved in the strategy we have chosen, as we discuss next.

First, MCMM/MM requires QM calculations on an isolated model system (as in the cluster-QM/MCMM that we discussed above), which are much simpler than doing QM/MM computation for the entire system. For example, finding the QM/MM transition-state structure for the entire system is in general much harder than finding the QM transition-state structure for the isolated model system.

Second, we note that MCMM/MM does not need to handle large Hessian matrices when constructing V_{12}, while in QM/MM-MCMM, unless one uses a reduced-dimensional representation for V_{12}, it is necessary to manipulate a full-dimensional Hessian matrix, whose size increases as N^2 (N is the number of the atoms). Not only is such a big Hessian matrix expensive in calculations but also it presents a problem in storage. Moreover, most electronic-structure programs provide analytic derivatives for the PS atoms but not for the SS atoms, which are treated as background point charges in an electrostatic embedding QM/MM computation. Thus, MCMM/MM is less computationally demanding for large-size systems.

Finally, we add a note on MCMM parametrization. It is always a great advantage if the effort required for parametrization is limited. When using MCMM/MM we actually implicitly assume that the MCMM potential is transferable between different conformations of the entire system. (This may be also true for QM/MM-MCMM employing a reduced-dimensional effective V_{12}.) A greater degree of transferability could be achieved by regarding the PS as a reacting functional group, which is present in various parts of a system or in different systems, just like a residue in proteins interacting with the solvent water molecules. Thus, one might develop a set of MCMM potentials for selected reactions and use them as reactive MM potentials for standard MD simulations, e.g., to simulate the hydrogen exchange between water molecules and a solvated protein, the exchange taking place for various residues at the same time. Here, one would assume that the reactions are taking place at the same time at different locations without interference between each other, which is equivalent to approximating a large (valence-bond theory) matrix by several sub-blocks neglecting the off-diagonal elements among the sub-blocks. This kind of transferability would permit MCMM/MM and/or QM/MM-MCMM with a reduced-dimensional V_{12} to be treated as generalized MM potentials for simulating reactive biomolecules.

Acknowledgment. The authors are grateful to Jerry Mohrig for valuable discussions. This work is supported in part by the National Science Foundation and the Office of Naval Research.

Supporting Information Available: Detailed description of the MM parameters used for reaction R2, the optimized geometries for the reactant and products for reaction R2 at the QM, MM, and QM/MM levels of theory, the redundant internal coordinates used for generalized normal-mode analyses for reaction R2, and the vibrational frequencies that are computed at the QM and QM/MM levels for reactions R1 and R2. This material is available free of charge via the Internet at http://pubs.acs.org.

References

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

(2) Warshel, A.; Karplus, M. *Chem. Phys. Lett.* **1975**, *32*, 11.

(3) Leforestier, C. *J. Chem. Phys.* **1978**, *68*, 4406.

(4) Truhlar, D. G.; Duff, J. W.; Blais, N. C.; Tully, J. C.; Garrett, B. C. *J. Chem. Phys.* **1982**, *72*, 764.

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

(6) Baldridge, K. K.; Gordon, M. S.; Steckler, R.; Truhlar, D. G. *J. Phys. Chem.* **1989**, *93*, 5107.

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

(8) Garrett, B. C.; Melius, C. F.; Page, M. In *Theoretical and Computational Models for Organic Chemistry*; Formosinho, S. J., Csizmadia, I. G., Arnaut, L. G., Eds.; Kluwer: Dordrecht, 1991; p 35.

(9) Uggerud, E.; Helgaker, T. *J. Am. Chem. Soc.* **1992**, *114*, 4265.

(10) Woo, T. K.; Margl, P. M.; Ziegler, T.; Blöchl, P. E. *J. Phys. Chem. B* **1997**, *101*, 7877.

(11) Bolton, K.; Hase, W. L.; Peslherbe, G. H. In *Modern Methods for Multidimensional Dynamics Calculations in Chemistry*; Thompson, D. C., Ed.; World Scientific: Singapore, 1998; p 143.

(12) Li, X.; Millam, J. M.; Schlegel, H. B. *J. Chem. Phys.* **2000**, *113*, 10062.

(13) Kim, Y.; Corchado, J. C.; Villa, J.; Xing, J.; Truhlar, D. G. *J. Chem. Phys.* **2000**, *112*, 2718.

(14) Albu, T. V.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 8465.

(15) Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 5048.

(16) Lin, H.; Pu, J. Z.; Albu, T. V.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 4112.

(17) Kim, K. H.; Kim, Y. *J. Chem. Phys.* **2004**, *120*, 623. Kim, Y.; Kim, Y. *J. Phys. Chem. A* **2006**, *110*, 60.

(18) Olson, W. K.; Flory, P. J. *Biopolymers* **1972**, *11*, 25.

(19) van Gunsteren, W. F.; Berendsen, H. J. C.; Geurtsen, R. G.; Zwinderman, H. R. J. *Ann. N. Y. Acad. Sci*. **1986**, *482*, 287.

(20) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(21) Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566.

(22) Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576.

(23) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. *J. Phys. Chem.* **1990**, *94*, 8897.

(24) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skid, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

(25) Casewit, C. J.; Colwell, K. S.; Rappé, A. K. *J. Am. Chem. Soc.* **1992**, *114*, 10035. Casewit, C. J.; Colwell, K. S.; Rappé, A. K. *J. Am. Chem. Soc.* **1992**, *114*, 10046.

(26) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E. I.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. *Comput. Phys. Commun.* **1995**, *91*, 1.

(27) Rappé, A. K.; Colwell, K. S.; Casewit, C. J. *Inorg. Chem.* **1993**, *32*, 3438.

(28) 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.

(29) Hill, J.-R.; Sauer, J. *J. Phys. Chem.* **1994**, *98*, 1238. Hill, J.-R.; Sauer, J. *J. Phys. Chem.* **1995**, *99*, 9536.

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

(31) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 520. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 552.

(32) Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, *17*, 587.

(33) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 616.

(34) Sierka, M.; Sauer, J. *Faraday Discuss.* **1997**, *106*, 41.

(35) Schröder, K.-P.; Sauer, J. *J. Phys. Chem.* **1996**, *100*, 11043.

(36) Landis, C. R.; Cleveland, T.; Firman, T. K. *J. Am. Chem. Soc.* **1998**, *120*, 2641.

(37) MacKerell, A. D., Jr.; 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., III.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(38) McDonald, N. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 8049.

(39) Jorgensen, W. L.; McDonald, N. A. *THEOCHEM* **1998**, *424*, 145.

(40) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827.

(41) Ewig, C. S.; Thacher, T. S.; Hagler, A. T. *J. Phys. Chem. B* **1999**, *103*, 6999.

(42) Allinger, N. L.; Durkin, K. A. *J. Comput. Chem.* **2000**, *21*, 1229.

(43) Firman, T. K.; Landis, C. R. *J. Am. Chem. Soc.* **2001**, *123*, 11728.

(44) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.

(45) Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. *J. Comput. Chem.* **2001**, *22*, 1340.

(46) Ren, P.; Ponder, J. W. *J. Comput. Chem*. **2002**, *23*, 1497.

(47) Ponder, J. W.; Case, D. A. *Adv. Protein Chem*. **2003**, *66*, 27.

(48) London, F. *Z. Elektrochem.* **1929**, *35*, 552.

(49) Eyring, H.; Polanyi, M. *Z. Phys. Chem. B* **1931**, *12*, 279.

(50) Kimball, G. E.; Eyring, H. *J. Am. Chem. Soc.* **1932**, *54*, 3876.

(51) Coulson, C. A.; Danielsson, U. *Arkiv Fysik* **1954**, *8*, 245.

(52) Sato, S. *J. Chem. Phys*. **1955**, *23*, 2465.

(53) Porter, R. N.; Karplus, M. *J. Chem. Phys*. **1964**, *40*, 1105.

(54) Parr, C. A.; Truhlar, D. G. *J. Phys. Chem.* **1971**, *75*, 1844.

(55) Raff, L. M. *J. Chem. Phys.* **1974**, *60*, 2220.

(56) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218.

(57) Chang, Y.-T.; Minichino, C.; Miller, W. H. *J. Chem. Phys*. **1992**, *96*, 4341.

(58) Minichino, C.; Voth, G. A. *J. Phys. Chem. B* **1997**, *101*, 4544.

(59) Okuyama-Yoshida, N.; Nagaoka, M.; Yamabe, T. *J. Phys. Chem. A* **1998**, *102*, 285.

(60) Lefohn, A. E.; Ovchinnikov, M.; Voth, G. A. *J. Phys. Chem. B* **2001**, *105*, 6628.

(61) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.

(62) Schlegel, H. B.; Sonnenberg, J. L. *J. Chem. Theory Comput.* **2006**, *2*, 905.

(63) Whitlock, P. A.; Muckerman, J. T.; Fisher, E. R. *J. Chem. Phys*. **1982**, *76*, 4468.

(64) Bernardi, F.; Olivucci, M.; Robb, M. A. *J. Am. Chem. Soc.* **1992**, *114*, 1606.

(65) Halvick, P.; Truhlar, D. G. *J. Chem. Phys*. **1992**, *96*, 2895.

(66) Caltaneo, P.; Persico, M. *Theor. Chem. Acc*. **2000**, *103*, 390.

(67) Ischtwan, J.; Collins, M. A. *J. Chem. Phys.* **1994**, *100*, 8080.

(68) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. *J. Chem. Phys.* **1995**, *103*, 5522.

(69) Truhlar, D. G.; Brown, F. B.; Steckler, R.; Isaacson, A. D. In *The Theory of Chemical Reaction Dynamics*; Clary, D. C., Ed.; NATO ASI Series C, Reidel: Dordrecht, 1986; Vol. 170, p 285.

(70) Marcus, R. A.; Coltrin, M. E. *J. Chem. Phys.* **1977**, *67*, 2609.

(71) Skodje, R. T.; Truhlar, D. G.; Garrett, B. C. *J. Phys. Chem.* **1981**, *85*, 3019.

(72) Skodje, R. T.; Truhlar, D. G.; Garrett, B. C. *J. Chem. Phys*. **1982**, *78*, 1213.

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

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

(75) Babamov, V. K.; Marcus, R. A. *J. Chem. Phys.* **1981**, *74*, 1790.

(76) Garrett, B. C.; Truhlar, D. G.; Wagner, A. F.; Dunning, T. H., Jr. *J. Chem. Phys.* **1983**, *78*, 4400.

(77) Bondi, D. K.; Connor, J. N. L.; Garrett, B. C.; Truhlar, D. G. *J. Chem. Phys.* **1983**, *78*, 5981.

(78) Garrett, B. C.; Abusalbi, N.; Kouri, D. J.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *83*, 2252.

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

(80) Ishida, T.; Schatz, G. C. *Chem. Phys. Lett.* **1999**, *314*, 369.

(81) Collins, M. A.; *Theor. Chem. Acc.* **2002**, *108*, 313.

(82) Tishchenko, O.; Truhlar, D. G., to be published.

(83) Garrett, B. C.; Truhlar, D. G. *J. Phys. Chem.* **1979**, *83*, 1079. Garrett, B. C.; Truhlar, D. G. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 4755.

(84) Garrett, B. C.; Truhlar, D. G.; Grev, R. S.; Magnuson, A. W. *J. Phys. Chem.* **1980**, *84*, 1730.

(85) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771.

(86) Fernandez-Ramos, A.; Ellingson, B. A.; Garret, B. C.; Truhlar, D. G. *Rev. Comput. Chem.* In press.

(87) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(88) Singh, U. C.; Kollmann, P. A. *J. Comput. Chem.* **1986**, *7*, 718.

(89) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(90) Thery, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.

(91) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J. Comput. Chem.* **1995**, *16*, 113.

(92) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.

(93) Thompson, M. A. *J. Phys. Chem.* **1995**, *99*, 4794.

(94) Barnes, J. A.; Williams, I. H. *Biochem. Soc. Trans.* **1996**, *24*, 263.

(95) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chem. Phys.* **1996**, *211*, 313.

(96) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100.

(97) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum. Chem.* **1996**, *60*, 1189.

(98) Bersuker, I. B.; Leong, M. K.; Boggs, J. E.; Pearlman, R. S. *Int. J. Quantum. Chem.* **1997**, *63*, 1051.

(99) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.

(100) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.

(101) Eichler, U.; Kölmel, C. M.; Sauer, J. *J. Comput. Chem.* **1996**, *18*, 463.

(102) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.

(103) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.

(104) Gao, J. *Rev. Comput. Chem.* **1996**, *7*, 119.

(105) Cummins, P. L.; Gready, J. E. *J. Comput. Chem.* **1997**, *18*, 1496.

(106) Sinclair, P. E.; de Vries, A.; Sherwood, P.; Catlow, C. R. A.; van Santen, R. A. *J. Chem. Soc., Faraday Trans*. **1998**, *94*, 3401. 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.

(107) Burton, N. A.; Harrison, M. J.; Hart, J. C.; Hillier, I. H.; Shepppard, D. W. *Faraday Discuss. Chem. Soc*. **1998**, *110*, 463.

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

(109) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.

(110) Tongraar, A.; Liedl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1998**, *102*, 10340.

(111) Woo, T. K.; Cavallo, L.; Ziegler, T. *Theor. Chem. Acc.* **1998**, *100*, 307.

(112) Monard, G.; Merz, K. M., Jr. *Acc. Chem. Res.* **1999**, *32*, 904.

(113) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452.

(114) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46.

(115) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468.

(116) Shoemaker, J. R.; Burggraf, L. W.; Gordon, M. S. *J. Phys. Chem. A* **1999**, *103*, 3245.

(117) Lyne, P. D.; Hodoscek, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462.

(118) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.

(119) Turner, A. J.; Moliner, V.; Williams, I. H. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323.

(120) Dapprich, S.; Komiromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *THEOCHEM* **1999**, *461*−*462*, 1.

(121) Hillier, I. H. *THEOCHEM* **1999**, *463*, 45.

(122) Röthlisberger, U.; Carloni, P.; Doclo, K.; Parrinello, M. *J. Biol. Inorg. Chem.* **2000**, *5*, 236.

(123) Sierka, M.; Sauer, J. *J. Chem. Phys.* **2000**, *112*, 6983.

(124) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.

(125) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656.

(126) Sushko, P. V.; Shluger, A. L.; Catlow, C. R. A. *Surf. Sci.* **2000**, *450*, 153.

(127) Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J. *J. Chem. Phys.* **2001**, *115*, 62.

(128) Poteau, R.; Ortega, I.; Alary, F.; Solis, A. R.; Barthelat, J. C.; Daudey, J. P. *J. Phys. Chem. A* **2001**, *105*, 198.

(129) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.

(130) Truhlar, D. G.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sanchez, M. L.; Villa, J. *Acc. Chem. Res.* **2002**, *35*, 341.

(131) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

(132) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* **2002**, *355*, 257.

(133) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578.

(134) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534.

(135) Mordasini, T.; Curioni, A.; Andreoni, W. *J. Biol. Chem.* **2003**, *278*, 4381.

(136) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *J. Phys. Chem. A* **2003**, *107*, 2578.

(137) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjovoll, M.; Fahmi, A.; Schafer, A.; Lennartz, C. *THEOCHEM* **2003**, *632*, 1.

(138) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43.

(139) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.

(140) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632.

(141) Riccardi, D.; Li, G. H.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467.

(142) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2.

(143) Lin, H.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 3991.

(144) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2006**, in press.

(145) Heyden, A.; Truhlar, D. G. To be published.

(146) Atkinson, R.; Arey, J. *Acc. Chem. Res*. **1998**, *31*, 574.

(147) Reissell, A.; Arey, J.; Atkinson, R. *Int. J. Chem. Kinet*. **2001**, *33*, 56.

(148) Atkinson, R.; Arey, J. *Atmos. Environ.* **2003**, *37*, 1997.

(149) Truhlar D. G., Garret, B. C. *Annu. Rev. Phys. Chem.* **1984**, *35*, 139.

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

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

(152) Marcus, R. A. *J. Chem. Phys*. **1966**, *45*, 4493.

(153) Wyatt, R. E. *J. Chem. Phys*. **1969**, *51*, 3489.

(154) Duff, J. W.; Truhlar, D. G. *J. Chem. Phys*. **1975**, *52*, 2744.

(155) Miller, W. H.; Handy, N. C.; Adams, J. E. *J. Chem. Phys.* **1980**, *72*, 99.

(156) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.

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

(158) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.

(159) Perdew, J. P. In *Proceedings of the 21st Annual International Symposium on the Electronic Structure of Solids*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991.

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

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

(162) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.

(163) Frisch, M. J. et al. *Gaussian03*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(164) Ponder, J. W. *TINKER; Version 4.2;* Washington University: St. Louis, MO, 2004.

(165) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(166) Albu, T. V.; Corchado, J. C.; Kim, Y.; Villà, J.; Xing, J.; Lin, H.; Truhlar, D. G. *MC-TINKER-version 1.1*; University of Minnesota: Minneapolis, MN 55455, 2004.

(167) Lin, H.; Truhlar, D. G. *QMMM; Version 1.0*; University of Minneapolis: Minneapolis, 2004.

(168) Schlegel, H. B. *J. Comput. Chem.* **1982**, *3*, 214.

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

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

(171) Albu, T. V.; Corchado, J. C.; Kim, Y.; Villà, J.; Xing, J.; Lin, H.; Truhlar, D. G. *MC-TINKERATE; Version 9.1*; University of Minnesota: Minneapolis, MN, 2003.

(172) Schöneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. *J. Am. Chem. Soc*. **2002**, *124*, 8142.

CT600171U

JCTC Journal of Chemical Theory and Computation

Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant

Yuqing Deng[†] and Benoît Roux*[,†,‡]

Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Building 221, Argonne, Illinois 60439, and Institute of Molecular Pediatric Sciences, Gordon Center for Integrative Sciences, University of Chicago, 929 East 57th Street, Room W323B, Chicago, Illinois 60637

Received February 1, 2006

Abstract: Calculations of the binding free energy of various nonpolar aromatic ligands with the L99A mutant of T4 lysozyme using molecular dynamics (MD) simulation are presented. To ensure better convergence, biasing potentials are used to restrain the ligand orientation and center-of-mass movement relative to the binding site when the ligand is decoupled from its environment in the binding pocket. The bias introduced by the restraint potentials is removed once the ligand fully interacts with the rest of the system and the calculated binding free energy is independent of the applied restraints. To decrease the computational cost, the simulations are generated with a reduced system in which protein and water atoms within a 15 Å-radius sphere around the ligand are included explicitly, while the rest of the system is treated with the generalized solvent boundary potential (GSBP). For all the ligands, the precision of the calculated free energy is less than 0.5 kcal/mol. For small nonpolar ligands such as benzene, toluene, and ethylbenzene, the calculated binding free energies are within 1.1 kcal/mol of the experimental values. For larger ligands, the computed binding free energies are slightly more favorable than the experimental values. The nonbinding polar molecule, phenol, has a calculated binding free energy of −0.88 kcal/mol. The simulation protocol presented here provides a way to calculate the binding free energy of small molecules to receptors at moderate computational cost.

Introduction

"*Corpora non agunt nisi fixata*"[1]—A substance is not (biologically) active unless it is fixed (bound) by a receptor. Nearly a century ago, this Latin maxim from Paul Ehrlich (1913) epitomized the central role of molecular recognition in chemistry and biology. Such a view is currently more pertinent then ever, as the rational design of small ligands binding to macromolecules with high affinity and specificity has become one of the cornerstones in the development of new drugs.[2−6] Theoretical and computational tools, if they were systematically able to predict binding free energies with quantitative accuracy, would have a big impact on drug design. Developing and improving these methods is, thus, an important goal of computational chemistry and biophysics.

Calculations of binding free energies employ a wide variety of methods, ranging from simple empirical scoring functions to thermodynamic free energy simulations with explicit solvent and full atomic details.[7] While it may be expected that fast and simple scoring schemes will remain the method of choice to sieve through a large database comprising millions of compounds, more accurate treatments are clearly necessary in order to accurately rank-order similar and potentially active ligands. With the growth in computational resources, advances in simulations methodologies, and improvement of all-atom force fields, approaches to calculate standard (absolute) binding free energies based on rigorous statistical mechanical treatments become increasingly practical.[8−11] It is of interest to take advantage of these recent advances in order to make rigorous calculations of binding free energies tracktable computationally to the point that they could be utilized routinely.

---

* Corresponding author e-mail: roux@uchicago.edu.

[†] Argonne National Laboratory.

[‡] University of Chicago.

10.1021/ct060037v CCC: $33.50

Fundamentally, the binding free energy can be expressed from the reversible thermodynamic work to separate the ligand from its receptor in solution.[12−20] Alchemical free energy perturbation (FEP) simulations,[21,22] in which the ligand is annihilated (decoupled from its surrounding), offer a practical route to carry out such computations.[23] Alternatively, it is possible to integrate the Boltzmann factor of the potential of mean force (PMF) as a function of the ligand−receptor distance.[12,18,20] While the latter may be advantageous in the case of highly charged ligands bound to the surface of a receptor (e.g., phosphotyrosine peptide pYEEI binding to an SH2 domain),[20] the former is particularly useful in the case of a ligand bound deeply in the interior of a macromolecule. The main advantage of alchemical FEP is that a realistic physical path for the association/dissociation of the ligand from the binding pocket is not required. In other words, the calculations can proceed by annihilating the ligand via unphysical intermediate states as along as the final endpoint states are correctly constructed. A straightforward double-annihilation scheme[23] is, however, impractical because the decoupled (noninteracting) ligand can drift away from the binding site and wander anywhere in the simulated system, leading to an ill-defined reference endpoint state (see the discussion in ref 19). These problems are solved in the double-decoupling method (DDM), in which biasing restraint potentials are applied to the ligand translational degrees of freedom relative to the receptor. Handling these restraints properly enables a rigorous treatment of the endpoint decoupled state.[13−17,19,24]

In the context of this theoretical framework, it is possible to carry detailed FEP molecular dynamics (MD) simulations to calculate the standard (absolute) binding free energy of a ligand to a receptor. Conducting all-atom FEP/MD free energy simulations for an entire protein−solvent system with periodic boundary conditions is, however, often prohibitive, and it is important to seek ways to reduce the computational cost of the calculations. In the present calculations, all the simulations are performed on reduced atomic systems using the spherical solvent boundary potential (SSBP)[25] for bulk solvent and the generalized solvent boundary potential (GSBP) for the binding site.[26] In these treatments, all the atomic details near the ligand are simulated explicitly, while the influence of the rest of the system is incorporated implicity via a mean-field continuum electrostatic approximation. Free energy calculations using both SSBP[27] and GSBP[28] have been shown to yield accurate results at a much reduced computational cost. With these approximations, the number of atoms that are simulated explicitly typically ranges from 1500 to about 3000, which reduces significantly the computational effort for FEP simulations.

Additional considerations may also help to construct a robust and efficient method for calculating binding free energies. For example, intermolecular forces are typically dominated by short-range harsh repulsive interactions and more slowly varying van der Waals attraction and electrostatic interactions; it is consequentially advantageous to stage the decoupling alchemical FEP simulations accordingly with a proper choice of intermediate states by exploiting the drastically different characters in spatial range and magnitude of these various contributions. A staged decoupling FEP scheme based on the Weeks−Chandler−Andersen (WCA)[29] separation of the Lennard-Jones (LJ) 12−6 potential has been shown to provide a clear separation of the competing repulsive and dispersive forces in the solvation process and to greatly enhance the efficiency of solvation free energy computations.[27] The WCA separation leads to the interesting observation that, while the total contribution to the solvation free energy from nonpolar interactions is relatively small, it typically results from the cancellation of much larger repulsive and dispersive contributions. It may be expected that a similar separation of the binding free energy will help clarify important aspects of the binding process which might otherwise remain hidden.

Our goal with this paper is to present a detailed examination of theoretical and computational issues related to the calculation of the absolute binding free energy using FEP simulations. As a test system, we chose to study the binding of aromatic ligands to the engineered cavity inside the L99A T4 lysozyme mutant, for which there are many available crystallographic structures.[30,31] The availability of accurate structure of the complexes allows us to avoid the challenging step of determining the ligand position and configuration (*docking*) in its binding site. From this point of view, the present effort is primarily concerned with the evaluation of the absolute binding free energy associated with a given ligand position and configuration (*scoring*) using a rigorous but computationally inexpensive FEP simulation protocol. The paper is organized as follows. The theoretical formulation of the binding free energy calculation forms the next section. What follows is all computational details. The results and discussion are then presented, and the paper is concluded with a brief summary of the most important points.

II. Theory

A. Equilibrium Binding Constant. For the sake of simplicity, we consider a single receptor molecule (R) in thermodynamic equilibrium with a dilute solution containing flexible ligand molecules (L). The equilibrium constant, K_b, of the binding reaction $L + R \rightleftharpoons LR$ is defined as $K_b = [LR]/([L][R])$, where [L], [R], and [LR], are the equilibrium concentrations of the ligand, receptor, and complex, respectively. It is assumed that the *binding site* of the receptor and the *bulk* can be distinguished clearly as two distinct spatial regions. This implies that the binding site of the receptor molecule can either be empty or occupied by a ligand. A rigorous expression for the equilibrium binding constant can be derived directly from population configurational ensemble averages.[14,18,20] This derivation, which can be traced back to Bjerrum,[32] is arguably clearer and more direct than the more traditional treatment that consists of equating the chemical potentials of the three species, L, R, and LR.[15,19] Let the probability to find the receptor with no ligand bound be $\mathscr{P}_0$, and the probability to find the receptor with one ligand bound be $\mathscr{P}_1$;[14] by normalization, $\mathscr{P}_0 + \mathscr{P}_1 = 1$. The ratio of these occupancy probabilities, $\mathscr{P}_1/\mathscr{P}_0 = K_b[L]$, is related to the thermodynamic reversible work to take one ligand molecule in the bulk solution and insert it into the binding site.[14,33,20] It follows that K_b can be expressed in terms of a ratio of configurational integrals as

$$K_{\rm b} = \frac{1}{[{\rm L}]} \frac{N\int_{\rm site} d(\mathbf{1}) \int_{\rm bulk} d(\mathbf{2}) \cdots \int_{\rm bulk} d(\mathbf{N}) \int d\mathbf{X}\, e^{-\beta U}}{\int_{\rm bulk} d(\mathbf{1}) \int_{\rm bulk} d(\mathbf{2}) \cdots \int_{\rm bulk} d(\mathbf{N}) \int d\mathbf{X}\, e^{-\beta U}} \qquad (1)$$

where U is the total potential energy of the system; $(\mathbf{1})$, $(\mathbf{2})$, ..., (N) and $\mathbf{X}$ are the coordinates of the N ligand molecules and the remaining atoms (solvent and receptor), respectively; $\beta \equiv 1/k_{\rm B}T$, where $k_{\rm B}$ is the Boltzmann constant, and T is the temperature. The factor of N accounts for the fact that any of the N identical ligands could be chosen to occupy the binding site. Because the bulk region is homogeneous and invariant by translation, this expression can be rewritten as

$$K_{\rm b} = \frac{1}{[{\rm L}]} \times \frac{N\int_{\rm site} d(\mathbf{1}) \int_{\rm bulk} d(\mathbf{2}) \cdots \int_{\rm bulk} d(\mathbf{N}) \int d\mathbf{X}\, e^{-\beta U}}{V_{\rm bulk} \int_{\rm bulk} d(\mathbf{1})\, \delta\, (\mathbf{r}_1 - \mathbf{r}^*) \int_{\rm bulk} d(\mathbf{2}) \cdots \int_{\rm bulk} d(\mathbf{N}) \int d\mathbf{X}\, e^{-\beta U}} \qquad (2)$$

where $\boldsymbol{r}_1$ is the position of the center of mass (COM) of ligand 1, and $\boldsymbol{r}^*$ is some arbitrary location in the bulk region. In the limit of an excess ligand concentration, the binding constant is

$$K_{\rm b} = \frac{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta U}}{\int_{\rm bulk} d\,(\mathbf{1}) \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta U}} \qquad (3)$$

where the integrals over the $(N - 1)$ remaining ligand in the bulk have been omitted for the sake of clarity. Eq 3 is our starting point for designing effective binding free energy simulation strategies.[14] It should be noted that the delta function in the denominator of eq 3 implies that $K_{\rm b}$ has dimension of volume.

The general idea to compute the equilibrium binding constant from eq 3 can be expressed by inserting unity as the ratio of configurational integrals, Z_i/Z_i, $i = 1...n$, corresponding to suitably chosen intermediate states

$$K_{\rm b} = \frac{\int_{\rm site} d\mathbf{1} \int d\mathbf{X}\, e^{-\beta U}}{Z_n} \times \frac{Z_n}{Z_{n-1}} \times \cdots \frac{Z_3}{Z_2} \times \frac{Z_2}{Z_1} \times \frac{Z_1}{\int_{\rm bulk} d\mathbf{1}\, \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X}\, e^{-\beta U}} \qquad (4)$$

Here, the numerator of the leftmost ratio and the denominator of the rightmost ratio may be considered as *initial* and *final* states of the binding process: the ligand bound to the receptor and the ligand held with its COM at $\mathbf{r}_1^*$, far away from the receptor. The remaining configurational integrals, Z_i, represent intermediate states chosen for the sake of computational effectiveness and convenience. Wise choices of those intermediates are the key to the design of a practical scheme— the contribution of each of the intermediates should be amenable to efficient computer simulations. One possible strategy is to insert intermediates with different fixed distances between the ligand and the receptor, so that the binding constant is expressed as an integration over a potential of mean force (PMF)[12,18,33,34,20] along the chosen pathway. An alternative strategy is to annihilate the ligand in the binding site and re-create it in the bulk solution.[13−19] This strategy corresponds to the double-annihilation method,[23] where a single intermediate configurational integral Z_1 is introduced in which the ligand is completely decoupled from its surroundings (the protein receptor or bulk solvent). A straightforward application of the double-annihilation scheme is, however, impractical because the noninteracting ligand can drift away from the binding site and wander anywhere in the simulation box (see the discussion in ref 19). In fact, a simulation has to be exceedingly (impossibly) long to sample the decoupling of the ligand from the receptor in a true double-annihilation computation. Such difficulties are avoided very simply by introducing one additional configurational integral Z_2, an intermediate state of a decoupled ligand confined in the binding site by an extra potential, $u_{\rm t}$, which restrains the position of the ligand relative to the receptor.[13,14,16,17,19] Similarly, additional restraining potentials can also be introduced to bias the orientation[16,19] and conformation[20] of the ligand to accelerate the convergence further. To facilitate further development, it is useful to define U_1 as the total potential energy of the fully interacting system and U_0 as the total potential energy of a fictitious system in which the ligand does not interact with the rest of the system. The binding constant can then be expressed as

$$\begin{aligned} K_{\rm b} = {} & \frac{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta U_1}}{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}]}} \times \frac{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}]}}{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}+u_{\rm t}]}} \times \\ & \frac{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}+u_{\rm t}]}}{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}+u_{\rm t}+u_{\rm r}]}} \times \frac{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}+u_{\rm t}+u_{\rm r}]}}{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_{\rm c}+u_{\rm t}+u_{\rm r}]}} \times \\ & \frac{\int_{\rm site} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_{\rm c}+u_{\rm t}+u_{\rm r}]}}{\int_{\rm bulk} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_{\rm c}+u_{\rm t}]}} \times \frac{\int_{\rm bulk} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_{\rm c}+u_{\rm t}]}}{\int_{\rm bulk} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_0+u_{\rm c}]}} \times \\ & \frac{\int_{\rm bulk} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_0+u_{\rm c}]}}{\int_{\rm bulk} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}]}} \times \frac{\int_{\rm bulk} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}]}}{\int_{\rm bulk} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta U_1}} \end{aligned} \qquad (5)$$

where $u_{\rm c}$ is a restraint potential acting on the internal configuration of the ligand, $u_{\rm t}$ is a restraint potential acting on the position of the COM (relative to the receptor) of the ligand, and $u_{\rm r}$ is a restraint potential acting on the rotational orientation (relative to the receptor) of the ligand. Typically, the restraint potentials are chosen as harmonic potentials centered around some reference values. The configurational restraint potential, u_c, for example, is a harmonic potential with respect to the root-mean-square deviation (RMSD) of the ligand relative to a reference bound configuration.[20] The formulation of the binding constant developed here is quite similar to the PMF formulation in ref 20. The difference is that in the PMF formulation, the interaction between the ligand and the receptor is never turned off, and the binding is achieved by the change in the distance between the ligand and the protein. In terms of free energies, the binding constant can be written as

$$K_b = e^{+\beta\Delta G_c^{\mathrm{site}}} \times e^{+\beta\Delta G_t^{\mathrm{site}}} \times e^{+\beta\Delta G_r^{\mathrm{site}}} \times e^{-\beta\Delta G_{\mathrm{int}}^{\mathrm{site}}} \times F_r \times F_t \times e^{+\beta\Delta G_{\mathrm{int}}^{\mathrm{bulk}}} \times e^{-\beta\Delta G_c^{\mathrm{bulk}}} \tag{6}$$

where the various free energies are defined as what follows

$$e^{-\beta\Delta G_c^{\mathrm{site}}} = \frac{\int_{\mathrm{site}} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}}{\int_{\mathrm{site}} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta U_1}} = \langle e^{-\beta u_c}\rangle_{(\mathrm{site};\, U_1)} \tag{7}$$

$$e^{-\beta\Delta G_t^{\mathrm{site}}} = \frac{\int_{\mathrm{site}} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t]}}{\int_{\mathrm{site}} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}} = \langle e^{-\beta \mathrm{u_t}}\rangle_{(\mathrm{site};U_1+u_c)} \tag{8}$$

$$e^{-\beta\Delta G_r^{\mathrm{site}}} = \frac{\sigma\int_{\mathrm{site}} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t+u_r]}}{\int_{\mathrm{site}} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t]}} = \langle e^{-\beta u_r}\rangle_{(\mathrm{site};U_1+u_{c+}u_t)} \tag{9}$$

$$e^{-\beta\Delta G_{\mathrm{int}}^{\mathrm{site}}} = \frac{\int_{\mathrm{site}} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_c+u_t+u_r]}}{\int_{\mathrm{site}} d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t+u_r]}} = \langle e^{-\beta[U_1-U_0]}\rangle_{(\mathrm{site};U_0+u_c+u_t+u_r)} \tag{10}$$

$$e^{-\beta\Delta G_{\mathrm{int}}^{\mathrm{bulk}}} = \frac{\int_{\mathrm{bulk}} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}}{\int_{\mathrm{bulk}} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_0+u_c]}} = \langle e^{-\beta[U_1-U_0]}\rangle_{(\mathrm{bulk};U_0+u_c)} \tag{11}$$

and

$$e^{-\beta\Delta G_c^{\mathrm{bulk}}} = \frac{\int_{\mathrm{bulk}} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_1+u_c]}}{\int_{\mathrm{bulk}} d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta U_1}} = \langle e^{-\beta u_c}\rangle_{(\mathrm{bulk};U_1)} \tag{12}$$

where $\mathbf{r}^*$ is the fixed position of the ligand COM in the bulk region. Note that the symmetry number of the ligand, σ, has been included in eq 9 because the configurational integral for $\Delta G_r^{\mathrm{site}}$ is restricted to a single distinct orientation (σ is the number of rotational operations in the point group of the ligand). All those ΔG's can be calculated from separate free energy simulations according to a standard methodology.[21,22,35]

In contrast to those contributions above, the translational factor F_t

$$F_t = \frac{\int d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t]}}{\int d\,(\mathbf{1})\, \delta(\mathbf{r}_1 - \mathbf{r}^*) \int d\mathbf{X}\, e^{-\beta[U_0+u_c]}} \tag{13}$$

and the rotational factor F_r

$$F_r = \frac{\int d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t+u_r]}}{\int d\,(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_0+u_c+u_t]}} \tag{14}$$

can be evaluated directly with numerical integration schemes. The position of the ligand with respect to the receptor is restrained with the potential $u_t(\mathbf{r}_1)$, where $\mathbf{r}_1$ is the relative vector between the COM of the ligand and the receptor. Since the ligand molecule does not interact with its surroundings in the uncoupled state with potential U_0, the translational factor F_t can be expressed as

$$F_t = \int d\mathbf{r}_1\, e^{-\beta \mathrm{u_t}(\mathbf{r}_1)} \tag{15}$$

(all other terms cancel out). Likewise, the orientation of the ligand molecule (relative to the receptor) is restrained via the potential $u_r(\Omega_1)$ where Ω_1 is the set of three angles for the rigid body rotation

$$F_r = \frac{\int d\Omega_1\, e^{-\beta u_r(\Omega_1)}}{\int d\Omega_1} \tag{16}$$

Similar treatment of restraining potentials has been discussed previously.[13−20] The simplification of the translational and rotational factors lies in the fact that there are six translational and rotational degrees of freedom of the bound complex. As long as the restraint potentials are placed on coordinates relative to the receptor, it is always possible to separate the ligand degrees of freedom from that of the proteins, thereby reducing the factors to products of one-dimensional integrals. It should be noted that the translational factor F_t (and hence the binding constant K_b) has a dimension of volume (with a natural unit of Å^3 in atomistic simulations). This dimensionality of the binding constant gives rise to the standard free energy, defined relative to the standard concentration 1 mol/L

$$\Delta G_b^\circ \equiv -k_B T \ln[K_b C^\circ] \tag{17}$$

where $C^\circ = 1$ mol·L^{-1} (=1/1660 Å^{-3}). This the statistical mechanical basis for the definition of the thermodynamic standard state in the binding free energy.[14,16,19]

The standard binding constant ($K_b C^\circ$) can be expressed as $K_b^\circ \equiv \exp[-\beta\Delta G_b^\circ]$, where the standard binding free energy is

$$\Delta G_b^\circ = \Delta\Delta G_{\mathrm{int}} + \Delta\Delta G_t^\circ + \Delta\Delta G_r + \Delta\Delta G_c \tag{18}$$

where $\Delta\Delta G_{\mathrm{int}} = [\Delta G_{\mathrm{int}}^{\mathrm{site}} - \Delta G_{\mathrm{int}}^{\mathrm{bulk}}]$, $\Delta\Delta G_c = [\Delta G_c^{\mathrm{bulk}} - \Delta G_c^{\mathrm{site}}]$, $\Delta\Delta G_r = [-\Delta G_r^{\mathrm{site}} - k_B T \ln(F_r)]$, and $\Delta\Delta G_t^\circ = [-\Delta G_t^{\mathrm{site}} - k_B T \ln(F_t C^\circ)]$. Each contribution has a well-defined meaning: $\Delta G_{\rm int}^{\rm bulk}$ corresponds to the free energy to introduce the interaction between the ligand (restrained in its bound conformation by the potentials $u_{\rm c}$) and the surrounding bulk solution; $\Delta G_{\rm int}^{\rm site}$ corresponds to the free energy to introduce the interaction of the ligand with its binding pocket (restrained in its conformation, translation, and rotation by the potentials $u_{\rm c}$, $u_{\rm t}$, and $u_{\rm r}$, respectively); $\Delta\Delta G_{\rm int}$ corresponds to the free energy difference associated with removing the ligand from the bulk and inserting it in the binding site; $\Delta\Delta G_{\rm t}$ and $\Delta\Delta G_{\rm r}$ correspond to the free energy cost associated with application and removal of the translational and rotational restraints on the ligand; $\Delta\Delta G_{\rm c}$ corresponds to the free energy cost associated with restricting the conformational freedom of the ligand upon binding. The free energies associated with the introduction of the conformational restraint potential can be computed, for example, by using the RMSD relative to the conformation of the bound ligand[36]

$$\Delta G_{\rm c}^{\rm site} = -k_{\rm B}T \ln\frac{\int d\zeta\ \rho_{\rm site}(\zeta)e^{-\beta u_c}}{\int d\zeta\ \rho_{\rm site}(\zeta)} \tag{19}$$

$$\Delta G_{\rm c}^{\rm bulk} = -k_{\rm B}T \ln\frac{\int d\zeta\ \rho_{\rm bulk}(\zeta)e^{-\beta u_c}}{\int d\zeta\ \rho_{\rm bulk}(\zeta)} \tag{20}$$

where $\rho_{\rm site}(\zeta)$ and $\rho_{\rm bulk}(\zeta)$ are the unbiased distributions of RMSD. One may note that the contribution from the conformational restraint, $-\beta\Delta\Delta G_{\rm c}$, is necessarily unfavorable because it is related to the reduction in the number of accessible conformations of the ligand in the binding site relative to the bulk, i.e., this may be formally represented as $\mathcal{N}_{\rm c}^{\rm site}/\mathcal{N}_{\rm c}^{\rm bulk}$. Equation 18 is the basis for all of the computations presented in this work. Apart from the additional conformational restraint $u_{\rm c}$, the present framework is equivalent to the scheme of Boresch et al.,[19] though it is derived using a different route.

B. Translational and Rotational Restriction in the Bound State. With appropriate choices of the translational and orientational restraint potentials, it is possible to relate $\Delta\Delta G_{\rm t}°$ and $\Delta\Delta G_{\rm r}$ directly to the reduction in the translational and rotational freedom of the bound ligand. The free energy associated with the translation of the ligand is

$$\begin{aligned} e^{-\beta\Delta\Delta G_{\rm t}°} &= C° \times \frac{\int_{\rm site} d\,(\mathbf{1}) \int d\boldsymbol{X}\, e^{-\beta[U_1+u_{\rm c}]}}{\int_{\rm site} d\,(\mathbf{1}) \int d\boldsymbol{X}\, e^{-\beta[U_1+u_{\rm c}+u_{\rm t}]}} \times \\ &\qquad \int d\mathbf{r}_1\, e^{-\beta u_{\rm t}(\mathbf{r}_1)} \\ &= C° \times \frac{\int_{\rm site} d\mathbf{r}_1\, P^{\rm site}(\mathbf{r}_1)}{\int_{\rm site} d\mathbf{r}_1\, P_{\rm t}^{\rm site}(\mathbf{r}_1) e^{-\beta u_{\rm t}(\mathbf{r}_1)}} \times \\ &\qquad \int d\,\mathbf{r}_1\, e^{-\beta u_{\rm t}(\mathbf{r}_1)} \end{aligned} \tag{21}$$

where $P_{\rm t}^{\rm site}$ is the probability distribution of the ligand position in the binding site (this distribution depends implicitly on $u_{\rm c}$). In the limit of strong restraint potential, $u_{\rm t}(\mathbf{r}_1)$, the translational restraint acts essentially as a delta function

$$\frac{e^{-\beta u_{\rm t}(\mathbf{r}_1)}}{\int_{\rm site} d\,\mathbf{r}_1\, e^{-\beta u_{\rm t}(\mathbf{r}_1)}} \approx \delta(\mathbf{r}_1 - \mathbf{r}_{\rm m}) \tag{22}$$

where $\mathbf{r}_{\rm m}$ is the energy minimum in the translational restraint. Assuming that $\mathbf{r}_{\rm m}$ is set to match the most probable position of the COM of the ligand in the binding site (i.e., the maximum of $P_{\rm t}^{\rm site}$), the translational contribution becomes

$$\begin{aligned} e^{-\beta\Delta\Delta G_{\rm t}°} &\approx C° \times \int_{\rm site} d\,\mathbf{r_1} \frac{P_{\rm t}^{\rm site}(\mathbf{r_1})}{P_{\rm t}^{\rm site}(\mathbf{r}_{\rm m})} \\ &= C°\Delta V \end{aligned} \tag{23}$$

where ΔV is the effective accessible volume of the COM of the ligand in the binding site. Similarly, the free energy associated with the orientation of the ligand is

$$\begin{aligned} e^{-\beta\Delta\Delta G_{\rm r}} &= \frac{\int_{\rm site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}+u_{\rm t}]}}{\int_{\rm site} d(\mathbf{1}) \int d\mathbf{X}\, e^{-\beta[U_1+u_{\rm c}+u_{\rm t}+u_{\rm r}]}} \times \frac{\int d\mathbf{\Omega_1}\, e^{-\beta {\rm u}_{\rm r}(\Omega_1)}}{\int d\mathbf{\Omega_1}} \\ &= \frac{\int d\mathbf{\Omega_1}\, P_{\rm r}^{\rm site}(\mathbf{\Omega_1})}{\int d\mathbf{\Omega_1}\, P_{\rm r}^{\rm site}(\mathbf{\Omega_1}) e^{-\beta {\rm u}_{\rm r}(\Omega_1)}} \times \frac{\int d\mathbf{\Omega_1}\, e^{-\beta {\rm u}_{\rm r}(\Omega_1)}}{\int d\mathbf{\Omega_1}} \end{aligned} \tag{24}$$

where $P_{\rm r}^{\rm site}$ is the distribution of the orientation angles (this distribution depends implicitly on $u_{\rm c}$ and $u_{\rm t}$). Here we drop the symmetry factor for the moment and will add it in the final binding constant formulation. In the limit of strong $u_{\rm r}$(Ω), the orientational restraint acts essentially as a delta function

$$\frac{e^{-\beta u_{\rm r}(\Omega_1)}}{\int d\mathbf{\Omega_1}\, e^{-\beta u_{\rm r}(\Omega_1)}} \approx \delta(\Omega_1 - \Omega_{\rm m}) \tag{25}$$

which is sharply peaked at $\Omega_{\rm m}$, chosen to match the most probable orientation of the ligand in the binding site (i.e., the maximum of $P_{\rm r}^{\rm site}$). For a nonlinear ligand, it follows that

$$\begin{aligned} e^{-\beta\Delta\Delta G_{\rm r}} &\approx \frac{1}{\int_{\rm site} d\Omega_1} \int_{\rm site} d\Omega_1 \frac{P_{\rm r}^{\rm site}(\Omega_1)}{P_{\rm r}^{\rm site}(\Omega_{\rm m})} \\ &= \frac{\Delta\Omega}{8\pi^2} \end{aligned} \tag{26}$$

Because the factor $\Delta\Omega/8\pi^2$ is necessarily smaller than (or equal to) 1, the change in orientational degrees of freedom makes an unfavorable contribution to binding free energy. For both the orientational and translational probabilities, a Gaussian approximation can be made to relate the accessible orientation and volume to fluctuations in the angles and position.[18,37] In the following, we will use the limit of strong restraints to compute the accessible volume and orientation from FEP simulations.

With the definition of the accessible volume and orientation of the ligand in the binding site, the standard binding constant $K_{\rm b}^{\circ}$ can be expressed in the familiar form

$$K_{\rm b}^{\circ} = C°\Delta V\left(\frac{\sigma\Delta\Omega}{8\pi^2}\right)\left(\frac{\mathcal{N}_{\rm c}^{\rm site}}{\mathcal{N}_{\rm c}^{\rm bulk}}\right)e^{-\beta[\Delta G_{\rm int}^{\rm site} - \Delta G_{\rm int}^{\rm bulk}]} \tag{27}$$

In eq 27 the conventional interpretation of the binding process becomes self-evident: a ligand in a single conformation is desolvated from the bulk solution and inserted to the binding site [$\Delta G_{\mathrm{int}}^{\mathrm{site}} - \Delta G_{\mathrm{int}}^{\mathrm{bulk}}$], and the contribution from conformation change as well as reduction in translational and orientational freedom are accounted for by $\mathcal{N}_{\mathrm{c}}^{\mathrm{site}}/\mathcal{N}_{\mathrm{c}}^{\mathrm{bulk}}$, $C°\Delta V$ (with the proper standard state) and $\Delta\Omega/8\pi^2$, respectively. Conceptually, this offers a useful decomposition of the binding constant in terms of various contributions. Although the statistical mechanical definition of all the terms is unambiguous, each term depends (conditionally) on the details of how the restraint potentials are implemented. In comparison with other methods, only a quantitative comparison of the final binding constant values will be meaningful in general.

III. Computational Details

A. Free Energy Perturbation. The computation of $\Delta G_{\mathrm{int}}^{\mathrm{site}}$ and $\Delta G_{\mathrm{int}}^{\mathrm{bulk}}$ is done with a step-by-step series of alchemical transformations. The interaction free energies ($\Delta G_{\mathrm{int}}^{\mathrm{site}}$ and $\Delta G_{\mathrm{int}}^{\mathrm{bulk}}$) are decomposed into repulsive, dispersive, and electrostatic contributions[27]

$$\Delta G_{\mathrm{int}}^{\mathrm{a}} = \Delta G_{\mathrm{rep}}^{\mathrm{a}} + \Delta G_{\mathrm{dis}}^{\mathrm{a}} + \Delta G_{\mathrm{elec}}^{\mathrm{a}} \tag{28}$$

where a = site, bulk. The free energy components are computed with free energy perturbation (FEP)[21,22,35] techniques. To this end, an auxiliary simulation potential energy is constructed with couplings parameters controlling the interactions between the ligand and its environment: λ_{rep}, and λ_{dis} for the repulsive and dispersive part of the Lennard-Jones potential, and λ_{elec}, for electrostatics. In addition, when the ligand is in the binding site, translational and rotational restraints are applied. Their magnitude is controlled by the coupling parameters λ_{t} and λ_{r}, respectively (when λ_{t} and λ_{r} go to zero, all translational and orientational restraints are removed). The total auxiliary potential energy is

$$U(\lambda_{\mathrm{rep}},\lambda_{\mathrm{dis}},\lambda_{\mathrm{elec}},\lambda_{\mathrm{t}},\lambda_{\mathrm{r}}) = U_0 + U^{\mathrm{rep}}(\lambda_{\mathrm{rep}}) + \lambda_{\mathrm{dis}}U^{\mathrm{dis}} + \lambda_{\mathrm{elec}}U^{\mathrm{elec}} + \lambda_{\mathrm{t}}\mathrm{u}_{\mathrm{t}} + \lambda_{\mathrm{r}}u_{\mathrm{r}} \tag{29}$$

where U_0 is the potential of the system with the noninteracting ligand; λ_{rep}, λ_{dis}, λ_{elec}, λ_{t},λ_{r} $\in[0, 1]$; $U^{\mathrm{rep}}(\lambda_{\mathrm{rep}})$ is the shifted Weeks−Chandler−Andersen (WCA)[29] repulsive interaction[27] between the ligand and its environment (when $\lambda_{\mathrm{rep}} = 0$ the repulsive interaction is completely turned off, when $\lambda_{\mathrm{rep}} = 1$, it is fully switched on); U^{dis} and U^{elec} are the WCA dispersive interaction and electrostatic interaction between the ligand and its environment, respectively. The total repulsive interaction is the sum over the shifted repulsive LJ-WCA pair potential[27]

$$u_{ij}^{\mathrm{rep}}(r;\lambda_{\mathrm{rep}}) = \begin{cases} \epsilon_{ij}\left\{\dfrac{(R_{ij}^*)^{12}}{[r^2+(1-\lambda_{\mathrm{rep}})^2(R_{ij}^*)^2]^6} - 2\dfrac{(R_{ij}^*)^6}{[r^2+(1-\lambda_{\mathrm{rep}})^2(R_{ij}^*)^2]^3} + 1\right\} & r \le R_{ij}^*\sqrt{1-(1-\lambda_{\mathrm{rep}})^2} \\ 0 & r > R_{ij}^*\sqrt{1-(1-\lambda_{\mathrm{rep}})^2} \end{cases} \tag{30}$$

where r and R_{ij}^* are the distance and minimal energy distance between atom type i and j, respectively. The dispersion interaction is the sum over the attractive LJ-WCA pair potential

$$u_{ij}^{\mathrm{dis}}(r) = \begin{cases} -\epsilon_{ij} & r \le R_{ij}^* \\ \epsilon_{ij}\left[\left(\dfrac{R_{ij}^*}{r}\right)^{12} - 2\left(\dfrac{R_{ij}^*}{r}\right)^6\right] & r \le R_{ij}^* \end{cases} \tag{31}$$

The insertion of the ligand into the binding pocket ($\Delta G_{\mathrm{int}}^{\mathrm{site}}$) is done in three steps, with the help of the coupling parameters controlling the nonbonded interaction (λ_{rep}, λ_{dis}, and λ_{elec}). Two additional steps are used to remove the translational and orientational restraints. The repulsive contribution $\Delta G_{\mathrm{rep}}^{\mathrm{site}}$ corresponds to the process

$$U(\lambda_{\mathrm{rep}} = 0,\lambda_{\mathrm{dis}} = 0,\lambda_{\mathrm{elec}} = 0,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 1) \rightarrow U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 0,\lambda_{\mathrm{elec}} = 0,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 1) \tag{32}$$

the dispersive contribution $\Delta G_{\mathrm{dis}}^{\mathrm{site}}$ corresponds to the process

$$U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 0,\lambda_{\mathrm{elec}} = 0,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 1) \rightarrow U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 1,\lambda_{\mathrm{elec}} = 0,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 1) \tag{33}$$

and the electrostatic contribution $\Delta G_{\mathrm{elec}}^{\mathrm{site}}$ corresponds to the process

$$U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 1,\lambda_{\mathrm{elec}} = 0,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 1) \rightarrow U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 1,\lambda_{\mathrm{elec}} = 1,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 1) \tag{34}$$

The free energies, $\Delta G_{\mathrm{r}}^{\mathrm{site}}$ and $\Delta G_{\mathrm{t}}^{\mathrm{site}}$, correspond respectively to the process

$$U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 1,\lambda_{\mathrm{elec}} = 1,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 1) \rightarrow U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 1,\lambda_{\mathrm{elec}} = 1,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 0) \tag{35}$$

and

$$U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 1,\lambda_{\mathrm{elec}} = 1,\lambda_{\mathrm{t}} = 1,\lambda_{\mathrm{r}} = 0) \rightarrow U(\lambda_{\mathrm{rep}} = 1,\lambda_{\mathrm{dis}} = 1,\lambda_{\mathrm{elec}} = 1,\lambda_{\mathrm{t}} = 0,\lambda_{\mathrm{r}} = 0) \tag{36}$$

The insertion of the ligand into the bulk sovent is calculated according to the same protocol but without translational and orientation restraints ($\lambda_{\mathrm{t}} = 0$ and $\lambda_{\mathrm{r}} = 0$). It should be noted that the values of $\Delta G_{\mathrm{rep}}^{\mathrm{site}}$, $\Delta G_{\mathrm{dis}}^{\mathrm{site}}$, and $\Delta G_{\mathrm{elec}}^{\mathrm{site}}$ depend on the translational and orientational restraints (thus, they are *conditional* free energies), though a comparison with the values in the bulk, $\Delta G_{\mathrm{rep}}^{\mathrm{bulk}}$, $\Delta G_{\mathrm{dis}}^{\mathrm{bulk}}$, and $\Delta G_{\mathrm{elec}}^{\mathrm{bulk}}$, is nonetheless expected to be informative. Table 1 lists all the values of the coupling parameters used in the step-by-step alchemical free energy calculations.

B. MD Simulations. All the molecular dynamics calculations were carried out with the CHARMM program.[38] For the translational/orientational restraint potentials (with the

Table 1: Values of the Coupling Parameters Used in the Free Energy Calculations

λ_{rep}	0.0	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0					
λ_{dis}	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0[a]				
λ_{elec}	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0				
λ_{t}	0.0	0.0025	0.005	0.0075	0.01	0.02	0.04	0.06	0.08	0.1	0.2	0.4	0.6	0.8	1.0
λ_{r}	0.0	0.0025	0.005	0.0075	0.01	0.02	0.04	0.06	0.08	0.1	0.2	0.4	0.6	0.8	1.0

[a] See Results and Discussion section and ref 27 for the choice of the optimal number of windows in the dispersive free energy calculation.

ligand in the binding site), two trajectories of 40 ps were generated starting from different initial velocities with values of the coupling parameters λ_t and λ_r equal to [0.0, 0.0025, 0.0050, 0.0075, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0]. The last 20 ps were collected and processed with the weighted histogram analysis method (WHAM).[39−41] For the repulsion contribution, two MD trajectories of 110 ps were run for values of the coupling parameter λ_{rep} equal to [0.0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]. The last 100 ps were processed with WHAM for each pair of adjacent coupling constants. For the dispersion contribution, values of the coupling parameter λ_{dis}, equal to [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], were used in the simulation (see also Results and Discussion). For the electrostatic contribution, two MD trajectories of 110 ps with values of the coupling parameter λ_{elec} equal to [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0] were generated, and the last 100 ps were processed with WHAM.

All the simulations were generated on reduced atomic systems using the generalized solvent boundary potential (GSBP)[26] for the bound ligand and using the spherical solvent boundary potential (SSBP)[25,27] for free ligand inbulk solvent. For the bound ligand simulations, only the atoms within 15 Å from the center of the binding pocket were allowed to move, while the influence of the rest was incorporated with GSBP. A basis set of 16 spherical harmonic functions was used to approximate generalized reaction field. The electrostatic interactions between explicit atoms were not truncated within the simulation sphere, but those beyond 12 Å were represented using an extended electrostatic method.[42] The ligand in solution was simulated using SSBP[25,27] with 400 explicit water molecules. The coupling parameters λ_{rep}, λ_{dis}, and λ_{elec} were the same as for the bound ligand simulations. Those nonbonded coupling parameters affect also the internal interactions of the ligand, thus an additional calculation in vacuo was performed for each ligand which was subtracted from simulations of the solvated and bound ligand to enable the evaluation of solvation free energy. The length of each window simulation was 20 ps equilibration followed by 40 ps of sampling. Langevin dynamics with a friction constant of 5 ps^{-1} was used to ensure thermalization and keep the temperature constant at 300 K. All bonds involving hydrogen atoms were fixed with the SHAKE algorithm.[43] The integration time-step of the Langevin dynamics was 2 fs. For comparison, the free energy calculations with the ligand in bulk solution were repeated with periodic boundary conditions (PBC) at constant pressure using particle mesh Ewald[44] summation for the electrostatics. The PBC simulations comprise the ligand and 125 TIP3P water molecules placed in a cubic box. The Ewald coefficient was set to 0.34 Å^{-1}, and a sixth-order B-spline interpolation was used with a grid of 32 × 32 × 32. The pressure was kept constant at 1.0 atm by using the Langevin piston[45] method with a pressure piston mass of 400.0 amu and a piston collision frequency of 20 ps^{-1}. The temperature was held at 300 K with a mass 3000.0 kcal•ps^2 Nose-Hoover thermostat.[46] The integration time-step was 2 fs. Two 30 ps trajectories were run for each value of the coupling parameters, and the last 20 ps were used in WHAM analysis to compute the free energies. For both SSBP and PBC calculations, the COM of the ligand was weakly restrained to the center of the simulation region using a harmonic potential with a force constant 2.0 kcal/mol/Å^2 to prevent the ligand from drifting.

The binding free energy of benzene, toluene, *o*-xylene, *p*-xylene, ethylbenzene, benzofuran, indene, indole, isobutylbenzene, and *n*-butylbenzene is computed without the conformational restraint. The binding free energy of a polar molecule phenol that does not bind experimentally[47] was also computed for comparison. All of the binding free energy calculations start from the bound structure (with the exception that the benzene bound structure were used for phenol and toluene). The computing time for each ligand binding free energy comprises 156 independent simulations, for a total of about 130 CPU h on a 2.4 GHz Xeon processor.

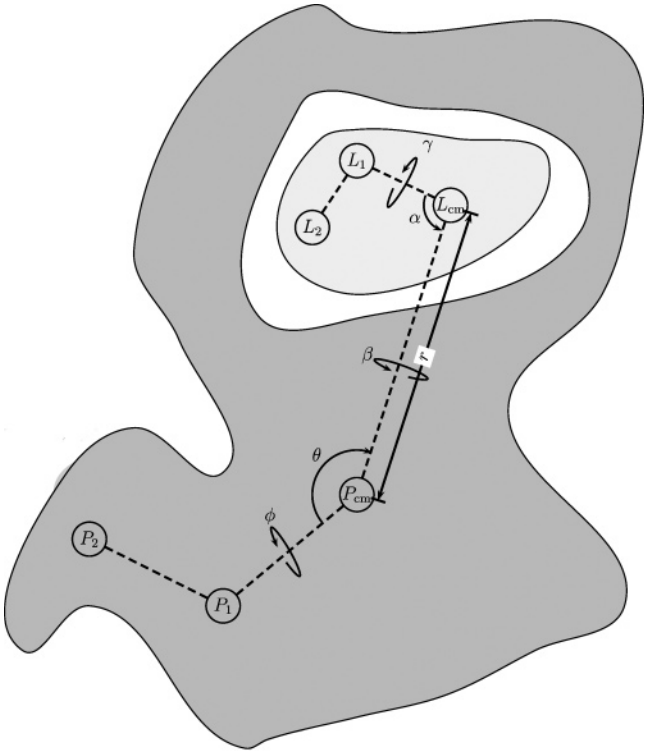


Figure 1. Restraint on the ligand. P_{cm} and L_{cm} are the COM of the protein and ligand, respectively. L_1 and L_2 are ligand atoms. P_1 and P_2 are protein atoms.

C. Restraint Factors. In the present computations, the translational restraint potential is implemented in polar coordinates

$$u_t(r,\theta,\phi) = \frac{1}{2}k_{\rm dist}(r - r_0)^2 + \frac{1}{2}k_{\rm ang}(\theta - \theta_0)^2 + \frac{1}{2}k_{\rm ang}(\phi - \phi_0)^2 \quad (37)$$

where r is the distance between the COM of the ligand and protein; θ is the angle $\angle P_1P_{\rm cm}L_{\rm cm}$; and ϕ is the dihedral angle between plane $P_2P_1P_{\rm cm}$ and $P_1P_{\rm cm}L_{\rm cm}$ (cf. Figure 1). The translational factor F_t is thus

$$F_t = \int_0^\infty dr\, r^2 \int_0^\pi d\theta \sin\theta \int_{-\pi}^{\pi} d\phi\, e^{-\beta u_t(r,\theta,\phi)} \approx r^2{}_0 \sin\theta_0 \sqrt{\frac{(2\pi k_B T)^3}{k_{\rm dist} k_{\rm ang}{}^2}} \quad (38)$$

where the rigid rotor approximation[19] has been used to derive the last equation. The rotational restraint potential has the following form (cf. Figure 1)

$$u_r = \frac{1}{2}k_r(\alpha - \alpha_0)^2 + \frac{1}{2}k_r(\beta - \beta_0)^2 + \frac{1}{2}k_r(\gamma - \gamma_0)^2 \quad (39)$$

where k_r is the force constant; α is the angle $\angle P_{\rm cm}L_{\rm cm}L_1$; β is the dihedral angle between plane $L_1L_{\rm cm}P_{\rm cm}$ and $P_1P_{\rm cm}L_{\rm cm}$; γ is the dihedral angle between plane $L_2L_1L_{\rm cm}$ and $L_1L_{\rm cm}P_{\rm cm}$. The rotational factor is

$$F_r = \frac{1}{8\pi^2}\int_0^\pi d\alpha \int_{-\pi}^{\pi} d\beta \int_{-\pi}^{\pi} d\gamma\, \sin(\alpha) e^{-\beta u_r(\alpha,\beta,\gamma)} \approx \frac{1}{8\pi^2}\left(\frac{2\pi k_B T}{k_r}\right)^{3/2} \quad (40)$$

where the last equation assumed a rigid rotor approximation. All these integrals can also be computed numerically with high accuracy using Simpson's rule. In the present applications, the rigid rotor approximation for F_t and F_r yields essentially identical results to direct numerical integration.

D. Preparation of Protein Structure. Crystal structures of the complexes in the Protein Data Bank were used to generate the all-atom structures. The explicit simulation spheres were henceforth constructed from those all-atom structures. Take the benzene-T4 lysozyme complex, for example. The 15 Å explicit sphere was defined around the center of the binding pocket. Grand canonical Monte Carlo simulation (GCMC)[36] was run to put 107 water molecules in the sphere. Almost all the water molecules were distributed on the surface of the protein. Even though four water molecules could be inserted in binding pocket, they all drifted out within several picoseconds of a subsequent MD run. This is consistent with the observation that the hydrophobic engineered binding pocket in the L99A mutant does not contain water.[16,17] The anchor atoms in the ligand and the protein were assigned automatically by selecting the first non-hydrogen atom in the ligand or protein and then another non-hydrogen atom randomly. For example, for the benzene-bound lysozyme, P_1 was the backbone nitrogen atom of MET1; P_2 was the backbone carbon atom of TYR161. L_1 and L_2 were the two carbon atoms in the benzene ring in meta-position to each other. The corresponding initial values were $r_0 = 5.71568$ Å, $\theta_0 = 120.785°$, $\phi_0 = 151.906°$, $\alpha_0 = 74.9721°$, $\beta_0 = -93.3057°$, and $\gamma_0 = 145.364°$. For the toluene and phenol calculations, the protein coordinates were taken from the crystal structure of the benzene-T4 lysozyme complex (PDB 181L). For the rest of the ligands (*o*-xylene, *p*-xylene, ethylbenzene, isobutylbenzene, *n*-butylbenzene, benzofuran, indole, and indene), the initial simulation spheres were constructed from the correspond crystal structures (PDB 188L, 187L, 1NHB, 184L, 186L, 182L, 185L, and 183L, respectively).

E. Ligand Force Field. Apart for benzene, the binding free energies of phenol, toluene, *o*-xylene, *p*-xylene, ethylbenzene, benzofuran, indene, indole, isobutylbenzene, and *n*-butylbenzene were computed. The partial charges of the ligands were generated from the CHARMM PARAM22 models,[48] except for benzofuran and indene. The partial charges of *o*-xylene and *p*-xylene were created by replacing a hydrogen with a methyl group on the aromatic ring of toluene and setting the partial charge on the carbon attached to the methyl group to zero. Similarly, ethylbenzene, isobutylbenzene, and *n*-butylbenzene models were created by replacing a hydrogen on the benzene ring with ethyl and butyl groups and setting the charge on the corresponding carbon to zero. The model of phenol was constructed by replacing the methyl group of *p*-cresol with a hydrogen of partial charge 0.115*e* and setting the corresponding carbon partial charge to −0.115*e*. For benzofuran and indene, the partial charges were computed by the CHELPG[49] method implemented in Gaussian 98.[50] The rest of the ligand force field parameters were taken from the PARAM22 force field of CHARMM.[48]

IV. Results and Discussion

Table 2 lists the computed binding free energies for nine compounds with available crystal structures in the engineered lysozyme binding pocket. For small ligands such as benzene, toluene, and ethylbenzene, the agreement with experimental results is excellent. For larger ligands, the agreement is reasonable, although the binding free energies appear to be overestimated by a few kilocalories for the xylenes and butylbenzenes. In contrast, the computed free energy is several kilocalories less favorable in the case of indene. In the rest of this section, we examine the details of the binding free energy calculations, analyze the origin of the discrepancies, and conclude with the discussion of the issues that need to be addressed in order to improve the methodology

A. Independence from the Restraining Potentials. According to eq 5, the calculated binding free energy must be formally independent of the strength of the translational and orientational restraining potentials and of the choice of protein anchor atoms used to define those restraints (see also ref 14 or 19 for discussion). The binding of benzene was used as a test system to illustrate the independence of the free energy calculation on the restraints. First, a series of free energy calculations with different translational restraint constants while the rotational restraint is held constant is considered. The results in Table 3 show that the final binding

Table 2: Values of the Binding Free Energy of Aromatic Compounds with Lysozyme[a]

ligand	ΔG_{int}^{site}	ΔG_{int}^{bulk}	$\Delta\Delta G_r + \Delta\Delta G_t^\circ$	ΔV	$\Delta\Omega/8\pi^2$	σ	ΔG_b°	exp[b]
benzene	-11.52 ± 0.09	-0.14 ± 0.12	5.42 ± 0.10	0.87	0.018	12	-5.96 ± 0.19	-5.19 ± 0.16
toluene	-11.72 ± 0.29	-0.25 ± 0.08	7.01 ± 0.06	1.57	3.1×10^{-3}	2	-4.46 ± 0.31	-5.52 ± 0.04
o-xylene	-15.74 ± 0.10	-0.20 ± 0.10	7.95 ± 0.12	0.95	1.4×10^{-3}	2	-7.59 ± 0.19	-4.60 ± 0.06
p-xylene	-16.04 ± 0.02	0.14 ± 0.20	7.12 ± 0.06	0.85	2.9×10^{-3}	4	-9.06 ± 0.21	-4.67 ± 0.06
ethylbenzene	-13.49 ± 0.24	-0.12 ± 0.22	8.33 ± 0.09	0.80	1.8×10^{-3}	1	-5.04 ± 0.34	-5.76 ± 0.07
benzofuran	-16.25 ± 0.02	-1.54 ± 0.20	9.09 ± 0.02	0.90	4.3×10^{-4}	1	-5.62 ± 0.20	-5.46 ± 0.03
indene	-11.02 ± 0.16	-0.28 ± 0.17	8.27 ± 0.03	0.60	2.0×10^{-3}	1	-2.47 ± 0.24	-5.46 ± 0.03
indole	-17.68 ± 0.13	-3.96 ± 0.11	9.48 ± 0.01	1.11	1.6×10^{-4}	1	-4.24 ± 0.17	-4.89 ± 0.06
isobutylbenzene	-18.45 ± 0.18	0.43 ± 0.33	9.21 ± 0.03	0.52	6.2×10^{-4}	1	-9.67 ± 0.38	-6.51 ± 0.06
n-butylbenzene	-18.31 ± 0.31	-1.08 ± 0.18	8.48 ± 0.02	0.57	1.9×10^{-3}	1	-8.75 ± 0.36	-6.70 ± 0.02
phenol	-13.11 ± 0.07	-4.68 ± 0.07	7.55 ± 0.12	1.35	1.9×10^{-3}	2	-0.88 ± 0.13	decoy

[a] The unit of the free energies is kcal/mol. The translational restraint force constants are 10 kcal·mol^{-1}·Å^{-2} for the distant and 200 kcal·mol^{-1}·rad^{-2} for the orientation. The rotational restraint force constant is 200 kcal·mol^{-1}·rad^{-2}. The unit of ΔV is Å^3. The bulk free energies are computed with PBC. The errors are the standard deviation of three simulations from different initial velocities. [b] Taken from ref 31.

Table 3: Values of the Binding Free Energy of Benzene with Different Translational Restraint[a]

	1:50	5:100	10:200	20:400	40:800	100:2000	200:4000
ΔG_{rep}^{site}	9.47 ± 0.11	8.73 ± 0.14	8.49 ± 0.10	8.30 ± 0.08	8.04 ± 0.09	8.18 ± 0.10	8.17 ± 0.16
ΔG_{dis}^{site}	-18.78 ± 0.05	-18.79 ± 0.04	-18.83 ± 0.02	-18.86 ± 0.02	-18.77 ± 0.03	-18.78 ± 0.02	-18.79 ± 0.05
ΔG_{elec}^{site}	-1.19 ± 0.01	-1.18 ± 0.00	-1.18 ± 0.02	-1.18 ± 0.01	-1.19 ± 0.01	-1.2 ± 0.00	-1.19 ± 0.01
ΔG_{int}^{site}	-10.50 ± 0.10	-11.24 ± 0.10	-11.52 ± 0.09	-11.73 ± 0.08	-11.91 ± 0.10	-11.79 ± 0.10	-11.81 ± 0.20
$\Delta\Delta G_r + \Delta\Delta G_t^\circ$	3.96 ± 0.28	4.95 ± 0.11	5.42 ± 0.10	5.65 ± 0.16	5.70 ± 0.25	6.19 ± 0.13	5.98 ± 0.16
$-k_BT\ln(F_tC^\circ)$	3.58	4.48	5.10	5.72	6.34	7.16	7.78
$-k_BT\ln(F_tC^\circ)$ RRA	3.58	4.48	5.10	5.72	6.34	7.15	7.78
ΔV	5.53	1.65	0.87	0.52	0.54	0.30	0.28
$\Delta\Omega/8\pi^2$	0.033	0.021	0.018	0.020	0.025	0.014	0.022
ΔG_b°	-6.40 ± 0.30	-6.15 ± 0.15	-5.96 ± 0.13	-5.94 ± 0.18	-6.07 ± 0.27	-5.46 ± 0.16	-5.69 ± 0.26

[a] The unit of the free energies is kcal/mol. The units of the force constants k_{dist} and k_{ang} are kcal·mol^{-1}·Å^{-2} and kcal·mol^{-1}·rad^{-2}, respectively. The translational restraints are written as k_{dist}: k_{ang}. The rotational restraint of benzene is kept fixed at 200 kcal·mol^{-1}·rad^{-2}. Total is defined as a sum of ΔG_{int}^{site} and $\Delta\Delta G_r + \Delta\Delta G_t^\circ$(the symmetry factor 12 is included in $\Delta\Delta G_r$). The unit of ΔV is Å^3. The solvation free energy ΔG_{bulk} is -0.14 kcal·mol^{-1} and is used to compute ΔG_b°. The errors are the standard deviations from five simulations with different initial velocity. The value of $-k_BT\ln F_t$ computed with numerical integration and its rigid rotor approximation (RRA) are also listed.

free energy is indeed independent of the restraint. In Figure 2, the dependence of ΔG_t^{site},ΔG_r^{site}, ΔG_{int}^{site}, and $-k_BT$ ln(F_tC°) with respect to the strength of the translational restraint is plotted. Despite significant variations in the different components, the sum of $-k_BT\ln(F_tC^\circ)$, $-\Delta G_t^{site}$, and ΔG_{int}^{site} remains nearly invariant with respect to the change of the restraint. At weak restraint, there are small deviations, but the sum reaches a stable value quickly at moderate and strong restraints. This suggest that sampling is less efficient if the restraints are not sufficiently strong. To avoid adding strong artificial forces in the simulation system, we chose the translational restraint force constants to be 10 kcal·mol^{-1}·Å^{-2} for the distance and 200 kcal·mol^{-1}·rad^{-2} for the angles.

Second, a series of free energy calculations with different rotational restraint constants with constant transational restraint are considered. Again, the results in Table 4 show that the final binding free energy is indeed independent of the restraint. In Figure 3, the free energies with respect to the change in rotational restraint are plotted. It is observed that the variations in $-\Delta G_r^{site}$ and $-k_BT\ln(F_r)$ nearly cancel each other, and the resulting binding free energy is independent of the restraint. For the remaining calculations, we chose the rotational restraint force constant to be 200 kcal·mol^{-1}·rad^{-2} for all the rotational angles. The present results do not display the dependence on the restraint force constants encountered by Hamelberg and McCammon in binding free energy computations of water in the interior of proteins.[10] Most likely, the convergence of the free energy results is sensitive to various sampling issues, which can be system-dependent.

The information in Tables 3 and 4 can also be exploited to estimate the accessible volume and orientational freedom of the bound ligand acccording to eqs 23 and 26. The accessible volume, ΔV, of benzene is about 0.5 Å^3; the values at low restraint strength should be discarded because eq 23 holds only in the strong restraint limit. The orientational freedom, $\Delta\Omega/8\pi^2$, is about 1%. These values confirm that benzene does not have much freedom to translate or rotate in the binding pocket. In a previous study, Hermans and Wang[16] also estimated the free energies corresponding to the loss of translational and rotational freedom of benzene bound to T4 lysozyme to be around 5.1 and 4.8 kcal/mol, respectively. However, an exact match is difficult because the definition of $\Delta\Delta G_t^\circ$and $\Delta\Delta G_r$ depends on the applied restraining potentials. In particular, although the final results for the binding free energy do not depend on the applied restraints, a meaningful identification of the contributions associated with the loss of translation or rotation evaluated via eqs 22−26 requires that the applied restraints be centered on the most probable position and orientation of the ligand in the binding pocket. Such estimates may also be related to the results obtained from a quasi-harmonic approximation

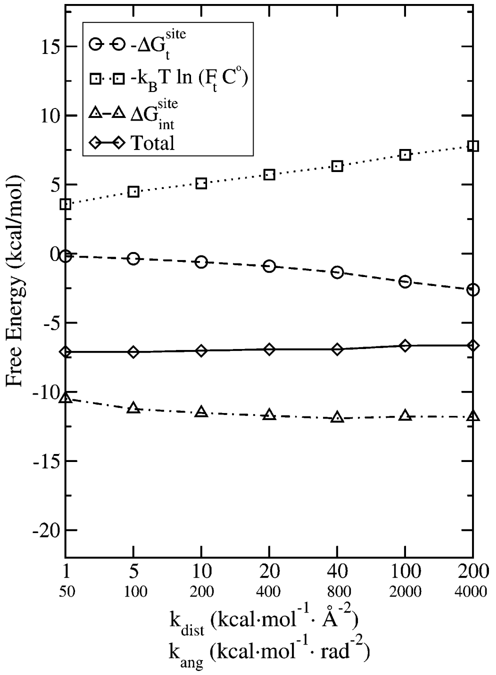


Figure 2. Independence of the binding free energy of benzene on the strength of the translational restraining potential. Total is the sum of ΔG_{int}^{site}, $-\Delta G_t^{site}$, and $-k_B T \ln(F_t C^\circ)$. The top row tick labels of the *x*-axis show the distance force constants k_{dist}. The lower row tick labels show the angular force constants k_{ang}. The restraint on the rotation is fixed at 200 kcal·mol^{-1}·rad^{-2}.

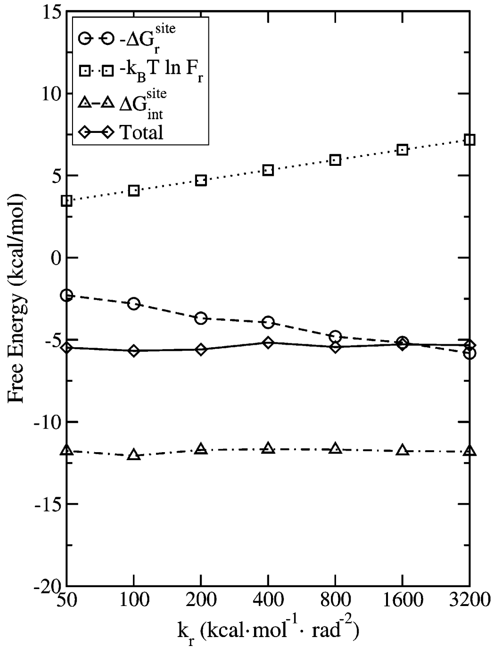


Figure 3. Independence of the binding free energy of benzene on the strength of the orientational restraining potential. Total is the sum of ΔG_{int}^{site}, $-\Delta G_r^{site}$, and $-k_B T \ln F_r$. The translational restraint is held fixed at 10 kcal·mol^{-1}·Å^{-2} for the distance and 200 kcal·mol^{-1}·rad^{-2} for the angles.

by assuming that the position and orientation distribution of the bound ligand is Gaussian. Generally, the accessible volumes for different ligands in the binding pocket have similar values, though there can be considerable reduction of orientational freedom for the largest ligands.

The computed standard binding free energy of benzene is very close to the −5.19 kcal/mol experimental value for all of the restraint strength in both Tables 3 and 4 except the weakest restraint. The final computed standard free energy is comparable to the former calculated binding free energy of −5.14 kcal/mol (the experimental solvation free energy of benzene was used in that calculation).[16]

B. Decomposition of Binding Free Energy. The present free energy calculations distinctively use the WCA separation[27] to separate the LJ potential contributions into strict repulsive and dispersive contributions. Since the free energy computed in the binding site contains the effect from the restraint potentials, it should be noted that the resulting values for ΔG_{rep}^{site}, ΔG_{dis}^{site}, and ΔG_{elec}^{site} are *conditional* free energies. Nonetheless, the WCA decomposition is useful in clarifying the origin of the free energy change occurring upon ligand binding. In the following, we examine the importance of the different interaction components in the case of benzene.

In Figure 4, the progression of the repulsive free energy is shown as a function of the coupling parameter λ_{rep}. It is clear that the variation in the free energy with respect to λ_{rep} is quite different in the binding site and in bulk water. In general, the free energy contribution from the repulsive part of the LJ potential is smoother in the binding site than in the solvent. In particular, a cusp at $\lambda_{rep} = 0.2$ in the solvent environment is not observed in the binding site environment. Clearly, the L99A T4 lysozyme mutant has a preformed

Table 4: Values of the Binding Free Energy Values of Benzene with Rotational Restraint[a]

	50	100	200	400	800	1600	3200
ΔG_{rep}^{site}	8.3	8.01	8.32	8.43	8.40	8.31	8.34
ΔG_{dis}^{site}	−18.92	−18.89	−18.87	−18.89	−18.88	−18.93	−18.98
ΔG_{elec}^{site}	−1.14	−1.18	−1.16	−1.20	−1.20	−1.16	−1.16
ΔG_{int}^{site}	−11.76	−12.06	−11.71	−11.66	−11.68	−11.78	−11.8
$\Delta\Delta G_r + \Delta\Delta G_t^\circ$	5.68	5.79	5.42	5.89	5.64	5.90	5.87
$-k_B T \ln(F_r)$	3.47	4.09	4.71	5.33	5.95	6.57	7.18
$-k_B T \ln(F_r)$ RRA	3.46	4.08	4.70	5.32	5.94	6.56	7.18
ΔV	0.87	0.87	0.87	0.87	0.87	0.87	0.87
$\Delta\Omega/8\pi^2$	0.012	0.0096	0.018	0.0081	0.012	0.0072	0.0084
ΔG_b°	−5.94	−6.13	−6.15	−5.63	−5.90	−5.74	−5.79

[a] The unit of the free energies is kcal/mol. The unit of the rotational restraint force constant is kcal·mol^{-1}·rad^{-2}. All three rotational degrees of freedom are restrained with the same force constant. The translational restraint force constants are fixed at 10 kcal·mol^{-1}·Å^{-2} and 200 kcal·mol^{-1}·rad^{-2}. Total is defined as a sum of ΔG_{int}^{site} and $\Delta\Delta G_r + \Delta\Delta G_t^\circ$(the symmetry factor 12 is included in $\Delta\Delta G_r$). The unit of ΔV is Å^3. The solvation free energy ΔG_{bulk} is −0.14 kcal·mol^{-1}. The value of $-k_B T \ln(F_r)$ computed with numerical integration and its rigid rotor approximation (RRA) are also listed.

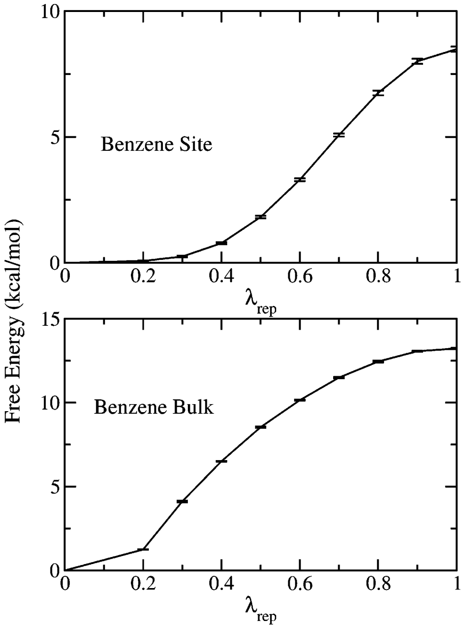


Figure 4. Progression of ΔG_{rep} with respect to the coupling parameter λ_{rep} for benzene in the binding site (top) and in bulk water (bottom).

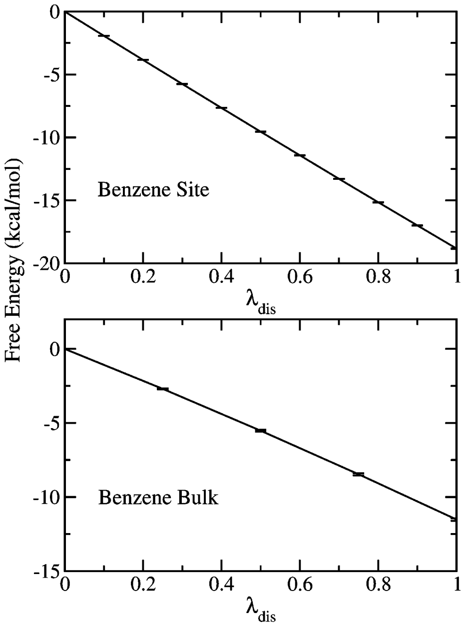


Figure 5. Progression of ΔG_{dis} with respect to the coupling parameter λ_{dis} for benzene in the binding site (top) and in bulk water (bottom).

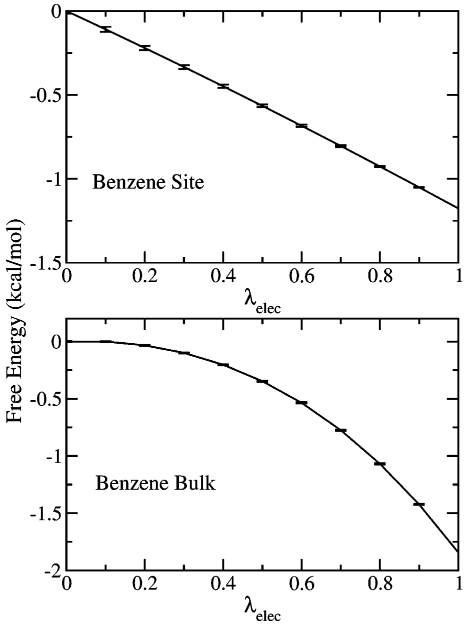


Figure 6. Progression of ΔG_{elec} with respect to the coupling parameter λ_{elec} for benzene in the binding site (top) and in bulk water (bottom).

cavity of the appropriate size for benzene, hence the early stages of repulsion are introduced without significant free energy cost. As a consequence, the free energy contribution from the repulsive part of the LJ potential favors ligand binding. The overall gain is about 5 kcal/mol in favor of ligand binding.

In Figure 5, the progression of the dispersive free energy is plotted as a function of the coupling parameter λ_{dis}. In both bulk solvent and in the binding site, the dispersive free energy increases linearly with λ_{dis}, though the slope is different in the two environments. The origin of the difference can be directly traced back to the number density of van der Waals interaction centers per unit volume surrounding the ligand. This number is invariably larger in a protein compared with bulk water, which is a fairly *loose* and extended liquid. A similar observation was made by Levy and co-workers in developing implicit solvent models.[51] An important consequence is that the free energy contribution from the dispersive van der Waals interaction favors ligand binding. This is also related to trends previously observed with the linear interaction energy (LIE) approximation.[52] The linear progression of the dispersive free energy with λ_{dis} reflects the fact that there is no significant structural change as this interaction is switched-on during the simulation. That also leads to a computational advantage that this contribution converges very easily with a reduced number of simulation windows.[27] For the remaining calculations with ligands in the binding site, values of [0, 0.25, 0.5, 0.75, 1.0] were used for the coupling parameter λ_{dis} with trajectories of 100 ps.

In Figure 6, the progression of the electrostatic free energy of benzene is plotted with respect to the coupling parameter, λ_{elec}. The free energy varies linearly with the coupling parameter for benzene in the binding site, whereas it displays a typical quadratic behavior for benzene in bulk solvent. This indicates that the protein environment does not respond (electrostatically) to the charging of benzene. The ligand–protein interaction is dominated by a constant static field. This can be compared with the progression of the electrostatic free energy contribution of two polar ligands. The variation of the electrostatic free energy of indole and phenol as a function of λ_{elec} is shown in Figures 7 and 8, respectively. In the case of indole (a binder), the electrostatic free energy in the protein site is linear with respect to the coupling parameter λ_{elec}, whereas for phenol (a nonbinder) the progression of the electrostatic free energy in the binding site is not linear anymore. The difference results from the stronger polarity of phenol, which induces an electrostatic response in the protein. The simple progression of the

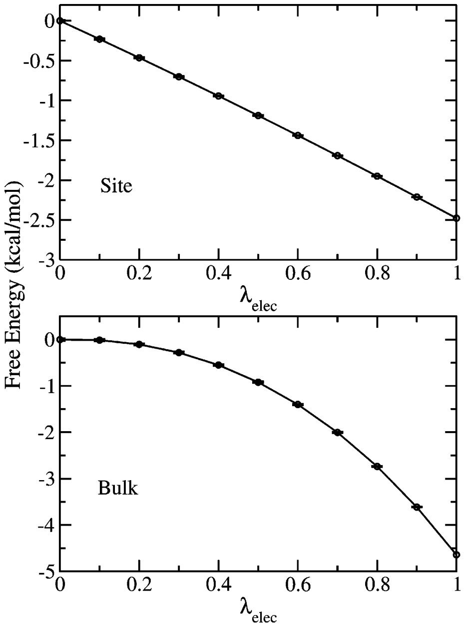


Figure 7. Progression of ΔG_{elec} with respect to the coupling parameter λ_{elec} for indole in the binding site (top) and bulk water (bottom).

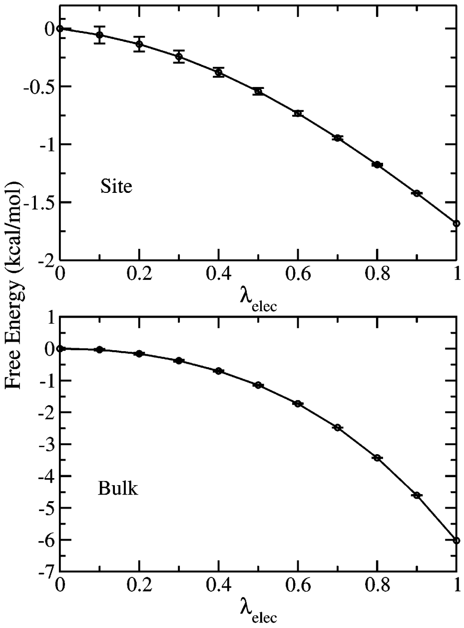


Figure 8. Progression of ΔG_{elec} with respect to the coupling parameter λ_{elec} for phenol in the binding site (top) and in bulk water (bottom).

electrostatic free energy in the binding site also suggests that it might be possible to reduce the number of sampling windows considerably without loss of precision. These observation suggest that the average electrostatic energy used in the linear interaction energy approximation (LIE)[52] may yield accurate results in the case of nonpolar ligands.

In Table 6, the free energy contributions in the binding site and bulk water are listed. The decomposition of solvation free energies computed with SSBP and PBC is also shown. As previously noted,[27] calculations with SSBP tend to yield slightly more unfavorable solvation free energy for the nonpolar molecules compared with PBC computations. For the final results of the present study, the values from PBC are used. Generally, both the repulsive and dispersive contributions favor the binding process, whereas the electrostatic contribution does not. For all of the ligands except indole, electrostatics contributes about 1 kcal/mol of unfavorable free energy. For the more polar indole, the electrostatic contribution to binding is 2.3 kcal/mol. The favorable repulsive contribution varies from molecule to molecule. The dispersive free energy has a clear dependence on the ligand size. As their size increases, the dispersive contribution becomes more and more favorable for binding. Ligands of similar size have similar dispersive contributions. This does seem to support the empirical observation that the *addition* of atoms improves the binding affinity (as long as they do not cause steric clashes).[53] For all of the ligands, most of the binding free energy comes from the nonpolar contribution. Even for the nonbinding molecule, phenol, the nonpolar contribution is comparable to that of toluene; it is the 4.4 kcal/mol of unfavorable electrostatic contribution that causes the unfavorable binding free energy for phenol. It is worth emphasizing that nonpolar contributions always exist independent of electrostatic contributions. Therefore, given the structural arrangement for the formation of the binding pocket buried in a protein, it is likely that nonpolar contributions always favor binding, for the molecules that do bind. The results of water binding free energy in different receptors[10] indeed show favorable nonpolar contributions, although the electrostatic contribution dominates.

C. Force Field and Free Energy Calculation. Accurate free energy calculations require highly refined force fields.[48,54] The atomic model affects the binding free energy in two ways: the solvation free energy in the bulk solvent and the free energy in the binding site. Consequently, the atomic model must be accurate for the ligand, protein, and water. The quality of the models in water solution can be evaluated by comparing computed solvation free energy with experimental values. In the present results, the calculated solvation free energies are about 1 kcal/mol more unfavorable. This gives the possibility of more favorable computed binding free energy, although, to arrive at this kind of conclusion, the deviation of the site free energy is ignored. The compatibility with the protein is more difficult to test directly. Here we resort to a less rigorous test of the ligand models by investigating how much the binding free energy changes when the ligand partial charges are changed.

First, the charges of the compounds are perturbed from CHARMM charges to AM1-CM2 AMSOL[55] charges. The simulation windows and sampling length are the same as the site and bulk electrostatic free energy calculation, respectively. The results are shown in Table 7. For all compounds except phenol, AMSOL charges give more favorable solvation and site free energies. The AMSOL electrostatic free energy of phenol is slightly more positive than that of the CHARMM charges. The AMSOL charges for indole and isobutylbenzene give 3 kcal/mol more favorable solvation free energies. In general, the site and solvation free energy change in the same direction when the ligand charges are changed. Therefore, for most of the compounds, the difference in the binding free energy with

Table 5: Values of the Solvation Free Energy of the Aromatic Ligand Compounds Computed with SSBP and PBC[a]

ligand		$\Delta G_{\text{rep}}^{\text{bulk}}$	$\Delta G_{\text{dis}}^{\text{bulk}}$	$\Delta G_{\text{elec}}^{\text{bulk}}$	$\Delta G_{\text{int}}^{\text{bulk}}$	$\Delta G^{\text{exp } b}$
benzene	SSBP	15.19	−13.30	−1.86	0.03	−0.87
	PBC	13.22 ± 0.04	−11.49 ± 0.11	−1.87 ± 0.01	−0.14 ± 0.12	
toluene	SSBP	17.27	−15.27	−1.73	0.27	−0.76
	PBC	14.59 ± 0.23	−13.17 ± 0.13	−1.67 ± 0.02	−0.25 ± 0.08	
o-xylene	SSBP	19.48	−17.05	−1.64	0.78	−0.90
	PBC	16.14 ± 0.10	−14.72 ± 0.05	−1.62 ± 0.01	−0.20 ± 0.10	
p-xylene	SSBP	19.64	−17.06	−1.56	1.02	−0.81
	PBC	16.33 ± 0.02	−14.64 ± 0.20	−1.55 ± 0.03	0.14 ± 0.20	
ethylbenzene	SSBP	19.05	−16.98	−1.86	0.22	−0.80
	PBC	16.38 ± 0.18	−14.65 ± 0.08	−1.84 ± 0.04	−0.12 ± 0.22	
benzofuran	SSBP	18.40	−17.31	−2.51	−1.42	
	PBC	15.46 ± 0.13	−14.81 ± 0.11	−2.19 ± 0.02	−1.54 ± 0.20	
indene	SSBP	18.99	−16.97	−1.92	0.10	
	PBC	16.16 ± 0.19	−14.57 ± 0.12	−1.87 ± 0.01	−0.28 ± 0.17	
indole	SSBP	18.41	−17.72	−4.68	−3.99	
	PBC	15.88 ± 0.15	−15.19 ± 0.12	−4.64 ± 0.02	−3.96 ± 0.11	
isobutylbenzene	SSBP	23.12	−20.67	−1.78	0.67	−0.44
	PBC	19.51 ± 0.14	−17.37 ± 0.20	−1.70 ± 0.03	0.44 ± 0.33	
n-butylbenzene	SSBP	22.20	−20.64	−1.98	−0.42	−0.40
	PBC	18.37 ± 0.14	−17.39 ± 0.15	−2.07 ± 0.02	−1.08 ± 0.18	
phenol	SSBP	15.92	−14.37	−6.18	−4.63	−6.62
	PBC	13.76 ± 0.04	−12.42 ± 0.05	−6.03 ± 0.03	−4.68 ± 0.07	

[a] The unit of free energies is kcal/mol. $\Delta G_{\text{int}}^{\text{bulk}}$ is defined as a sum of all the free energy components (repulsion, dispersion, and electrostatics). The errors are the standard deviation of three simulations from different initial velocities. [b] Taken from ref 65.

Table 6: Decomposition of the Interaction Free Energy in Bulk Solution and in the Binding Site for the Various Ligand Compounds[a]

ligand		ΔG_{rep}	ΔG_{dis}	ΔG_{elec}	ΔG_{int}	$\Delta G^{\text{exp}b}$
benzene	site	8.49 ± 0.10	−18.83 ± 0.02	−1.18 ± 0.02	−11.52 ± 0.09	
	bulk	13.22 ± 0.04	−11.49 ± 0.11	−1.87 ± 0.01	−0.14 ± 0.12	−0.87
toluene	site	11.40 ± 0.17	−22.19 ± 0.11	−0.93 ± 0.01	−11.72 ± 0.29	
	bulk	14.59 ± 0.23	−13.17 ± 0.13	−1.67 ± 0.02	−0.25 ± 0.08	−0.76
o-xylene	site	10.60 ± 0.13	−25.65 ± 0.03	−0.69 ± 0.01	−15.74 ± 0.10	
	bulk	16.14 ± 0.10	−14.72 ± 0.05	−1.62 ± 0.01	−0.20 ± 0.10	−0.90
p-xylene	site	9.11 ± 0.03	−24.46 ± 0.04	−0.69 ± 0.01	−16.04 ± 0.02	
	bulk	16.33 ± 0.02	−14.64 ± 0.20	−1.55 ± 0.03	0.14 ± 0.20	−0.81
ethylbenzene	site	13.53 ± 0.23	−26.02 ± 0.09	−0.99 ± 0.00	−13.49 ± 0.24	
	bulk	16.38 ± 0.18	−14.65 ± 0.08	−1.84 ± 0.04	−0.12 ± 0.22	−0.80
benzofuran	site	11.87 ± 0.02	−25.95 ± 0.03	−2.17 ± 0.00	−16.25 ± 0.02	
	bulk	15.46 ± 0.13	−14.81 ± 0.11	−2.19 ± 0.02	−1.54 ± 0.20	
indene	site	16.27 ± 0.13	−25.80 ± 0.03	−1.48 ± 0.01	−11.02 ± 0.16	
	bulk	16.16 ± 0.19	−14.57 ± 0.12	−1.87 ± 0.01	−0.28 ± 0.17	
indole	site	11.69 ± 0.10	−26.90 ± 0.03	−2.48 ± 0.01	−17.68 ± 0.13	
	bulk	15.88 ± 0.15	−15.19 ± 0.12	−4.64 ± 0.02	−3.96 ± 0.11	
isobutylbenzene	site	14.49 ± 0.22	−32.04 ± 0.04	−0.90 ± 0.00	−18.45 ± 0.18	
	bulk	19.51 ± 0.14	−17.37 ± 0.20	−1.70 ± 0.03	0.44 ± 0.33	−0.44
n-butylbenzene	site	15.15 ± 0.22	−32.45 ± 0.13	−1.00 ± 0.04	−18.31 ± 0.31	
	bulk	18.37 ± 0.14	−17.39 ± 0.15	−2.07 ± 0.02	−1.08 ± 0.18	−0.40
phenol	site	9.18 ± 0.06	−20.61 ± 0.05	−1.68 ± 0.08	−13.11 ± 0.07	
	bulk	13.76 ± 0.04	−12.42 ± 0.05	−6.03 ± 0.03	−4.68 ± 0.07	−6.62

[a] The solvation free energies are computed with PBC. The unit of the free energies is kcal/mol. The translation restraint force constants are 10 kcal·mol^{-1}·Å^{-2} for the distant and 200 kcal·mol^{-1}·rad^{-2} for the orientation. The rotational restraint force constant is 200 kcal·mol^{-1}·rad^{-2}. ΔG_{int} is defined as a sum of all the free energy components (repulsion, dispersion, and electrostatics). The errors are the standard deviation of three simulations from different initial velocities. [b] Taken from ref 65.

CHARMM and AMSOL charges are not very pronounced. For indole, the deviation in solvation free energy is expected because the CHARMM solvation free energy is about 2 kcal/mol less favorable (this is estimated with the experimental solvation free energy of 3-methylindole) than the experimental value. The indole model was recently revised to have a more realistic charge distribution and a more accurate solvation free energy in liquid water.[56] However, the binding free energy of −3.18 kcal/mol computed with the revised model is not as good in agreement with experiment than the

Table 7: Free Energy Differences from CHARMM Charges to AMSOL Charges[a]

ligand	$\Delta\Delta G_{\text{elec}}^{\text{site}}$	$\Delta\Delta G_{\text{elec}}^{\text{bulk}}$	$\Delta\Delta G_{\text{b}}^{\circ}$
benzene	−0.08	−0.36	0.28
ethylbenzene	−0.88	−0.89	0.01
isobutylbenzene	−1.27	−2.62	1.35
indole	−1.11	−3.23	2.12
n-butylbenzene	−0.64	−0.73	0.09
o-xylene	−1.14	−1.52	0.38
phenol	0.17	0.49	−0.32
p-xylene	−0.13	−1.53	1.40
toluene	−0.51	−0.91	0.40

[a] Free energies in kcal/mol.

Table 8: Free Energy Differences from CHARMM Charges to CHELPG Charges[a]

ligand	$\Delta\Delta G_{\text{elec}}^{\text{site}}$	$\Delta\Delta G_{\text{elec}}^{\text{bulk}}$	$\Delta\Delta G_{\text{b}}^{\circ}$
benzene	0.16	0.55	−0.39
ethylbenzene	0.26	0.43	−0.17
isobutylbenzene	0.74	−3.09	3.83
indole	0.22	0.23	−0.01
n-butylbenzene	0.48	−0.02	0.50
o-xylene	0.27	0.30	−0.03
phenol	−0.23	0.16	−0.39
p-xylene	−0.56	0.25	−0.81
toluene	0.58	0.39	0.19

[a] Free energies in kcal/mol.

original model. The decomposition of the free energies indicates that the difference results mostly from changes in the electrostatic contribution in the bulk solvent, which varied from −4.6 kcal/mol to nearly −5.8 kcal/mol, while the other terms have a similar magnitude. The interaction free energy in the binding site remain almost unchanged compared with the original model. The revised model is better solvated by improving the electrostatic interactions. The total nonpolar contribution to hydration for the revised model is 0.32 kcal/mol. This suggests that the balance of nonpolar (dispersion) and polar (electrostatic) contribution for the new model may not be optimal; even though the solvation free energy is more accurate, the binding free energy of the ligand inside a nonpolar protein cavity is not well described. This analysis illustrates (but does not resolve) the difficulties in designing a nonpolarizable force field with fixed partial charges that is simultaneously accurate in different environments such as liquid water and a nonpolar cavity in the interior of a protein.

Table 8 lists the changes of free energies from CHARMM to CHELPG charges. In this case, the solvation free energies computed with CHELPG charges are very close to those computed with CHARMM charges except for isobutylbenzene. The CHELPG charges overestimate the dipole moment for isobutylbenzene, yielding a large difference in solvation free energy. Table 9 lists the change of free energies from CHELPG to AMSOL charges for benzofuran and indene (there are no CHARMM partial charges for these molecules). The difference in solvation free energy is close to 3 kcal/mol. This large difference implies that the computed solvation free energy might not be accurate for benzofuran and indene, although we have found no experiment solvation free energy for these molecules. These tests with different compound force field models clearly indicate that binding free energy calculations place a very high demand on the force field, although, for many of the test compounds, the variations between different parameter sets are acceptable.

Table 9: Free Energy Differences from CHELPG Charges to AMSOL Charges[a]

ligand	$\Delta\Delta G_{\text{elec}}^{\text{site}}$	$\Delta\Delta G_{\text{elec}}^{\text{bulk}}$	$\Delta\Delta G_{\text{b}}^{\circ}$
benzofuran	−0.83	−3.07	2.24
indene	−0.38	−2.97	2.59

[a] Free energies in kcal/mol.

D. Flexibility and Conformational Sampling. In the present study, the alchemical FEP simulations were started from the experimentally determined structure of the protein−ligand complex. It obviously is advantageous to make use of any available high-resolution structural information about the protein−ligand complex, whenever this is possible. In the case of small ligands such as benzene, the change in the protein conformation is quite small. As the ligand vanishes in the alchemical simulation, the mobile atoms of the protein in the GSBP inner region relax to their average position in the apo structure, and the alchemical free energy simulation converges with no problems. The flexibility of a binding pocket makes an important contribution to the binding free energy.[57] In this context, the current FEP simulations protocol with the GSBP approximation is able to incorporate the influence of the local angstrom-scale thermal fluctuations of the protein and ligand as well as the loss of translational and orientational freedom into the calculated binding free energy. Nonetheless, it should noted that the finite GSBP simulation system gives rise to some artifacts. The average fluctuations of the mobile backbone atoms in the inner region of the GSBP simulation are slightly smaller than those from the fully solvated protein with PBC. There is an obvious compromise in choosing the size of the inner GSBP region, reducing the computational cost, and realistically simulating the dynamics of the atoms surrounding the ligand.

There can also be additional difficulties to sample ligands that are highly flexible in bulk solution but become restricted in the bound state.[20] This can be illustrated with the case of *n*-butylbenzene, as the movement of the butyl group becomes restricted upon binding. In this case, brute force sampling of the butyl group internal rotation might be inefficient. To address this issue, we introduce a conformational restraint on the butyl group and recomputed the free energies. The conformational restraint potential is $k_c\zeta^2$, where $k_c = 2$ kcal·mol·Å^{-2} and ζ is the RMSD from the bound conformation. The conformational factor (free energy) is computed by introducing a reference RMSD in the restraint potential, $k_c(\zeta - \zeta_0)^2$. The RMSD is sampled at each ζ_0 values of 0, 0.2, 0.4, 0.6, 0.8, ..., 1.8, 2.0 Å, for 60 ps, through which the unbiased distribution of RMSD at $\zeta_0 = 0$ Å, $\rho(\zeta)$, is computed.[58] The comparisons of the calculated free energies are listed in Table 10. In the particular case of *n*-butylbenzene, the forced sampling with the conformational restraint does not change significantly the resulting computed binding free energy. The system is dominated by a single bound conformation for the ligand, and only the local fluctuation

Table 10: Decomposition of the Binding Free Energy for *n*-Butylbenzene Binding Computed with Conformational Restraint[a]

ligand		ΔG_{rep}	ΔG_{dis}	ΔG_{elec}	ΔG_{int}	ΔG_{c}	$\Delta\Delta G_{\mathrm{r}} + \Delta\Delta G_{\mathrm{t}}^{\circ}$	$\Delta G_{\mathrm{b}}^{\circ}$
n-butylbenzene with conf. restr.	site	14.95	−32.22	−0.95	−18.22	0.23	8.61	−7.99
	bulk	21.51	−20.90	−1.66	−1.05	0.80		
n-butylbenzene no conf. restr.	site	15.15	−32.45	−1.00	−18.31	0	8.48	−8.75
	bulk	18.37	−17.39	−2.07	−1.08	0		

[a] The unit of the free energies is kcal/mol. The translation restraint force constants are 10 kcal·mol^{-1}·Å^{-2} for the distant and 200 kcal·mol^{-1}·rad^{-2} for the orientation. The rotational restraint force constant is 200 kcal·mol^{-1}·rad^{-2}. The conformational restraint is $k_c\zeta^2$, where $k_c = 2$ kcal·mol·Å^{-2}, and ζ is the RMSD of *n*-butylbenzene from the bound conformation.

around it contributes to the binding free energy. Therefore, the conformational restraint is not essential in this case. However, for highly flexible ligands with many rotable groups the conformational restraint can help to achieve a better sampling.[59]

Of more significance are the difficulties encountered if the protein conformation in the complex differs significantly from that of the apo state, as in the case of the larger ligands. For example, there is considerable shift in the backbone position and large B-factor in helix F of the T4 lysozyme L99A mutant upon binding the largest ligands;[31] the RMSD per residue between the bound and apo structures for isobutylbenzene reaches a maximum of 2.5 Å, centered mostly around helix F (residues 109−113). An insufficiently sampled alchemical simulation started from the protein−ligand complex may underestimate the reversible work needed to distort and adapt the protein conformation to the ligand (the *strain* energy). The magnitude of the strain energy associated with distortion of the protein in the region of the binding site of T4 lysozyme has been estimated to be on the order of 2 kcal/mol by NMR experiments.[60] Because of the insufficient simulation time in the present calculations, the side chain atoms in helix F do not relax to their position in the apo structure and the calculated binding free energy is slightly too favorable for the largest ligands. This problem could be addressed by generating longer simulations, or alternatively, by generalizing the present theoretical formulation to introduce a conformational restriction potential u_c for biasing the conformation of the protein. Following this treatment, the influence of the bias could be rigorously removed and unbiased to obtain binding free energies without generating prohibitively long simulations. Further work along those lines is being pursued.

E. Importance of Accurate Starting Structure of the Complex. The calculated binding free energy of indene is several kcal/mol more unfavorable than the experimental value. Because the result for this molecule seems unaccountably inaccurate compared to the other ligands, it is particularly informative to understand the origin of this failure. According to the decomposition of the free energy, the difference arises mainly from the repulsive contribution in the binding site. This is indicative that, somehow, there is not enough space in the binding pocket for indene. The crystal structure shows that the binding of indene in the engineered lysozyme site induces increased disorder (as indicated by the temperature factors) and positional displacement of the helix F (residues 109−113).[31] Similar changes from the apo structure are observed for other ligands with a similar shape. But the other ligands (which are quite similar) do not display an unfavorable repulsive contribution as indene does. Because the displacement of helix F between the indene and *o*-xylene bound lysozyme is quite similar, we decided to compute the binding free energy of indene using the crystallographic structure of lysozyme with *o*-xylene bound (PDB 188L). As seen in Table 11, using this protein conformation, there is a significant reduction of the repulsive free energy (7 kcal/mol) upon binding. The resulting total binding free energy ends up being −8.10 kcal/mol (compared to −2.47 kcal/mol when indene is bound to its own crystal structure), which is 3 kcal/mol more favorable than the experimental value for indene. What could give rise to such discrepancy? Upon close examination, it is found that the determining factor is the χ_1 dihedral angle of Val111. When the side chain of Val111 in the indene bound crystal structure (PDB 183L) is set to the rotamer observed in the *o*-xylene bound crystal structure (PDB 188L), the computed free energy of indene becomes −8.06 kcal/mol. The unfavorable binding free energy of indene computed with crystal structure PDB 183L results directly from the rotamer conformation of Val111.

Table 11: Decomposition of the Binding Free Energy for Indene Computed from a Different Starting Structure of Lysozyme[a]

lysozyme structure	$\Delta G_{\mathrm{rep}}^{\mathrm{site}}$	$\Delta G_{\mathrm{dis}}^{\mathrm{site}}$	$\Delta G_{\mathrm{elec}}^{\mathrm{site}}$	$\Delta G_{\mathrm{int}}^{\mathrm{site}}$	$\Delta\Delta G_{\mathrm{r}} + \Delta\Delta G_{\mathrm{t}}^{\circ}$	$\Delta G_{\mathrm{b}}^{\circ}$
PDB 183L	16.27	−25.80	−1.48	−11.02	8.27	−2.47
PDB 188L	9.04	−25.37	−0.82	−17.15	8.77	−8.10
PDB 183L*	9.12	−24.93	−0.95	−16.79	8.42	−8.06

[a] PDB 183L is the lysozyme crystal structure with indene bound. PDB 188L is the lysozyme crystal structure with *o*-xylene bound. PDB 183L* is the modified PDB 183L lysozyme crystal structure in which the χ_1 of Val111 changed to the value −45.48°, as in PDB 188L, a *o*-xylene bound structure. The unit of the free energies is kcal/mol.

In the *o*-xylene bound crystal structure, the backbone dihedral angles of Val111 are $\phi = -88.67°$ and $\psi = 1.14°$. Given the backbone dihedral angles, the rotamer library[61] predicts a single possible χ_1 of −60° (g$^-$ configuration). The side chain dihedral angle in the crystal structure confirms the prediction with $\chi_1 = -45.58°$. Similarly, in the indene bound crystal structure, backbone dihedral angles of Val111 are $\phi = -82.22°$ and $\psi = 5.68°$, which leads to a predicted χ_1 angle of −60° in the same g$^-$ configuration. However, the crystal structure has a χ_1 value of 171.42°. Analysis of all the other residues within 4 Å of all the nine complexes shows no such deviation from the rotamer library prediction in χ_1. In the indene bound crystal structure (PDB 183L), the backbone carbonyl oxygen of Val111 is only 2.74 Å away from the closest γ carbon atom. Examination of the electron density map of the indene bound structure[31] reveals that only one of the γ carbon atom of the valine side chain is clearly resolved. In fact, both side chain dihedral χ_1 values would be compatible with the experimental density map. These structural observations support the idea that the Val111 side chain dihedral angle in the indene bound crystal structure (PDB 183L) might be incorrect.

The observation that the rotameric state of a single residue (in the crystal structure) in direct contact with the ligand can lead to several kcal/mol difference in the binding free energy is alarming but not unexpected. Clearly, this type of side chain barrier crossing, which takes place on a relatively slow time scale (typically, the barriers are several kcal/mol, e.g., see ref 62), could be sampled during very long FEP simulations.[63] However, having to generate prohibitively long trajectories and, in effect, trying to locate the correct conformation of the complex using brute force MD is obviously inefficient. An approach relying on a systematic search with a side chain rotamer library would be more efficient. Furthermore, it is antithetical to the spirit of the current methodological effort aimed at evaluating the binding free energy corresponding to a given state of the complex (scoring). These considerations highlight the great importance of an accurate starting structure (either from experiments or from docking followed by careful side chain optimization) for binding free energy calculation. They also show that much caution should be used to adjust empirical scoring functions on the basis of static X-ray protein−ligand structural complexes.

V. Conclusions

In this paper, we have presented a general statistical mechanical formulation of the standard binding constant and its application to the calculation of the binding free energy of aromatic ligands to an engineered nonpolar binding pocket in T4 lysozyme. The calculation of the free energy is done with several restraint potentials on the ligand translational and rotational (relative to the protein) degrees of freedom to improve the sampling. Those restraints are used to hold the uncoupled ligand in the binding pocket. We have shown that the final result (binding free energy) is independent of the choice of restraint potential. The definition of standard states appears naturally in the theoretical formulation. To decrease computational time, the MD simulations in the binding site were carried out with a reduced system including only atoms in a 15-Å sphere around the binding pocket, while the rest of the system is treated with a generalized solvent boundary potential (GSBP).[26] The solvation free energy calculation was done with both spherical solvent boundary potential (SSBP)[25] and periodic boundary potential. With GSBP the binding free energy calculation for a single ligand is achieved in 130 CPU h on a 2.4 GHz Xeon processor.

According to the different character of molecular interactions, the nonbonded interaction free energy was separated into repulsive, dispersive, and electrostatic contributions. Generally, the repulsive and dispersive interaction contribute significantly to the binding free energy, while the electrostatic interaction is slightly unfavorable. The repulsive free energy contribution is generally smaller in the binding site than in the bulk solvent, which is consistent with the existence of a preformed binding cavity in the protein. Furthermore, the dispersive van der Waals free energy contributes favorably to binding because the protein environment is more dense than bulk water. As expected, it is the large unfavorable contribution in the electrostatic free energy, insufficiently compensated by the nonpolar protein cavity, that that makes binding unfavorable in the case of the polar molecule phenol.

The agreement of calculated binding free energies with experimental values is generally satisfactory. For small ligands such as benzene, toluene, and ethylbenzene, the results are in excellent agreement with experiment, while for larger ligands, the agreement is reasonable though the computed free energies are somewhat more favorable than the experimental ones. Beyond specific successes and failures, the current study provides a good opportunity to highlight many of the key issues needed to predict accurate binding free energy using atomic models with realistic interactions and, in many ways, outlines the challenges that lie ahead. Those are as follows: (1) correct statistical mechanical formulation of the binding free energy, (2) accuracy and transferability of the atomic potential function, (3) access to accurate structure of the protein−ligand complex, and finally, (4) adequate conformational sampling of the molecular systems.

First and foremost, a valid statistical mechanical treatment of the binding process should establish the basis of any computational treatment. In particular, a translational restraint for the uncoupled ligand is an unavoidable ingredient of a sound double decoupling method (DDM) based on alchemical free energy. Trying to assess the convergence and precision of extensively long alchemical free energy simulations generated without translational restraint for the uncoupled state[64] only adds to the confusion because the significance of such calculations is unclear.

Calculations of absolute binding free energies place high demand on the accuracy of a force field. To examine how much the force field could change the computed results, we computed the change in binding free energies when partial charges of the ligands are changed. While the difference is not pronounced in general, for indene and benzofuran, the change in partial charges gives rise to differences up to 3 kcal/mol in the bulk electrostatic free energy. Even though the influence of the partial charge was certainly limited by the nonpolar nature of the ligands and binding pocket in the present case, such differences in binding free energies cannot generally be ignored.

Issues of protein conformation, flexibility, and thermal fluctuations are also particularly important. Having a good starting structure is essential for accurate binding free energy computations. This was dramatically illustrated in the case of indene, which yielded an incorrect free energy by several kcal/mol due to a single side chain misconfiguration in a high resolution crystallographic structure. Nonetheless, assuming that a good starting structure is available, the current binding free energy calculation method can be precise. The good news is that the small local angstrom-scale thermal fluctuations of the ligand and protein are sampled very well during relatively modest MD simulations. Simulation of a reduced system with GSBP may have limited the fluctuation of F-helix, which is somewhat displaced upon binding of the larger ligands. The loss of conformational freedom of a flexible ligand upon binding requires special attention to get results of quantitative accuracy. A PMF approach based on the RMSD of the ligand relative to its bound conformation is used to obtain an efficient sampling of the ligand conformation restriction. This approach, which was illustrated with the case of *n*-butylbenzene, could be generalized to include biasing restraints on the conformation of the receptor as well as the ligand. Sampling the thermal fluctuations and motions in the system is absolutely essential; it simply cannot be avoided. It seems highly unlikely that simple effective scoring schemes based on a static (fixed) snapshot of the protein−ligand complex would ever be able to incorporate the effects of those fluctuations correctly.

The present effort shows that computationally tractable methods yielding first-principle absolute binding free energies are promising. With the development of more accurate force field and efficient sampling of the protein flexibility with conformational restraints, binding free energy predictions could soon become a useful tool in pharmaceutical studies.

Acknowledgment. We thank Brian Shoichet for helpful discussions and for providing AMSOL charges. We are grateful to Hyung-June Woo for providing the initial coordinates of the benzene-T4 lysozyme complex. We also thank Yanxiang Zhao for help with the analysis of the lysozyme crystal structure and Alexander D. MacKerell, Jr. for providing the new indole model. This work was supported by the National Science Foundation through Grant MCB-0415784.

References

(1) Ehrlich, P. Address in pathology on chemotherapeutics: scientific principles, methods, and results. *Lancet* **1913**, *182*, 445−451.

(2) Vindigni, A. Energetic dissection of specificity in serine proteases. *Comb. Chem. High Throughput Screening.* **1999**, *2*, 139−153.

(3) Cheng, A. C.; Calabro, V.; Frankel, A. D. Design of rna-binding proteins and ligands. *Curr. Opin. Struct. Biol.* **2001**, *11* (4), 478−484.

(4) Garvie, C. W.; Wolberger, C. Recognition of specific dna sequences. *Mol. Cell.* **2001**, *8* (5), 937−946.

(5) Smith, P. E.; Pettitt, B. M. Modeling solvent in biomolecular systems. *J. Phys. Chem.* **1994**, *98*, 9700−9711.

(6) Orozco, M.; Luque, F. J. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **2000**, *100*, 4187−4225.

(7) Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligand to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644−2676.

(8) Pearlman, D. A.; Charifson, P. S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **2001**, *44*, 3417−3423.

(9) Dixit, S. B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.

(10) Hamelberg, D.; McCammon, J. A. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *J. Am. Chem. Soc.* **2004**, *126*, 7683−7689.

(11) Clark, M.; Guarnieri, F.; Shkurko, I.; Wiseman, J. Grand canonical Monte Carlo simulation of ligand-protein binding. *J. Chem. Inf. Model.* **2006**, *46*, 231−242.

(12) Jorgensen, W. L. Interactions between amides in solution and the thermodynamics of weak binding. *J. Am. Chem. Soc.* **1989**, *111*, 3770−3771.

(13) Hermans, J.; Subramaniam, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* **1986**, *27*, 225−227.

(14) Roux, B.; Nina, M.; Pomès, R.; Smith, J. C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.* **1996**, *71*, 670−681.

(15) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermaldynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **1997**, *72*, 1047−1069.

(16) Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.

(17) Mann, G.; Hermans, J. Modeling protein-small molecule interactions: Structure and thermodynamics of noble gases binding in a cavity in mutant phage T4 lysozyme L99A. *J. Mol. Biol.* **2000**, *302*, 979−989.

(18) Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10399−10404.

(19) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.

(20) Woo, H.-J.; Roux, B. Calculation of absolute protein−ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825−6830.

(21) Straatsma, T. P.; McCammon, J. A. Computational alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

(22) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(23) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulations. application to the methane dimer in water. *J. Chem. Phys.* **1988**, *89*, 3742−3746.

(24) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraint and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* In press.

(25) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

(26) Im, W.; Bernèche, S.; Roux, B. Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.* **2000**, *114*, 2924−2937.

(27) Deng, Y.; Roux, B. Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.

(28) Banavali, N. K.; Im, W.; Roux, B. Electrostatic free energy calculation using the generalized solvent boundary potential method. *J. Chem. Phys.* **2002**, *117*, 7381−7388.

(29) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* **1971**, *54*, 5237−5247.

(30) Morton, A.; Baase, W. A.; Matthews, B. W. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* **1995**, *34*, 8564−8575.

(31) Morton, A.; Matthews, B. W. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34*, 8576−8588.

(32) Bjerrum, N. Untersuchungen über Ionenassoziation. I. der Einfluss der Ionenassoziation auf die Akativität der Ionen bei Mittleren Assoziationsgraden. *K. Dan. Vidensk. Selsk., Mat.-Fys. Medd.* **1926**, *7*, 1−48.

(33) Mihailescu, M.; Gilson, M. K. On the theory of noncovalent binding. *Biophys. J.* **2004**, *87*, 23−36.

(34) Allen, T. W.; Andersen, O. S.; Roux, B. Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 117−122.

(35) Zwanzig, R. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(36) Woo, H.-J.; Dinner, A. R.; Roux, B. Grand canonical Monte Carlo simulation of water in protein environments. *J. Chem. Phys.* **2004**, *121*, 6392−6400.

(37) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, *86*, 67−74.

(38) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM − A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(39) Ferrenberg, A. M.; Swendsen, R. H. New monte carlo technique for studying phase transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.

(40) Ferrenberg, A. M.; Swendsen, R. H. Optimized monte carlo data analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.

(41) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(42) Stote, R. H.; States, D. J.; Karplus, M. On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys. Phys.-Chim. Biol.* **1991**, *88*, 2419−2433.

(43) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(44) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N•log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(45) Feller, S. E.; Zhang, Y.; Paster, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulations: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.

(46) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.

(47) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* **2002**, *322*, 339−355.

(48) MacKerell, A. J.; Bashford, D.; Bellot, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., D. J.-M.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chems. B* **1998**, *102*, 3586−3616.

(49) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.

(50) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revision A.9*; Gaussian, Inc.: Pittsburgh, PA, 1998,

(51) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.* **2003**, *125*, 9523−9530.

(52) Marelius, J.; Hansson, T.; Åqvist, J. Calculation of ligand binding free energies from molecular dynamics simulations. *Int. J. Quantum Chem.* **1998**, *69*, 77−88.

(53) Shaikh, S. A.; Ahmed, S. R.; Jayaram, B. A molecular thermodynamic view of DNA-drug interactions: a case study of 25 minor-groove binders. *Arch. Biochem. Biophys.* **2004**, *429*, 81−99.

(54) MacKerell, A. D., Jr. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.

(55) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. New class IV charge model for extracting accurate partial charges from wave function. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.

(56) Macias, A. T.; MacKerell, A. D., Jr. Ch/*pi* interactions involving aromatic amino acids: Refinement of the CHARMM tryptophan force field. *J. Comput. Chem.* **2005**, *26*, 1452−1463.

(57) Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* **2004**, *337*, 1161−1182.

(58) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations. *Comput. Phys. Comm.* **2001**, *135*, 40−57.

(59) Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraint potentials. *Biophys. J.* In press.

(60) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **2001**, *8*, 932−935.

(61) Dunbrack, R. L., Jr.; Karplus, M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.* **1994**, *1*, 334−340.

(62) Petrella, R. J.; Karplus, M. The energetics of off-rotamer protein side-chain conformations. *J. Mol. Biol.* **2001**, *312*, 1161−1175.

(63) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(64) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005**, *123*, 084108.

(65) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

CT060037V

1274 J. Chem. Theory Comput. **2006,** *2,* 1274−1281

JCTC Journal of Chemical Theory and Computation

Time-Dependent Properties of Liquid Water: A Comparison of Car−Parrinello and Born−Oppenheimer Molecular Dynamics Simulations

I-Feng W. Kuo,† Christopher J. Mundy,† Matthew J. McGrath,‡ and J. Ilja Siepmann*,‡

Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, Livermore, California 94550, and Departments of Chemistry and of Chemical Engineering and Material Science, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455-0431

Received June 5, 2006

Abstract: A series of 30 ps first principles molecular dynamics simulations in the microcanonical ensemble were carried out to investigate transport and vibrational properties of liquid water. To allow for sufficient sampling, the thermodynamic constraints were set to an elevated temperature of around 423 K and a density of 0.71 g cm^{-3} corresponding to the saturated liquid density for the Becke−Lee−Yang−Parr (BLYP) representation of water. Four simulations using the Car−Parrinello molecular dynamics (CPMD) technique with varying values of the fictitious electronic mass (μ) and two simulations using the Born−Oppenheimer molecular dynamics (BOMD) technique are analyzed to yield structural and dynamical information. At the selected state point, the simulations are found to exhibit nonglassy dynamics and yield consistent results for the liquid structure and the self-diffusion coefficient, although the statistical uncertainties in the latter quantity are quite large. Consequently, it can be said that the CPMD and BOMD methods produce equivalent results for these properties on the time scales reported here. However, it was found that the choice of μ affects the frequency spectrum of the intramolecular modes, shifting them slightly to regions of lower frequency. Using a value of $\mu = 400$ au results in a significant drift in the electronic kinetic energy of the system over the course of 30 ps and a downward drift in the ionic temperature. Therefore, for long trajectories at elevated temperatures, lower values of this parameter are recommended for CPMD simulations of water.

Introduction

Water's ubiquity on Earth leads to a vital role in most biological and environmental processes, either directly or indirectly. As a result, water has been studied by experimental, theoretical, and computational methods for many years. However, water deviates in many important ways from simple fluids to the extent of becoming a grand challenge for both liquid state theory and molecular simulation.[1,2] Originally, particle-based simulations employed empirical potentials for the determination of energies and forces,[3] but some of water's physical (e.g. a large dipole moment and large polarizability) and chemical (such as self-dissociation) properties have made it difficult to use an empirical force field to accurately reproduce experimental data over a wide range of state points.

The advent of the Car−Parrinello (CP) method in 1985[4] enabled one to use a quantum mechanical description of the electronic degrees of freedom combined with a classical phase space trajectory involving a fictitious electronic mass parameter. The first CP molecular dynamics simulation for liquid water followed in 1993,[5] a significant advance despite that computational resources limited this simulation to 32 molecules and less than 4 ps for the combined equilibration and production periods. Since then, the increase in available

---

* Corresponding author e-mail: siepmann@chem.umn.edu.

† Lawrence Livermore National Laboratory.

‡ University of Minnesota.

10.1021/ct6001913 CCC: $33.50

computing power and the refinement of algorithms has led to an increased ability to perform these simulations, and the CP method has now been applied to many aqueous systems at a variety of conditions (for examples, see refs 6−10). The information obtained from these studies has substantially advanced our understanding of the physical and chemical properties of aqueous systems.

Recently, the first principles simulation protocols employed in studies of liquid water have been examined critically.[11−15] It has been suggested that the appropriate selection of the fictitious electronic mass parameter will result in consistent structural properties.[12−14] The effect of the fictitious electronic mass on dynamical properties (such as diffusion constant and vibrational frequencies) has been examined for both crystalline systems[16,17] and liquid water,[13] but a comprehensive comparison between Born−Oppenheimer (BO) and CP molecular dynamics for time-dependent properties of liquid water has been lacking. Although there is clear evidence that vibrational properties are affected by the fictitious mass parameter, it is not certain whether transport properties are sensitive to this parameter as long as reasonable values are used (which may or may not coincide with the upper range suggested by previous works for consistent structural properties). It is the goal of this study to explore this issue more thoroughly by using a range of fictitious mass parameters for CP molecular dynamics simulations and comparing the results to BO simulations that quench the electronic density at every step to a given tolerance.

The main thermophysical quantity in question is the self-diffusion coefficient. To precisely compute this quantity the molecules must move a significant distance from their starting positions, i.e., many molecular diameters. This can be achieved by increasing the simulation length, increasing the average velocity of the molecules (i.e. raising the temperature), or lowering the barrier for diffusive transport (i.e. lowering the density). Given the expense of ab initio simulations, the nuclear kinetic temperature of this study was raised to 423 K, and the density was set to the saturated liquid density for water calculated previously[18] for the Becke−Lee−Yang−Parr (BLYP) exchange and correlation functionals.[19,20] This enabled the molecules in the simulation to move on average between four and five molecular diameters from their origin for each 30 ps trajectory. A rigorous statistical analysis is presented in an attempt to determine the nature and magnitude of inconsistencies for both dynamical and structural properties, which leads to the conclusion that not all dynamical properties are affected equally by the change in fictitious electronic mass.

Simulation Methods and Details

A series of six ab initio molecular dynamics (MD) simulations was performed using the publicly available software suite CPMD.[21] Four of these simulations were based on the extended Lagrangian approach developed by Car and Parrinello.[4] The CP approach introduces a fictitious electronic kinetic energy term into the Lagrangian to efficiently propagate the electronic density (i.e. it makes the integration schemes used feasible). This allows one to use density functional theory to provide a quantum mechanical description of the electronic component of the system. In this work, the Kohn−Sham formulation of density functional theory[22] is used with the gradient corrected exchange functional of Becke[19] and the Lee−Yang−Parr correlation functional.[20] The choice of these functionals as well as the norm-conserving Martins−Troullier pseudopotentials[23] with Kleinman−Bylander transformation to the fully nonlocal form[24] to describe the core electronic states and a plane wave cutoff of 85 Ry for the Kohn−Sham orbitals follow those used in previous CPMD simulations.[14,25,26]

The use of the CP method implies that a value of the fictitious electronic mass parameter must be selected. For previous simulations of liquid water near ambient conditions and a duration of 20 ps or less, a value of $\mu = 400$ au appears to be a reasonable choice.[12,14] The goal of this work is to examine the effect of this parameter on dynamic properties of water, and, therefore, three additional choices of the fictitious mass ($\mu = 100$, 200, and 300 au) were explored. All runs using the Car−Parrinello method are labeled CP-μ. The CPMD program is also capable of quenching the electronic density after every time step instead of propagating it with the CP method. This leads to direct sampling from the Born−Oppenheimer surface. Two simulations were run using this method (quenching the energy to 10^{-7} Hartrees) in order to provide another level of comparison. These runs are labeled BO-1 and BO-2.

The initial configurations for all simulations except BO-2 were taken from the saturated liquid phase of an equilibrated Monte Carlo simulation at $T = 423$ K in the Gibbs ensemble and scaled to the computed average density of $\rho = 0.71$ g cm^{-3},[18] which corresponds to a supercell of $L = 13.9176$ Å for 64 water molecules. BO-2 was started from a configuration taken from the middle of BO-1 and rescaling the velocities to generate a slightly higher temperature. An equilibration period of about 5 ps in the canonical (*NVT*) ensemble was used to equilibrate each system and provide a random initial velocity distribution to ensure divergence of trajectories. Subsequently, production periods were run in the microcanonical (*NVE*) ensemble for 30 ps over which all reported properties were computed (unless otherwise stated). The time steps vary between runs (from 0.0484 to 0.484 fs) and are reported in Table 1.

Results and Discussion

It is well-known that all simulations utilizing the CP method suffer from some degree of nonadiabaticity between the nuclear and electronic degrees of freedom,[4,16,17,27] which is an effect that grows with increasing values of the fictitious electronic mass. This adiabaticity breakdown results in CP forces on the ions becoming significantly different than the BO forces, which necessarily impacts the trajectory. [16,17,27] More quantitative results for the system of interest (water) are shown in refs 12 and 14. In particular, ref 14 shows a simulation in the microcanonical ensemble of 64 molecules and $\mu = 800$ au being terminated after 4.4 ps due to the kinetic energy of the fictitious electronic degrees of freedom drifting by around 0.05 au. This is accompanied by an ionic temperature drop of around 30 K, which clearly demonstrates

Table 1. Simulation Details (Fictitious Mass and Time Step) and Selected Results (Average Nuclear Kinetic Temperature, Height of the First Peak in the Oxygen−Oxygen Radial Distribution Function and Diffusion Coefficient) for the Car−Parrinello (CP) and Born−Oppenheimer (BO) Molecular Dynamics Simulations[a]

run	μ (au)	δt (fs)	T_{nuc} (K)	$g_{\text{OO}}(r_{\max})$	$D_{2,7}$[b] (10^{-5} cm^2 s^{-1})	$D_{7,27}$[b] (10^{-5} cm^2 s^{-1})
BO-1		0.484	417(5)	2.37(0.06)	8.3(2.7)	8.0
BO-2		0.484	426(5)	2.28(0.12)	12.4(6.5)	12.3
CP-100	100	0.048	428(7)	2.20(0.15)	12.8(5.4)	12.7
CP-200	200	0.048	422(6)	2.20(0.10)	12.3(6.6)	9.9
CP-300	300	0.073	418(10)	2.14(0.04)	12.5(3.3)	14.0
BO−SGGG[c]		0.24	393(n/a)	3.10(n/a)	1.2(n/a)	
CP−SGGG[c]	340	0.07	399(n/a)	2.60(n/a)	2.2(n/a)	

[a] The values reported are the averages of three 10 ps blocks, with the numbers in parentheses being the 95% confidence interval assuming a normal distribution. [b] $D_{2,7}$ and $D_{7,27}$ denote diffusion coefficients calculated for the 2−7 ps region of the three blocks and for the 7 to 27 ps region of the full production period, respectively. [c] CP-SGGG and BO-SGGG are simulations for the PBE functional carried out by Schwegler et al.[13]

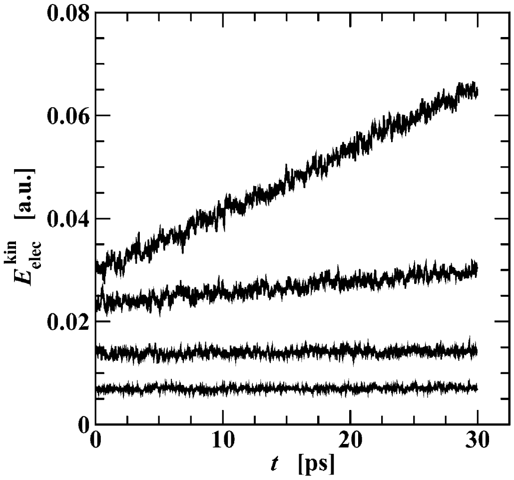


Figure 1. The instantaneous fictitious electronic kinetic energy as a function of time for all four Car−Parrinello molecular dynamics simulations. From top to bottom: CP-400, CP-300, CP-200, and CP-100.

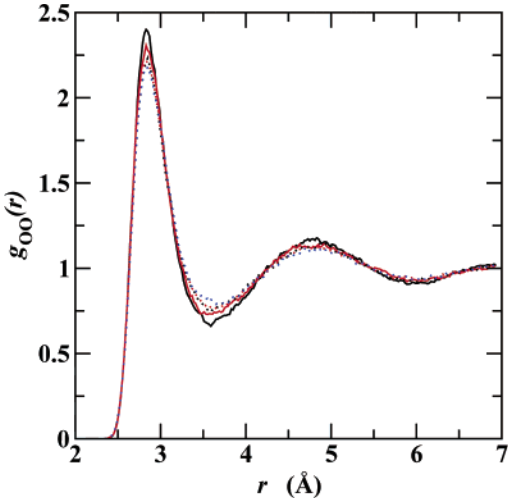


Figure 2. The oxygen−oxygen radial distribution functions obtained from molecular dynamics simulation of water using the BLYP exchange/correlation energy functional at $T = 423$ K and $\rho = 0.71$ g cm^{-3}. The lines correspond to (refer to text for descriptions) the following: BO-1 (black, solid), BO-2 (red, solid), CP-100 (black, dot), CP-200 (red, dot), and CP-300 (blue, dot).

the nonadiabaticity of the system. Figure 1 shows the electronic kinetic energy of all four CP simulations reported here as a function of time. It should be noted that one of the reported simulations (CP-400, with $\mu = 400$ au) shows a drift over the full 30 ps that is comparable to the drift for the 4.4 ps trajectory with $\mu = 800$ au from ref 14. Consequently, using a fictitious electronic mass of 400 au for long simulations of water at elevated temperature in the microcanonical ensemble cannot be recommended, and no results are shown for this run. It has been shown previously that control of the electronic degree of freedom through the use of thermostats can prevent drifting of the fictitious electronic kinetic energy,[14,28] but caution needs to be applied to remove spurious thermostating effects from dynamic properties.

Table 1 gives a summary of structural and transport properties computed for the current simulations. Errors are estimated by breaking the full trajectories into three blocks of equal length (i.e. 10 ps), computing the property of interest, and calculating the 95% confidence interval assuming a normal distribution. Additional analysis (see below) indicates that the resulting block length of 10 ps is shorter than the correlation time for the system, and consequently the confidence intervals reported in Table 1 are probably underestimated. It can be seen from this table that the average nuclear kinetic temperatures range from 417 K (for BO-1) to 428 K (for CP-100). Despite this, the structural properties are fairly consistent, showing a spread in the first peak height of the oxygen−oxygen radial distribution function of 0.23, which is reasonable when compared to previous studies.[11−14] However, it should be noted that the 95% confidence limits for BO-1 and CP-300 do not overlap; presumably because these error limits are most likely underestimated (see below). Another possible explanation might be some type of systematic fictitious mass effect, e.g., as discussed recently by Tangney.[17]

The oxygen−oxygen and oxygen−hydrogen radial distribution functions are depicted in Figures 2 and 3. Again, agreement between the five simulations is quite good but with the (low-temperature) BO-1 simulation showing slightly more structural ordering. In comparison to other first principles simulations of liquid water, we note that simulations of supercritical water[29,30] carried out at a similar density (≈0.7 g cm^{-3}) and much higher temperatures (653 K) lead to water with significantly less structure than observed here and that simulations of dense water[13] ($\rho = 1.0$ g cm^{-3}) at elevated temperature (≈400 K) yield significantly more structure.

To compare transport properties, the self-diffusion coefficient was computed from the Einstein relation

$$2tD = \frac{1}{3}\langle|\mathbf{r}(t) - \mathbf{r}(0)|^2\rangle \tag{1}$$

where the slope is found from a linear least-squares analysis

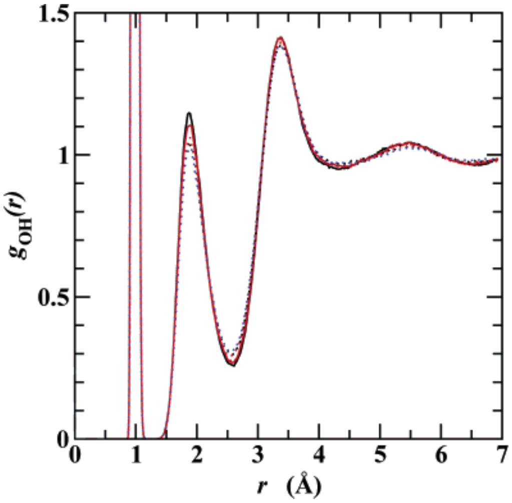


Figure 3. The oxygen−hydrogen radial distribution functions obtained from molecular dynamics simulation of water using the BLYP exchange/correlation energy functional at $T = 423$ K and $\rho = 0.71$ g cm^{-3}. The line styles and colors are the same as Figure 2.

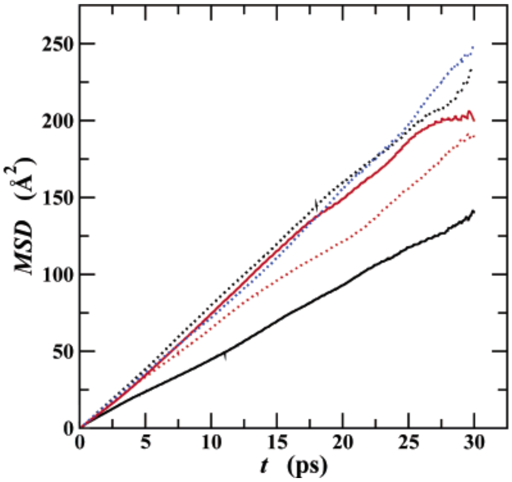


Figure 4. The mean squared displacements obtained from molecular dynamics simulation of water using the BLYP exchange/correlation energy functional at $T = 423$ K and $\rho = 0.71$ g cm^{-3}. The line styles and colors are the same as Figure 2.

to fit the 2.0−7.0 ps region of the mean square displacement (MSD). To improve statistics, origins were taken every 0.5 ps, i.e., seven time origins are used for each of the 10 ps blocks. The self-diffusion coefficients for all five simulations show overlapping confidence limits and range from 8.3 (for BO-1) to 12.8 × 10^{-5} cm^2 s^{-1} (for CP-100), suggesting that reasonable values of μ with regards to structural properties also produce consistent diffusion coefficients. A comparison of the MSD for the five simulations (see Figure 4) shows graphically the large but statistically insignificant deviations between the five simulations. In addition, we have also computed the diffusion coefficient from a least-squares analysis fit to the 7−27 ps region of the full production period. The results obtained for this longer time interval agree with the 10 ps block averages to within statistical uncertainties (see Table 1).

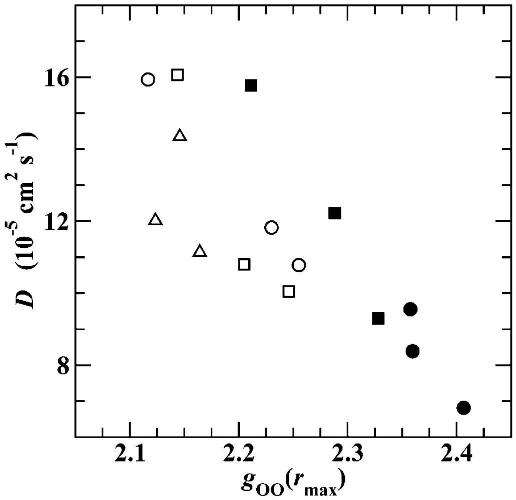


Figure 5. The diffusion coefficient as a function of the peak height of the first maximum in the oxygen−oxygen radial distribution function. Data points were obtained by dividing the simulations into three parts of equal length and computing the values for these segments. The runs are as follows: BO-1 (solid circle), BO-2 (solid square), CP-100 (open circle), CP-200 (open square), and CP-300 (open triangle up).

Assuming that the experimental correlation of high-temperature self-diffusion coefficients given in eq 4 of ref 31 can be extended to the state point investigated here (which would be inside the two-phase loop for real water), one obtains a self-diffusion coefficient of 31.8 × 10^{-5} cm^2 s^{-1}, which is approximately a factor of 3 higher than the values observed here. Thus, the BLYP representation appears to yield a liquid phase with sluggish diffusive properties, whereas it yields a vapor−liquid coexistence curve that is shifted downward.[18]

Experimentally, the self-diffusion coefficient of water changes by almost a factor of 2 for a temperature change of 20 K at lower temperatures,[32] and consequently the varying temperatures present in microcanonical simulations should be considered when comparing diffusion properties. The spread of temperatures reported here is 11 K, but there does not appear to be any strong relation between the temperature and diffusion coefficient, suggesting that other factors may influence the simulation more. For instance, one must consider how the transient changes in structure affect diffusion. Movement in a more structured fluid results in a higher barrier for diffusive transport (escape from the surrounding solvation shell), which in turn produces a slower self-diffusion coefficient. Figure 5 shows a plot of the self-diffusion coefficient as a function of the height of the first peak in the oxygen−oxygen radial distribution function for all 15 10-ps blocks used in the statistical analysis in Table 1. Using an unweighted least-squares method to fit a straight line to these points gives a correlation coefficient of −0.80, indicating a moderately negative linear correlation between peak height and self-diffusion coefficient.

Table 1 also includes self-diffusion coefficients and peak heights reported by Schwegler et al.[13] for BO and CP molecular dynamics simulations using the Perdew-Burke-Ernzerhof exchange/correlation functional. Because these runs are at a different state point (slightly lower temperature

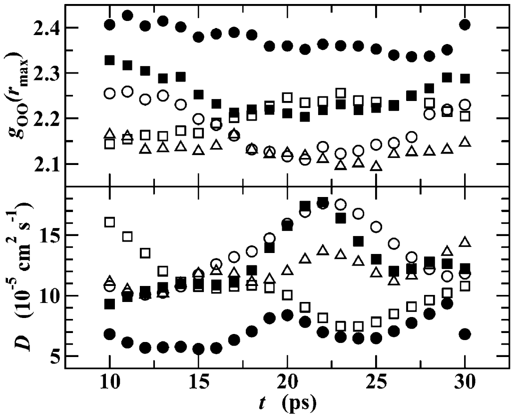


Figure 6. Evolution of the peak height of the first maximum in the oxygen−oxygen radial distribution function (top) and of the diffusion coefficient (bottom). Data points are given for the peak height and diffusion coefficient calculated over the previous 10 ps. The symbols are the same as in Figure 5.

but significantly higher density) and use a different functional, they cannot be compared directly to the simulations performed in this work. However, an attempt can be made to explain the reported difference using insight acquired from this work. The difference in the reported height of the first peak of the oxygen−oxygen radial distribution function between CP and BO runs performed by Schwegler et al. near the same nuclear kinetic temperature (400 K) is 0.5, and the difference in self-diffusion coefficient is around a factor of 2. By using a linear-least-squares analysis and fitting a line to the data in Figure 5, one can compute the expected diffusion coefficients for a change in peak height from 2.0 to 2.5 (which includes the range of values found here and over which the data appears to be linear). This results in a change of diffusion coefficient of a factor of 3, providing evidence that the reported difference in ref 13 could be largely due to the difference in structure (peak heights of 2.6 and 3.1). Furthermore, one might expect larger statistical uncertainties for the simulations by Schwegler et al. because they were run only for 20 ps and yield substantially lower self-diffusion coefficients (presumably because of the higher density), i.e., the trajectories reached much smaller MSDs (and, hence, fewer escapes from the solvation shell) than those reported here.

Looking at the evolution of the system's properties over time is a worthwhile exercise in several respects. It can give information on the extent to which properties fluctuate and correlate. Unfortunately the self-diffusion constant (being the long-time limit of the Einstein relation) is not obtainable from an instantaneous snapshot or even a short time interval. Thus, as above, we determined this value here over the 2.0−7.0 ps range of 10 ps blocks but repeat this procedure for multiple initial times along the trajectory to follow the evolution. Figure 6 shows a plot of the first maximum in the oxygen−oxygen radial distribution function and of the self-diffusion coefficient as a function of simulation time with the former averaged over the previous 10 ps and the latter obtained from an Einstein plot of the full 10 ps. This produces data points that are necessarily correlated (each point overlaps 90% with its neighbors), but it still provides insight into the evolution of the system. It appears that for all five simulations these two properties undergo fluctuations (versus a drift that would signal poorly equilibrated samples). It is evident that the correlation times for these systems are on a scale approaching the simulation length since no system undergoes more than two complete fluctuations. The figure also depicts the same trend as is seen in Figure 5, namely that a more structured liquid-phase yields a smaller self-diffusion coefficient as well as showing significant overlap between the BO and CP simulations.

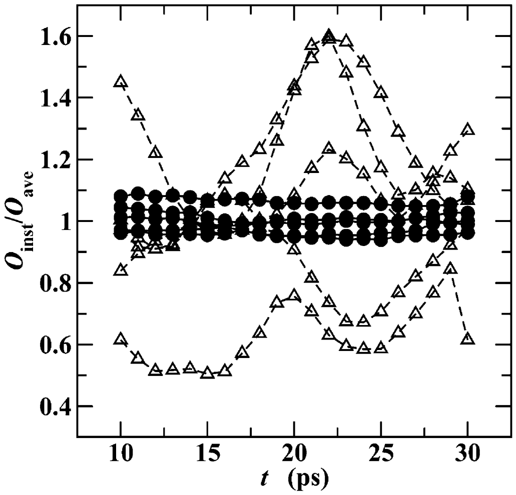


Figure 7. The ratio of the instantaneous peak height (filled circles) and self-diffusion coefficient (open triangles) to the average value across all simulations as a function of time. Data points are given for properties computed over the previous 10 ps. Lines are drawn as guides for the eye.

Figure 7 graphically illustrates another interesting feature of these simulations. The ratio of the 10 ps interval values in Figure 6 with the respective average value calculated from all five simulations is plotted as a function of time. These results demonstrate that the fluctuations of the self-diffusion coefficient are significantly greater than that of the first peak height of the oxygen−oxygen radial distribution function. Specifically, no value of the peak height differs from the average by more than 10%. On the other hand, some values of the instantaneous mean-square-displacement are found 50% higher or lower than the mean value. This is also demonstrated in Table 1 by the much larger confidence intervals attached to the self-diffusion coefficient. Therefore small uncertainties in the structure of a simulation can lead to even larger uncertainties in the diffusion coefficient, and one should be very cautious in comparing the transport properties of simulations with significant differences in structure. Although the magnitude of the fluctuations in the structure is small (and might therefore be overlooked), the correlation times are very long, i.e., an estimation of the statistical uncertainties from block averages needs to be viewed with caution.

It has been found that under some conditions, water described by certain density functionals may exist in a glassy state even at elevated temperatures.[33,34] Although the MSDs from this work (see Figure 4) appear to be quite linear (and there is no indication that they would reach a plateau), two tests were carried out to ensure that water under these

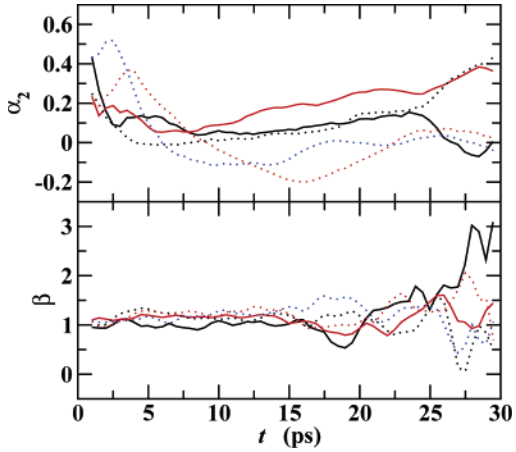


Figure 8. The values of the α_2 (top) and β (bottom) parameters as a test of glassy diffusion. The line styles and colors are the same as Figure 2.

conditions is truly in a liquid state. The first indicator, α_2, is given by

$$\alpha_2(t) = \frac{\langle|\mathbf{r}(t) - \mathbf{r}(0)|^4\rangle}{3\langle|\mathbf{r}(t) - \mathbf{r}(0)|^2\rangle^2} - 1 \tag{2}$$

and decays to zero for random walk statistics.[35] The second indicator of diffusive dynamics, β, is given by

$$\beta(t) = \frac{\mathrm{d}\ln(\langle|\mathbf{r}(t) - \mathbf{r}(0)|^2\rangle)}{\mathrm{d}\ln(t)} \tag{3}$$

A value of unity for this parameter indicates that the system is diffusing in a liquid regime. Figure 8 shows plots of both these indicators over the full trajectories for all five simulations. As for the MSDs, these indicators were computed using multiple origins spaced 0.5 ps apart. Because β is a derivative, this indicator would be artificially altered close to these origins (examination of Figure 4 reveals discontinuities here). Consequently, it was decided to compute

$$\beta(t) = \frac{\ln(\langle|\mathbf{r}(t+0.5) - \mathbf{r}(0)|^2\rangle) - \ln(\langle|\mathbf{r}(t-0.5) - \mathbf{r}(0)|^2\rangle)}{\ln(t+0.5) - \ln(t-0.5)} \tag{4}$$

With this finite-difference method the discontinuities do not generate large jumps in β. Origins taken every 0.05 ps did not give qualitatively different results, so 0.5 ps is used here for consistency.

The indicator $\beta(t)$, as shown in Figure 8, is equal to unity for all five runs in the region of good statistics (time scales up to about 15 ps). Poor statistics after this time lead to fluctuations around unity. The plot of α_2 as a function of time is a little less clear. One expects this parameter to tend to zero at long times, if the molecular distances adopt a Gaussian distribution. For all simulations, α_2 hovers around zero for intermediate times, but then it begins to diverge again. At long times, this can be explained by poor statistics, just like for β and the MSD. For short times the configurations are still correlated, so motion is not completely random, and α_2 is not anticipated to be zero. A combination of Figures 4 and 8 leads to the conclusion that water represented by the BLYP exchange/correlation functional is not glassy at $T = 423$ K and $\rho = 0.71$ g cm^{-3}.

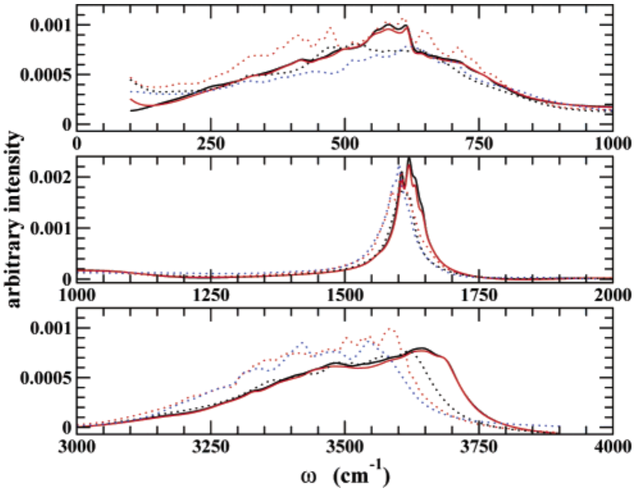


Figure 9. Power spectra computed from the regularized resolvent transform of the total velocity autocorrelation function.[36,37] The spectrum is divided into three regions (from top to bottom: librations, bending motion, and stretching modes) for clarity. The line styles and colors are the same as Figure 2.

Dynamical information of water interacting with its local environment can be gleaned from the vibrational power spectrum. It was found here that the Fourier transform (FT) of the total velocity autocorrelation function could not adequately resolve all low-frequency modes (less than 1000 cm^{-1}) of these systems. Therefore, the power spectra were generated with the regularized resolvent transform (RRT) method.[36,37] Like FT, RRT is also a direct transform of the data from the time domain into the frequency domain and is exactly equivalent to discrete FT when the frequency spectra can be represented by a sum of sinusoids. The resulting spectra are shown in the three sections of Figure 9. The regions shown correspond to the molecular librations, the H−O−H bending motion, and the O−H stretching modes. Whereas the self-diffusion coefficient shows no statistically significant fictitious mass effects, the power spectra in the mid-infrared region do show signs of dynamic retardation due to larger masses. This is in agreement with previous works[16,17,38−40] and further validates the use of a scaling parameter when analyzing the mid-infrared region to extrapolate the frequency to the limit of an infinitely small fictitious mass. The effect of the electronic mass parameter on the librational modes of water is less clear. It should also be noted that the BO trajectories appear to yield “smoother” power spectra than the CP simulations despite that an equal time span was covered in the trajectories.

The peak “position” is not well defined for the broad peaks obtained here for the librational and stretching modes (shown in the top and bottom panels of Figure 9). Consequently, Table 2 lists the median values of each peak in the RRT power spectra computed by integrating from an arbitrary intensity value (close to zero) on one side of the peak to the same value on the other side. These median peak positions show a systematic red-shift for the bending and stretching modes in the CP trajectories, with the magnitude of the shift

Table 2. Median Locations for the Three Vibrational Modes (in units of cm^{-1})

run	libration	bend	stretch
BO-1	551	1617	3522
BO-2	553	1617	3525
CP-100	498	1605	3491
CP-200	495	1599	3455
CP-300	529	1595	3446

relative to the BO value being smallest for CP-100 and largest for CP-300. The median values for the librational peak of the CP simulations show a red-shift when compared to the BO runs, but increasing the magnitude of the fictitious electronic mass does not necessarily lead to a further shift in the peak position. The evident, but not systematic, shift suggests large uncertainties in the low-frequency region of the spectra, a concern supported by the short simulation length but obfuscated by the agreement between the BO simulations (see below).

One striking feature present in both Figure 9 and Table 2 is the close agreement obtained for both BO runs. The variations between the two spectra are minimal, despite the difference in structure and diffusivity highlighted in Figures 2−4. This agreement lends credence to the belief that the differences seen in the CP simulations are not artifacts of the analysis procedure and reaffirms conclusions from earlier work that certain properties (in particular, those directly related to forces and momentum via changes of the mass) are influenced by the use of the CP method.[16,17,27]

Finally, one should note that the librational part of the spectrum observed here does not show the distinct features observed in CP molecular dynamics simulations for the water spectrum at ambient conditions.[41] The most likely explanation is that the tetrahedral hydrogen-bonded network is very fluctuational at the high-temperature, low-density conditions investigated here. Thus, interchange between different local configurations is fast, and their lifetimes are too short to establish specific librational features. This explanation is supported by recent experimental analysis of vibrational lifetimes in liquid water.[42]

Conclusions

A series of long molecular dynamics trajectories for liquid water near 423 K obtained from BO molecular dynamics and CP molecular dynamics with varying values of the fictitious mass parameters, μ, has allowed us to look at the structural and transport properties of water using density functional theory with BLYP exchange and correlation. All five simulations show similar structural properties with a range in the height of the first peak in the oxygen−oxygen radial distribution function of about 0.2. It has been concluded before,[12,14] and shall be reiterated here, that reasonable values of μ (in other words, values that do not lead to a large drift in the electronic kinetic energy and a subsequent lowering of ionic temperature) do not produce statistically significant differences in structural properties on the observed simulation time scales.

All five simulations show large fluctuations and hence large uncertainties for the self-diffusion coefficient. Again, on the simulation time scales used here there does not appear to be a statistically significant difference in transport properties between BO and CP molecular dynamics simulations as long as reasonable values of μ are used. If one assumes that the statistical uncertainties for the self-diffusion coefficient are solely a function of the distance a molecule travels on average, simulations near ambient temperature and density would require around 500 ps to obtain a diffusion coefficient with the same statistical uncertainty (about 50%) as found in this work. Consequently, it appears that only order of magnitude estimates for transport properties are reliable using current technology and algorithms for first principles simulations at ambient conditions (with the use of higher temperatures and lower densities enabling somewhat more precise values).

From this it can be concluded that, for reasonable values of the fictitious mass parameter, several physical properties (including structural information and the self-diffusion coefficient) of CP and BO simulations are indistinguishable on time scales of tens of picoseconds. However, it should be emphasized that even the CP trajectories with $\mu = 100$ or 200 au show a small increase of the fictitious electronic energy over the 30 ps trajectory. Thus, for these fictitious mass parameters there will be a much longer time scale over which this increase will lead to a noticeable decrease of the ionic temperature.

One observable that is clearly influenced by the choice of μ is the power spectrum, in which using CP dynamics with fictitious electronic mass shifts the spectrum into regions of lower frequency, even for small values of this parameter. A likely explanation for the manifestation of fictitious mass effects in the power spectrum is that this observable is governed directly by the instantaneous forces and the reduced mass of a given normal mode. References 16, 17, and 27 point out that the instantaneous forces of CP simulations with any value of the electronic mass are different than the BO forces for an identical system. On the other hand, properties that depend on free energies (e.g. the radial distribution function) or are governed by long-time processes involving free energy barriers (e.g., diffusive transport) should be significantly less affected given a reasonable choice of μ because instantaneous differences in the forces can cancel upon averaging.

Acknowledgment. We thank Juerg Hutter and Ed Maginn for many helpful discussions and Larry Fried and Charlie Westbrook for their ongoing support of this work. Financial support from the National Science Foundation Grants CTS-0138393 and ITR-0428774, a 3M Foundation Graduate Fellowship (M.J.M.), and a Department of Energy Computational Science Graduate Fellowship (M.J.M.) are gratefully acknowledged. Part of this work was performed under the auspices of the U.S. Department of Energy by the University of California Lawrence Livermore National Laboratory (LLNL) under contract No. W-7405-Eng-48. J.I.S. and M.J.M. thank the University Relations Program (LLNL) for hosting their sabbatical visit. Computer resources were provided by LLNL Computing and the Minnesota Supercomputing Institute.

References

(1) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: New York, NY, 1986.

(2) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford University Press: Oxford, England, 1987.

(3) Barker, J. P.; Watts, R. O. *Chem. Phys. Lett.* **1969**, *3*, 144.

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

(5) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9080.

(6) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. *Nature* **1999**, *397*, 601.

(7) Cavazzoni, C.; Chiarotti, G. L.; Scandolo, S.; Tosatti, E.; Bernasconi, M.; Parrinello, M. *Science* **1999**, *283*, 44.

(8) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. *Science* **2001**, *291*, 2121.

(9) Kuo, I.-F. W.; Mundy, C. J. *Science* **2004**, *303*, 658.

(10) Tateyama, Y.; Blumberger, J.; Sprik, M.; Tavernelli, I. *J. Chem. Phys.* **2005**, *122*, 234505.

(11) Asthagiri, D.; Pratt, L. R.; Kress, J. D. *Phys. Rev. E* **2003**, *68*, 041505.

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

(13) Schwegler, E.; Grossman, J. C.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *121*, 5400.

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

(15) McGrath, M. J.; Siepmann, J. I.; Kuo, I.-F. W.; Mundy, C. J.; VandeVondele, J.; Hutter, J.; Mohamed, F.; Krack, M. *ChemPhysChem* **2005**, *6*, 1894.

(16) Tangney, P.; Scandolo, S. *J. Chem. Phys.* **2002**, *116*, 14.

(17) Tangney, P. *J. Chem. Phys.* **2006**, *124*, 044111.

(18) McGrath, M. J.; Siepmann, J. I.; Kuo, I.-F. W.; Mundy, C. J.; VandeVondele, J.; Mohamed, F.; Krack, M. *J. Phys. Chem. A* **2006**, *110*, 640.

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

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

(21) CPMD, Version 3.9.2, Copyright IBM Corp. 1990−2001, Copyright MPI für Festkörperforschung. Stuttgart 1997−2005. http://www.cpmd.org.

(22) Kohn, W.; Sham, L. A. *Phys. Rev. A* **1965**, *140*, A1133.

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

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

(25) Silvestrelli, P. L. *Phys. Rev. B* **1999**, *59*, 9703.

(26) Chen, B.; Ivanov, I.; Park, J. M.; Parrinello, M.; Klein, M. L. *J. Phys. Chem. B* **2002**, *106*, 12006.

(27) Pastore, G.; Smargiassi, E.; Buda, F. *Phys. Rev. A* **1991**, *44*, 6334.

(28) Mantz, Y. A.; Chen, B.; Martyna, G. J. *Chem. Phys. Lett.* **2005**, *405*, 294.

(29) Tassaing, T.; Bellissent-Funel, M.-C.; Guillot, B.; Guissani, Y. *Europhys. Lett.* **1998**, *42*, 265.

(30) Boero, M.; Terakura, K.; Ikeshoji, T.; Liew, C. C.; Parrinello, M. *J. Chem. Phys.* **2001**, *115*, 2219.

(31) Lamb, W. J.; Hoffman, G. A.; Jonas, J. *J. Chem. Phys.* **1981**, *74*, 6875.

(32) Mills, R. *J. Phys. Chem.* **1973**, *77*, 685.

(33) Sit, P. H.-L.; Marzari, M. *J. Chem. Phys.* **2005**, *122*, 204510.

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

(35) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.

(36) Mandelshtam, V. A. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *38*, 159.

(37) Chen, J. H.; Shaka, A. J.; Mandelshtam, V. A. *J. Magn. Reson.* **2000**, *147*, 129.

(38) Tobias, D. J.; Jungwirth, P.; Parrinello, M. *J. Chem. Phys.* **2001**, *114*, 7036.

(39) Kuo, I.-F. W.; Tobias, D. J. *J. Phys. Chem. A* **2002**, *106*, 10969.

(40) Goncharov, A. F.; Goldman, N.; Fried, L. E.; Crowhurst, J. C.; Kuo, I.-F. W.; Mundy, C. J.; Zaug, J. M. *Phys. Rev. Lett.* **2005**, *94*, 125508.

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

(42) Eaves, J. D.; Loparo, J. J.; Fecko, C. J.; Roberts, S. T.; Tokmakoff, A.; Geissler, P. L. *Proc. Natl. Acad. Sci.* **2005**, *102*, 13019.

CT6001913

JCTC Journal of Chemical Theory and Computation

Geometries of Transition-Metal Complexes from Density-Functional Theory

Michael Bühl* and Hendrik Kabrede

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1,
D-45470 Mülheim an der Ruhr, Germany

Received March 31, 2006

Abstract: Several levels of density functional theory, i.e., various combinations of exchange-correlation functionals and basis sets, have been employed to compute equilibrium geometries for a diverse set of 32 metal complexes from the first transition row, for which precise gas-phase geometries are known from electron diffraction or microwave spectroscopy. Most DFT levels beyond the local density approximation can reproduce the 50 metal−ligand bond distances selected in this set with reasonable accuracy, as assessed by mean and standard deviations of optimized vs observed values. The ranking of some popular functionals, ordered according to decreasing standard deviation, is BLYP ≈ HCTH > B3LYP > BP86 > TPSS ≈ TPSSh. Together with its hybrid variant, the recently introduced meta-GGA functional TPSS performs best of all tested functionals, with mean and standard deviations of −0.5 and 1.4 pm, respectively. Even smaller errors are found for a more compact but less diverse set of transition-metal mono- and dihalides, for which experimentally derived equilibrium geometries are available.

Introduction

Accurate molecular geometries are prerequisite for reliable quantum-chemical computations. The predictive power of high-level ab initio calculations for molecular properties relies on the quality of the geometries for which these properties are evaluated. For molecules composed of lighter main-group atoms, geometrical parameters can usually be calculated reliably at intermediate ab initio levels such as MP2. More sophisticated methods such as CCSD(T) in conjunction with large basis sets afford bond lengths and angles that can even be more accurate than those derived by experiment. The performance of these and other ab initio methods in the calculation of equilibrium bond distances has nicely and instructively been demonstrated by Helgaker and co-workers,[1,2] who assessed each method in terms of mean and standard deviation between optimized and observed bond lengths and visualized these results in the form of normalized Gaussian distributions with the same characteristics. This was done for a set of 19 small molecules (28 distances in total), for which accurate experimental data are available.

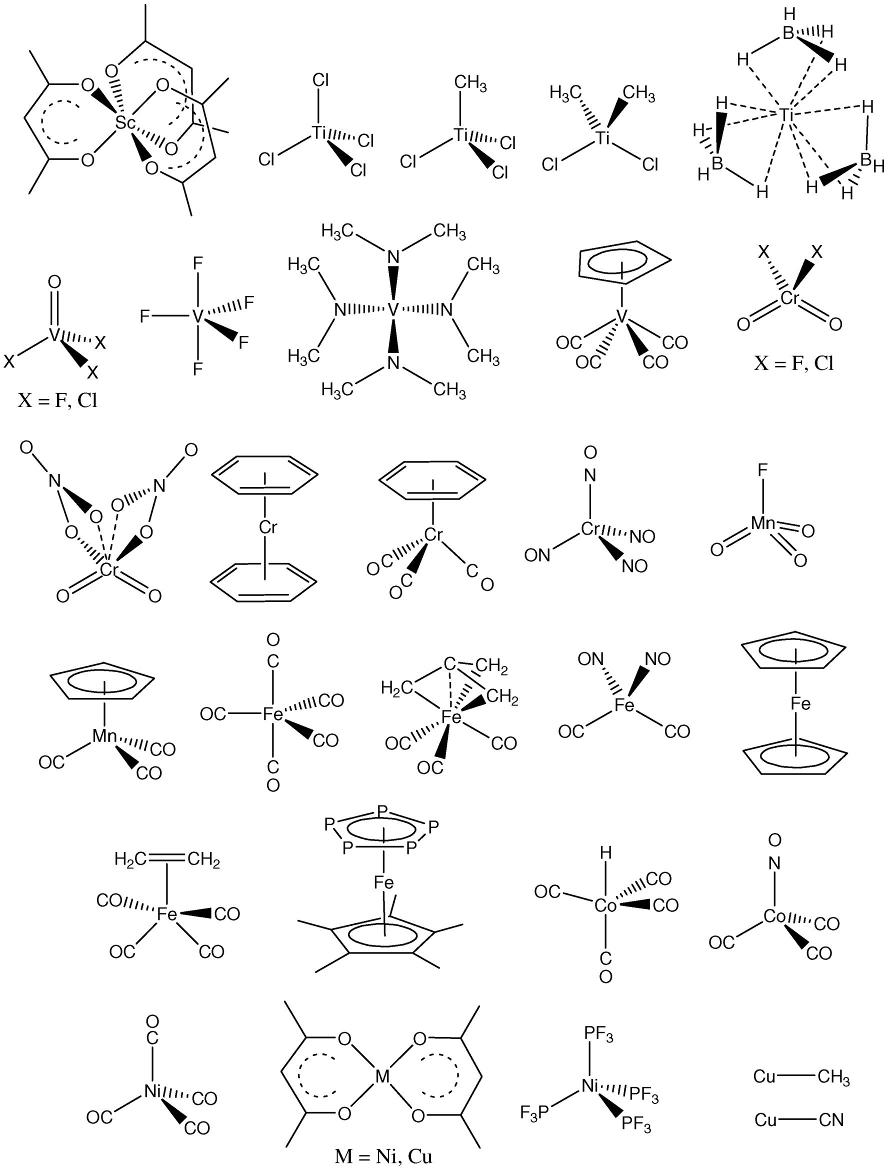
This test set contained only small molecules including main-group atoms up to the first row. It would be highly desirable to be able to achieve a similar accuracy for the more challenging class of transition-metal complexes. These are frequently more demanding than main-group molecules not only in terms of the level of electron correlation that has to be covered but also in terms of size. Thus, CCSD(T) or MR-CI geometry optimizations with sufficiently large basis sets are extremely, usually prohibitively, expensive. For this reason, computational transition-metal chemistry remains a stronghold of methods rooted in modern density functional theory (DFT), as these can account for a large fraction of electron correlation at an affordable cost. The price to be paid is a lower accuracy, compared to that of sophisticated ab initio methods, and the difficulty to improve it systematically. In practice, the density functional best suited for a given problem has to be identified by careful validation against experiment. As far as experimental structures are concerned, this validation is hampered by a dearth of accurate structure determinations in the gas phase, to which the overwhelming majority of DFT applications would refer. Quite frequently, parameters optimized for pristine molecules are compared to those obtained from X-ray crystallography

---

* Corresponding author fax: +(0)208-306 2996; e-mail: buehl@mpi-muelheim.mpg.de.

10.1021/ct6001187 CCC: $33.50

Chart 1



or neutron diffraction, that is, for structures in the solid with unknown effects from packing forces and intermolecular interactions. The overall experience with DFT-optimized geometries is that most gradient-corrected (GGA) or hybrid functionals perform reasonably well, albeit with a tendency to overestimate metal−ligand bond distances by several pm, and with deviations typically increasing from metal−C to metal−P bonds.[3]

Occasionally, newly developed functionals are also tested against gas-phase geometries but usually only for a small number of complexes (see refs 4−8 for a few illustrative examples). We have now selected a much larger test set comprising all 3d metals from Sc to Cu, for which quite precise experimental data are available from gas-phase electron diffraction (GED) and/or microwave spectroscopy (MW). This test set, depicted in Chart 1, should be diverse enough to cover a wide range of bonding situations, from complexes of high-valent early transition metals with electronegative ligands to electron-rich organometallic compounds of middle or late transition metals. Drawing from a large compilation of gas-phase structures,[9] we chose complexes for which at least one metal−ligand bond length was determined with a precision better than 1 pm, affording a final set of 32 molecules with 50 individual bond distances with that precision, which should be sufficient for reasonable statistics. We have then optimized these structures with popular local, gradient-corrected, hybrid, and meta-GGA functionals, together with a variety of basis sets, and assessed each combination, in the spirit of Helgaker and co-workers,[1,2] in terms of mean and standard deviation from the corresponding experimental reference values. As it turns out, the BP86 functional performs best among the traditional GGAs and hybrid functionals, and a recent meta-GGA, TPSS, is slightly superior.

Computational Details

Geometries were fully optimized in the given symmetry (as given in Table 1) using Gaussian 03[10] and several local (LSDA)[11] and gradient-corrected density functional combinations as implemented therein. Most functionals are composed of one of several exchange parts, namely Becke[12] (B), Becke hybrid[13] (B3), OPTX[14] (O), or OPTX hybrid[15] (O3), together

Table 1: Geometrical Parameters (Bond Lengths r in pm, Bond Angles a in Degrees) of First-Row Transition-Metal Complexes in the Gas Phase[a,b]

compd (mult) *sym*	parameter	[bond no.]	$(r/a)_{\mathrm{a}/\alpha}$	$(r/a)_{\mathrm{z/0/av/g}}$	ref
$\mathrm{Sc(acac)_3}$ (1) D_3	$r(\mathrm{Sc{-}O})$	[1]	207.6(4)	208.3(4)	36
$\mathrm{TiCl_4}$ (1) T_d	$r(\mathrm{Ti{-}Cl})$	[2]	216.9	217.0(2)	37
$\mathrm{TiMeCl_3}$ (1) C_{3v}	$r(\mathrm{Ti{-}C})$	[3]	204.7(6)		38
	$r(\mathrm{Ti{-}Cl})$	[4]	218.5(3)		
	$a(\mathrm{Cl{-}Ti{-}C})$		105.6(2)		
$\mathrm{TiMe_2Cl_2}$ (1) C_{2v}	$r(\mathrm{Ti{-}C})$	[5]	205.8(4)		39
	$r(\mathrm{Ti{-}Cl})$	[6]	219.6(3)		
	$a(\mathrm{C{-}Ti{-}C})$		102.8(9)		
	$a(\mathrm{Cl{-}Ti{-}Cl})$		117.3(3)		
$\mathrm{Ti(BD_4)_3}$ (2) C_{3h}	$r(\mathrm{Ti{-}B})$	[7]	217.5(4)		40
	r(Ti-D^{br})	[8]	198.4(5)		
$\mathrm{VOF_3}$ (1) C_{3v}	$r(\mathrm{V{=}O})$	[9]		157.0(5)	41
	$r(\mathrm{V{-}F})$	[10]		172.9(2)	
	$a(\mathrm{O{-}V{-}F})$		107.7(4)		
$\mathrm{VF_5}$ (1) D_{3h}	$r(\mathrm{V{-}F}^{ax})$	[11]	173.4(7)		42
	$r(\mathrm{V{-}F}^{eq})$	[12]	170.8(5)		
$\mathrm{VOCl_3}$ (1) C_{3v}	$r(\mathrm{V{=}O})$	[13]	157.3(8)	156.8(5)	43
	$r(\mathrm{V{-}Cl})$	[14]	213.8(2)	213.8(2)	
	$a(\mathrm{Cl{-}V{-}Cl})$		111.4(4)	111.3(2)	
$\mathrm{V(NMe_2)_4}$ (2) S_4	$r(\mathrm{V{-}N})$	[15]	187.9(4)		44
	$a(\mathrm{N{-}V{-}N})$		100.6(5)		
	$a(\mathrm{N{-}V{-}N'})$		114.1(1)		
$\mathrm{V(Cp)(CO)_4}$ (1) C_s	$r(\mathrm{V{-}C^{CO}})$	[16]	196.3(7)		45
$\mathrm{CrO_2F_2}$ (1) C_{2v}	$r(\mathrm{Cr{=}O})$	[17]	157.4(2)	157.5(2)	46
	$r(\mathrm{Cr{-}F})$	[18]	171.9(2)	172.0(2)	
	$a(\mathrm{O{=}Cr{=}O})$		107.8(8)		
	$a(\mathrm{F{-}Cr{-}F})$		111.9(9)		
$\mathrm{CrO_2Cl_2}$ (1) C_{2v}	$r(\mathrm{Cr{=}O})$	[19]	157.7(2)	158.1(2)	47
	$r(\mathrm{Cr{-}Cl})$	[20]	212.2(2)	212.6(2)	
	$a(\mathrm{O{=}Cr{=}O})$		108.5(4)		
	$a(\mathrm{Cl{-}Cr{-}Cl})$		113.3(3)		
$\mathrm{CrO_2(NO_3)_2}$ (1) C_2	$r(\mathrm{Cr{=}O})$	[21]	158.4(2)		48
	$r(\mathrm{Cr{-}O})$	[22]	195.4(5)		
$\mathrm{Cr(C_6H_6)_2}$ (1) D_{6h}	$r(\mathrm{Cr{-}C})$	[23]		215.0(2)	49
$\mathrm{Cr(C_6H_6)(CO)_3}$ (1) C_{3v}	$r(\mathrm{Cr{-}C^{Ar}})$	[24]	220.8(6)		50
	$r(\mathrm{Cr{-}C^{CO}})$	[25]	186.3(5)		
	$a(\mathrm{C^{CO}\text{-}Cr{-}C^{CO}})$			87.4(6)[50b]	
$\mathrm{Cr(NO)_4}$ (1) T_d	$r(\mathrm{Cr{-}N})$	[26]	175.0(2)	176.3(2)	51
$\mathrm{MnO_3F}$ (1) C_{3v}	$r(\mathrm{Mn{=}O})$	[27]		158.6(5)	52
	$r(\mathrm{Mn{-}F})$	[28]		172.4(5)	
	$a(\mathrm{O{=}Mn{-}F})$			108.5(1)	
$\mathrm{MnCp(CO)_3}$ (1) C_1	$r(\mathrm{Mn{-}C^{Cp}})$	[29]	214.7(3)		53
	$r(\mathrm{Mn{-}C^{CO}})$	[30]	180.6(3)		
$\mathrm{Fe(CO)_5}$ (1) D_{3h}	$r(\mathrm{Fe{-}C})^{\mathrm{mean}}$	[31]		182.9(2)	54
	$(r(\mathrm{Fe{-}C^{ax}})$		180.9	181.0(16))	
	$(r(\mathrm{Fe{-}C^{eq}})$		184.1	184.2(11))	
$\mathrm{Fe(CO)_3(tmm)}$ (1) C_{3v}	$r(\mathrm{Fe{-}C^{CO}})$	[32]		181.0(3)	55
	$r(\mathrm{Fe{-}C^{cent}})$	[33]		193.8(5)	
	$r(\mathrm{Fe{-}C^{CH2}})$	[34]		212.3(5)	
	$a(\mathrm{Fe{-}C^{cent}{-}C^{CH2}})$			76.4(2)	
$\mathrm{Fe(CO)_2(NO)_2}$ (1) C_{2v}	$r(\mathrm{Fe{-}C})$	[35]	187.2	188.3(3)	51
	$r(\mathrm{Fe{-}N})$	[36]	167.4	168.8(3)	
$\mathrm{FeCp_2}$ (1) D_{5h}	$r(\mathrm{Fe{-}C})$	[37]		206.4(3)	56
$\mathrm{Fe(C_2H_4)(CO)_4}$ (1) C_{2v}	$r(\mathrm{Fe{-}C}^{et})$	[38]		211.7(4)	57
	$r(\mathrm{Fe{-}C}^{ax})$	[39]		181.5(2)	
	$r(\mathrm{Fe{-}C}^{eq})$	[40]		180.6(9)	
	$a(\mathrm{C}^{eq}\text{-}\mathrm{Fe{-}C}^{eq})$			111.7(9)	
$\mathrm{Fe(C_5Me_5)}(\eta^5\text{-}\mathrm{P_5})$ (1) C_5	$r(\mathrm{Fe{-}P})$	[41]	237.7(5)		58
$\mathrm{CoH(CO)_4}$ (1) C_{3v}	$r(\mathrm{Co{-}C}^{eq})$	[42]	181.8(3)	179.8(2)	59
	$a(\mathrm{H{-}Co{-}C}^{eq})$		80.3(6)	81.4(11)	
$\mathrm{Co(CO)_3(NO)}$ (1) C_{3v}	$r(\mathrm{Co{-}N})$	[43]	165.8(6)	167.1(6)	60
	$r(\mathrm{Co{-}C})$	[44]	183.0(3)	184.3(3)	
$\mathrm{Ni(CO)_4}$ (1) T_d	$r(\mathrm{Ni{-}C})$	[45]	182.5(2)	183.8(2)	61
$\mathrm{Ni(acac)_2}$ (1) D_{2h}	$r(\mathrm{Ni{-}O})$	[46]		187.6(5)	62
$\mathrm{Ni(PF_3)_4}$ (1) T_d	$r(\mathrm{Ni{-}P})$	[47]	209.9(3)		63
$\mathrm{CuCH_3}$ (1) C_{3v}	$r(\mathrm{Cu{-}C})$	[48]		188.41(2)	64
CuCN(1) $C_{\infty v}$	$r(\mathrm{Cu{-}C})$	[49]		183.231(7)	65
$\mathrm{Cu(acac)_2}$ (2) D_{2h}	$r(\mathrm{Cu{-}O})$	[50]		191.4(2)	66

---

[a] Where available, $(r/a)_{\mathrm{a}/\alpha}$ values were taken as reference. [b] (In parentheses: multiplicity) acac = acetylacetonato, ax = axial, br = bridging, cent = central, Cp = cyclopentadienyl; eq = equatorial. et = ethylene, tmm = trimethylenemethane [in brackets: running number of bonds].

with one of several correlation parts, namely P86[16] PW91,[17] or LYP[18] (in parentheses: symbols used in combined forms). Other functionals comprise HCTH/407 (denoted HCTH)[6,19] as well as the meta-GGAs BMK,[20] VSXC,[21] TPSS,[22] and TPSS hybrid (denoted TPSSh).[23] A fine integration grid (75 radial shells with 302 angular points per shell) has been used, except for VSXC, which has been shown to require finer grids[24] (here we used 99 radial shells with 590 angular points).

The following basis sets were used: AE1 denotes the Wachters basis (augmented by two diffuse p and one diffuse d sets) on the metals[25] (8s7p4d, full contraction scheme 62111111/3311111/3111) and 6-31G* on the ligands;[26] AE2 stands for the same augmented Wachters basis on the metals, polarized by an additional set of f-functions (exponents 0.60, 0.69, 0.78, 0.87, 0.96, 1.05, 1.17, 1.29, 1.44 for Sc−Cu), and 6-311+G* for the ligands.[27] svp,[28] tzvp,[29] and qzvp[30] are the polarized split-valence basis sets from Ahlrichs and co-workers; the latter qzvp basis has been adopted as in a recent study by Furche and Perdew (i.e. in the 11s6p5d3f1g contraction for the metals).[6] In addition to these all-electron basis sets, small-core effective core potentials (ECPs) with the corresponding valence basis sets were also employed on the metals, namely nonrelativistic LANL2DZ[31] (with [3s3p2d] valence basis and Dunning's double-ζ basis[32] on the ligands), and SDD,[33] i.e., the relativistic Stuttgart-Dresden ECP (together with the [6s5p3d1f] valence basis and 6-31G* basis on the ligands—note that the latter is not supplied with the SDD keyword in the Gaussian program). For essentially all levels, the minimum character of all optimized structures was verified by evaluation of the harmonic vibrational frequencies. Closed- and open-shell species were treated with restricted and unrestricted formalisms, respectively.

Results and Discussion

Selection of Reference Values. In addition to the precision criterion mentioned in the Introduction, we limited our selection to molecules measured at room temperature or slightly above. Thus, the large body of GED data for simple transition-metal di- and trihalides[34] has not been included because typical experimental temperatures reach or exceed 1000 K (see below for selected halides where equilibrium distances have been derived). In many cases, not all degrees of freedom have been refined experimentally, and only mean values for formally nonequivalent distances are known. In those cases, we evaluated and assessed the same average of the corresponding optimized parameters, even though full geometry optimizations were performed. For $Fe(CO)_5$, only the mean value of equatorial and axial Fe−C bond lengths is known with the required precision, and the difference between them is associated with a much larger uncertainty;[54] again, only the mean value was evaluated and assessed.

The final selected experimental parameters are collected in Table 1. Most distances are r_a or r_α values determined from GED (third column in Table 1). When both sets were given in the original papers, r_a values were chosen, as these should be closer to the equilibrium values, r_e. In some cases, only r_z and/or r_0 geometries are known from MW spectroscopy (fourth column in Table 1).[35] When both GED and MW data are available, both sets are given in Table 1, but only the former (third column) has been taken as reference values. In some GED studies, averaged structures (r_{av} or r_g) were derived from force fields and refined amplitudes; when no MW data are known, these values are included in the fourth column of Table 1 for comparison. In most cases, the various data sets are in good mutual accord, with differences on the order of 1 pm. When larger deviations were encountered (e.g. more than 5 pm between r_a and r_g in $Mn(CO)(NO)_3$),[51] these parameters were excluded from the reference set.

Looking at the final set in Chart 1, it seems that high-valent early and low-valent middle and late transition-metal complexes are present in a fairly well-balanced manner. It should be noted, however, that hydride and phosphine complexes are rather under-represented. The only metal−hydrogen distance included in the test set refers to that involving a bridging borane, namely the Ti−D distance in $Ti(BD_4)_3$ (metal−H distances in hydride complexes are ususally associated with large uncertainties, e.g. almost 2 pm for the Co−H bond in $CoH(CO)_4$). In light of the importance of aliphatic and aromatic phosphines as versatile ligands in transition-metal chemistry, the lack of any gas-phase structures of a 3d-metal with these moieties is particularly deplorable. The only metal−P distances included in the test set are those involving PF_3 and cyclic P_5 ligands. It would be highly desirable to have more, accurate reference structures for the important class of phosphine complexes.

Performance of the Models. The distances optimized with selected density-functional/basis-set combinations are given as Supporting Information. The resulting statistical assessment, that is, the mean deviations from the reference data in Table 1, and the standard deviations from those mean values[67] are summarized in Table 2. Deviations are defined as $r_{calc} - r_{exp}$, so that positive mean deviations denote overestimation of the bond lengths by DFT. In addition, the (unsigned) maximum errors are included in Table 1. First, all functionals were tested with the medium-sized basis set denoted AE1, which consists of the augmented all-electron Wachters basis on the metal and 6-31G* basis on the ligands. Second, other basis sets were employed for selected functionals, notably BP86 and B3LYP. Truly large basis sets, such as the generally contracted ANO bases by Roos and co-workers used in benchmark calculations,[68] are quite expensive for routine applications to larger molecules and have therefore not been used. The largest basis employed in this study is the Ahrichs-type qzvp basis from ref 30. The following conclusions can be drawn from our results:

1. LSDA (entry 1) produces much too short bonds, consistent with the well-known tendency for overbinding at that level.[69] This functional should not be used for computations involving 3d-transition metals.

2. Among the "pure" GGAs, BP86 appears to be slightly superior over the other ones, at least in conjunction with the medium-sized AE1 basis (entries 2−5). Most of these GGAs produce quite similar geometries, with the exception of BLYP (entry 3), which tends to overestimate the bond distances considerably (cf. the mean deviation of more than 2 pm).

Table 2: Statistical Assessment of Various Density-Functional/Basis-Set Combinations in Terms of Mean Signed Deviations from the Reference Distances *r* in Table 1,[a] Standard Deviations from These Mean Signed Values, and Absolute Maximum Errors[b,c]

entry	functional	basis set	mean (MAD)	standard	maximum
1	LSDA	AE1	−3.80 (3.82)	2.07	9.46 [35]
2	BP86	AE1	0.35 (1.46)	1.76	4.90 [35]
3	BLYP	AE1	2.31 (2.43)	2.04	6.94 [24]
4	HCTH	AE1	−0.21 (1.71)	2.16	5.50 [50]
5	OLYP	AE1	0.43 (1.71)	2.13	5.97 [50]
6	B3LYP	AE1	0.54 (1.61)	1.87	4.52 [24]
7	O3LYP	AE1	−0.38 (1.64)	1.95	5.31 [35]
8	B3P86	AE1	−1.36 (1.65)	1.57	5.94 [35]
9	BMK	AE1	0.02 (1.89)	2.30	5.42 [41]
10	VSXC	AE1	1.56 (1.83)	1.71	6.20 [24]
11	TPSS	AE1	0.13 (1.12)	1.48	4.44 [35]
12	TPSSh	AE1	−0.51 (1.14)	1.39	4.86 [35]
13	BP86	LANL2DZ	1.05 (3.10)	3.77	12.93 [47]
14	BP86	SDD	−0.63 (1.55)	1.88	5.82 [35]
15	BP86	AE2	0.11 (1.40)	1.75	4.98 [35]
16	BP86	svp	−0.40 (1.58)	1.89	5.73[35]
17	BP86	tzvp	0.43 (1.58)	1.91	5.10 [35]
18	BP86	qzvp	0.02 (1.40)	1.75	5.14 [35]
19	B3LYP	LANL2DZ	1.19 (3.01)	3.98	14.30 [47]
20	B3LYP	SDD	−0.46 (1.73)	2.06	4.41 [35]
21	B3LYP	AE2	0.31 (1.63)	1.99	4.13 [24]
22	B3LYP	svp	−0.21 (1.67)	1.96	4.42 [35]
23	B3LYP	tzvp	0.62 (1.68)	2.00	4.85 [24]
24	B3LYP	qzvp	0.28 (1.61)	1.95	4.14 [24]
25	TPSS	svp	−0.57 (1.32)	1.57	5.16 [35]
26	TPSS	tzvp	0.20 (1.23)	1.57	4.57 [35]
27	TPSS	qzvp	−0.20 (1.17)	1.48	4.70 [35]

[a] In parentheses: mean absolute deviations. [b] In brackets: number of the corresponding bond as identified in Table 1. [c] All values in pm.

3. Inclusion of exact exchange tends to result in decreased bond lengths and slight improvement of the standard deviation with respect to the corresponding nonhybrid functional (compare entries 2 vs 8, 3 vs 6, 5 vs 7, or 11 vs 12). The errors of BLYP are so large, however, that the corrections brought about by the corresponding hybrid functional are not sufficient, and B3LYP is worse than BP86 (with virtually any basis set). B3P86 is superior to BP86 in terms of the standard deviation but produces significantly too short bonds (entry 8).

4. Of the meta-GGAs that have been tested (entries 10 and 11), TPSS affords the best accord with experiment, slightly better even than the best GGA, BP86. The hybrid-variant TPSSh (entry 12) furnishes the lowest standard deviation of all functionals studied here, below 1.4 pm, but is also quite prominent in underestimating the bond lengths. The BMK functional produces a large scatter (cf. the large standard deviation in entry 9), presumably due to the large amount of Hartree−Fock exchange (42%),[20] which, unlike for some reaction barriers, does not appear to be beneficial in our case.

5. The choice of basis set affects mostly the mean deviation from experiment and not the scatter. With BP86, for instance, the standard deviation is virtually the same for the all-electron bases (1.7−1.9 pm, cf. entries 2 and 15−18), whereas the mean error varies between ca. −0.4 and +0.5 pm. These "extreme" values refer to the Ahlrich-type svp and tzvp bases, respectively; the much larger qzvp basis has a mean error between, i.e., close to zero. Similar observations are made with B3LYP (entries 7 and 21−24), with mean deviations increased by ca. 0.2−0.3 pm with respect to the corresponding BP86 values.

6. The relativistic Stuttgart-Dresden pseudopotential (included in SDD, entries 14 and 20) affords shorter bonds and a somewhat larger scatter than obtained with similar all-electron basis sets (compare e.g. entries 7 and 20). The Hay-Wadt pseudopotential was tested only in conjunction with a very small DZ basis on the ligands, a popular combination for routine applications in the literature. Clearly, this basis is much too small, and the resulting LANL2DZ level has the lowest precision of all, judged by the large standard deviation of almost 4 pm (entries 13 and 19). The shortcoming of this small basis is particularly pronounced for distances to the second-row element P, which are overestimated by more than 10 pm (cf. maximum errors for entries 13 and 19 in Table 2).

Compared to the distances, less attention was paid to the bond angles. In the whole test set, there are only 15 angles about the metal centers that are not dictated by symmetry and that have been refined with a precision better than 1°. These angles *a* cover the range between ca. 77° and 117° and are included in Table 1. At most levels studied, these angles are reproduced within ±0.1° (mean error) and a standard deviation around ca. 1.2°−1.5°, with little discrimination between the various functional/basis combinations.

To conclude this section, most GGAs are quite robust in reproducing geometries of transition-metal complexes, and effects of exact exchange or basis sets are usually small or moderate. The TPSS meta-GGA as well as its hybrid variant furnishes the most accurate metal−ligand distances, as judged from the reasonably small standard deviation from experiment, ca. 1.4 pm. The best GGA is BP86, slightly superior to B3LYP. The performance of these three functionals is shown schematically in Figure 1a, a plot of normalized Gaussian distributions using the corresponding data from Table 1 (analogous to the presentation by Helgaker et al.).[1,2] Figure 1b illustrates the basis-set dependence for one particular density functional, BP86, where increase of the basis results in but small shifts of the normal distribution, and hardly affects its width.

When assessing the results in terms of mean deviations from experiment, it should be kept in mind that calculated equilibrium geometries are compared to mean or effective observed ones. Thus, perfect agreement (i.e. zero mean deviation) is not to be expected, and the optimized distances should be systematically shorter than the experimental ones. In principle, equilibrium distances can be deduced from GED experiments, provided accurate force fields and refined vibrational amplitudes are available (see section on dihalides below).[70] Such analyses are rather involved, however, and

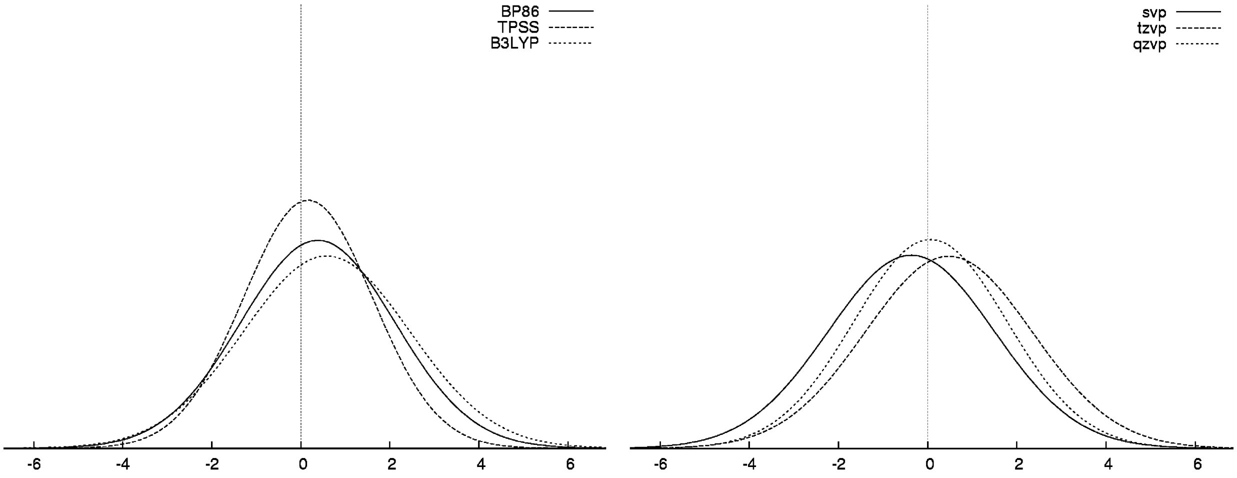


Figure 1. Normal distributions for the errors in the optimized bond distances for the test set in Chart 1. The distributions have been calculated from the mean and standard deviations in Table 2 and are all normalized to one: (a) left, dependence on the density functional using AE1 basis and (b) right, dependence on the basis set for the BP86 functional.

therefore rare. A recent example comprises ferrocene, for which an equilibrium r_e(Fe−C) distance of 205.4(3) pm has been inferred,[71] 1 pm shorter than the r_a value in Table 1. Rovibrational corrections to equilibrium bond distances can also be computed quantum-mechanically (from anharmonic force fields).[72] We have recently computed zero-point effects, which contribute significantly to these corrections, for a number of Ti-, V-, Mn-, Fe-, and Co-complexes.[73] In the vast majority of cases, such zero-point corrections tend to increase bond distances from their r_e values, by up to ca. 1 pm.[74] This value, which is comparable to the precision of the chosen experimental parameters, thus marks a lower limit for the degree of agreement that one can hope to achieve in our analysis.

In addition, GED results for transition-metal complexes, even when reasonably precise, need not necessarily be highly accurate. If any decomposition reactions during vaporization of the samples go undetected, the observed radial distributions and, thus, the structural parameters derived thereof may be affected noticeably. In view of these limitations, the performance of most DFT methods is actually quite satisfactory. However, an accuracy as that achievable with highly sophisticated ab initio methods for equilibrium bond distances of light main-group compounds (eg. mean and standard deviation around 0.2 and 0.3 pm, respectively, at CCSD(T)/cc-pVQZ)[1,75] appears to be out of reach, or at least undetectable, for transition-metal complexes.

Performance for Simple Halides. As mentioned above, experimental structures determined at very high temperatures *T* were not included in the test set, because the shift of thermally averaged parameters from their equilibrium values is expected to increase with *T*. For a number of transition-metal dihalides MX_2 (M = Mn, Fe, Co, Ni, X = F, Cl), however, facilitated by their high symmetry, such high-temperature GED studies in combination with vibrational spectroscopy have been used to deduce equilibrium bond distances. These r_e values (collected in Table 3), which can be directly compared with the DFT optimizations, are different in nature from the effective or averaged parameters in Table 1. Hence, we decided to evaluate the results for these dihalides, together with those from the diatomics CuF and CuCl (for which accurate equlibrium geometries are also known), separately. Mean and standard deviations for DFT geometries[76] are collected in Table 4.

Table 3: Experimental Equilibrium Bond Lengths (r_e in pm) of First-Row Transition-Metal Halides in the Gas Phase[a]

compd (mult)	[bond no.]	r_e(M−X)	ref
MnF_2 (6)	[51]	179.7(6)	79
FeF_2 (5)	[52]	175.5(6)	79
CoF_2 (4)	[53]	173.8(6)	79
NiF_2 (3)	[54]	171.5(7)	79
CuF (1)	[55]	174.4922(21)	80
$MnCl_2$ (6)	[56]	218.4(5)	81
$FeCl_2$ (5)	[57]	212.8(5)	81
$CoCl_2$ (4)	[58]	209.0(5)	81
$NiCl_2$ (3)	[59]	205.6(5)	81
CuCl (1)	[60]	205.11778(8)	82

[a] In parentheses: multiplicity; in brackets: running number of bonds.

Evidently, with just 10 distances this subset is smaller and less diverse than the larger test set of Chart 1. However, very similar conclusions can be drawn regarding the performance of the various exchange-correlation functionals. For instance, BP86 (together with the related BPW91) performs best among the GGAs, clearly better than B3LYP, and slightly inferior only to TPSS (see Figure 2 for a graphical representation).[77] Basis-set effects tend to be similar for the halides and the larger test set as far as the mean error is concerned, but for the halides the standard deviation is somewhat more sensitive to basis-set enlargement, which affords a noticeable improvement (compare e.g. the svp-tvzp-qzvp triads in Table 4, such as entries 30−32). LANL2DZ is very poor throughout, in keeping with similar findings in the literature.[78]

Again, TPSS and its hybrid variant produce the best results, with mean and standard deviations around ca. ±0.2 and 0.8 pm, respectively. These functionals thus emerge as very promising tools for first-principles calculations of transition-metal complexes. Other recent studies, which also

Table 4: Statistical Assessment of the Various Density-Functional/Basis-Set Combinations in Terms of Mean, Signed Deviations from the Reference Distances in Table 3,[a] Standard Deviations from These Mean Signed Values, and Absolute Maximum[b,c]

entry	functional	basis set	mean (MAD)	standard	maximum
1	LSDA	AE1	−4.37 (4.36)	0.74	5.29 [57]
2	BP86	AE1	−0.21 (0.88)	1.09	2.17 [60]
3	BLYP	AE1	1.17 (1.33)	1.47	4.16 [60]
4	HCTH	AE1	0.63 (0.88)	1.16	3.43 [60]
5	OLYP	AE1	1.15 (1.35)	1.37	4.05 [60]
6	BPW91	AE1	0.07 (0.79)	1.09	2.49 [60]
7	B3LYP	AE1	1.07 (1.49)	1.66	4.20 [60]
8	O3LYP	AE1	1.03 (1.34)	1.50	4.15 [60]
9	B3P86	AE1	−0.21 (0.95)	1.23	2.40 [60]
10	B3PW91	AE1	0.36 (1.05)	1.31	3.03 [60]
11	BMK	AE1	1.75 (2.04)	1.99	5.07 [60]
12	VSXC	AE1	1.42 (1.53)	1.43	4.20 [60]
13	TPSS	AE1	−0.02 (0.87)	1.09	2.17 [60]
14	TPSSh	AE1	0.19 (0.98)	1.21	2.55 [60]
15	BP86	LANL2DZ	4.08 (4.08)	2.53	8.91 [60]
16	BP86	SDD	−0.32 (1.33)	1.25	1.78 [51]
17	BP86	AE2	0.54 (0.89)	1.24	2.80 [59]
18	BP86	svp	−0.06 (0.99)	1.23	2.17 [60]
19	BP86	tzvp	0.87 (0.96)	1.14	3.22 [60]
20	BP86	qzvp	−0.46 (0.93)	0.98	1.69 [55]
21	B3LYP	LANL2DZ	6.44 (6.44)	2.73	10.11 [60]
22	B3LYP	SDD	1.82 (2.62)	2.65	5.39 [59]
23	B3LYP	AE2	2.04 (2.04)	1.33	4.83 [60]
24	B3LYP	svp	1.08 (1.54)	1.70	4.01 [60]
25	B3LYP	tzvp	2.27 (2.27)	1.54	5.54 [60]
26	B3LYP	qzvp	1.24 (1.24)	1.14	3.59 [60]
27	TPSS	LANL2DZ	4.69 (4.69)	2.64	9.50 [60]
28	TPSS	SDD	0.74 (2.08)	1.71	3.86 [54]
29	TPSS	AE2	0.70 (0.82)	1.02	2.58 [60]
30	TPSS	svp	0.24 (1.12)	1.36	2.42 [60]
31	TPSS	tzvp	1.00 (1.00)	1.03	3.18 [60]
32	TPSS	qzvp	−0.16 (0.60)	0.73	1.17 [55]
33	TPSSh	LANL2DZ	4.71 (4.71)	2.63	9.63 [60]
34	TPSSh	SDD	0.87 (1.67)	1.72	3.51 [54]
35	TPSSh	AE2	0.88 (0.91)	1.05	2.92 [60]
36	TPSSh	svp	0.47 (1.20)	1.46	2.76 [60]
37	TPSSh	tzvp	1.18 (1.18)	1.12	3.57 [60]
38	TPSSh	qzvp	0.06 (0.58)	0.77	1.46 [60]

[a] In parentheses: mean absolute deviations. [b] In brackets: number of the corresponding bond as identified in Table 3. [c] All values in pm.

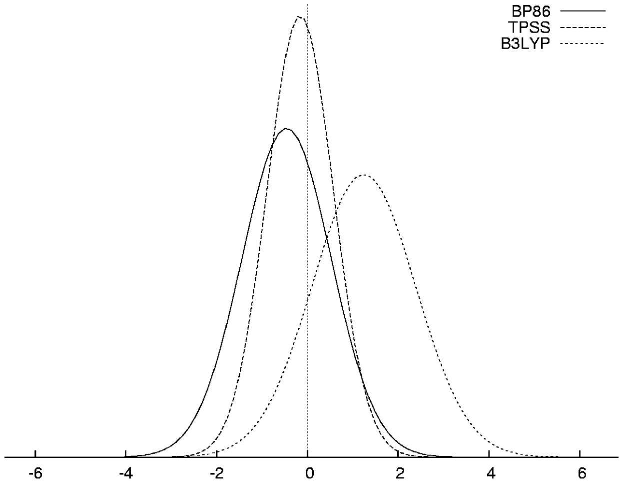


Figure 2. Normal distributions for the errors in the equilibrium bond distances in mono- and dihalides. The distributions have been calculated from the mean and standard deviations in Table 4, normalized to one, and drawn to scale with the plots in Figure 1. Selected density functionals have been employed, together with qzvp basis.

included assessments of computed binding and reaction energies,[6,83] reached the same conclusion.

Conclusions

We have tested a number of density-functional/basis-set combinations for their ability to reproduce experimental bond distances of transition-metal complexes in the gas phase. For this purpose, two test sets were selected from the literature, a larger one containing 50 effective or averaged metal−ligand distances, and a smaller one with 10 equilibrium metal−halide bond lengths. Care has been taken to include only parameters that have been determined with sufficient precision (all better than 1 pm, most better than 0.5 pm). Among the traditional GGAs, BP86 performs best, better than the popular B3LYP combination. In general, hybrid functionals furnish only minor improvements over their nonhybrid variants. The best results are obtained with a recently introduced meta-GGA, TPSS, with or without exact exchange included. Our findings thus corroborate other recent studies that had attested great potential of this functional for computational transition-metal chemistry. We have now established a solid ground for the assessment of molecular geometries in this area, upon which not only performance tests of other contemporary or future exchange-correlation functionals but also high-level ab initio methods can be based.

Acknowledgment. M.B. wishes to thank Prof. W. Thiel, the MPI Mülheim, and the Deutsche Forschungsgemeinschaft for continuing support. We thank H. Lenk for technical assistance. Computations were performed on an Intel Xeon and Opteron PC cluster at the MPI Mülheim.

Supporting Information Available: Bond distances of the test set in Table 1, optimized at selected levels, and BP86/qzvp optimized Cartesian coordinates. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Helgaker, T.; Gauss, J.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **1997**, *106*, 6430−6440.

(2) Bak, K. L.; Gauss, J.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Stanton, J. F. *J. Chem. Phys.* **2001**, *114*, 6548−6556.

(3) See for instance: Bray, M. R.; Deeth, R. J.; Paget, V. J.; Sheen, P. D. *Int. J. Quantum Chem.* **1997**, *61*, 85−91.

(4) Rosa, A.; Ehlers, A. W.; Baerends, E. J.; Snijders, J. C.; te Velde, G. *J. Phys. Chem.* **1996**, *100*, 5690−5696.

(5) Filatov, M.; Thiel, W. *Phys. Rev. A* **1998**, *57*, 189−199.

(6) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264−6271.

(7) (a) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 4388−4403. (b) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 11127−11143.

(8) Furche, F.; Perdew, J. P. *J. Chem. Phys*. **2006**, *124*, 044103.

(9) *Landolt-Börnstein, Structure Data of Free Polyatomic Molecules*; Kuchitsu, K., Ed.; Springer-Verlag: Berlin, 1998; New Series, Vol. II/25.

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

(11) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys*. **1980**, *58*, 1200−1211; functional III of that paper has been used.

(12) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(13) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5642.

(14) Handy, N. C.; Cohen, A. J. *Mol. Phys*. **2001**, *99*, 403−412.

(15) Cohen, A. J.; Handy, N. C. *Mol. Phys*. **2001**, *99*, 607−615.

(16) (a) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824. (b) Perdew, J. P. *Phys. Rev. B* **1986**,*34*, 7406.

(17) (a) Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P., Eischrig, H., Eds.: Akademie Verlag: Berlin, 1991. (b) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244−13249.

(18) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(19) Boese, A. D.; Handy, N. C. *J. Chem. Phys*. **2001**, *114*, 5497−5503.

(20) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405−3416.

(21) Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400−410.

(22) (a) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett*. **2003**, *91*, 146401. (b) Perdew, J. P.; Tao, J.; Staroverov, V. N.; Scuseria, G. E. *J. Chem. Phys*. **2004**, *120*, 6898−6911.

(23) (a) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys*. **2003**, *119*, 12129-12137. (b) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys*. **2004**, *121*, 11507.

(24) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett*. **2004**, *394*, 334−338.

(25) (a) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033−1036. (b) Hay, P. J. *J. Chem. Phys.* **1977**, *66*, 4377−4384.

(26) (a) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257−2261. (b) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(27) (a) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys*. **1980**, *72*, 650−654. (b) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem*. **1983**, *4*, 294−301.

(28) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys*. **1992**, *97*, 2571−2577.

(29) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys*. **1994**, *100*, 5829−5835.

(30) Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys*. **2003**, *119*, 12753−12762.

(31) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299−310.

(32) Dunning, T. H.; Hay, P. J. In *Modern Theoretical Chemistry;* Schaefer, H. F., Ed.; Plenum Press: New York, 1977; Vol. 4, pp 1−27.

(33) Dolg, M.; Wedig, U.; Stoll, H.; Preuss H. *J. Chem. Phys*. **1987**, *86*, 866.

(34) Review: Hargittai, M. *Chem. Rev*. **2000**, *100*, 2233−2301.

(35) r_e is the equilibrium distance between the positions of the nuclei on the potential energy surface, r_a is the effective internuclear distance as derived from electron scattering intensities, r_α is the distance between average nuclear positions in the thermal equilibrium at temperature T, r_z is the distance between average nuclear positions in the ground vibrational state, r_0 is the effective internuclear distance obtained from the rotational constants, r_{av} and r_g are the average internuclear distance at temperature T; see e.g.: Hargittai, I. In *Stereochemical Applications of Gas-Phase Electron Diffraction, Part A: The Electron Diffraction Technique*; Hargittai, I., Hargittai, M., Eds.; VCH Publisher: Weinheim, 1988; pp 1−54.

(36) Belova, N. V.; Giricheva, N. I.; Girichev, G. V.; Shlykov, S. A.; Tverdova, N. V.; Kuz'mina, N. P.; Zaitseva, I. G. *J. Struct. Chem*. **2002**, *43*, 56−63.

(37) Morino, Y.; Uehara, H. *J. Chem. Phys*. **1966**, *45*, 4543−4550.

(38) Briant, P.; Green, J.; Haaland, A.; Møllendal, M.; Rypdal, K.; Tremmel, J. *J. Am. Chem. Soc.* **1989**, *111*, 3434−3436.

(39) McGrady, G. S.; Downs, A. J.; McKean, D. C.; Haaland, A.; Scherer, W.; Verne, H.-P.; Holden, H. V. *Inorg. Chem*. **1996**, *35*, 4713−4718.

(40) Dain, C. J.; Downs, A. J.; Goode, M. J.; Evans, D. G.; Nicholls, K. T.; Rankin, D. W. H.; Robertson, H. E. *J. Chem. Soc., Dalton Trans*. **1991**, 967−977.

(41) Almenningen, A.; Samdal, S. *J. Mol. Struct*. **1978**, *48*, 69−78.

(42) Hagen, K.; Gilbert, M. M.; Hedberg, L.; Hedberg, K. *Inorg. Chem*. **1982**, *21*, 2690−2693.

(43) (a) Karakida, K.-I.; Kuchitsu, K. *Inorg. Chim. Acta* **1975**, *13*, 113−119. Very similar GED data in (b) Oberhammer, H.; Strähle, J. *Z. Naturforsch*. **1975**, *30a*, 296−303.

(44) Haaland, A.; Rypdal, K.; Volden, H. V.; Andersen, R. A. *J. Chem. Soc., Dalton Trans*. **1992**, 891−895.

(45) Almond, M. J.; Page, E. M.; Rice, D. A.; Hagen, K. *J. Organomet. Chem*. **1996**, *511*, 303−307.

(46) French, C. R. J.; Hedberg, L.; Hedberg, K.; Gard, G. L.; Johnson, B. M. *Inorg. Chem*. **1983**, *22*, 892−895.

(47) Marsden, C. J.; Hedberg, L.; Hedberg, K. *Inorg. Chem.* **1982**, *21*, 1115−1118.

(48) Marsden, C. J.; Hedberg, K.; Ludwig, M. M.; Gard, G. L. *Inorg. Chem.* **1991**, *30*, 4761−4766.

(49) Haaland, A. *Acta Chem. Scand.* **1965**, *19*, 41−46.

(50) Assuming equal CC bond lengths: (a) Chiu, N.-S.; Schäfer, L. *J. Organomet. Chem.* **1975**, *101*, 331−346. Similar CrC distances (with somewhat larger uncertainties) have been inferred from microwave studies: (b) Sickafoose, S. M.; Breckenridge, S. M.; Kukolich, S. G. *Inorg. Chem.* **1994**, *33*, 5176−5179. For evidence for alternating CC bonds see e.g. (c) Kukolich, S. G. *J. Am. Chem. Soc.* **1995**, *117*, 5512−5514.

(51) Hedberg, L.; Hedberg, K.; Satija, S. K.; Swanson, B. I. *Inorg. Chem.* **1985**, *24*, 2766−2771: r_α values calculated from mean $<M_C,M_N>$ and difference $\Delta(M_C,M_N)$ distances.

(52) (a) Javan, A.; Engelbrecht, A. *Phys. Rev.* **1954**, *96*, 649−658. For a complete recent assignment of the microwave spectrum, see: (b) Bürger, H.; Weinrath, P.; Dressler, S.; Hansen, T.; Thiel, W. *J. Mol. Spectrosc.* **1997**, *183*, 139−150.

(53) Almond, M. J.; Page, E. M.; Rice, D. A.; Hagen, K.; Volden, H. V. *J. Mol. Struct.* **1994**, *319*, 223−230; $Mn-C^{Cp}$ calculated from $r_\alpha(Mn-X) = 1.773(3)$ Å and $r_a(CC) = 1.423(2)$ Å from ree rotation model.

(54) McClelland, B. W.; Robiette, A. G.; Hedberg, L.; Hedberg, K. *Inorg. Chem.* **2001**, *40*, 1358−1362.

(55) Almenningen, A.; Haaland, A.; Wahl, K. *Acta Chem. Scand.* **1969**, *23*, 1145−1150.

(56) Haaland, A.; Nilsson, J. E. *Acta Chem. Scand.* **1968**, *22*, 2653−2670.

(57) Droiun, B. J.; Kukolich, S. G. *J. Am. Chem. Soc.* **1999**, *121*, 4023−4030.

(58) Blom, R.; Brück, T.; Scherer, O. J. *Acta Chem. Scand.* **1989**, *43*, 458−462.

(59) (a) MW: Kukoloch, S. G.; Sickafoose, S. M. *J. Chem. Phys.* **1996**, *105*, 3466−3471. (b) GED: McNeill, E. A.; Scholer, F. R. *J. Am. Chem. Soc.* **1977**, *99*, 6243−6249. In this case, the MW parameters were taken as reference, because the relative sequence of $r(Co-C^{ax})$ and $r(Co-C^{eq})$ appears suspect in the GED structure; $r(Co-H)$ and $r(Co-C^{ax})$ have not been used because of large uncertainties in both studies.

(60) Hedberg, K.; Hedberg, L.; Hagen, K.; Ryan, R. R.; Jones, L. H. *Inorg. Chem.* **1985**, *24*, 2771−2774.

(61) Hedberg, L.; Iijima, T.; Hedberg, K. *J. Chem. Phys.* **1979**, *70*, 3224−3229.

(62) Shibata, S.; Ohta, M.; Tani, R. *J. Mol. Struct.* **1981**, *73*, 119−124.

(63) (a) Almenningen, A.; Andersen, B.; Astrup, E. E. *Acta Chem. Scand.* **1970**, *24*, 1579−1584. A similar distance with a larger uncertainty has been determined independently: (b) Marriott, J. C.; Salthouse, J. A.; Ware, M. J. *J. Chem. Soc., Chem. Commun.* **1970**, 595−596.

(64) MW, r_0 value: Grotjahn, D. B.; Halfen, D. T.; Ziurys, L. M.; Cooksy, A. L. *J. Am. Chem. Soc.* **2004**, *126*, 12621−12627.

(65) MW, r_0 value: Grotjahn, D. B.; Brewster, M. A.; Ziurys, L. M. *J. Am. Chem. Soc.* **2002**, *124*, 5895−5901.

(66) Shibata, S.; Sasase, T.; Ohta, M. *J. Mol. Struct.* **1983**, *96*, 347−352.

(67) Defined, e.g., in eqs 2 and 3 in ref 1.

(68) Pouamerigo, R.; Merchan, M.; Nebotgil, I.; Widmark, P. O.; Roos, B. O. *Theor. Chim. Acta* **1995**, *92*, 149−181.

(69) See for instance: Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH: Weinheim, 2001.

(70) See e.g.: *Stereochemical Applications of Gas-Phase Electron Diffraction, Part A: The Electron Diffraction Technique*; Hargittai, I., Hargittai, M., Eds.; VCH: Weinheim, 1988.

(71) Coriani, S.; Haaland, A.; Helgaker, T.; Jørgensen, P. *ChemPhysChem* **2006**, *7*, 245−249.

(72) See e.g.: Solomonik, V. G.; Stanton, J. F.; Boggs, J. E. *J. Chem. Phys.* **2005**, *122*, 094322.

(73) (a) Bühl, M.; Mauschick, F. T. *Magn. Reson. Chem.* **2004**, *42*, 737−744. (b) Grigoleit, S.; Bühl, M. *Chem. Eur. J.* **2004**, *10*, 5541−5552. (c) Grigoleit, S.; Bühl, M. *J. Chem. Theory Comput.* **2005**, *1*, 181−193.

(74) Computed at BP86/AE1; specific values for the compounds of this study are collected in Table S2 of the Supporting Information (9 complexes with 15 bonds in total). Here, the computed zero-point corrections to bond distances range from 0 to +0.4 pm for metal−ligand single bonds and from +0.6 to +0.7 pm for metal−carbon distances involving π-bonded ligands. Unfortunately, computation of these corrections for all molecules of the test set would be a formidable task beyond the scope of the present paper.

(75) Further, minor improvements for the correlated ab initio methods are brought about by basis sets augmented with diffuse and core-polarization functions, cf. ref 2. For the same set as employed in ref 1, we obtain mean and standard deviations of 0.7 and 0.6 pm, respectively, at TPSS/cc-pVQZ.

(76) At BP86/AE1, only linear minima could be located for all dihalides; thus all subsequent computations were done in $D_{\infty h}$ symmetry.

(77) For other classes of compounds, somewhat different performances of DFT methods may occur: e.g., for r_e values in coordinatively unsaturated diatomics involving first-row transition-metal atoms (including multiply bonded metal dimers), LSDA performs better than GGA or meta-GGA, and the quality of the results deteriorates upon inclusion of Hartree−Fock exchange, cf ref 7.

(78) Strassner, T.; Taige, M. A. *J. Chem. Theor. Comput.* **2005**, *1*, 848−855.

(79) Vogt, N. *J. Mol. Struct.* **2001**, *570*, 189−195.

(80) Hoeft, J.; Lovas, F. J.; Tiemann, E.; Törring, T. *Z. Naturforsch.* **1970**, *25a*, 35−39.

(81) Hargittai, M.; Subbotina, N. Y.; Kolonits, M.; Gershikov, A. G. *J. Chem. Phys.* **1991**, *94*, 7278−7286.

(82) Manson, E. L.; De Lucia, F. C.; Gordy, W. *J. Chem. Phys.* **1975**, *62*, 1040−1043.

(83) Quintal, M. M.; Karton, A.; Iron, M. A.; Boese, D.; Martin, J. M. L. *J. Phys. Chem. A* **2006**, *110*, 709−716.

CT6001187

J. Chem. Theory Comput. **2006,** *2,* 1291−1297

JCTC Journal of Chemical Theory and Computation

Electron Transmission through Aromatic Molecules

Matthias Ernzerhof,* Hilke Bahmann, Francois Goyer, Min Zhuang, and
Philippe Rocheleau

*Département de Chimie, Université de Montréal, C.P. 6128 Succursale A,
Montréal, Québec H3C 3J7, Canada*

Received March 7, 2006

Abstract: A prominent feature of aromatic compounds is the ring current that can be observed indirectly in nuclear magnetic resonance experiments. This current is generated by an external magnetic field. In molecular electronics, molecules serve as conductors, and they are connected to metallic contacts that act as electron sources and electron sinks. We show that ring currents can also be found in molecular electronic devices containing cyclic π-electron systems. The circular currents are related to interference phenomena that can render the molecule impenetrable to electrons. While only small currents pass through the molecule, large internal circular currents are stimulated. We conjecture that the internal currents should result in experimentally observable magnetic moments.

1. Introduction

To consider molecules as conductors[1−4] is a subject that has recently attracted much attention. Our understanding of electron transport through molecules is presently far from complete, and the mechanisms underlying molecular conductance are only now beginning to be explored. In this context theoretical studies of molecular conductors are a valuable tool that even allow us to conceive new experiments and predict their outcome. In fact, molecular electronic devices (MEDs) have first been suggested in a theoretical article.[5] It should, however, be noted that MEDs represent an enormous theoretical challenge and that our methods of modeling MEDs[6−13] are still in their infancy. In particular correlation effects are difficult to account for properly.[14]

An interesting question that arises in molecular electronics is whether established concepts of chemistry are still useful in the analysis of MEDs. Here, we consider the notion of the ring current that is a prominent feature of aromatic systems. Employing the simple Hückel model, we investigate the conductance of benzene and coronene, and we find indeed that ring currents are crucial to rationalize the transmission of electrons through aromatic molecules. We demonstrate that benzene and coronene exhibit prominent negative interference effects that are related to ring currents. Investigations similar to ours have already been performed for related systems.[15] In ref 16 the electron transmission probability for the benzene molecule with wires attached to it is discussed. An analysis of the molecular orbitals is presented to explain the interference phenomena observed. Maybe the most important advance made here, in comparison to earlier work, is that we analyze the internal current distribution in the molecule. We believe that this distribution is essential to fully understand the current transport through MEDs. From the molecular current distribution, we are able to calculate the magnetic moment generated by the ring currents. We predict that coronene, with a voltage applied to it, could generate a magnetic moment of several μ_B, where $\mu_B = (e\hbar/2m_e)$ is Bohr's magneton. The determining feature of the molecular conductors examined here is their circular delocalized π-electron system. This type of system continues to attract attention. Recently, the behavior of circular molecular conductors in magnetic fields has been studied (for review see ref 17). Three-terminal circular logic gates have been proposed,[18] and the dependence of the transmittance on the arrangement of the leads attached to a circular conductor has been investigated.[19]

In the remainder of this paper we use atomic units ($m_e = e^2 = \hbar = 1$), unless indicated otherwise.

2. Summary of Theoretical Concepts

A basic tool of molecular electronics is Landauer's formula[20] for the ballistic conductance $g(E)$

10.1021/ct600087c CCC: $33.50

$$g(E) = \frac{1}{2\pi} T(E) \qquad (1)$$

(Note that the conductance quantum (e^2/h) reduces to ($1/2\pi$) in atomic units.) If electron−electron interaction is considered through an effective single-particle approach, the transmission probability $T(E)$ can be obtained according to[6−8]

$$T(E) = Tr(\mathbf{\Gamma}_R \boldsymbol{G}^r \mathbf{\Gamma}_L \boldsymbol{G}^a) \qquad (2)$$

where $\boldsymbol{G}^{a(r)}(E)$ is the advanced (retarded) Green's function of the entire system. $\mathbf{\Gamma}_{L(R)}(E) = 2\pi \mathbf{M}_{L(R)} \delta(\mathbf{H}_{L(R)} - E)\mathbf{M}^{\ddagger}_{L(R)}$ contains the density of state operator $\delta(\mathbf{H}_{L(R)} - E)$ of the left (right) contact multiplied by the operator $\mathbf{M}_{L(R)}$ that couples the left (right) contact to the molecule. Numerous strategies are available to implement (see e.g., refs 10−12 and 21−23) this formula. Here we are interested in the properties of delocalized π systems, and it is known that the qualitative features of such systems can be reproduced with a simple Hückel model. Furthermore, we assume that the voltage applied to the system is small, such that its polarizing effect on the orbitals and thus on $T(E)$ can be neglected. Finally, we suppose that the contacts are identical, that they have a constant density of states (this is sometimes called the wide band limit), and that they couple to the molecule only through the molecular atom that is attached to the chain. In this case, $\mathbf{\Gamma}_{L(R)}(E)$ reduces[24] to a matrix with a single element. This element is treated as a parameter in our work, and we set it equal to the hopping matrix element used in the molecule. By varying the coupling strength, we ensured that the qualitative conclusions drawn here are not sensitive to the exact value of this parameter. We stress that in the present work we do not attempt to describe quantitative but rather qualitative effects that are robust with respect to the choice of the model parameters and with respect to the choice of the theoretical method employed to calculate $T(E)$. The simple approach just described makes it possible to quickly examine $T(E)$ of numerous structures. All the T versus E curves presented here have been generated with this scheme, the exception being Figure 9 described in the Appendix.

One of the main aims of our work is to use the internal current distribution as a tool to understand molecular conductance. All the current distributions presented in the following have been obtained with a method that we developed recently and that will be described in detail in a forthcoming publication.[25] Here we provide a short outline. We consider a finite model of a MED, the Hamiltonian of which contains a complex one-particle potential $\hat{\Sigma}$

$$\hat{H} = \hat{T} + \hat{v} + \hat{\Sigma} \qquad (3)$$

The Hamiltonian operator in eq 3 is non-Hermitian, nevertheless it permits the development of a consistent quantum mechanics.[25] To show that the Hamiltonian in eq 3 enables us to describe an MED with a stationary current passing through, we consider a one-dimensional contact along the x axis that extends from $-\infty$ to 0. This contact represents the left contact of a MED, and, at $x = 0$, we attach the molecule to it. In the left contact, sufficiently far away from the molecule, the wave function φ is a combination of a forward going Bloch wave $\varphi^+(x)$ and a backward going Bloch wave $\varphi^-(x)$

$$\varphi(x) \overset{\text{left contact}}{=} a\varphi^+(\mathrm{x}) + b\varphi^-(\mathrm{x}) \qquad (4)$$

Since we want our model system to be finite, we multiply $\varphi(x)$ by a damping factor $f^{\,L}(x)$. This function takes a value of one in the region that includes the molecule and beyond ($x \rightarrow \infty$), and it drops to zero inside of the left contact. We define

$$\phi(x) \overset{\text{left contact}}{=} f^{\,L}(x)\, \varphi(x) \qquad (5)$$

The damping factor $f^{\,L}(x)$ is of course not unique, but this does not pose a problem for our approach. Inserting ϕ into the stationary Schrödinger equation

$$\left(-\frac{1}{2}\Delta + \hat{v} + \hat{\Sigma}\right)\phi \overset{\text{left contact}}{=} E\phi \qquad (6)$$

and subsequent inversion of this equation yields an expression for the potential $\hat{\Sigma} = \hat{\Sigma}_L + \hat{\Sigma}_R$ in the left contact

$$\Sigma_L(x) \overset{\text{left contact}}{=} \frac{\Delta\phi(x)}{2\phi(\mathrm{x})} - v(x) + E \qquad (7)$$

Multiplication of ϕ by a constant does not change the value of the potential $\hat{\Sigma}_L$. This allows us to eliminate the coefficient a in eq 4, and we are left with one unknown complex parameter $r = b/a$ in $\hat{\Sigma}_L$. Note that for $|r| = 1$ there is no net current flow since forward and backward going Bloch wave have the same weight. In this case $\hat{\Sigma}_L$ is a real potential. If the molecule is transparent ($T = 1$), then $r = 0$. For $|r| \neq 1$, $\hat{\Sigma}_L$ is complex.

In the right contact that starts at $x = l$, we only have a forward going Bloch wave (φ^+) that we damp by a function $f^{\,R}(x)$ that drops to zero at some distance from the molecule. Similar to $f^{\,L}(x)$, $f^{\,R}(x)$ is equal to one in the physical region including the molecule and beyond ($x \rightarrow -\infty$). We have

$$\phi(x) \overset{\text{right contact}}{=} cf^{\,R}(x)\varphi^+(x) \qquad (8)$$

Inversion of the Schrödinger equation with the wave function in eq 8 yields $\hat{\Sigma}_R$. The parameter c cancels out of the equation for $\hat{\Sigma}_R$.

Equation 7 allows us to calculate $\hat{\Sigma}_L$ as a function of the parameter r that in turn depends on the energy E. r is directly related to $T(E)$ by

$$T(E) = 1 - |r(E)|^2 \qquad (9)$$

The remaining question is then how to obtain r? The condition that an eigenvalue of the model Hamiltonian in eq 3 is identical to the energy chosen in the potential $\hat{\Sigma}_{L(R)}$ suffices[13,25] to determine r. Note that the energy (and the eigenvalue of $\hat{H}$) of interest is real. The ideas just described are readily applied to the Hückel model.[25] The contacts, represented by infinite chains of equidistant atoms, are replaced by complex potentials. From the wave function that we obtain, we calculate the current distribution (using eq 10, given below). We verified that our model Hamiltonian

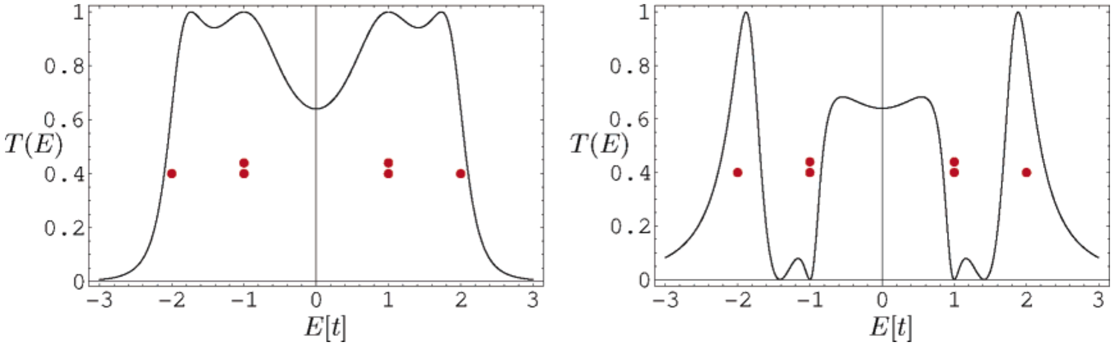


Figure 1. The transmission probability $T(E)$ of the benzene molecule with the leads attached in the para position (left) and in the ortho position (right). The dots indicate the positions of the orbital energies in the unperturbed molecule. The energy is given in units of the hopping parameter t.

yields the exact $T(E)$ of the underlying infinite system in which the contacts have not been replaced by $\hat{\Sigma}$.

As mentioned above, the $T(E)$ curves presented are obtained from eq 2 in the wide band limit. The features discussed here do not depend on which method we use, however. To demonstrate this, in the Appendix, $T(E)$ obtained from our non-Hermitian model Hamiltonian is compared to $T(E)$ obtained from eq 2.

In the Hückel model, that we employ in the present work, there are no continuous spatial variables. Instead, all equations are formulated in a finite-dimensional vector space. While the formulas given above can be transferred in a straightforward fashion to the Hückel model, it is useful to provide the expression for the current density. We find that the current density along a bond joining atoms k and l is

$$j_{k\rightarrow l} = i(c_k^* t c_l - c_l^* t c_k) \tag{10}$$

where the hopping matrix element between the atoms is t, and c_k is an orbital coefficient. Between two bound atoms, the Hückel model is effectively a one-dimensional model, and, in one dimension, the current density and the current are identical. The divergence of the current ($\nabla \mathbf{j}$) in the Hückel model is simply the difference between the current flowing in to and out of an atom. Apparently, $\nabla \mathbf{j}$ should be zero unless the atom in question has an imaginary matrix element. Finally, we list the Hückel parameters employed in the calculations. The on-site energies are set to zero. The hopping parameter in the contact chains is $1.4t$, where t is the hopping parameter in the molecule. The coupling matrix element between the contact and the molecule is also t.

3. Benzene Molecule

Using the tools discussed above, we now turn to the discussion of transmission probability through the prototypical aromatic molecule, namely benzene. It is well-known that the HOMO and LUMO of this molecule are doubly degenerate. The degeneracy originates from the fact that electrons with the same absolute value of momentum circulate left or right around the ring. If leads are added in the para position, the degeneracy is lifted, and we obtain two orbitals that are oriented such that one of them connects the two leads on opposite sides of the molecule and the second one is zero on the atoms connected to the leads. This behavior is isomorphic to the problem of two p orbitals: one with magnetic quantum number $m_z = 1(p_1)$ and one with magnetic quantum number $m_z = -1(p_{-1})$. An appropriate perturbation splits the degeneracy and yields a p_x- and a p_y-type orbital. The p_y-type orbital being zero along the direction of the p_x orbital. The transmission probability of the para-benzene is displayed in Figure 1. The curve can easily be understood in terms of the analysis presented above. Due to the symmetry of the perturbation, one of the molecular HOMOs is essentially excluded from the conductance path, and the other one yields a peak reaching up to one. We employ our method described above to generate a current flow through the molecule. A typical picture of the internal current distribution is shown in Figure 2. The arrows, indicating direction and magnitude of the current, have been normalized such that the longest arrow in each figure has a certain fixed length. Although this current distribution has been obtained for a particular energy ($E = -1.03$), it is representative for the entire energy range. After entering the molecule, the current splits symmetrically to pass through the π system. For the para-benzene, our findings agree essentially with ref 16, with the exception that we do not obtain the drop to zero in $T(E)$ at the position of the molecular HOMO and LUMO. These gaps in $T(E)$ are of zero width,[16] and they are probably eliminated by the level broadening inherent in the Green's function formalism (eq 2) used here to obtain $T(E)$. Features of zero width in $T(E)$ have probably no observable consequences.

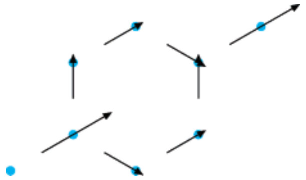


Figure 2. Current distribution in para-connected benzene at $E = -1.03$. At this energy, the molecule transmits well. This current distribution persists over the entire energy range except at $E = \pm1$, where the current distribution is zero. Note that this feature at $E = \pm1$ is not reproduced by the Green's function calculation of $T(E)$.

With the leads placed in the ortho or the meta position, the degeneracy of the molecular HOMO is also lifted, but there is no effective elimination of either resulting orbital from the conductance path. $T(E)$ touches zero when E passes through the molecular HOMO energy. This is an indication

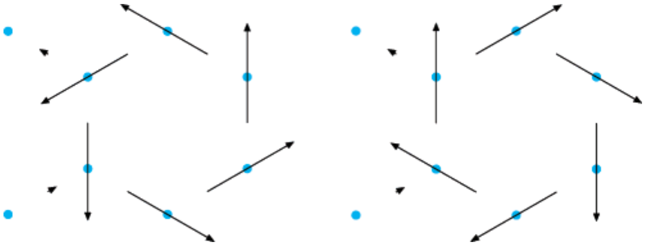


Figure 3. Current distribution in ortho-connected benzene. The two pictures show the current distribution slightly below (left) and above (right) $E = -1$. Note that the incoming and outgoing current is almost zero and that the direction of the current flow changes from one figure to the other.

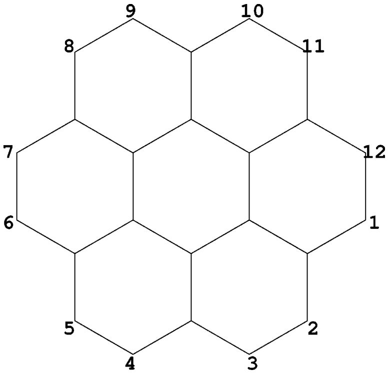


Figure 4. The carbon backbone of coronene. For the purpose of this article, the 1−3 position will be referred to as ortho, and the 1−7 position will be referred to as para.

of a strong negative interference effect that calls for a detailed investigation. The intramolecular current distribution slightly below and above $E = -1$ is shown in Figure 3. Below $E = -1$ there is a large counterclockwise circular current in the benzene molecule. This circular current yields only a small net transport through the molecule. Slightly above $E = -1$, we observe a similar phenomenon, but interestingly, the circular current is turning in the opposite direction. As a consequence of these opposing currents, there is a certain intermediate energy at which the circular currents cancel exactly such that no current flow through or inside of the molecule is obtained. At energies around that of the degenerate HOMOs, these orbitals essentially determine the internal current distribution. One of the HOMOs describes an electron moving clockwise, and the other one describes an electron moving counterclockwise. In the MED the degeneracy of the HOMOs is split, leading to the counter rotating electrons at different energies. We were not able to identify a simple criterion that would determine the orientation of the current circulation without having to perform the full calculation of the current distribution.

For energies close to ±1.4 there are additional zeros in $T(E)$. An inspection of the current distribution reveals that there is again a change in direction of a circular current, very similar to what is displayed in Figure 3. However, this interference is not associated with a resonance of a degenerate orbital, rather it is an interference between the tails of two resonances at different energies. We varied the parameters in our Hückel description, in particular the coupling between the contact and the molecule, to confirm that there are indeed additional interferences due to overlapping tails of resonances.

In the meta-benzene, that we do not discuss here, we also find strong interference effects similar to ortho-benzene, yet no new features are observed.

For the ortho-benzene, our results are also in agreement with the analysis of ref 16. These authors explain the interferences in terms of the phase and amplitude of the scattering wave function. This is a more general approach that does not bring to light the peculiar nature of the interference in the benzene, however.

4. Coronene Molecule

In Figure 4 we show the carbon backbone of the coronene molecule. For the purpose of this article, we define the 1−7 position to be para. Similar to benzene, coronene has a doubly degenerate HOMO and LUMO.

Connecting the leads in the para position leads to transmission features (Figure 5) analogous to the ones observed for the para-benzene. The degenerate molecular HOMO is split by the interaction with the contacts, and one of the resulting orbitals is eliminated from the conductance path, i.e., its orbital coefficients are zero in the 1- and 7-positions. The remaining orbital gives rise to a conductance channel that is completely open ($T(E) = 1$). The other doubly degenerate orbitals generate analogous transmission features. Interestingly, we also find transmission probabilities of zero for the para arrangement. To investigate this phenomenon, we study the current distribution at $E \approx -1.6$ (Figure 6).

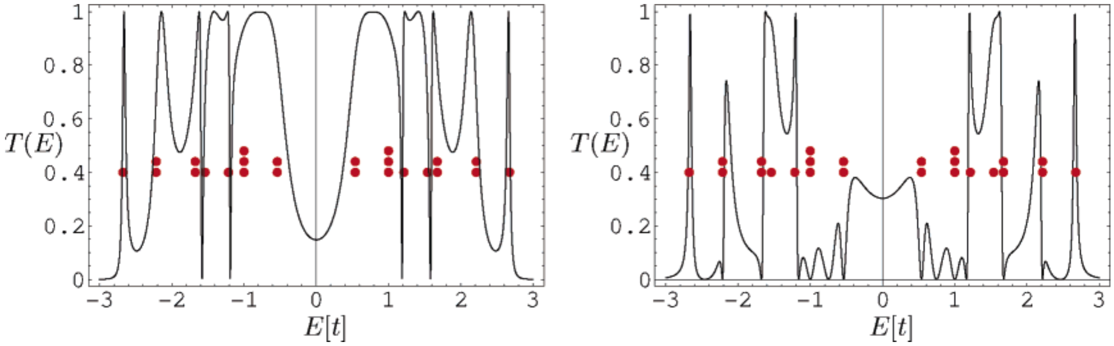


Figure 5. The transmission probability $T(E)$ of the coronene molecule with the leads attached in the para position (left) and in the ortho position (right). The dots indicate the positions of the orbital energies in the unperturbed molecule.

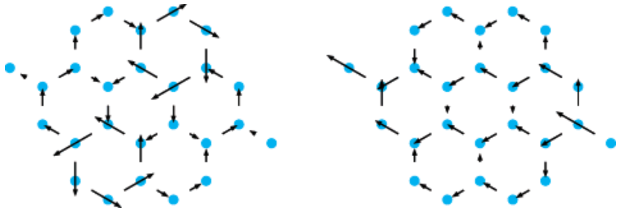


Figure 6. Intramolecular current distribution in the para-coronene for $E \approx -1.6$ (left) and $E \approx -0.9$ (right).

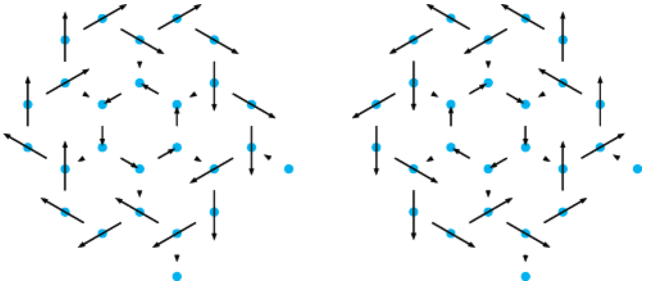


Figure 7. Shown are the current distributions slightly below (left) and above (right) $E \approx -0.6$. Note that the direction of the circular current changes from one figure to the other.

In the left panel of Figure 6, we see that there are two benzene-type substructures (upper right and lower left) that carry circular currents. The zero in $T(E)$ is due to the fact that at this particular energy the current is localized in a fragment of the molecule that does not include the atoms attached to the leads. The same explanation applies to the narrow feature at $E \approx \pm 1.2$.

To complete the discussion of the para-coronene, in Figure 6 we also show the current distribution at $E = \pm 0.9$. $T(E)$ is close to one for this energy. The entire molecule makes a coherent contribution to the electron transport. There is a slight preference for the current to pass through the antracene-type substructure joining the entry and exit point.

Similar to what we observed for the benzene molecule, there is a strong interference signature in $T(E)$ (Figure 5) of ortho coronene for energies around the molecular HOMO. There are again two pathways for electron transmission associated with orbitals slightly below and above the HOMO energy of the isolated molecule. As in the benzene, there is an intermediate energy at which the two pathways superimpose such as to annihilate each other. As a consequence, we observe complete destructive interference. To illustrate this point, in Figure 7 we show the current distribution at an energy slightly below and above the energy of the HOMO of the isolated molecule. There is a strong clockwise current flow in the outer ring and a counterclockwise current flow in the inner ring that is reversed upon passage through the HOMO energy. In the ortho-coronene there are not only resonant but also off-resonant destructive interferences. At $E \approx \pm 1.2$ we find such a feature that is associated with a change in direction of a circular current. For $E = \pm 1$ we find three degenerate orbitals in the molecule. At this energy, we observe another zero in $T(E)$. Inspection of $\mathbf{j}$ reveals that there is an isolated circular current localized in the center benzene ring that changes direction upon passage through $E = \pm 1$.

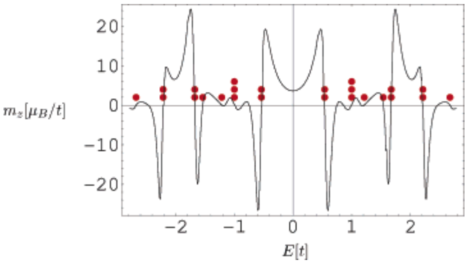


Figure 8. Magnetic moment generated by the internal current distribution of energy E. Three prominent oscillations are visible in each half of the figure. They are generated by circulating electrons changing direction. The corresponding states originate from doubly degenerate molecular orbitals, the position of which is indicated by red dots.

While we used the same Hückel parameters in the calculations for coronene and benzene, the peaks for the coronene are much narrower than the peaks for benzene. This is because the coronene is larger than the benzene, and the coronene orbitals are thus less perturbed by the addition of the contacts.

We considered all the other possible sites for wire attachments in the coronene molecule. In all cases we find similar interference patterns as the ones described here.

5. Observable Consequences of the Predicted Circular Currents

In the previous sections we showed that the concept of a ring current plays an important role in the interpretation of the transmission probabilities. We saw that large circular currents may be induced by the external contacts even if little current is flowing through the molecule. These circular currents should generate a magnetic moment **m**, normal to the molecular plane, that might be observable. To further elaborate on this issue, we calculate **m** as a function of the energy for the coronene. We are only interested in the contribution from the molecule, and we neglect contributions to **m** due to the contacts. The magnetic moment of a wire in a plane with current I passing through is[26]

$$\mathbf{m} = \frac{1}{2}\int \boldsymbol{r} \times d\mathbf{I} \tag{11}$$

where $d\mathbf{I}$ is a vector that points in the direction of the wire. The length of $d\mathbf{I}$ is scaled by I. Evaluation of this formula within the Hückel model yields[27]

$$\mathbf{m}(E) = \frac{1}{2}\sum_{i>k} j_{ik}(E)(\boldsymbol{r}_i \times \boldsymbol{r}_k) \tag{12}$$

$j_{ik}(E)$ is the current flowing along the bond (ik) in an orbital with energy E. $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ are the positions of atom i and j, respectively. Since coronene is planar, $\mathbf{m}(E)$ is perpendicular to the molecular plane, parallel to the z-axis. A plot of $m_z(E)$ is shown in Figure 8. In agreement with the observation of circular currents of rapidly changing orientation, there are strong variations in $m_z(E)$. For negative energies, we observe three large oscillations. These oscillations are generated by electrons circulating in the molecule in opposing directions, depending on their energy. Note that the large circular currents, i.e., large magnetic moments are associated with doubly degenerate molecular orbitals. As mentioned earlier, the 3-fold degenerate molecular level yields a circular current in a benzene-type substructure. Correspondingly, we observe a small oscillation in m_z at $E = \pm 1$. We should point out that the calculation of m_z is only approximately correct. Strictly speaking, the magnetic moment is only well defined for a closed charge distribution.[26] In our case, the external leads (that we neglected) also give rise to a contribution to the magnetic moment. Close to the points of destructive interference however, the current in the leads approaches zero, and thus the formula for the magnetic moment becomes correct.

If we assume that the voltage V applied to the molecule is small, such that $m_z(E)$ does not change as a function of V, then the total magnetic moment can be obtained from

$$m_z(V) = \int_{-E_1}^{E_1} dE m_z(E) \tag{13}$$

where $\Delta E = 2E_1 = eV$, and e is the electron charge.

In principle it is possible to assign values to the parameters in our Hückel treatment and to calculate $m_z(V)$. Instead, we give a simple estimate. We suppose that the integral in eq 13 extends from the HOMO to the LUMO. The magnetic moment of a circular current is given by

$$\mu = IA \tag{14}$$

where I is the current, and A is the surface area encircled by the current. The current can be obtained by

$$I = \frac{eh}{2\pi m r \lambda} \tag{15}$$

where e is the electron charge, m is the electron mass, r is the radius of the circle described by the current I, and λ is the electron wavelength. λ can be determined by inspection of the coronene π-electron system. The HOMO of coronene has an alternating bonding and antibonding structure. With this in mind, the electron wavelength in the HOMO is estimated to be four bond lengths (5.6 Å). By using eqs 14 and 15, we obtain a magnetic moment of about 4.1 μ_B for an electron circling on the outer ring of coronene. Because of spin degeneracy, this value has to be multiplied by a factor of 2. Furthermore, to estimate $m_z(V)$ in eq 13, where the moment is generated by the HOMO and the LUMO, we multiply by an additional factor of 2. Altogether, we obtain $m_z \approx 16.4\mu_B$. This is a magnetic moment that might be observable experimentally.

6. Discussion and Conclusion

For the representative aromatic molecules considered, we find that the most striking features in $T(E)$ are caused by negative interferences. By analyzing the internal current distribution, we demonstrate that often the interferences are related to opposing ring currents. Other interferences lead to localization of the current distribution in a substructure of the molecule that excludes atoms connected to the contacts.

One of the few studies of intramolecular current distributions has been performed for the C60 molecule.[27] Circular currents have also been observed in this system. In ref 27 and in our own studies on C60, it was found that for energies around the HOMO/LUMO there are no “global” current circulations, rather vortices are observed in five-membered rings. We conclude that the magnetic moments generated in C60 are much weaker than the ones reported here. (The moment is proportional to the area enclosed by the current.)

There is presently a renewed interest[28] in large circular π-electron systems because of their potential to exhibit the Aharonov-Bohm effect.[29] The giant circular molecule with ≈12 nm in diameter recently synthesized,[28] and systems such as Kekulene[30] might produce enormous magnetic moments in molecular conductance experiments. Although only small currents will pass through the molecule at the appropriate insertion energy, we predict that there will be large intramolecular currents. Until now, it appears that only para connected aromatic molecules have been investigated experimentally. The large molecules mentioned might permit a nonpara connection to the contacts that is difficult to realize with benzene and other small systems.

In the present work, we neglected the polarizing effect that an applied external field would have on the electronic structure. A finite applied voltage might considerably distort the molecular orbitals and hinder the appearance of circular currents. Furthermore, variations in the number of π-electrons (due to the contacts) can introduce geometry changes that also reduce the π-electron delocalization. However, recent progress[31] in achieving large gate fields might help in the experimental verification of the predicted circular currents. This is because gating allows a line up of the molecular orbitals with the Fermi level of the contacts. Thus it should not be necessary to apply a large voltage to the system in order to generate the predicted ring currents.

Acknowledgment. We would like to acknowledge the financial support provided by NSERC through Grant 238404-03.

Appendix: Comparison of Transmission Probabilities Obtained with Different Methods

To generate the $T(E)$ curves, we employed the Green's function method eq 2, describing the contacts by a single model parameter. To show that this approximation does not affect the qualitative features discussed here, we compare $T(E)$ (Figure 9) of a system with infinite chains of atoms as contacts (used to generate the internal current distribution) to the more approximate results.

If infinite one-dimensional contacts are used, then we have a band of states with a width of $2.8t$. The density of states of these chains is flat around the Fermi energy ($E = 0$), increases sharply towards the band edges to drop to zero precipitously beyond $E = \pm 2.8t$. This explains why the two curves deviate more in the outer left and right part of the figure. The approximate treatment of the contact used to generate the solid curve corresponds to a constant density of states. The zeros in $T(E)$ and its qualitative features are shared by both approaches.

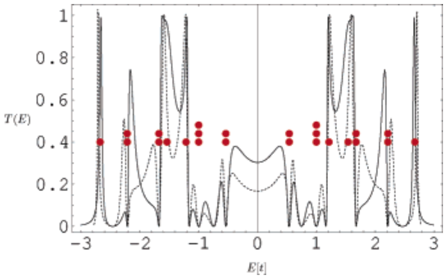


Figure 9. Comparison of transmission probabilities obtained from the Green's function approach (eq 2) with simplified contacts (solid curve) and from the non-Hermitian model Hamiltonian in eq 3 (dashed curve). The latter method yields results identical to those of the Green's function method with infinite one-dimensional contacts.

References

(1) Joachim, C.; Gimzewski, J. K.; Aviram, A. *Nature* **2000**, *408*, 541−548.

(2) Nitzan, A. *Annu. Rev. Phys. Chem.* **2001**, *52*, 681−750.

(3) Nitzan, A.; Ratner, M. A. *Science* **2003**, *300*, 1384−1389.

(4) Heath, J. R.; Ratner, M. A. *Phys. Today* **2003**, *56*, 43−49.

(5) Aviram, A.; Ratner, M. A. *Chem. Phys. Lett.* **1974**, *29*, 277−283.

(6) Meir, Y.; Wingreen, N. S. *Phys. Rev. Lett.* **1992**, *68*, 2512−2515.

(7) Mujica, V.; Kemp, M.; Ratner, M. A. *J. Chem. Phys.* **1994**, *101*, 6849−6855.

(8) Datta, S. *Electronic Transport in Mesoscopic Systems;* Cambridge University Press: 1995.

(9) Lang, N. D. *Phys. Rev. B* **1995**, *52*, 5335−5342.

(10) Xue, Y.; Datta, S.; Ratner, M. A. *J. Chem. Phys.* **2001**, *115*, 4292−4299.

(11) Taylor, J.; Guo, H.; Wang, J. *Phys. Rev. B* **2001**, *63*, 121104-(R)/1−4.

(12) Brandbyge, M.; Mozos, J.-L.; Ordejón, P.; Taylor, J.; Stokbro, K. *Phys. Rev. B* **2002**, *65*, 165401/1−17.

(13) Ernzerhof, M.; Zhuang, M. *J. Chem. Phys.* **2003**, *119*, 4134−4140.

(14) Zhuang, M.; Rocheleau, P.; Ernzerhof, M. *J. Chem. Phys.* **2005**, *122*, 154705/1−6.

(15) Walter, D.; Neuhauser, D.; Baer, R. *Chem. Phys.* **2004**, *299*, 139−145.

(16) Sautet, P.; Joachim, C. *Chem. Phys. Lett.* **1988**, *153*, 511−516.

(17) Hod, O.; Rabani, E.; Baer, R. *Acc. Chem. Res.* **2006**, *39*, 109−117.

(18) Yi, J.; Cuniberti, G. *Ann. N. Y. Acad. Sci.* **2003**, *1006*, 306−311.

(19) Yi, J.; Cuniberti, G.; Porto, M. *Eur. Phys. J. B* **2003**, *33*, 221−225.

(20) Landauer, R. *Philos. Mag.* **1970**, *21*, 863−867.

(21) Cuevas, J. C.; Levy Yeyati, A.; Martín-Rodero, A. *Phys. Rev. Lett.* **1998**, *80*, 1066−1069.

(22) Brandbyge, M.; Kobayashi, N.; Tsukada, M. *Phys. Rev. B* **1999**, *60*, 17064−17070.

(23) Ke, S. H.; Barranger, H. U.; Yang, W. *Phys. Rev. B* **2003**, *70*, 085410/1−12.

(24) Hall, L. E.; Reimers, J. R.; Hush, N. S.; Silverbrook, K. *J. Chem. Phys.* **2000**, *112*, 1510−1521.

(25) Goyer, F.; Ernzerhof, M. Manuscript in preparation.

(26) Jackson, J. D. *Classical Electrodynamics;* John Wiley & Sons: Ltd.: 1998.

(27) Nakanishi, S.; Tsukada, M. *Phys. Rev. Lett.* **2001**, *87*, 126801/1−4.

(28) Mayor, M.; Didschies, C. *Angew. Chem., Int. Ed.* **2003**, *42*, 3176−3179.

(29) Aharonov, Y.; Bohm, D. *Phys. Rev.* **1959**, *115*, 485−491.

(30) Diederich, F.; Staab, H. A. *Angew. Chem., Int. Ed.* **1978**, *17*, 372−374.

(31) Xu, B.; Xiao, X.; Yang, X.; Zang, L.; Tao, N. *J. Am. Chem. Soc.* **2004**, *127*, 2386-2387.

CT600087C

JCTC Journal of Chemical Theory and Computation

Reaction of H_2 with a Binuclear Zirconium Dinitrogen Complex − Evaluation of Theoretical Models and Hybrid Approaches

Brian F. Yates,*,[†] Harold Basch,[‡] Djamaladdin G. Musaev,[§] and Keiji Morokuma*,[§]

School of Chemistry, University of Tasmania, Private Bag 75, Hobart TAS 7001, Australia, Department of Chemistry, Bar Ilan University, Ramat Gan, Israel, and Cherry L Emerson Center for Scientific Computation and Department of Chemistry, Emory University, Atlanta, Georgia 30322

Received December 12, 2005

Abstract: Molecular orbital and hybrid ONIOM (both IMOMO and IMOMM) calculations have been carried out on the important reaction of H_2 with a binuclear zirconium dinitrogen complex to test the efficacy of several structural models of the ancillary ligand. The complete experimental ligand, $PhP(CH_2SiMe_2NSiMe_2CH_2)_2PPh$, in the zirconium complex has been treated at the IMOMM level, while two smaller approximations of the ligand, $HP(CH_2SiH_2NSiH_2CH_2)_2PH$ and $(PH_3)_2(NH_2)_2$, have received the full molecular orbital treatment. The mechanism of dihydrogen addition has been compared with our earlier study (Basch, Musaev, and Morokuma *J. Am. Chem. Soc.* **1999**, *121*, 5754−5761). We find that the substituent effects do cause some small changes in both the structures of the complexes studied and the activation energies of the transition structures. However for the most part the potential energy profiles are very similar to our earlier study and lend support to our use of simple theoretical models to represent moderately large experimental structures.

Introduction

The study of the activation of the N≡N triple bond in the dinitrogen molecule and its subsequent chemical reactions has remained a productive area of research over quite a number of years.[1−66] On one hand there is interest in understanding the biological mechanism of nitrogen fixation, while on the other the practical goal is to develop a catalytic cycle to enable nitrogen reduction to be carried out under mild conditions. Tuczek and Lehnert have stated that the most promising catalytic cycles are those "based on the mono- and binuclear transition metal N_2 compounds that give NH_3 or N_2H_4 upon protonation".[23] The first report of the addition of H_2 to a metal-coordinated N_2 was made by Fryzuk and co-workers.[15] In these experiments it was shown that the dihydrogen molecule added to the binuclear metal dinitrogen complex $[P_2N_2]Zr(\mu\text{-}\eta^2\text{-}N_2)\text{-}Zr[P_2N_2]$, where $P_2N_2 = PhP(CH_2SiMe_2NSiMe_2CH_2)_2PPh$. This complex contains a coordinated N_2 molecule in a side-on bridging arrangement. It reacts with H_2 to produce a new complex having bridging Zr···H···Zr and N−H bonds where the original H−H bond is broken while the N−N bond is conserved.

Using theoretical techniques and a model complex **1**, $[p_2n_2]Zr(\mu\text{-}\eta^2\text{-}N_2)Zr[p_2n_2]$, where $p_2n_2 = (PH_3)_2(NH_2)_2$, we showed previously[24,25] that the mechanism for this reaction proceeds in two steps (see Figure 1): (i) the activation of the H−H bond via transition structure **2** to form the intermediate diazenidohydride complex $[p_2n_2]Zr(\mu\text{-}\eta^2\text{-}N_2H)Zr[p_2n_2](\mu\text{-}H)$, **3**, and (ii) migration of the Zr-bonded hydride ligand to a position bridging the two Zr atoms to form the diazenido-μ-hydride complex $[p_2n_2]Zr(\mu\text{-}\eta^2\text{-}NNH)Zr[p_2n_2](\mu\text{-}H)$, **7**. The entire reaction was calculated to be exothermic by 13−15 kcal mol^{-1} with an activation energy of 21 kcal mol^{-1}. Although not observed experimentally, our theoretical studies also indicated that there are two further low-energy structures on the potential energy surface, $[p_2n_2]Zr(\mu\text{-}NNH_2)Zr[p_2n_2]$,

---

[*] Corresponding author fax: +61 3 6226−2858; e-mail: Brian.Yates@utas.edu.au.

[†] University of Tasmania.

[‡] Bar Ilan University.

[§] Emory University.

10.1021/ct050320e CCC: $33.50

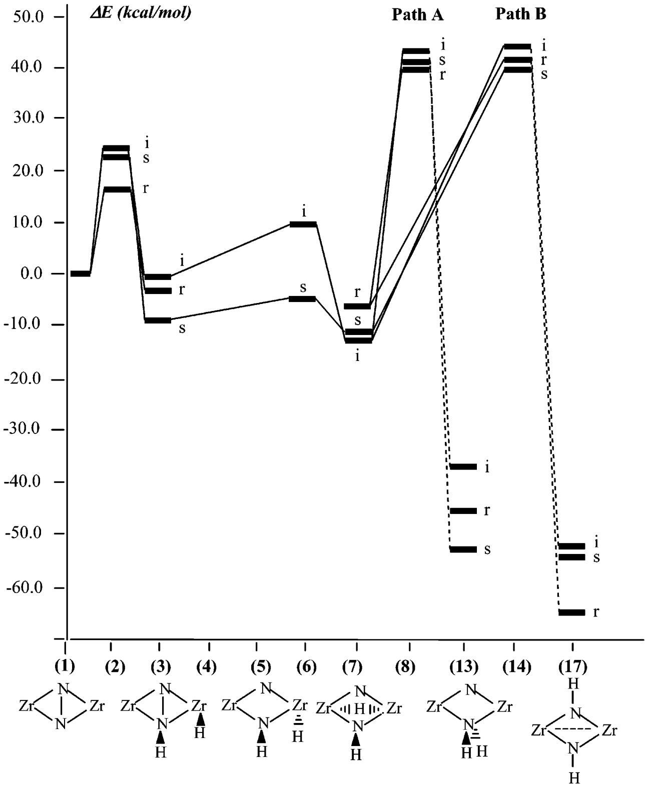
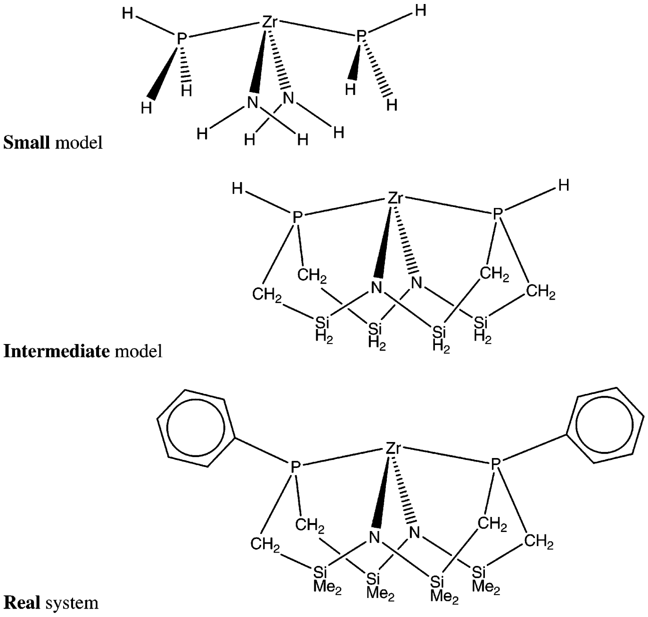


Figure 1. Potential energy surfaces calculated with the "c" (Zr−P constrained small model, s), "g" (intermediate model, i), and "j" (real system, r) levels of theory (see Tables 1 and 6). The dashed lines indicate that there are several intermediate minima and transition structures omitted between **8** and **13** and between **14** and **17**.

13, with a bridging NH_2, and $[p_2n_2]Zr(\mu\text{-}NHNH)Zr[p_2n_2]$, **17**, with two bridging NH units. These are about 40 kcal mol^{-1} more stable than **7** but are separated by barriers of nearly 60 kcal mol^{-1}.

The representation of the experimental macrocyclic $[P_2N_2]$ ligand by $(PH_3)_2(NH_2)_2$ is clearly a drastic approximation and one that we have been concerned to explore more fully. In our earlier work we noted that we could not use a more realistic representation of the ligand "because this would put the size of the calculations beyond current capabilities".[25] Although we have come a long way since the early calculations on bare Zr_2N_2 systems,[67] the straightforward molecular orbital calculation of the full experimental system still remains a difficult challenge for present-day computer facilities. Nevertheless, given the importance of the nitrogen fixation reaction and the current high level of interest in the activation of N_2 and other multiply bonded systems by organometallic complexes, it is important to assess the steric and electronic effects arising from the experimental $[P_2N_2]$ ligand. One of the difficulties we found in our earlier work with the model ligand was that one of the PH_3 ligands tended to escape from the first coordination shell to reduce the strain from overcrowding at the metal center. This produced unrealistically long Zr−P distances. These distances would not be possible in the chelated ligand, and to counteract this we fixed the Zr−P distance at 2.80 Å. We noted at the time that this was an oversimplification and that "it is highly desirable to reexamine the present results by using the real ligand".[25] Fryzuk and co-workers[68] have also carried out calculations using $(PH_3)_2(NH_2)_2$ as a model for the experimental $[P_2N_2]$ ligand in some tantalum complexes. They constrained several angles and a dihedral angle in an attempt to simulate the twist in the $[P_2N_2]$ ligand. However we must remember that the experimental ligand not only includes a structurally rigid aspect but also the steric and electronic factors arising from the phenyl groups on the phosphines.


Small model

Intermediate model

Real system

Figure 2. Definition of the different structural models used in this work.

One approach to using a more realistic ligand in the theoretical calculations is to use a larger model structure in the molecular orbital calculations. One might put substituents on the phosphorus or nitrogen to form, for example, $(PMe_3)_2(NMe_2)_2$, or one might include the basic skeleton of the macrocyclic ring, $P(C{-}Si{-}N{-}Si{-}C)_2P$, with just hydrogen substituents. Another approach is to use a hybrid theoretical technique such as the ONIOM method[69,70] which allows the complete experimental system to be included at least at a low level of theory (such as semiempirical MO or molecular mechanics), while a high level of molecular orbital theory is retained for the most important part, the reaction center.

In this paper we report our theoretical studies using both approaches. We have used a larger model for the experimental ligand in the MO calculations, and we have combined this with both IMOMO and IMOMM hybrid theoretical treatments. We have re-evaluated the reaction mechanism for complexes **1**−**7** corresponding to the experimentally observed process and in addition have recalculated the key structures (**8**, **13**, **14**, **17**) on the remainder of the potential energy surface.

Theoretical Methods

As stated in the Introduction, we have used a number of different structural models to represent the macrocyclic P_2N_2 ligand (see Figure 2). The model that we used in our original work,[25] ($(PH_3)_2(NH_2)_2$), we will refer to here as the "small model". The "intermediate model" incorporates the basic macrocyclic skeleton ($PH(CH_2SiH_2NSiH_2CH_2)_2PH$) without

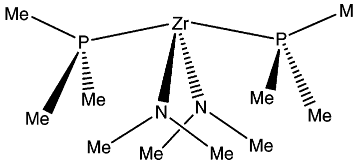


Figure 3. Alternative intermediate model.

the steric bulk of the extra methyl and phenyl groups. The third model structure used in this study corresponds to the real experimental system. These three structural models also correspond to the different layers that we have used in the hybrid theoretical methods discussed below.

In our molecular orbital calculations we have used several different basis sets. We began with the minimal LANL2MB basis set which incorporates the Hay and Wadt[71] small-core relativistic effective core potential and minimal valence basis set on all atoms. This resulted in 94 basis functions (bf) for the small model systems and 182 bf for the intermediate model systems. We also used the “SBK” basis set which incorporates the SKBJ small-core relativistic effective core potential[72,73] together with the standard double-ζ basis set on Zr, CEP-31G on N, P, and Cl, and CEP-4G on H, C, and Si. That is, a double-ζ basis set is used on the atoms directly bound to Zr and a minimal basis set on the other atoms. In addition, diffuse functions (with Gaussian exponent of 0.04395) were included on Cl in the monomer and d-type polarization functions (with Gaussian exponent of 0.8) on the bridging N_2 in structures **1**−**17**. In the separate calculations on the small model systems (i.e. the nonhybrid method), the CEP-31G basis set was used on hydrogen instead of CEP-4G. The SBK basis set combination as defined here resulted in 174 bf for the small model systems and 254 bf for the intermediate model systems. The small model systems were reoptimized with a slightly larger basis set (SBKBS1) in which polarization functions were added to all atoms connected to zirconium (CEP-31G* was used for N, P, and

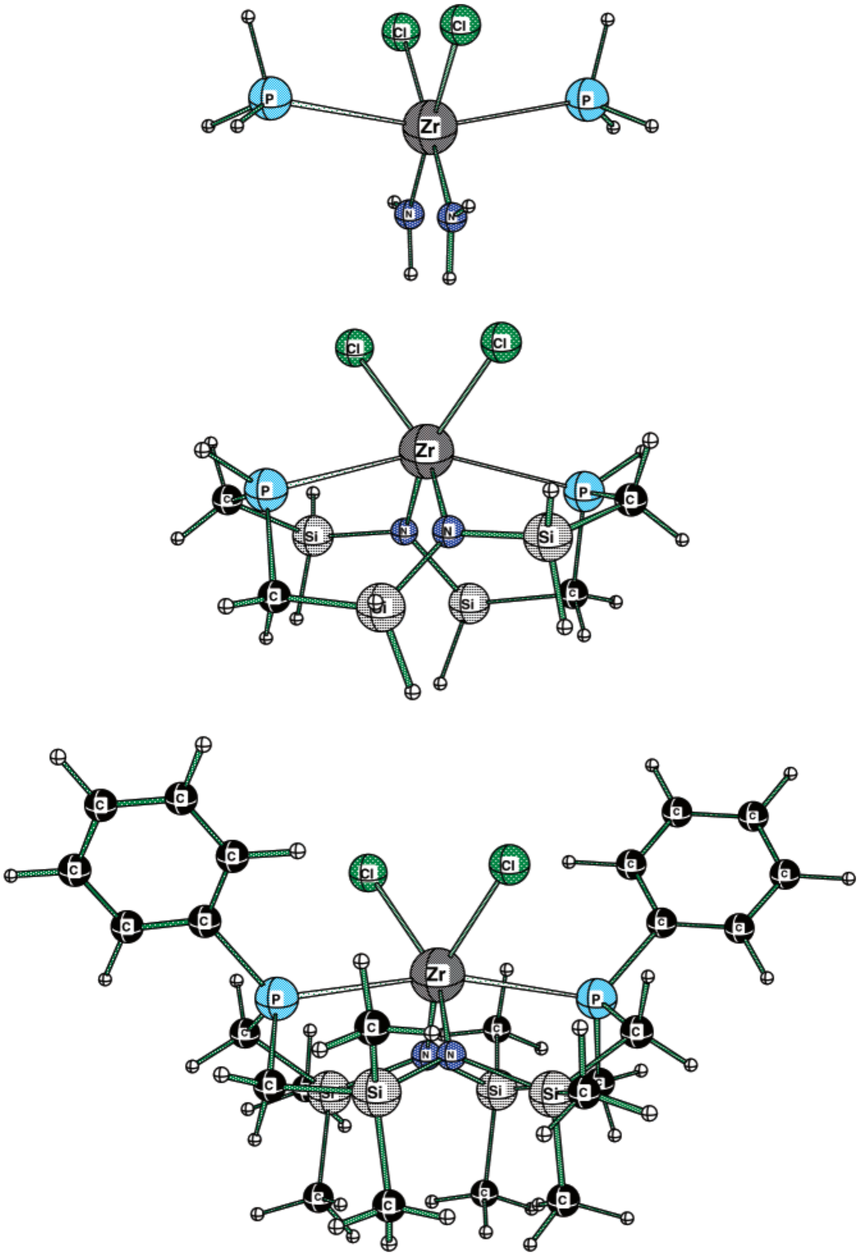


Figure 4. Monomer (small, intermediate, and real systems).

Table 1. Theoretical Methods Used in This Work[a]

structural model	theoretical method	
small model	a	B3LYP/SBK
	b	B3LYP/SBKBS1
	c	B3LYP/SBKBS2//“b”
intermediate model	d	RHF/LANL2MB
	e	IMOMO(B3LYP/SBK small:RHF/ LANL2MB intermediate)
	f	B3LYP/SBK
	g	B3LYP/SBKBS2//“f”
real system	h	IMOMM(RHF/LANL2MB intermediate:UFF real)
	i	IMOMM(B3LYP/SBK intermediate:UFF real)
	j	B3LYP/SBKBS2//“i”

[a] All structures were optimized at each level of theory, except for “c”, “g”, and “j” which represent single point calculations on geometries from “b”, “f”, and “i”, respectively.

Cl). Finally single point calculations were carried out on all systems (small, intermediate, and real) using a more flexible basis set (SBKBS2) which was valence-triple-ζ on zirconium, valence-triple-ζ plus polarization and diffuse functions on all atoms connected to zirconium, and valence-double-ζ plus polarization functions on all remaining atoms. This last basis set resulted in 333 bf for the small model, 572 bf for the intermediate model, and 1188 bf for the real system. Full details of the basis sets used are given in the Supporting Information.

RHF and density functional (B3LYP[74−76]) calculations were carried out at the molecular orbital level of theory. The initial SCF convergence in nearly all of the systems was problematic and required the judicious use of saved molecular orbital coefficients and level shifting. In some cases we employed stability tests[77,78] followed by reoptimization of the wave function (the “stable = (opt,rrhf)” keyword was used in the Gaussian program) to ensure that the correct wave function was used. Full geometry optimizations were carried out together with vibrational frequency calculations. Transition structures were characterized by the determination of only one imaginary frequency and by slightly distorting along the imaginary frequency normal mode in both directions and performing geometry minimizations using the transition structure force constants as the initial Hessian.

A number of hybrid theoretical methods was used. In this approach the most important part (reactive region) of the system is treated with a high level of theory, while the less important parts are treated with a low level of theory. The ONIOM method[69,70] was used to carry out IMOMO and IMOMM calculations on the intermediate and real model systems. In the IMOMO calculations we used B3LYP/SBK for the core region (small model, layer 1) and RHF/ LANL2MB for the outer region (intermediate model, layer 2). In the IMOMM calculations[79−82] we defined the intermediate model as the core region (layer 1) and the remainder of the real system as layer 2. We used either RHF/ LANL2MB or B3LYP/SBK to treat layer 1 and the UFF[83] or MM3[84,85] force fields for layer 2. In the MM3 calculations, the UFF van der Waals parameters by Rappé et al. were used for the Zr atoms,[83] while all other MM contributions

Table 2. Comparison of Selected Calculated Values with Experiment for Monomer

	model							
	small		intermediate			real		
	a[a]	b[a]	d[a]	e[a]	f[a]	h[a]	i[a]	X-ray[b]
Zr−N1	2.063	2.077	2.038	2.211	2.122	2.096	2.179	2.136(4)
Zr−N2	2.063	2.077	2.038	2.211	2.122	2.096	2.179	2.125(4)
Zr−P1	2.847	2.842	2.875	2.738	2.823	2.768	2.704	2.694(2)
Zr−P2	2.847	2.842	2.875	2.738	2.823	2.768	2.704	2.707(2)
Zr−Cl1	2.571	2.532	2.573	2.519	2.538	2.578	2.539	2.455(2)
Zr−Cl2	2.571	2.532	2.573	2.519	2.538	2.577	2.539	2.448(2)
N−Zr−N	86.6	85.5	95.1	104.0	92.8	104.2	105.8	96.8(2)
P−Zr−P	160.5	161.3	152.8	144.2	156.5	154.0	152.0	152.25(6)
Cl−Zr−Cl	98.1	98.8	91.1	87.2	91.1	89.0	89.0	82.57(7)
N1−Zr−Cl2	87.7	87.9	90.6	89.7	93.9	89.2	88.0	89.5(1)
N2−Zr−Cl1	87.7	87.9	90.6	89.7	93.9	89.2	88.0	91.3(1)

[a] Level of theory, see Table 1. [b] Fryzuk et al. *Organometallics* **1998**, *17*, 846−853. Atom numbering scheme in Figure 1 of that paper is used in the table above.

Table 3. Comparison of Selected Calculated Values with Experiment for Structure 1

	model							
	small		intermediate			real		
	a[a]	b[a]	d[a]	e[a]	f[a]	h[a]	i[a]	X-ray[b]
Zr1−N1	2.106	2.100	2.005	2.055	2.057	2.018	2.070	2.010(2)
Zr1−N3	2.116	2.121	2.118	2.251	2.171	2.166	2.209	2.203(4)
Zr1−P1	2.943	2.907	2.996	2.774	2.821	2.797	2.784	2.724(2)
N1−N2	1.530	1.518	1.507	1.539	1.508	1.498	1.506	1.43(1)
N3−Zr1−N4	118.9	117.1	116.4	119.0	114.3	112.4	110.4	108.4(2)
P1−Zr1−P2	159.6	161.0	137.6	139.2	142.0	134.6	136.2	140.90(7)
N1−Zr1−N2	43.2	42.7	44.0	44.0	42.9	43.3	42.4	41.8(3)
Zr1−N1−N2	70.3	67.3	68.3	67.9	68.9	69.0	69.7	69.1(2)
N1−N2−Zr1−Zr2	−46.8	−45.7	−17.1	−36.8	−44.2	−13.3	−15.0	

[a] Level of theory, see Table 1. [b] Fryzuk et al. *Science* **1997**, *275*, 1445−1447.

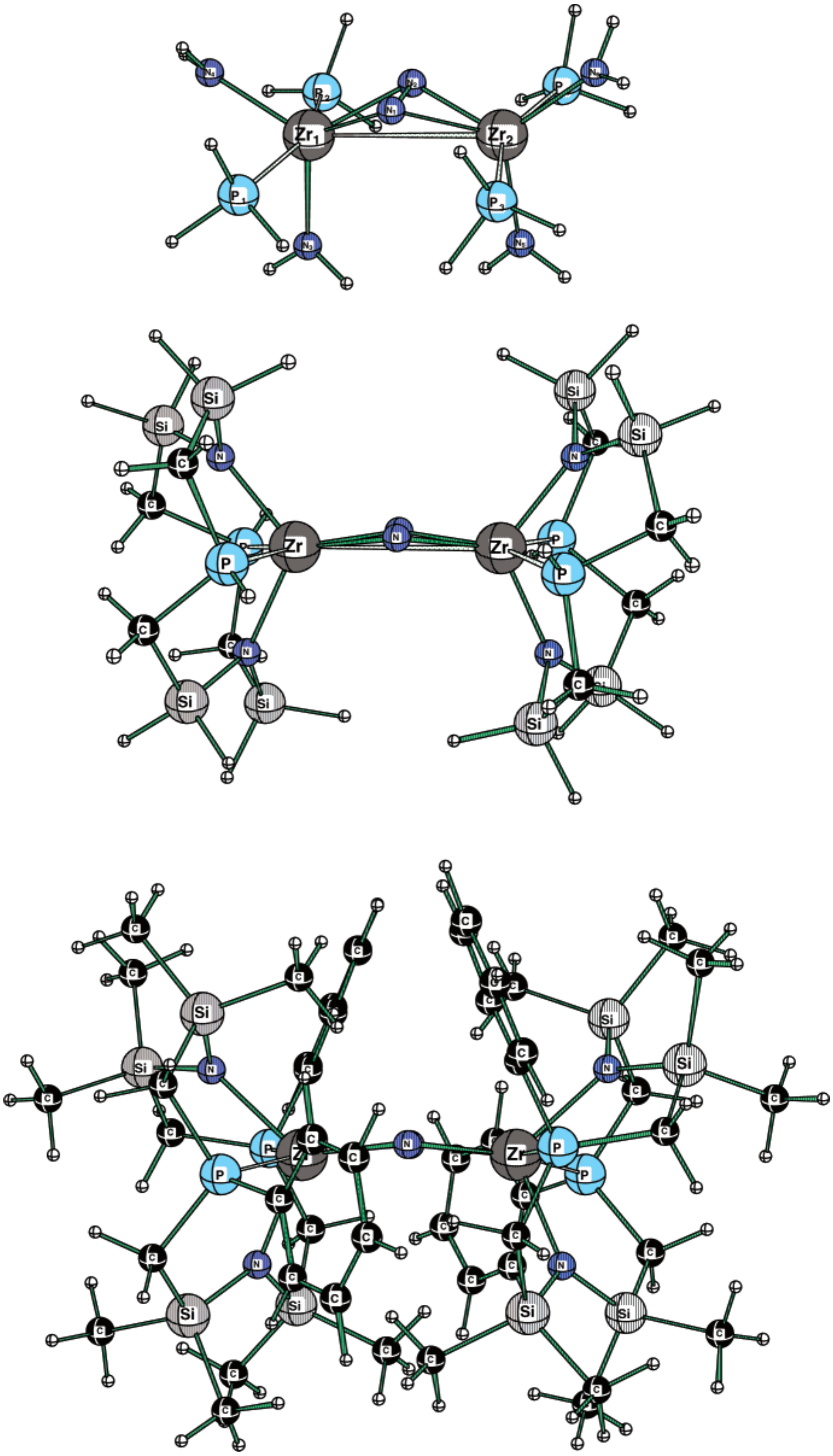


Figure 5. Structure **1** (reactant) (small, intermediate, and real systems).

involving the metal atoms were set to zero. The IMOMM link bonds were fixed at $R(\text{P}-\text{H}) = 1.435$ Å and $R(\text{Si}-\text{H}) = 1.490$ Å.

In our initial IMOMM calculations we used the small model (Figure 2) for the MO calculations and the real system for the MM calculations. However this lead to the Si−C distance in the calculated real system becoming too large. The Si−C bond corresponds to the nonbonded H···H distance in the small model, and it turns out that this leads to the calculation of an unrealistic repulsion in the small model. Our initial attempts to use the alternative intermediate model shown in Figure 3 for the MO level also lead to

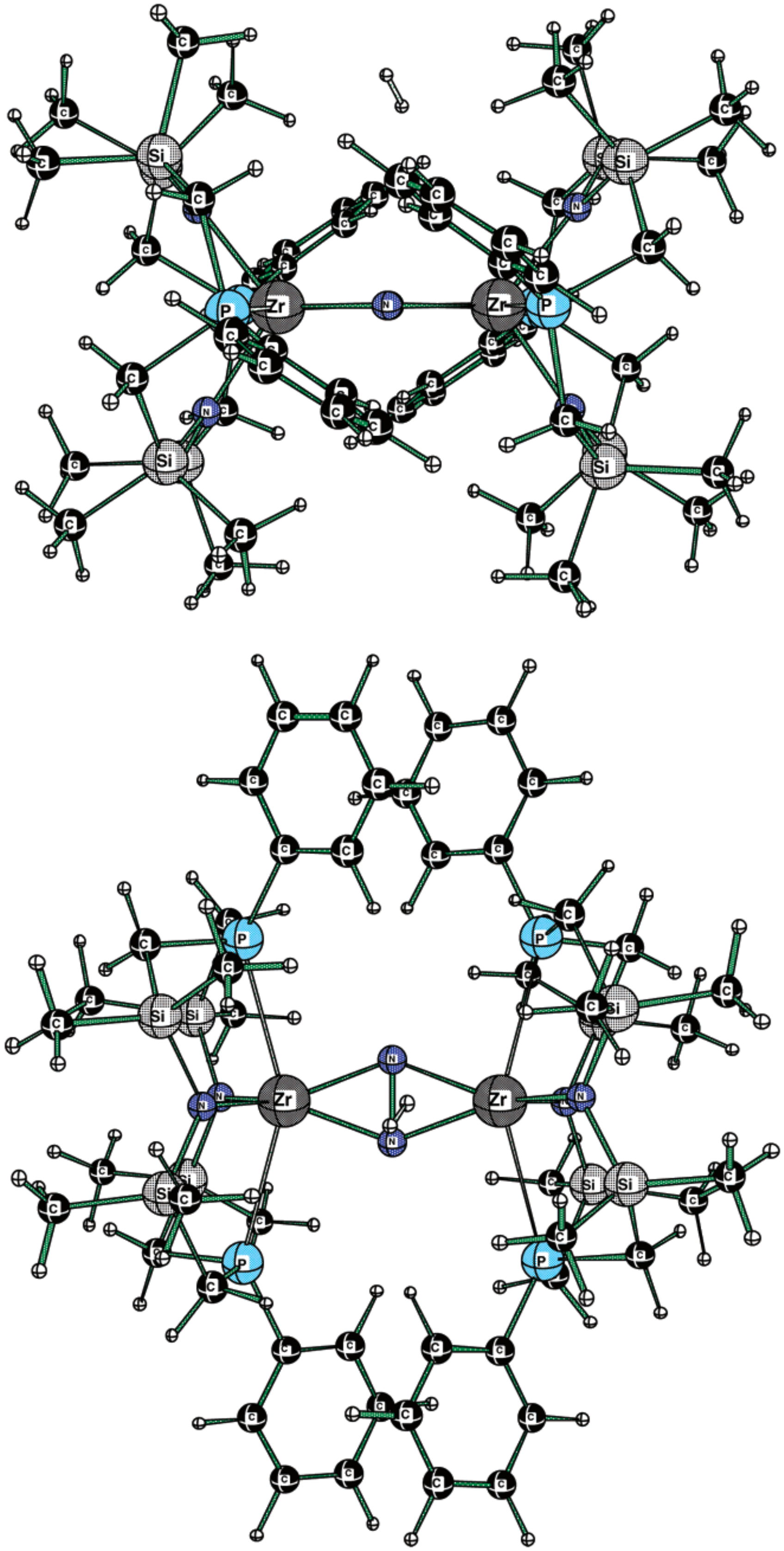


Figure 6. H_2 entering the coordination sphere: view from side-on (top) and from above (bottom).

unrealistic repulsions being calculated. Thus it is essential that the N−Si−C−P linkage be included in the MO portion of the calculation. We have studied this problem in detail previously for the *n*-butane system.[81]

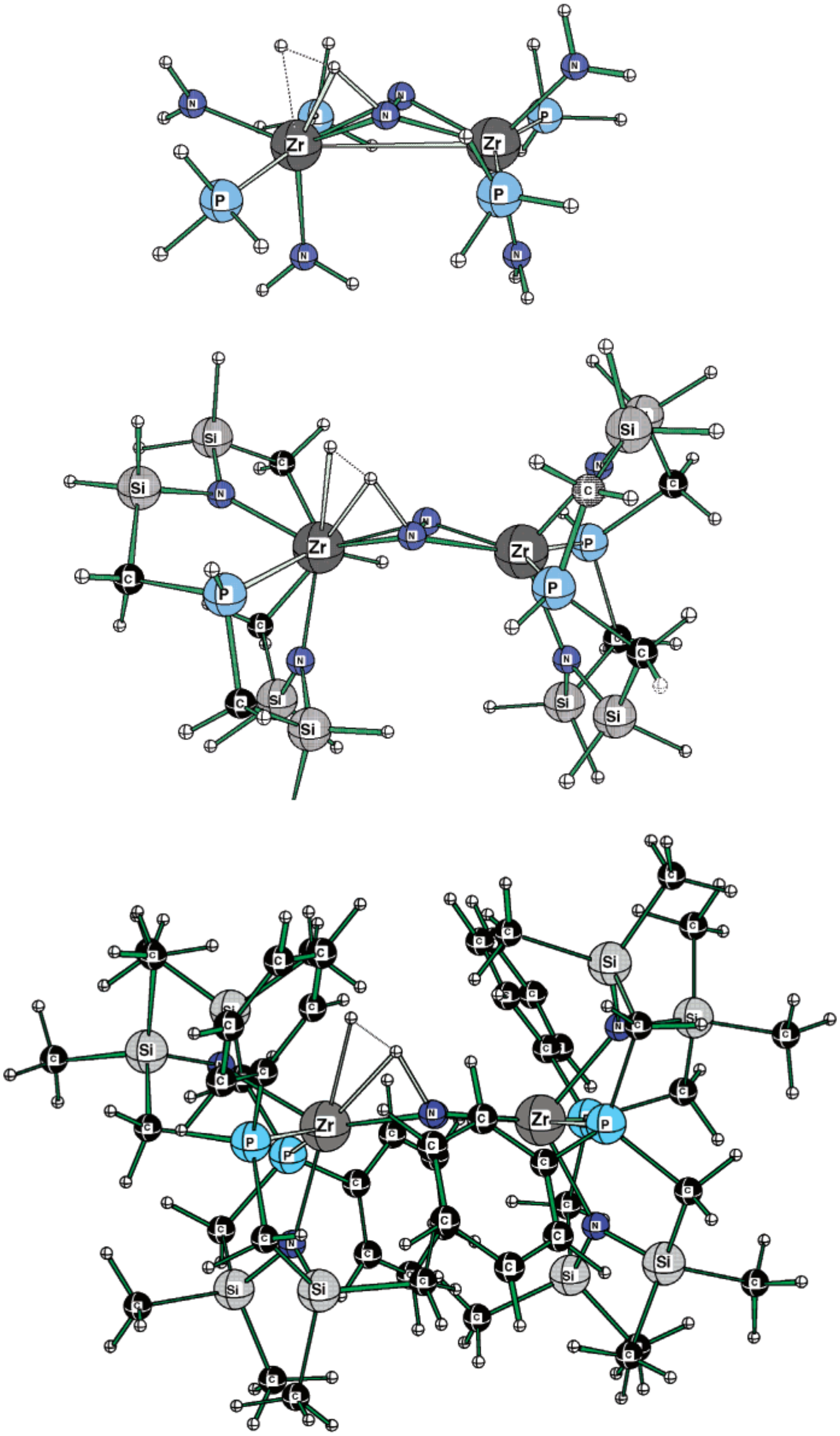


Figure 7. Structure **2** (TS) (small, intermediate, and real systems).

The theoretical methods used in this work are summarized in Table 1. This allows a number of comparisons to be made which assist in evaluating the theoretical approaches we have used and in separating out the different effects. In particular the effect of the level of theory can be assessed by comparing a with b (Table 1), d with f, e with f, and h with i. The effect of substituents can be assessed by comparing a with e, a with f, d with h, and f with i.

All calculations were carried out with the Gaussian 98[86] or Gaussian 03[87] programs, except for some IMOMM calculations which were performed with the IMOMM code[79] and the MM3(92)[84] and Gaussian92/DFT[88] programs.

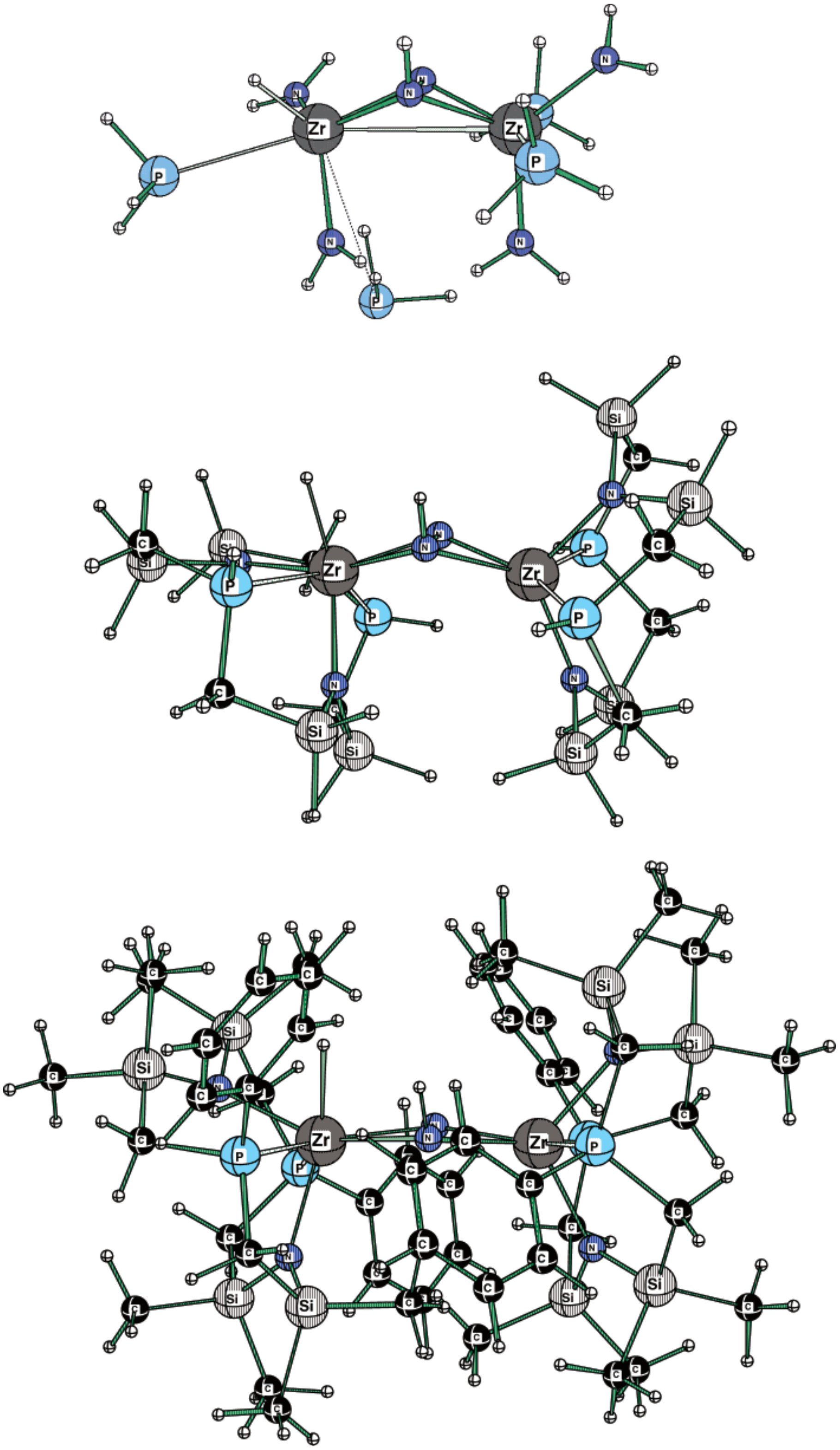


Figure 8. Structure **3** (small, intermediate, and real systems).

Results and Discussion

Geometries. Results are presented here for the monomer (one-half of the binuclear zirconium complex), the first part of the potential energy surface, **1**−**7**, and the key structures (**8**, **13**, **14**, **17**) on the remainder of the potential energy surface. In our previous work[25] we showed that **8** was connected to **13** (and **14** was connected to **17**) by a series of low-energy stationary points on the potential surface. These stationary points have not been considered in this work, but we have retained the same numbering system to allow comparison with our earlier study.

The structure of the monomer is shown in Figure 4 for the three model structures used in this work (small − intermediate − real, going from top to bottom). Table 2

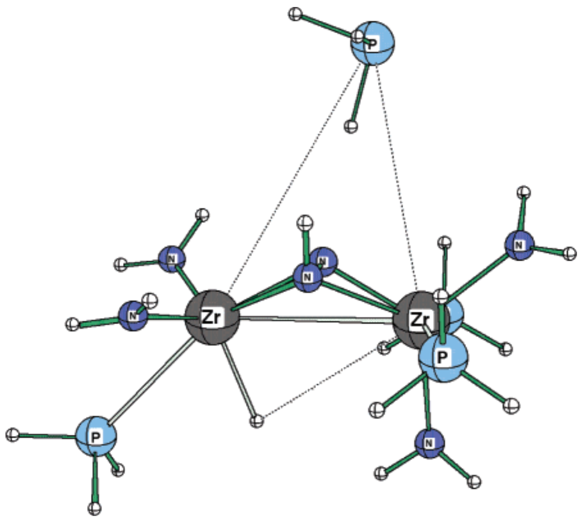


Figure 9. Structure **5** (small model).

includes values for the experimental geometry of the monomer as determined by X-ray analysis.[22] This shows that the N−Zr−N angles are calculated to be too small and the Cl−Zr−Cl angles too large in the small model. There is an improvement in these geometrical parameters on moving to the intermediate model. At the B3LYP/SBK level for the intermediate model (column f in Table 2) the calculated geometry agrees very well overall with the experimental geometry. On moving to the real system (column i) the Zr−P distances move even closer toward the experimental ones, accompanied by a slight opening of the N−Zr−N angle.

The structure of complex **1** is shown in Figure 5 for the three models used. It can be seen quite clearly that the Zr_2N_2 ring approaches planarity as one moves from the small model to the real system. This is shown more quantitatively in Table 3 where the dihedral angle for these four atoms is shown in the last row of the table. This is a significant result because we[25] and others have shown that the Zr_2N_2 core is strongly bent in model complexes, and a careful analysis of the electronic factors which give rise to the butterfly structure has recently been carried out.[50] The results presented here demonstrate that steric factors can overcome this electronic preference for bending and lead to a planar Zr_2N_2 core as the preferred geometry. What are these steric factors? When we noted in our earlier study[25] that the small model of **1** has a bent structure, we suggested that the extended macrocyclic ligand may prevent this bending. However it is clear from Table 3 that the basic skeleton of the macrocyclic ligand which is present in the intermediate model does not by itself confer planarity. In particular, a comparison of columns a and f shows that at the same level of theory the small and intermediate models are bent by nearly the same amount. Clearly the presence of the phenyl groups is required to achieve planarity.

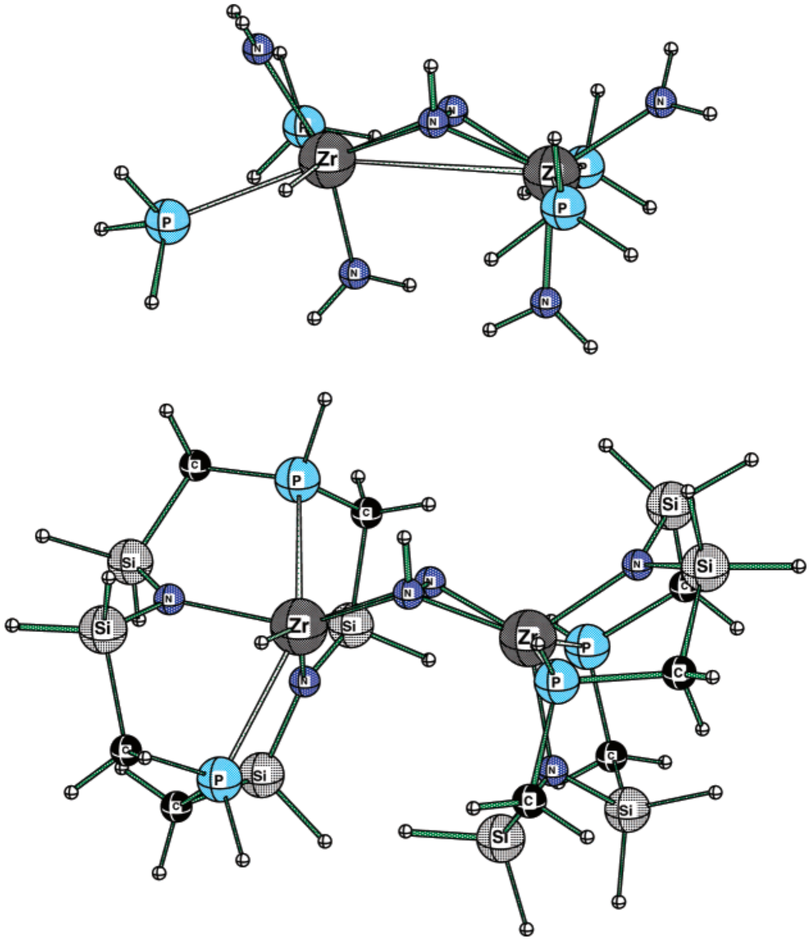


Figure 10. Structure **6** (small (constrained Zr−P) and intermediate models).

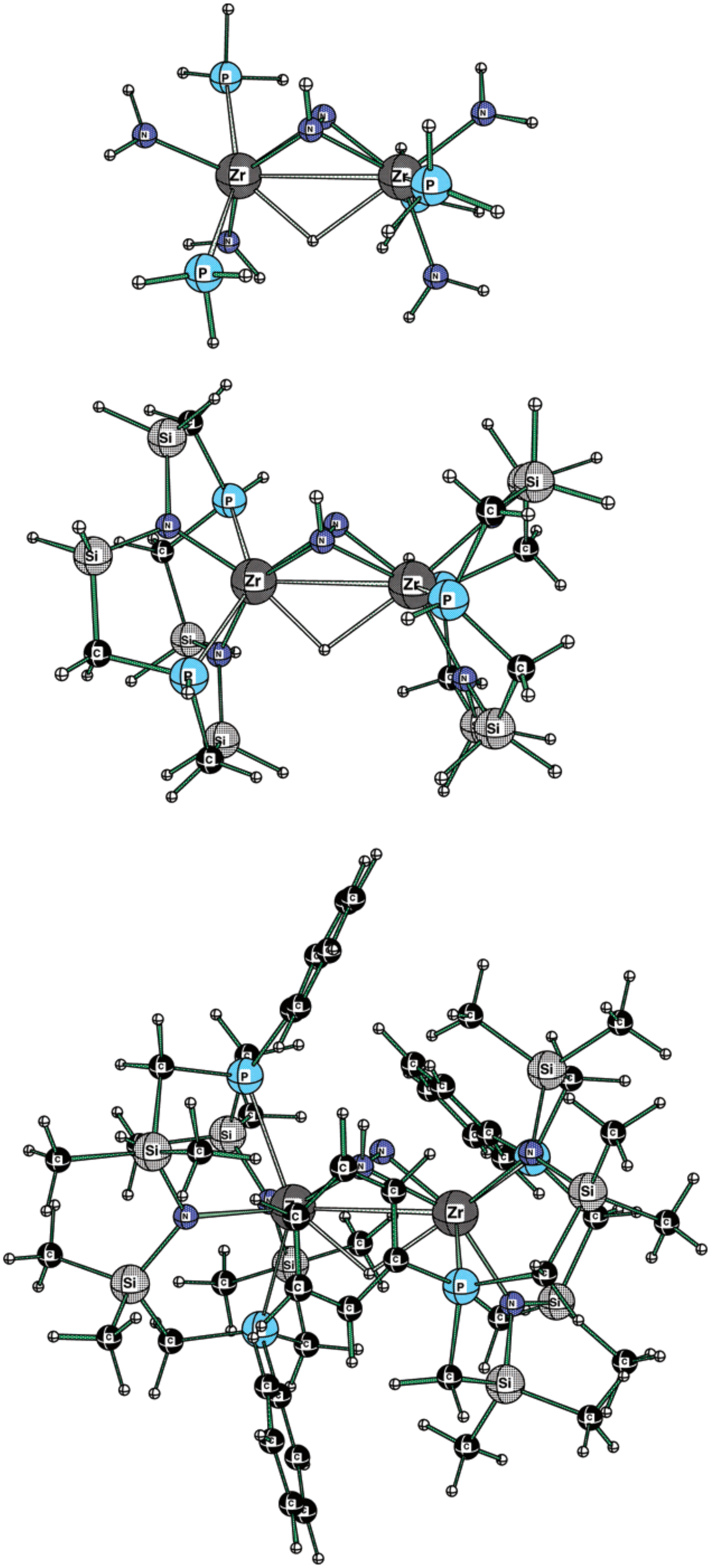


Figure 11. Structure **7** (small, intermediate, and real systems).

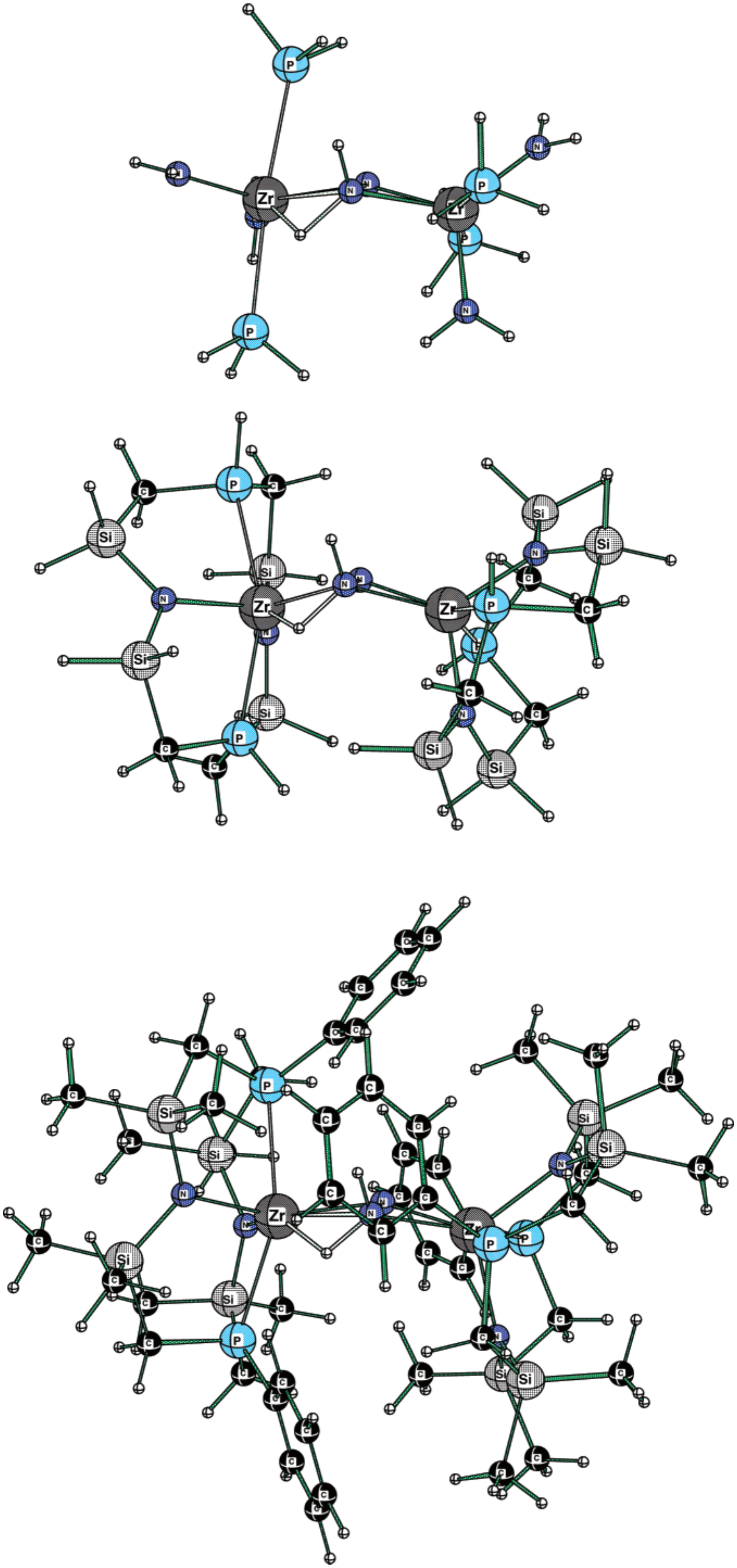


Figure 12. Structure **8** (small, intermediate, and real systems).

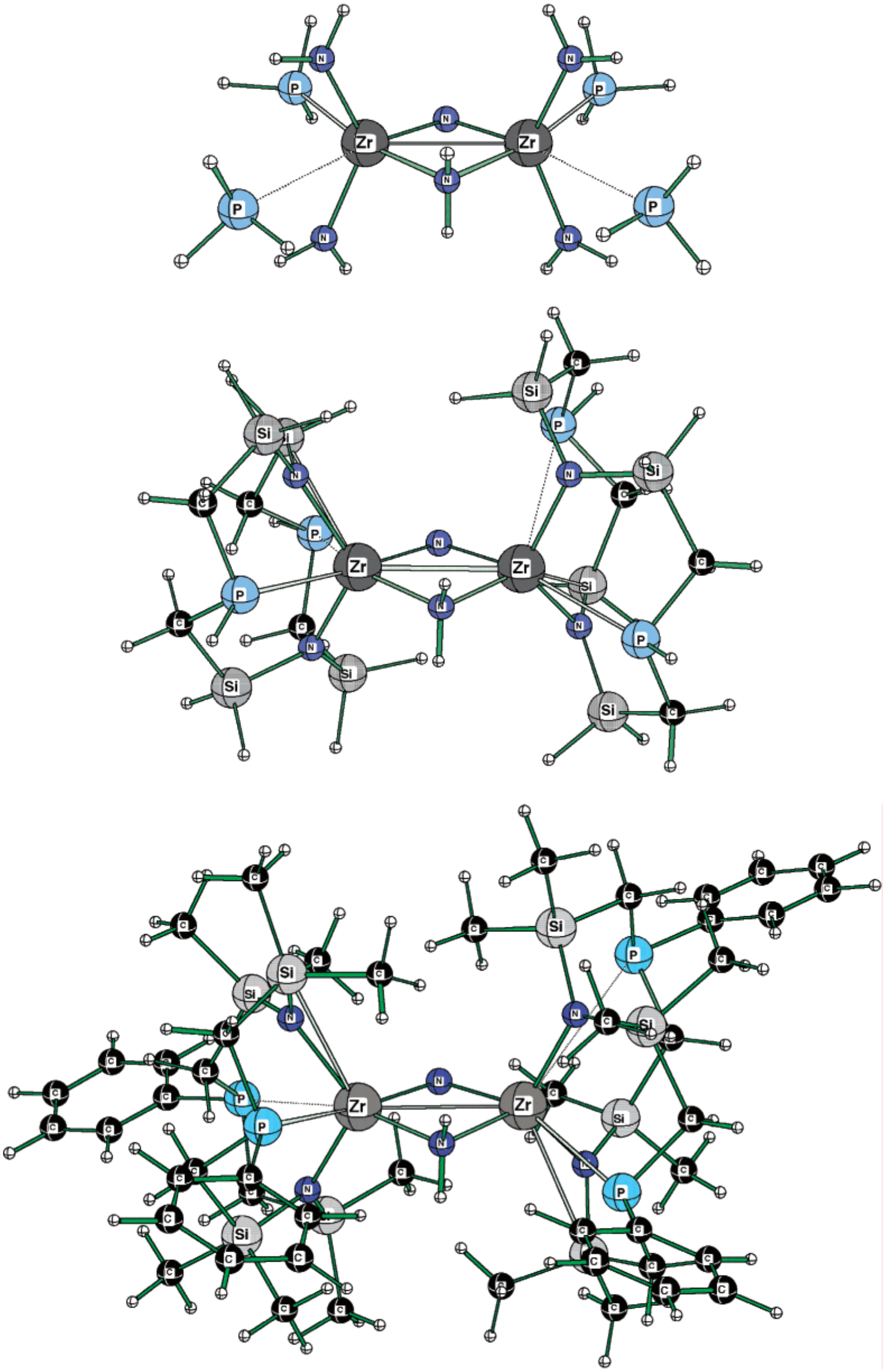


Figure 13. Structure **13** (small, intermediate, and real systems).

The experimental X-ray structure for **1** is also shown in Table 3.[15] This indicates that the Zr−P distances are once again overestimated in the small model and also at the RHF/LANL2MB level for the intermediate model (column d), but they are improved with the combination of the larger models and the B3LYP/SBK level of theory. Comparing columns a, f, and i it can be seen that the overestimation in the Zr−P distance decreases from 0.22 to 0.10 to 0.06 Å, respectively. The P−Zr−P angle is also overestimated with the small model but improves with the larger systems.

In Figure 6 we show a picture of the dihydrogen molecule entering the coordination sphere of the real system (this picture is obtained by minimization from transition structure **2** downhill toward the reactants). From the side-on perspective it appears that the H_2 would be prevented from reaching the nitrogens at the center of the complex by the shield of

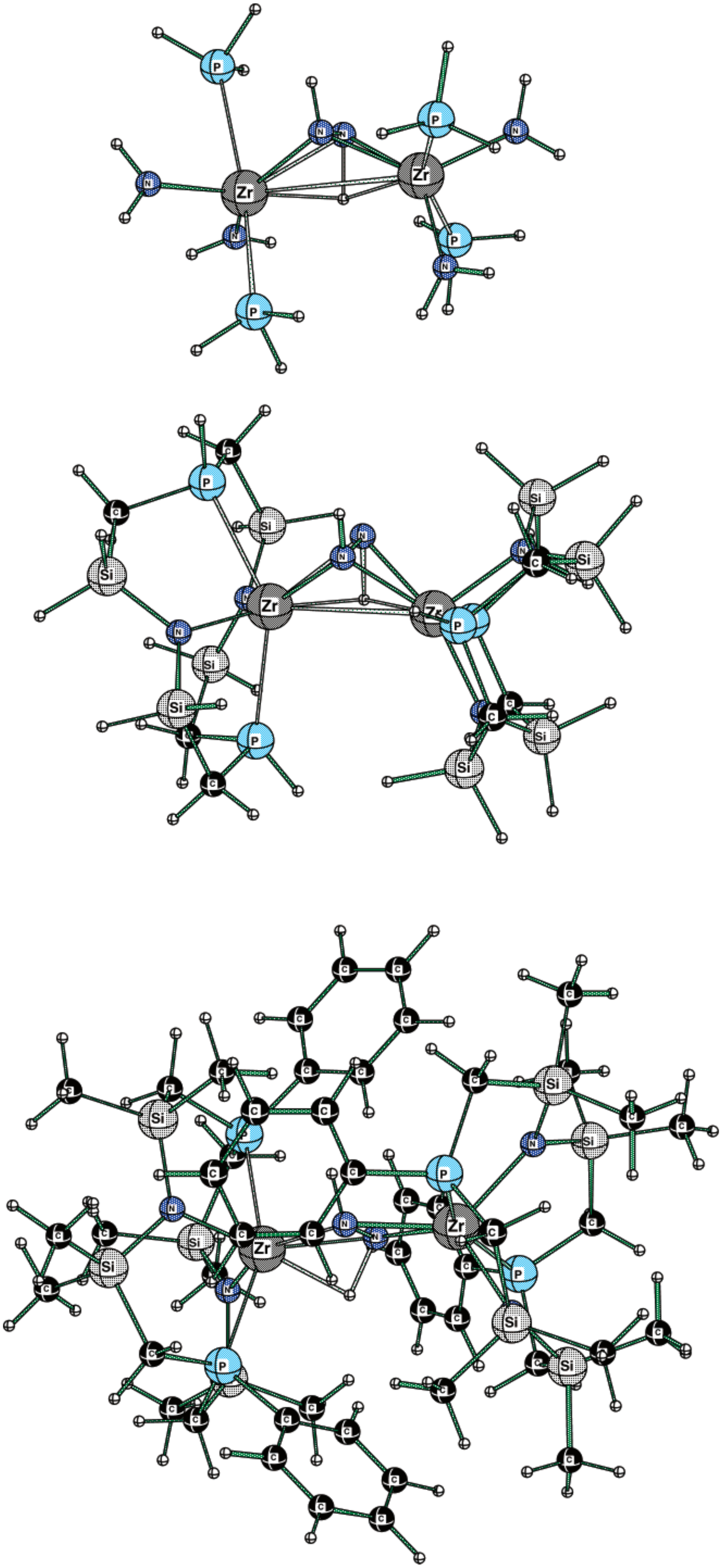


Figure 14. Structure **14** (small, intermediate, and real systems).

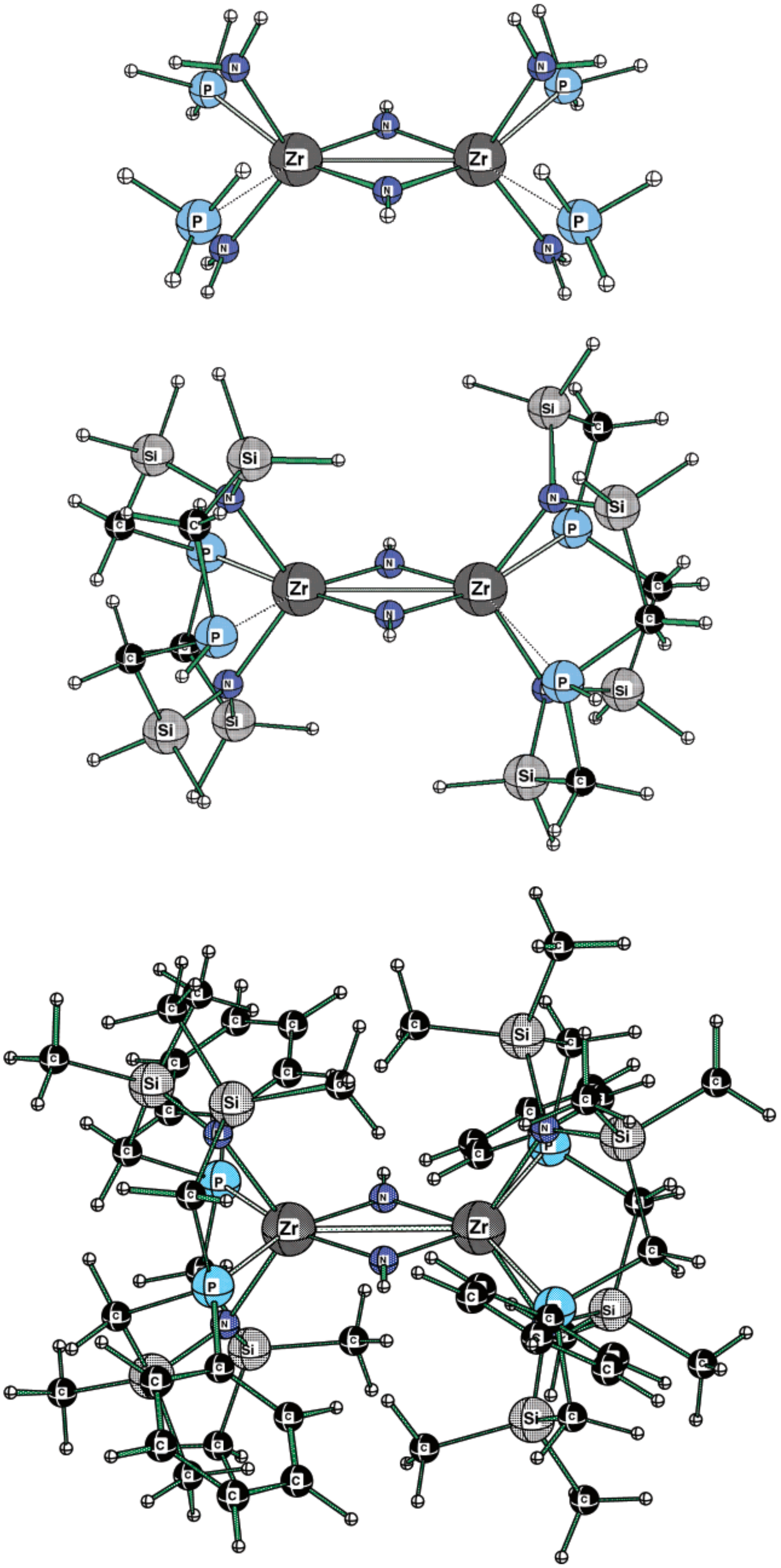


Figure 15. Structure **17** (small, intermediate, and real systems).

phenyl rings on the phosphines. But the view from above shows that phenyl rings are well clear of the reaction center and that the H_2 has a clear approach.

Figure 7 shows the actual transition structure for the coordination of the H_2. The key $Zr_2N_2H_2$ part of the geometry looks very similar for the three models used. The Zr−H distances decrease on going from the small model to the real system by 0.1−0.2 Å (see Supporting Information) which might indicate slightly greater binding in the real system; however, the H−H, N−N, and N−H distances remain almost the same.

The first product from this reaction, **3**, is shown in Figure 8. Once again we note the unrealistic loss of a phosphine in the small model. As expected, the phosphine is not dissociated in the intermediate model or the real system, and there is no need to fix the Zr−P distance in order to obtain a reasonable geometry. There appears to be a general tendency toward planarity as one moves to the real system, but the N−N−Zr−Zr portion is by no means completely flat. At the "a", "f", and "i" levels of theory the calculated N−N−Zr−Zr dihedral angle is −57.8, −49.2, and −13.5°, respectively.

Figure 9 shows the structure of complex **5** with the small model. We noted previously[25] that this is an artifact of the unconstrained ligand system, and indeed no corresponding structure could be found in this work when the intermediate model or real system was employed.

After much computational effort in searching the potential energy surfaces, we were finally able to locate a transition structure **6** which connects structures **3** and **7** at the small (constrained Zr−P) and intermediate models. The geometry of **6** is shown in Figure 10. We were unable to locate **6** for the unconstrained Zr−P small model, which is consistent with the comments made in the previous paragraph. Clearly the potential energy surface for **3** → **7** is different for the constrained and unconstrained small model. At the "f" level of theory transition structure **6** was shown to have a single very small imaginary frequency of just $27i$ cm^{-1}. Nevertheless, by carefully distorting the structure along the normal mode corresponding to this imaginary frequency we were able to verify that it connected up with structure **3** in one direction and structure **7** in the other. Transition structures such as this with very small imaginary frequencies are quite difficult to locate, and although we explored several different possibilities for the real system, we were unable to find a suitable structure at that level within a reasonable expenditure of computer time.

The experimentally observed product, **7**, is shown in Figure 11. An extensive comparison of the calculated geometries with the experimental neutron diffraction data[24] is included in the Supporting Information. The average absolute deviations from experiment for the "a", "f", and "i" levels of theory are 0.13, 0.07, and 0.07 Å, respectively, for bond distances and 12, 8, and 4° for bond angles. There is improvement in the Zr−P distances and in most of the angles on going from the small model to either the intermediate model or the real system. Structure **7** is quite bent around the Zr_2N_2 portion at all levels of theory (in contrast to **1** and **3**) which is expected due to the bridging hydrogen. The N−N−Zr−Zr dihedral angle is 66−70° at all levels of theory. The bridging hydrogen is fairly equidistant from both zirconiums at all levels of theory, although not to the same extent as indicated by the neutron diffraction data (see Supporting Information).

Transition structure **8** is shown in Figure 12. It adopts an almost planar Zr_2N_2 geometry in the intermediate and real systems, and the Zr−P distances are all quite reasonable (<2.8 Å). Minimizations along the normal mode corresponding to the imaginary frequency showed that **8** connects **7** with **9** for the intermediate model and real system. (Structure **9** was not investigated further here, but its geometry and relative energy were consistent with the corresponding structure **9** that we previously studied for the small model.) Previously we showed that with the small model, transition structure **8** connects **5** with **9** (rather than **7** with **9**), but once again this is an artifact of the unconstrained phosphine ligands in the small model; when Zr−P is constrained to remain at 2.80 Å, then this transition structure connects **7** with **9** as expected. This represents a slight change to our earlier mechanism in that now we conclude that the experimentally observed structure **7** is the common starting point for pathway A and pathway B (Figure 1) (leading to further low-energy structures).

The optimized geometries for structure **13** are shown in Figure 13. The unconstrained geometry for the small model adopted very nearly C_2 symmetry, and so this symmetry constraint was imposed for the remaining optimizations. This was verified by frequency calculations at the "f" level of theory. The core of the molecule exhibited very little change on going from the small to the intermediate model to the real system, with the Zr_2N_2 portion being planar at all levels of theory. Two partially dissociated phosphines are retained in all cases. This is an interesting observation since it implies that in the real system the macrocyclic ligand has enough flexibility to allow some lability in coordination to the zirconium.

Transition structure **14** is shown in Figure 14. It has a rather bent structure in the small and intermediate models and a fairly short N−N distance for the bridging nitrogens (1.49 Å). Minimizations along the normal mode corresponding to the imaginary frequency showed that **14** connects **7** with **15** for the small, intermediate, and real systems. (Structure **15** was not investigated further for the intermediate and real systems.)

Figure 15 shows the structure of product **17**. As with structure **13**, an unconstrained geometry optimization on the small model suggested a C_2 symmetric structure, and this point group was imposed for the remaining optimizations and verified by frequency calculations at the "f" level of theory. With the small model, two of the phosphines are dissociated. In the intermediate model two slightly elongated Zr−P distances are retained although overall the structure looks quite reasonable. In the real system there are no unbound phosphine ligands, and all four Zr−P distances are perfectly acceptable. The Zr_2N_2 portion is shown to be almost planar at all levels of theory.

As discussed by Fryzuk[33,50,56] and others,[28,89] the degree of activation of N_2 is often measured by the increase in the N−N bond length and decrease in the N−N stretching frequency compared to molecular nitrogen. In Table 4 is shown the N−N bond lengths, stretching frequencies, and force constants calculated at the B3LYP/SBK level of theory for the intermediate model of all the structures in this study.

Table 4. N−N Bond Lengths, Vibrational Frequencies, and Force Constants for the Intermediate Model Structures Calculated with B3LYP/SBK[a]

structure	$r_{\mathrm{N-N}}$ (Å)	$\nu_{\mathrm{N-N}}$ (cm^{-1})	$f_{\mathrm{N-N}}$ (mDyne/Å)
N_2	1.132	2445	49.33
1	1.508	795	3.03
2	1.523	722	0.84
3	1.505	834	0.86
6	1.499	859	1.43
7	1.482	872	1.86
8	1.494	846	1.29
13	2.808	480	0.76
14	1.485	854	0.94
17	2.656	645	0.95

[a] Fully optimized geometries from the “f” level of theory were used.

Much more sophisticated analyses have been carried out for similar model structures,[50,56] and the results in Table 4 represent a fairly simplistic treatment of the vibrational frequencies. Apart from the obvious activation in all the structures of N_2 compared to molecular nitrogen, this table shows that there is a remarkable similarity in the N−N stretching frequencies of structures **1**−**8**. At first sight this might suggest a similarity in the N−N bond strengths in these structures; however, the calculated force constants indicate that there is a subtle variation in the N−N bonds across the series.

Energies. The relative energies of the structures considered in the current study are shown in Table 5 at seven different levels of model/theory. For structure **2** it appears that the RHF/LANL2MB level of theory (columns d and h) leads to an overestimate of the activation energy. All of the activation energies calculated in the present work increase in a very systematic manner from “a” to “f” to “i” (22, 27, 35 for **2**, 55, 67, 71 for **8**, and 55, 62, 72 kcal mol^{-1} for **14**, respectively). The relative energy of **3** is fairly similar across the different levels of theory except for “i”. Since the geometry looks fine and since stability tests of the wave function indicate that the calculated energy is indeed for the lowest electronic state at this geometry, it is difficult to explain this discrepancy. Note that constraining the Zr−P distance in the ligand system in the small model has a pronounced effect in this case. This is also true of structures **13** and **17**, while for the remainder of the compounds it appears that constraining the Zr−P distance has little effect on the relative energies. The relative energies of **7**, **8**, and **14** are fairly similar across all levels of theory. The results for **8** and **14** correspond to full geometry optimizations. However for the intermediate and real systems we initially carried out calculations in which we froze the geometry of the central core of atoms at the transition structure geometries determined in our previous study for the small model, and then the remainder of the ligand system was minimized. This led to the activation energies being overestimated by 30−40 kcal mol^{-1} (see Supporting Information) and indicates the importance of allowing the core geometry to also be relaxed in the real system. The relative energies of **13** and **17** are similar across all the models we have used, except for column h which appears to overestimate their stability.

The effect of changing the level of theory in our calculations can be more clearly assessed by comparing column a with b in Table 5, d with f, e with f, and h with i. In the first case the energies are almost identical indicating that extra polarization functions on N and P had little effect. In the second case the energies are fairly similar except for **2**, **3**, **13**, and **17**, indicating that electron correlation is important for these systems. In the third case (e with f) the comparisons are quite excellent, indicating that the ONIOM treatment is a very effective model for the full B3LYP calculation. In the final case (h with i) the energies vary inconsistently indicating a possible problem with the RHF/LANL2MB component of the ONIOM calculations.

The relative energies from the singe point calculations at the B3LYP/SBKBS2 level of theory are shown in Table 6. Overall the consistency between the three different models is quite excellent (particularly if one focuses on the Zr−P constrained values in parentheses for the small model). This lends confidence to our conclusions about the mechanism of the dihydrogen activation (noting that **7** is now shown to be common to both pathways A and B). There are only minor differences in the relative energies between the three models, and the potential energy surfaces for the small, intermediate, and real systems are compared in Figure 1. A comparison of Table 6 with Table 5 also shows that the single point calculations did not change greatly the relative energies of the small and intermediate models; however, there are larger changes apparent for the real system. This might indicate

Table 5. Relative Energies (kcal mol^{-1}) for Optimized Geometries

	model						
	small		intermediate		real		
	a[a,b]	b[a,b]	d[a]	e[a]	f[a]	h[a]	i[a]
1	0 (0)	0 (0)	0	0	0	0	0
2	21.6(21.5)	22.2 (22.2)	43.0	24.5	26.7	49.5	34.6
3	−13.6(−8.4)	−14.0 (−10.2)	−10.1	3.9	3.1	7.7	19.0
6[c]	(−5.0)	(−6.2)	9.5	12.6	14.0		
7	−13.1(−13.0)	−13.7 (−13.6)	−13.9	−16.6	−15.3	3.0	7.9
8	43.4(42.0)	40.8 (40.3)	47.7	49.4	52.3	56.1	63.3
13	−65.5(−54.2)	−65.5 (−54.0)	−65.4	−47.9	−42.1	−110.9	−79.1
14	42.6(41.6)	40.0 (39.5)	39.6	44.7	46.9	17.3	64.3
17	−64.2(−51.1)	−65.3 (−52.5)	−74.7	−47.8	−56.9	−122.0	−102.3

[a] Level of theory, see Table 1. [b] Numbers in parentheses are calculated with Zr−P frozen at 2.80 Å. [c] Structure **6** could not be located at some levels of theory—see text.

Table 6. Relative Energies (kcal mol^{-1}) for Single-Point Calculations

	model		
	small c^{a,b}	intermediate g^{a}	real j^{a}
1	0 (0)	0	0
2	22.4(22.5)	24.3	16.3
3	−13.3(−9.2)	−0.7	−4.3
6	(−4.9)	9.6	
7	−12.5(−12.3)	−13.8	−7.2
8	41.1(40.5)	42.8	39.5
13	−66.3(−53.4)	−36.9	−44.6
14	40.1(39.6)	44.5	42.1
17	−68.3(−53.5)	−52.0	−64.2

[a] Level of theory, see Table 1. [b] Numbers in parentheses are calculated with Zr−P frozen at 2.80 Å.

Table 7. Substituent Values (kcal mol^{-1})^{a,b}

structure	Subst(RHF/LANL2MB)	Subst(B3LYP/SBK)
1	0(0)	0(0)
2	11.8 (7.0)	5.1(5.2)
3	36.7 (36.7)	16.7(11.5)
6		(9.0)
7	30.6 (6.6)	−2.2 (−2.3)
8	70.2 (46.4)	8.9(10.3)
13	27.6 (−1.2)	23.4 (12.1)
14	66.6 (39.9)	4.3(5.3)
17	6.5 (−36.8)	7.3 (−5.8)

[a] Here, Subst(method) = Rel E (method//method)$_{\text{intermediate}}$ − Rel E (method//method)$_{\text{small}}$ e.g. Subst(B3LYP/SBK) = Rel E (B3LYP/SBK//B3LYP/SBK)$_{\text{intermediate}}$ − Rel E (B3LYP/SBK//B3LYP/SBK)$_{\text{small}}$ = "f" − "a". [b] Numbers in parentheses are calculated with Zr−P frozen at 2.80 Å.

Table 8. Component Energies (kcal mol^{-1}) for IMOMM Calculations^{a}

structure	MM contribution	MO contribution	IMOMM rel E	MO destabilization^{b}
1	0	0	0	15.2(10.4)
2	1.3(1.7)	45.8(30.3)	47.1(32.0)	18.1(14.2)
3	2.6(6.7)	−2.5(20.4)	0.1(27.1)	22.9(29.0)
7	−2.1(4.8)	−16.0(−12.2)	−18.1(−7.4)	13.1(13.7)

[a] Results shown are for IMOMM(RHF/LANL2MB intermediate:MM3 real) and, in parentheses, for IMOMM(B3LYP/SBK intermediate:MM3 real). [b] MO destabilization = RHF/LANL2MB component of IMOMM calculation − RHF/LANL2MB//RHF/LANL2MB energy (from level "d"), and in parentheses MO destabilization = B3LYP/SBK component of IMOMM calculation − B3LYP/SBK//B3LYP/SBK energy (from level "f").

the limitations of using the UFF force field approach for calculating the energetics of the real system in Table 5.

As Table 6 shows, provided a large flexible basis set is used (SBKBS2) then the three ligand models give very similar relative energies. Thus from one point of view changing the nature of the substituents from the small to the intermediate to the real system has very little effect. However we have also looked at three other ways of investigating the substituent effects at the levels of theory used for the geometry optimizations.

First, substituent effects may be studied quantitatively by comparing the relative energies for the small and intermediate structural models at identical levels of theory. We have tabulated these "substituent values"[90] in Table 7. (To evaluate the substituent value at RHF/LANL2MB we have reoptimized the small model structures at this level of theory.) The results in Table 7 show that there is quite a variation in the substituent effect at the RHF/LANL2MB level, although to some extent this is due to the use of the unconstrained ligand. The values at B3LYP/SBK show much less variation; however, they still change sign and indicate a lack of consistency in the substituent effects. Of perhaps more interest is to compare the substituent values between the RHF and B3LYP methods for each structure. Levels of theory with similar substituent values may be combined in an ONIOM calculation with greater confidence. In this case Table 7 shows that for at least some of the structures the agreement between the levels of theory is rather poor.

Second, the effect of substituents can be further assessed by comparing a with e (Table 5), d with h, and f with i (each of these comparisons evaluates the effect of including substituents via the ONIOM method). The first of these comparisons (making use of the Zr−P constrained "a" results) shows that the substituents generally raise the relative energies by a small amount. The second comparison (d with h) shows that **2**−**8** are destabilized, while **13**, **14**, and **17** are stabilized considerably. Finally the comparison of f with i shows a consistent destabilization for most structures, but a substantial stabilization again for **13** and **17**.

Third, an analysis of the IMOMM component energies for the real system is shown in Table 8 (the relative energies for these levels of theory are shown in the Supporting Information). This shows a breakdown of the overall relative energy into contributions from the molecular orbital part and the molecular mechanics part of the IMOMM calculations. For both sets of IMOMM calculations the MO part dominates. This table also includes the "MO destabilisation energy" which is a measure of the absolute change in the MO portion of the molecule caused by the geometry distorting to accommodate the MM substituents (in this case, the SiMe$_2$ groups and the PPh group). The values for both sets of IMOMM calculations are fairly similar and indicate that the substituent destabilization is about the same for structures **1**, **2**, and **7** but somewhat larger for **3** at the B3LYP level.

Concluding Remarks

In this paper we have compared a number of different structural models and levels of theory for representing the reaction of a binuclear zirconium dinitrogen complex. First in terms of geometries, most of the structures for **1**−**17** look quite similar with the different approaches we have used. The problem of phosphine dissociation is fixed with the larger macrocyclic ligand. The phenyl groups on the phosphines, together with the macrocyclic ligand, are essential to get the correct planarity of the Zr$_2$N$_2$ portion of structures **1**, **2**, and **3**. Second in terms of energetics, our component analysis suggests that there are some moderately large effects on the absolute energies of the central region caused by the surrounding substituents. However the relative energetic effects of the substituents appear to be small (particularly with the use of the large flexible SBKBS2 basis set), and therefore the model we used in our earlier work does indeed appear to be a good model. Some of the activation energies may be larger than we originally calculated and we have now clarified the intermediacy of structure **7** for both pathways A and B, but otherwise our original potential energy surface and mechanism are quite satisfactory.

Finally we note that the molecular orbital calculations on the intermediate model and the ONIOM calculations on the real system are still quite time-consuming, particularly when detailed exploration of potential energy surfaces and investigation of subtle conformational effects is required. The work presented here lends support to our use of simple theoretical models to represent moderately large experimental structures. It is probably more efficient, without sacrificing reliability, to mimic experimental ligands by choosing simple hydrides with constrained distances (and angles) to obtain realistic structures. Such approaches have been used successfully by us and others for a variety of recent experimental systems.[34,68]

Acknowledgment. This work was supported by the Australian Research Council, the Australian Academy of Science, and in part by a grant (CHE-0209660) from the National Science Foundation. Acknowledgment is made for generous support of computing time at the Emerson Center of Emory University, the Maui High Performance Computing Center, and the Australian Partnership for Advanced Computing. B.F.Y. acknowledges the award of an Emerson Fellowship and is grateful to Dmitry Khoroshun for patiently answering many questions. We thank the anonymous referees for helping to strengthen this manuscript.

Supporting Information Available: Comparisons of theory with experiment for structure **7**, total energies and Cartesian coordinates of all structures, details of basis sets employed, and the complete references for refs 86−88. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Ertl, G. In *Catalytic Ammonia Synthesis*; Jennings, J. R., Ed.; Plenum: New York, 1991.

(2) Eady, R. R. *Perspectives on Bioinorganic Chemistry*; JAI Press: Greenwich, CT, 1991.

(3) Leigh, G. J. *Acc. Chem. Res.* **1992**, *25*, 177−181.

(4) Kim, J.; Rees, D. C. *Nature* **1992**, *360*, 553.

(5) Kim, J.; Rees, D. C. *Science* **1992**, *257*, 1667.

(6) Chan, M. K.; Kim, J.; Rees, D. C. *Science* **1993**, *260*, 792.

(7) Deng, H.; Hoffmann, R. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1062−1065.

(8) Ludden, P. W. In *Encyclopedia of Inorganic Chemistry*; Wiley: New York, 1994; p 2566.

(9) Coucouvanis, D. In *Encyclopedia of Inorganic Chemistry*; Wiley: New York, 1994; p 2557.

(10) Demadis, K. D.; Coucouvanis, D. *Inorg. Chem.* **1995**, *34*, 436.

(11) Howard, J. B.; Rees, D. C. *Chem. Rev.* **1996**, *96*, 2983.

(12) Burgess, B. K.; Lowe, D. J. *Chem. Rev.* **1996**, *96*, 2983.

(13) Eady, R. R. *Chem. Rev.* **1996**, *96*, 3013.

(14) Malinak, S. M.; Simeonov, A. M.; Mosier, P. E.; McKenna, C. E.; Coucouvanis, D. *J. Am. Chem. Soc.* **1997**, *119*, 1662.

(15) Fryzuk, M. D.; Love, J. B.; Rettig, S. J.; Young, V. G. *Science* **1997**, *275*, 1445−1447.

(16) Ferguson, R.; Solari, E.; Floriani, C.; Osella, D.; Ravera, M.; Re, N.; Chiesi-Villa, A.; Rizzoli, C. *J. Am. Chem. Soc.* **1997**, *119*, 10104−10115.

(17) Sellmann, D.; Sutter, J. *Acc. Chem. Res.* **1997**, *30*, 460−469.

(18) Neyman, K. M.; Nasluzov, V. A.; Hahn, J.; Landis, C. R.; Rösch, N. *Organometallics* **1997**, *16*, 995−1000.

(19) Campazzi, E.; Solari, E.; Floriani, C.; Scopelliti, R. *Chem. Commun.* **1998**, 2603−2604.

(20) Leigh, G. J. *Science* **1998**, *279*, 506.

(21) Nishibayashi, Y.; Iwai, S.; Hidai, M. *Science* **1998**, *279*, 540.

(22) Fryzuk, M. D.; Love, J. B.; Rettig, S. J. *Organometallics* **1998**, *17*, 846−853.

(23) Tuczek, F.; Lehnert, N. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 2636−2638.

(24) Basch, H.; Musaev, D. G.; Morokuma, K.; Fryzuk, M. D.; Love, J. B.; Seidel, W. W.; Albinati, A.; Koetzle, T. F.; Klooster, W. T.; Mason, S. A.; Eckert, J. *J. Am. Chem. Soc.* **1999**, *121*, 523−528.

(25) Basch, H.; Musaev, D. G.; Morokuma, K. *J. Am. Chem. Soc.* **1999**, *121*, 5754−5761.

(26) Musaev, D. G.; Cui, Q.; Svensson, M.; Morokuma, K. In *Transition State Modeling for Catalysis*; Truhlar, D. G., Morokuma, K., Eds.; American Chemical Society: Washington, DC, 1999; pp 198−207.

(27) Khoroshun, D. V.; Musaev, D. G.; Morokuma, K. *Organometallics* **1999**, *18*, 5653−5660.

(28) Lehnert, N.; Tuczek, F. *Inorg. Chem.* **1999**, *38*, 1671−1682.

(29) Ishino, H.; Tokunaga, S.; Seino, H.; Ishii, Y.; Hidai, M. *Inorg. Chem.* **1999**, *38*, 2489−2496.

(30) Clentsmith, G. K. B.; Bates, V. M. E.; Hitchcock, P. B.; Cloke, F. G. N. *J. Am. Chem. Soc.* **1999**, *121*, 10444−10445.

(31) Peters, J. C.; Cherry, J.-P. F.; Thomas, J. C.; Baraldo, L.; Mindiola, D. J.; Davis, W. M.; Cummins, C. C. *J. Am. Chem. Soc.* **1999**, *121*, 10053−10067.

(32) Mindiola, D. J.; Meyer, K.; Cherry, J.-P. F.; Baker, T. A.; Cummins, C. C. *Organometallics* **2000**, *19*, 1622−1624.

(33) Fryzuk, M. D.; Johnson, S. A. *Coord. Chem. Rev.* **2000**, *200−202*, 379−409.

(34) Basch, H.; Musaev, D. G.; Morokuma, K. *Organometallics* **2000**, *19*, 3393−3403.

(35) Caselli, A.; Solari, E.; Scopelliti, R.; Floriani, C.; Re, N.; Rizzoli, C.; Chiesi-Villa, A. *J. Am. Chem. Soc.* **2000**, *122*, 3652−3670.

(36) Dubé, T.; Ganesan, M.; Conoci, S.; Gambarotta, S.; Yap, G. P. A. *Organometallics* **2000**, *19*, 3716−3721.

(37) Guan, J.; Dubé, T.; Gambarotta, S.; Yap, G. P. A. *Organometallics* **2000**, *19*, 4820−4827.

(38) Fryzuk, M. D.; Johnson, S. A.; Patrick, B. O.; Albinati, A.; Mason, S. A.; Koetzle, T. F. *J. Am. Chem. Soc.* **2001**, *123*, 3960−3973.

(39) Solari, E.; Hesschenbrouck, J.; Scopelliti, R.; Floriani, C.; Re, N. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 932−934.

(40) Solari, E.; Da Silva, C.; Iacono, B.; Hesschenbrouck, J.; Rizzoli, C.; Scopelliti, R.; Floriani, C. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 3907−3909.

(41) Bielawa, H.; Hinrichsen, O.; Birkner, A.; Muhler, M. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 1061−1063.

(42) Evans, W. J.; Allen, N. T.; Ziller, J. W. *J. Am. Chem. Soc.* **2001**, *123*, 7927−7928.

(43) Chirik, P. J.; Henling, L. M.; Bercaw, J. E. *Organometallics* **2001**, *20*, 534−544.

(44) Roussel, P.; Errington, W.; Kaltsoyannis, N.; Scott, P. *J. Organomet. Chem.* **2001**, *635*, 69−74.

(45) Greco, G. E.; Schrock, R. R. *Inorg. Chem.* **2001**, *40*, 3861−3878.

(46) Hidai, M.; Mizobe, Y. *Pure Appl. Chem.* **2001**, *73*, 261−263.

(47) Smith, J. M.; Lachicotte, R. J.; Pittard, K. A.; Cundari, T. A.; Lukat-Rodgers, G.; Rodgers, K. R.; Holland, P. L. *J. Am. Chem. Soc.* **2001**, *123*, 9222−9223.

(48) Fryzuk, M. D.; Kozak, C. M.; Bowdridge, M. R.; Patrick, B. O.; Rettig, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 8389−8397.

(49) Musaev, D. G.; Basch, H.; Morokuma, K. In *Computational Modeling of Homogeneous Catalysis*; Maseras, F., Lledos, A., Eds.; Kluwer Academic Publ.: 2002; pp 325−361.

(50) Studt, F.; Morello, L.; Lehnert, N.; Fryzuk, M. D.; Tuczek, F. *Chem.--Eur. J.* **2003**, *9*, 520−530.

(51) Fryzuk, M. D.; MacKay, B. A.; Patrick, B. O. *J. Am. Chem. Soc.* **2003**, *125*, 3234−3235.

(52) Fryzuk, M. D. *Chem. Rec.* **2003**, *3*, 2−11.

(53) Shaver, M. P.; Fryzuk, M. D. *Adv. Synth. Catal.* **2003**, *345*, 1061−1076.

(54) Yandulov, D. V.; Schrock, R. R. *Science* **2003**, *301*, 76−78.

(55) Schrock, R. R. *Chem. Commun.* **2003**, 2389−2391.

(56) Studt, F.; MacKay, B. A.; Fryzuk, M. D.; Tuczek, F. *J. Am. Chem. Soc.* **2004**, *126*, 280−290.

(57) Pool, J. A.; Lobkovsky, E.; Chirik, P. J. *Nature* **2004**, *427*, 527−530.

(58) Fryzuk, M. D. *Nature* **2004**, *427*, 498−499.

(59) MacKay, B. A.; Fryzuk, M. D. *Chem. Rev.* **2004**, *104*, 385−401.

(60) Nishibayashi, Y.; Saito, M.; Uemura, S.; Takekuma, S.; Takekuma, H.; Yoshida, Z. *Nature* **2004**, *428*, 279−280.

(61) Musaev, D. G. *J. Phys. Chem. B* **2004**, *108*, 10012−10018.

(62) Evans, W. J.; Lee, D. S.; Rego, D. B.; Perotti, J. M.; Kozimor, S. A.; Moore, E. K.; Ziller, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 14574−14582.

(63) Pool, J. A.; Bernskoetter, W. H.; Chirik, P. J. *J. Am. Chem. Soc.* **2004**, *126*, 14326−14327.

(64) Kozak, C. M.; Mountford, P. *Angew. Chem., Int. Ed.* **2004**, *43*, 1186−1189.

(65) Hanna, T. E.; Lobkovsky, E.; Chirik, P. J. *J. Am. Chem. Soc.* **2004**, *126*, 14688−14689.

(66) Bobadova-Parvanova, P.; Wang, Q.; Morokuma, K.; Musaev, D. G. *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 7101−7103.

(67) Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Am. Chem. Soc.* **1993**, *115*, 6908−6915.

(68) Fryzuk, M. D.; Johnson, S. A.; Rettig, S. J. *Organometallics* **2000**, *19*, 3931−3941.

(69) Froese, R. D. J.; Morokuma, K. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley: Chichester, 1998; pp 1245−1257.

(70) Dapprich, S.; Komaroni, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. (THEOCHEM)* **1999**, *461−462*, 1−21.

(71) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299−310.

(72) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026.

(73) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612.

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

(75) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(76) Stephens, P. J.; Devlin, J. F.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(77) Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1977**, *66*, 3045−3050.

(78) Bauernschmitt, R.; Ahlrichs, R. *J. Chem. Phys.* **1996**, *104*, 9047−9052.

(79) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170−1179.

(80) Matsubara, T.; Maseras, F.; Koga, N.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 2573.

(81) Matsubara, T.; Sieber, S.; Morokuma, K. *Int. J. Quantum Chem.* **1996**, *60*, 1101−1109.

(82) Maseras, F.; Lledos, A.; Clot, E.; et al. *Chem. Rev.* **2000**, *100*, 601−636.

(83) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(84) Allinger, N. L. *MM3(92)*; Quantum Chemistry Program Exchange: Indiana University, 1992.

(85) Aped, A.; Allinger, N. L. *J. Am. Chem. Soc.* **1992**, *114*, 4.

(86) Frisch, M. J. et al. *Gaussian 98 A.1, Revision A.1*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(87) Frisch, M. J. et al. *Gaussian 03, Revision B.03*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(88) Frisch, M. J. et al. *Gaussian 92/DFT, Revision F*; Gaussian, Inc.: Pittsburgh, PA, 1993.

(89) Deeth, R. *J. Organomet. Chem.* **2001**, *635*, 165−172.

(90) Morokuma, K.; Musaev, D. G.; Vreven, T.; Basch, H.; Torrent, M.; Khoroshun, D. V. *IBM J. Res. Dev.* **2001**, *45*, 367−395.

CT050320E

J. Chem. Theory Comput. **2006,** *2,* 1317−1324

JCTC Journal of Chemical Theory and Computation

Performance Evaluation of the Three-Layer ONIOM Method: Case Study for a Zwitterionic Peptide

Keiji Morokuma,*,† Qingfang Wang,† and Thom Vreven‡

Cherry L. Emerson Center for Scientific Computation and Department of Chemistry,
Emory University, Atlanta, Georgia 30322, and Gaussian, Inc., 340 Quinnipiac Street,
Building 40, Wallingford, Connecticut 06492

Received April 12, 2006

Abstract: The performance of the three-layer ONIOM method was systematically investigated by comparing the optimized geometries and calculated deprotonation energy of a zwitterionic peptide molecule, $NH_3^+-CH^nBu-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$, using all possible combinations of B3LYP/6-31G* as the high-level quantum (HQ), AM1 as the low-level quantum (LQ), and Amber as the MM method. Results show that the three-layer ONIOM(HQ:LQ:MM) method, which includes a medium-level quantum method in the middle system to take into account the electronic effects of the middle layer and to keep the problematic QM-MM boundary away from the action region, is more reliable and more stable than the QM:MM:MM or generic QM/MM method and is the best compromise between accuracy and computational cost.

I. Introduction

As theoretical/computational chemistry has gained a major role in studies of chemical problems in the last few decades, it became a challenge for theoreticians to accurately treat large molecular systems such as encountered in biochemistry. No single theoretical method is able to provide both the accuracy and acceptable computational cost that are required for the investigation of such chemical processes. Accurate ab initio quantum mechanics (QM) methods either scale nonlinearly with the size of the system, or have large pre-factors that prevent them to be applied to large systems. Low cost molecular mechanics (MM) methods are widely used and scale linearly but have the obvious weakness of the poor description of bond breaking and forming processes. Meanwhile, in many biological systems the actual reaction only localizes in a relatively small region.

A breakthrough came with the realization that it is not necessary to use a single computational method, which resulted in the development and application of a variety of hybrid methods. These methods utilize the localization feature of large reaction systems, and most combine the merit of accuracy of a QM method with the low cost of MM methods.[1−3] Our ONIOM method is a versatile and popular hybrid method.[4−15] ONIOM divides the system into several onion-like layers, treating the active center with the highest level ab initio QM method, while outer layers can be treated with less expensive methods, such as low-level ab initio QM, semiempirical QM, or MM methods.

Figure 1 illustrates the basic concept of the multilayered ONIOM method, taking three-layer ONIOM as an example. The three-layer ONIOM(high:medium:low) energy is an approximation to the energy at the high level for the real system, $E_{\text{real}}^{\text{high}}$, referred to as the target, and is given by

$$E_{\text{real}}^{\text{ONIOM}} = E_{\text{model}}^{\text{high}} + E_{\text{middle}}^{\text{medium}} + E_{\text{real}}^{\text{low}} - E_{\text{model}}^{\text{medium}} - E_{\text{middle}}^{\text{low}} \quad (1)$$

where *high*, *medium*, and *low* refer to the high-, medium-, and low-level theoretical methods, respectively, while *model*, *middle*, and *real* refer to the model, middle, and real systems, respectively. The middle system is a part cut from the real system, and the model system is a part cut from the middle system. The model and middle systems are mended by *link atoms* to satisfy the valencies if covalent bonds are cut. With the term "energy", we typically refer to a relative energy, such as the binding energy or the barrier height.

The accuracy of the ONIOM method is defined as the error of ONIOM relative to the target calculation (the high-level

---

* Corresponding author e-mail: morokuma@emory.edu.
† Emory University.
‡ Gaussian, Inc.

10.1021/ct600135b CCC: $33.50

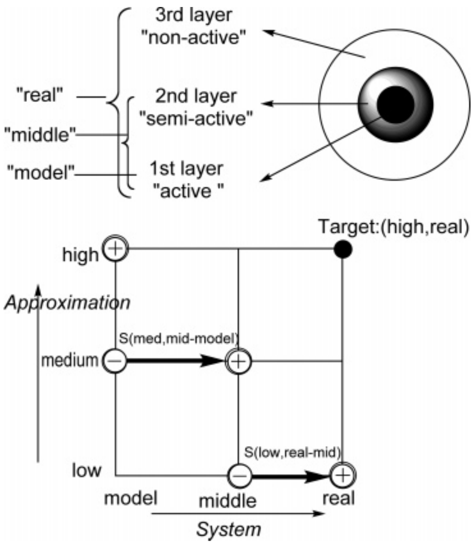


Figure 1. The layer partition and energy extrapolation scheme of the three-layer ONIOM method.

result for the real system):

$$\mathrm{Err}_{\mathrm{real}}^{\mathrm{ONIOM}} = E_{\mathrm{real}}^{\mathrm{ONIOM}} - E_{\mathrm{real}}^{\mathrm{high}} \tag{2}$$

$\mathrm{Err}_{\mathrm{real}}^{\mathrm{ONIOM}}$ depends critically on the partitioning of the system into middle and model systems and the reliability of medium- and low-level methods used in ONIOM. The accuracy of any ONIOM combination can be tested using the *S*-values,[11] which are defined as

$$\begin{aligned} S_{\mathrm{real-middle}}^{\mathrm{low}} &= E_{\mathrm{real}}^{\mathrm{low}} - E_{\mathrm{middle}}^{\mathrm{low}} \\ S_{\mathrm{middle-model}}^{\mathrm{medium}} &= E_{\mathrm{middle}}^{\mathrm{medium}} - E_{\mathrm{model}}^{\mathrm{medium}} \\ S_{\mathrm{real-model}}^{\mathrm{high}} &= E_{\mathrm{real}}^{\mathrm{high}} - E_{\mathrm{model}}^{\mathrm{high}} \end{aligned} \tag{3}$$

With these definitions of the *S*-values, the ONIOM energy can be written as

$$E_{\mathrm{real}}^{\mathrm{ONIOM}} = E_{\mathrm{model}}^{\mathrm{high}} + S_{\mathrm{middle-model}}^{\mathrm{medium}} + S_{\mathrm{real-middle}}^{\mathrm{low}} \tag{4}$$

Therefore

$$\begin{aligned} \mathrm{Err}_{\mathrm{real}}^{\mathrm{ONIOM}} &= S_{\mathrm{real-model}}^{\mathrm{high}} - S_{\mathrm{middle-model}}^{\mathrm{medium}} - S_{\mathrm{real-middle}}^{\mathrm{low}} \\ &= [S_{\mathrm{middle-model}}^{\mathrm{high}} - S_{\mathrm{middle-model}}^{\mathrm{medium}}] + \\ &\qquad [S_{\mathrm{real-middle}}^{\mathrm{high}} - S_{\mathrm{real-middle}}^{\mathrm{low}}] \\ &= \Delta S_{\mathrm{middle-model}}^{\mathrm{high-medium}} + \Delta S_{\mathrm{real-middle}}^{\mathrm{high-low}} \end{aligned} \tag{5}$$

Thus, the error of the three-layer ONIOM approximation is zero if both $S_{\mathrm{middle-model}}^{\mathrm{high}} = S_{\mathrm{middle-model}}^{\mathrm{medium}}$ and $S_{\mathrm{real-middle}}^{\mathrm{high}} = S_{\mathrm{real-middle}}^{\mathrm{low}}$, namely, if the energy difference between the middle and the model at the medium level reproduces that of the high level and at the same time the energy difference between the real and the middle systems at the low level reproduces that of the high level. The *S*-value is therefore a useful tool in the calibration of hybrid methods. Of course one can benchmark hybrid methods also by calculating the error simply from the target result and the hybrid method result, but the *S*-value test has some advantages. First, it allows for a systematic and clear investigation of different method combinations and partitions, as we will show in the current paper. Second, for the three-layer ONIOM, it allows for investigation of the performance of the low and medium levels of theory individually. From eq 5 it is clear that the error in three-layer ONIOM can be zero when $\Delta S_{\mathrm{middle-model}}^{\mathrm{high-medium}}$ is the negative of $\Delta S_{\mathrm{real-middle}}^{\mathrm{high-low}}$, but neither is zero. Such cancellation of errors is undesirable and may be missed with a simple calibration based on only the target and hybrid method results.

Below we will systematically apply the three-layer, two-layer, and single-level (non-ONIOM, conventional) calculations for a sample test molecule. We just mention standard (one-layer) and two-layer calculations can be considered to be a special case of three-layer ONIOM in which two or more levels happen to be equal. For instance, ONIOM(high:low:low) is equivalent to the two-layer ONIOM(high:low) method using the real and the model system, and the same type of *S*-value analysis can be performed.

Two different approaches exist for the treatment of electrostatic interaction between the QM layers and the MM layer.[10,16] In the default ONIOM scheme, the electrostatic interaction is included in the MM calculations, which is referred to as *mechanical embedding* (ONIOM-ME). The energy expression for the three-layer method with MM as the low-level, is then

$$E_{\mathrm{real}}^{\mathrm{ONIOM-ME}} = E_{\mathrm{model}}^{\mathrm{high}} + E_{\mathrm{middle}}^{\mathrm{medium}} + E_{\mathrm{real}}^{\mathrm{MM}} - E_{\mathrm{model}}^{\mathrm{medium}} - E_{\mathrm{middle}}^{\mathrm{MM}} \tag{6}$$

Here $E_{\mathrm{real}}^{\mathrm{MM}}$ includes the molecular mechanics electrostatic interaction in the full (real) system ($\mathrm{ES}_{\mathrm{real}}$), and $E_{\mathrm{middle}}^{\mathrm{MM}}$ includes the molecular mechanics electrostatic interaction for the middle system ($\mathrm{ES}_{\mathrm{middle}}$). Terms not involving the interactions between the QM parts and the MM part mostly cancel out between these two ES terms. $\mathrm{ES}_{\mathrm{real}}$ and $\mathrm{ES}_{\mathrm{middle}}$ are given as

$$\mathrm{ES}_{\mathrm{system}} = \sum_{J<K}^{\mathrm{systems}} \frac{s_{JK} q_J q_K}{r_{JK}} \tag{7}$$

where q_J is the partial charge on the atom *J*, r_{JK} is the distance between atoms *J* and *K*, and summation *J* and *K* are for all the atoms in the *system* (real or middle). Here s_{JK} is the scale factor of electrostatic interaction introduced in most force field definitions (independent of the ONIOM or QM/MM applications). For instance in the Amber force field, it is 0, 1/1.2, and 1 for *JK* atom pairs separated by one and two bonds (first and second neighbors), three bonds (third neighbors), and more than three bonds (fourth and further neighbors), respectively. The same scale factor is usually used for the generic QM/MM-ME or ONIOM-ME schemes.

In the electronic embedding ONIOM-EE scheme,[10] one includes the electrostatic interaction between the QM charges and the MM part as additional scaled electrostatic Hamiltonians H' of the QM systems

$$H'_{\text{system}} = \sum_{i}^{\text{system}} \sum_{J}^{MM} \frac{\sigma_J q_J}{r_{iJ}} \tag{8}$$

where system refers to model and middle systems, and *i* refers to the electrons in that system. There is a fundamental difference between the scaling factors in the two eqs 7 and 8. In eq 7 individual interactions are scaled based on the number of covalent bonds between the atom pairs. This cannot be converted to a QM Hamiltonian. Therefore, in the screened Hamiltonian, eq 8, individual MM atomic charges are scaled by factor σ_J. In the present paper, we use $\sigma_J = 0$ for the first and second classical neighbor of the quantum link atom host (atom replaced by the link atom in the QM model system) and $\sigma_J = 1$ for the third and further classical neighbor. In addition to including the electrostatic interaction energy between the layers (which ONIOM-ME also does), electronic embedding also polarizes the wave function of the QM region by the MM atomic charges. It is therefore supposed to be better in describing a very polar system. However, the EE scheme is more expensive than the ME scheme, because a self-consistent charge iteration is required. Although the basic ideas of the QM/MM-EE and ONIOM-EE are the same, there are some subtle differences in cancellation of terms, and a detailed discussion on the comparison of ONIOM(QM:MM) and generic QM/MM methods has recently been published.[10]

Regardless of whether the ME or EE scheme is used, the arbitrary screening of an otherwise strong short-range electrostatic interaction, possible contributions of link atoms, and other coupling terms (for example the cancellation problem, discussed in detail in ref 10) between the QM and the MM regions is one of the weakest points of either the generic QM/MM or the ONIOM(QM:MM) scheme and affects their performance depending on the scaling and coupling schemes adopted. The most important advantage of the three-layer ONIOM(QM:QM:MM) method is that the inexpensive low-level QM layer pushes the problematic QM-MM boundary far away from the high-level QM region, and the arbitrariness in electrostatic scaling and coupling does not affect very much the final results for the events (or properties) that take place in the model region. Furthermore, often the ME and EE schemes would give similar results, since the polarization by long-distance MM charges does not significantly change the wave function in the vicinity of the action, in which case one would be able to avoid the somewhat more expensive EE scheme.

Currently two- and three-layer ONIOM methods are implemented in the Gaussian program.[17] The purpose of this paper is to present a systematic case study of the performance of the three-layer ONIOM(HQ:LQ:MM), where HQ is the high-level QM method and LQ is the low-level QM method, in comparison with single-level methods and various two-layer ONIOM combinations. We chose the proton dissociation energy of the zwitterionic form of peptide $NH_3^+-CHR_1-CO-NH-CH_2-CO-NH-CHR_2-COO^-$ to form $NH_2-CHR_1-CO-NH-CH_2-CO-NH-CHR_2-COO^-$ as our test system. The intramolecular charge separation makes this a good candidate to evaluate the performance of various ONIOM method combinations for very polar systems.

II. Computational Methods

Peptides can exist either in the neutral form or the zwitterionic form (protonated at N-terminus and deprotonated at C-terminus). Usually for small peptides, the zwitterionic form is stabilized by solvent molecules or crystal force, while in the gas phase the neutral form is favored.[18] Our calculation at the B3LYP/6-31G* level for $NH_3^+-CHR_1-CO-NH-CH_2-CO-NH-CHR_2-COO^-$ in the gas phase shows that when both R_1 and R_2 are *n*-butyl (^{n}Bu) groups, the zwitterionic form is a local minimum structure, and we use this as our case study system. The deprotonation energy of the zwitterionic form $NH_3^+-CH^nBu-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$ at the H_3N^+ position, defined as $\Delta E = E(NH_2\cdots COO^-) + E(H^+) - E(NH_3^+\cdots COO^-)$, was calculated by one-layer, two-layer, and three-layer ONIOM methods using all possible combinations of B3LYP/6-31G* as high-level QM (called HQ throughout the paper), AM1 as low-level QM (LQ), and Amber as the MM method. The geometry of the zwitterionic form $NH_3^+\cdots COO^-$ was fully optimized by the respective ONIOM methods, and the geometry of $NH_2\cdots COO^-$ was assumed unchanged from that of $NH_3^+\cdots COO^-$ optimized geometry, removing the proton at H4, the proton closest to the Bu group on NH_3^+. All the calculations were performed using the Gaussian 03 package.[17]

Three partition schemes, depicted in Figure 2, were adopted. Partition C adopts a minimal model system consisting of only the reaction site H_3N^+L (H: link H atom) as well as the minimal middle system consisting of the reaction site plus the next residue $H_3N^+CHL-CO-NHL$. This partition is just examined as to how small one can make the models and is not likely to be adopted in practical calculations. In addition, QM:MM:MM calculations with this partition will suffer from the aforementioned cancellation problem. Partition B, on the other hand, adopts the reaction site plus the next residue $H_3N^+CHL-CO-NHL$ as the model system for high-level calculation, and the middle system $NH_3^+-CH^nBu-CO-NH-CH_2-CO-NHL$ contains one more residue. Partition A adopts the same model system, $H_3N^+CHL-CO-NHL$ as in partition B, but uses a middle system $(H_3N^+)CHL-CO-NH-CH_2-CO-NH-CHL-(COO-)$ that contains all the ionic components of the system. Even when the real system is much larger as in the case of polypeptides, these model and middle systems would be reasonable choices for practical ONIOM applications. Most of the calculations were performed with the mechanical embedding (ME). Electronic embedding (EE) calculations were performed only for the DFT:MM combinations. We did not perform the EE calculations involving the semiempirical AM1 method, since the one-electron integral corresponding to the QM-MM electrostatic Hamiltonian, eq 8, has not been implemented.

III. Results and Discussions

1. Optimized Geometry in Partition B. At first we want to compare the optimized geometries of the present zwitterionic peptide, $NH_3^+-CH^nBu-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$, using different ONIOM combinations in partition B. The fully optimized geometries are given in the Supporting Information, and Table 1 gives the root-mean-

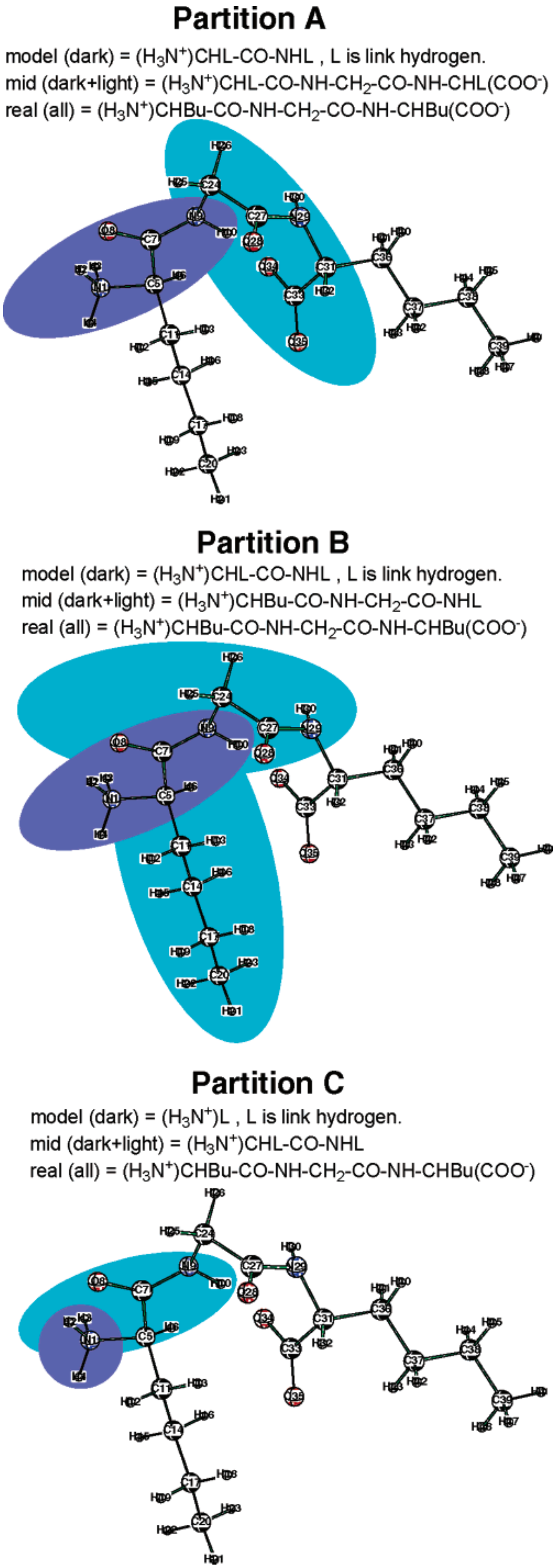


Figure 2. Three partition schemes used for the three-layer ONIOM calculations

square (RMS) deviations of all the bond distances, bond angles, and dihedral angles of the model system (excluding link hydrogens) and the real system, respectively, from those of the target calculation, HQ:HQ:HQ, which is nothing but the pure B3LYP/6-31G* calculation. One notices that both the pure Amber MM method, MM:MM:MM, and the pure semiempirical AM1 method, LQ:LQ:LQ, give very large deviations in geometry optimization of the entire peptide as well as those of the model system; among the two, Amber seems to do a little better than AM1. Among various ONIOM combinations, HQ:HQ:MM gives the smallest error. All the ONIOM combinations using HQ in the model system give rather small deviations in the model part of the geometry, which is expected but not necessarily automatic. An interesting finding is that this is true even for the geometry of the entire peptide, the real system; the correct geometry of the model system seems to dictate the errors in the rest of the system.

Among the methods using MM, the use of the Mulliken charge seems to give somewhat better geometries than the RESP charge; this is opposite to the trend in deprotonation energy to be discussed later. It is noted that when the MM is adopted already in the middle system, i.e., in HQ:MM:MM, the electronic embedding (EE) gives a substantially better geometry than the mechanical embedding (MM), in particular in the geometry of the model system. However, when the MM is used only in the real system, i.e., HQ:HQ:MM, the differences between EE and ME are negligible, because, as suggested earlier, the problematic boundary between QM and MM layers is now on the outside peripheral of the middle system and is located far away from the model system, and the polarization of the QM layer due to the MM charges becomes less important.

2. Deprotonation Energy in Partition B. Results of the deprotonation energies are shown in Table 2. The deprotonation energy of 322.09 kcal/mol at the pure B3LYP/6-31G* level is the target result which ONIOM approximations are trying to reproduce. At first we pay attention to the combinations without MM. The most expensive HQ:HQ:LQ combination, the two-layer HQ:LQ method with the large HQ region (up to the mid system), as expected gives the smallest error of only -4.4 kcal/mol or only 1.3% of the deprotonation energy. If one can afford a large HQ region (HQ:HQ) as well as LQ for the entire system, obviously this is an excellent approximation. The next level of approximation, HQ:LQ:LQ, gives a little larger error of -6.9 kcal/mol, with a smaller cost. The LQ:LQ:LQ or pure semiempirical AM1 calculation is not worth considering as this method is unable to describe the deprotonation reaction even qualitatively, with an absolute error of over 80 kcal/mol. These combinations are all quantum calculations and are likely to remain to be too expensive (see the rough estimated cost in Table 1) in the near future as tools for exploring potential energy surfaces of reactions of most very large (> thousands of atoms) biological systems.

In most real biomolecular calculations, one will have to use MM as the lowest level method for at least a part of the very large system. Of the methods that use MM, HQ:HQ:MM with the RESP[19] charges in the mechanical embedding

Table 1. RMS Errors of ONIOM Optimized Geometries (Relative to the Target Calculation HQ:HQ:HQ) of the $NH_3^+-C^nBuH-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$ System with Partition B

		atoms in the model system only			all atoms in the real system		
combination[a]	estimated cost[b]	9 bond lengths (Å)	13 bond angles (deg)	10 dihedral angles (deg)	47 bond lengths (Å)	87 bond angles (deg)	98 dihedral angles (deg)
HQ:HQ:HQ	10000						
HQ:HQ:LQ	200	0.009	1.34	9.04	0.014	1.27	6.32
HQ:HQ:MM	100	0.017	1.20	2.87	0.010	1.24	5.38
		(0.008)	(1.13)	(2.67)	(0.009)	(1.29)	(4.98)
HQ:HQ:MM (EE)	500	0.017	1.19	2.85	0.010	1.24	5.38
		(0.009)	(1.17)	(2.77)	(0.010)	(1.31)	(5.08)
HQ:LQ:LQ	110	0.012	1.62	14.27	0.022	1.70	8.04
HQ:LQ:MM	21	0.019	1.51	4.62	0.018	1.62	6.10
		(0.016)	(1.32)	(4.97)	(0.018)	(1.70)	(5.58)
HQ:MM:MM	10	0.018	1.54	2.28	0.013	1.55	6.49
		(0.013)	(2.26)	(4.68)	(0.013)	(1.94)	(6.19)
HQ:MM:MM(EE)	50	0.015	1.38	2.97	0.012	1.69	8.05
		(0.012)	(2.11)	(5.30)	(0.011)	(2.23)	(9.23)
LQ:LQ:LQ	100	0.035	3.74	28.40	0.025	2.29	15.10
LQ:LQ:MM	11	0.033	3.77	23.87	0.022	2.22	11.62
		(0.027)	(3.44)	(19.64)	(0.021)	(3.42)	(10.26)
LQ:MM:MM	2	0.032	3.87	17.41	0.018	2.20	9.33
		(0.028)	(4.12)	(16.59)	(0.017)	(3.27)	(9.31)
MM:MM:MM	1	0.031	3.87	27.85	0.017	2.06	12.10
		(0.022)	(2.96)	(18.85)	(0.015)	(3.64)	(10.51)

[a] HQ = B3LYP/6-31G*, LQ = AM1, MM = Amber. For the combinations involving the MM method, B3LYP/6-31G* RESP and Mulliken charges are adopted in the Amber calculation for values without and with parentheses, respectively, and mechanical embedding (ME) is used unless specified as electronic embedding (EE). [b] Very rough estimate of relative cost for a very large system, based on assumed cost: MM = $(10^{-3},10^{-2},1)$, LQ = $(1,10,10^2)$, HQ = $(10,10^2,10^4)$ for (model,mid,real) systems for ME, respectively. For EE the time for HQ and LQ calculations was multiplied by a factor of 5 to reflect the charge-iteration process.

Table 2. One-, Two-, and Three-Layered ONIOM Calculations for the Deprotonation Energy (in kcal/mol) of $NH_3^+-C^nBuH-CO-NH-CH_2-Co-NH-CH^nBu-COO^-$ System with Partition B Using the Optimized Geometries by the Respective Methods

combination[a]	E(high/model)	E(med/mid)	E(med/model)	E(low/real)	E(low/mid)	E(ONIOM)	S(med/ mid-model)	S(low/real-mid)
HQ:HQ:HQ	241.70	245.51	241.70	322.09	245.51	322.09	3.81	76.59
HQ:HQ:LQ	244.04	248.18	244.04	244.96	175.48	317.66	4.15	69.48
HQ:HQ:MM	243.25 (238.86)	243.11 (243.25)	238.68 (238.86)	107.13 (25.80)	68.50 (−10.70)	281.75 (279.75)	4.43 (4.39)	38.63 (36.51)
HQ:HQ:MM (EE)		244.47 (231.47)		69.56 (25.88)	107.14 (−20.67)	282.05 (278.02)		37.58 (46.55)
HQ:LQ:LQ	243.61	174.55	172.43	244.03	174.55	315.21	2.11	69.48
HQ:LQ:MM	239.82 (239.29)	170.29 (170.05)	168.31 (168.08)	104.93 (24.75)	67.25 (−10.99)	279.47 (277.00)	1.98 (1.96)	37.68 (35.74)
HQ:MM:MM	238.50 (236.02)	67.46 (−9.68)	70.29 (28.53)	106.68 (28.35)	67.46 (−9.68)	274.89 (235.83)	−2.83 (−38.21)	39.22 (38.03)
HQ:MM:MM(EE)	258.13 (242.20)		84.27 (38.44)	109.22 (30.70)		283.09 (234.46)		
LQ:LQ:LQ	158.24	160.56	158.24	238.08	160.56	238.08	2.32	77.52
LQ:LQ:MM	158.69 (158.97)	160.83 (161.21)	158.69 (158.97)	108.44 (27.02)	66.72 (−12.34)	202.55 (200.57)	2.14 (2.24)	41.72 (39.36)
LQ:MM:MM	158.99 (158.76)	68.03 (−10.33)	69.33 (26.22)	110.17 (30.85)	68.03 (−10.33)	199.82 (163.40)	−1.30 (−36.55)	42.13 (41.19)

[a] See footnote *a* of Table 1.

(ME) has an error in deprotonation energy of −40 kcal/mol, followed by HQ:LQ:MM of −43 kcal/mol and then HQ:MM:MM of −47 kcal/mol. The HQ:HQ:MM contains a large HQ region and is expensive. The three-layered HQ:LQ:MM method, which is inexpensive because the mid layer is calculated by the inexpensive LQ method, lost only 2.3 kcal/mol over the more expensive HQ:HQ:MM. The standard ONIOM(QM:MM) or generic QM/MM method corresponds to QM:MM:MM, which shows a larger error by 6.9 kcal/mol. The electronic embedding in QM:MM:MM reduces the error from −47 kcal/mol in ME to −38 kcal/mol but is expensive because it has to iterate QM calculations to converge the polarized charges. Thus one can conclude clearly for this example that HQ:LQ:MM is an excellent approximation to the impractical HQ:HQ:MM method and is the method of choice which improves over QM:MM:MM or the standard QM/MM with very little additional cost of semiempirical calculation for the middle system. Again the combinations that use AM1 for the highest level, LQ:LQ:MM and LQ:MM:MM, are in error over 120 kcal/mol and are not worth considering.

One notes that the results of QM:MM:MM depend sensitively on the choice of the charges used in the Amber calculation. The use of the Mulliken charges, for instance,

Table 3. One-, Two-, and Three-Layered ONIOM Calculations for the Deprotonation Energy (in kcal/mol) of $NH_3^+{-}C^nBuH{-}CO{-}NH{-}CH_2{-}Co{-}NH{-}CH^nBu{-}COO^-$ System with Partition A

combination[a]	*E*(high/ model)	*E*(med/ mid)	*E*(med/ model)	*E*(low/ real)	*E*(low/ mid)	*E*(ONIOM)	*S*(med/ mid-model)	*S*(low/ real-mid)
HQ:HQ:HQ	241.70	318.06	241.70	322.09	318.06	322.09	76.36	4.04
HQ:HQ:LQ	245.04	318.67	245.04	248.06	242.11	324.62	73.63	5.95
HQ:HQ:MM	243.47	321.57	243.47	108.56	108.84	321.29	78.10	−0.28
HQ:HQ:MM (EE)		332.66		108.53	118.89	322.30		−10.36
HQ:LQ:LQ	243.61	236.53	172.43	244.03	236.53	315.21	64.10	7.50
HQ:LQ:MM	237.81	235.91	166.04	110.17	110.43	307.41	69.86	−0.27
HQ:MM:MM	238.50	106.72	70.29	106.68	106.72	274.89	36.43	−0.04
HQ:MM:MM(EE)	258.13		84.27	109.22		283.09		
LQ:LQ:LQ	158.24	232.51	158.24	238.08	232.51	238.08	74.27	5.57
LQ:LQ:MM	156.63	234.04	156.63	122.22	122.67	233.59	77.41	−0.46
LQ:MM:MM	158.99	110.12	69.33	110.17	110.12	199.82	40.79	0.05

[a] See footnote *a* of Table 1.

in QM:MM:MM increases the error from −47 kcal/mol with RESP charges to −86 kcal/mol. This implies that the results will also depend sensitively on how to arbitrarily "scale" the near-border charges for the QM-MM interaction, because the problematic QM-MM boundary is very close to the reaction center. This is also reflected in the comparison of ME and EE schemes as well as RESP and Mulliken charges in the QM:QM:MM method. Here the differences among these choices are small because the problematic QM-MM boundary is far away from the action region, and the choice of EE vs ME or the choice of charges has small effects on the calculated deprotonation energies.

The performance of different methods can be evaluated more systematically by examining the *S*-values for the deprotonation energy. Looking at the *S*-value between the middle and model systems, *S*(med/mid-model) in Table 2, one sees that the target *S*-value is 3.8−4.4 kcal/mol for the HQ level, derived from HQ:HQ:HQ, HQ:HQ:LQ, and HQ:HQ:MM calculations. The *S*-value for LQ is in the range of 2.0−2.3 kcal/mol from a variety of combinations involving LQ for the middle and model system. This implies that the LQ in this middle system introduces an error of 2−4 kcal/mol in the deprotonation energy, suggesting that AM1 in the middle system is a good choice of the method. The *S*-value for MM is −2.8 to −1.3 kcal/mol with the RESP charges and −36 to −38 kcal/mol with the Mulliken charges. MM with RESP is not bad at least in this region but is very sensitive to the choice of charges.

Now we switch our attention to the *S*-value between the real and middle systems, *S*(low/real-mid). The target *S*-value for the deprotonation energy is 77.6 kcal/mol for the HQ level from the HQ:HQ:HQ calculation. The *S*-value for LQ is in the range of 69−78 kcal/mol from HQ:HQ:LQ to LQ:LQ:LQ. LQ is not bad for the real system but is too costly. The *S*-value for MM is 38−42 kcal mol/mol with RESP and 35−41 kcal/mol with Mulliken. Here we find a large source of error in deprotonation energy by using the Amber method for the real system. The present results clearly show that MM, even used as the low-level method in the outermost region of a large system, can introduce a substantial error in the energetics.

3. Deprotonation Energy in Partition A. Partition scheme A includes all the polar groups in the model and middle systems. From RESP or Mulliken analysis, it appears that there is a charge transfer of about half an electron from the CO_2 to the NH_3 group. It is therefore likely that the small model in partition scheme B, with an integer charge, does not describe the deprotonation accurately, while AM1 is not able to recover the charge transfer quantitatively. Therefore partition A would be a more desired choice of the partition for a polar system. In this partition, all the combinations including large HQ calculations up to the middle system, HQ:HQ:LQ, HQ:HQ:MM, and HQ:HQ:MM(EE), reproduce the target results with an error of 2−3 kcal/mol, which is an excellent result. However, the large HQ calculation is relatively expensive. On the other hand, the combination with small HQ calculation, HQ:MM:MM-ME, gives a large error of 47 kcal/mol. Electronic embedding, HQ:MM:MM-EE, improves the results by 8 kcal/mol but still has a large error of 39 kcal/mol. The method with LQ as the highest method is again totally wrong, with an error of 90 kcal/mol or so. The three-layer ONIOM method, HQ:LQ:MM-ME, performs well, with the error of 15 kcal/mol. Considering the estimated cost of the method, this is a remarkable achievement. As discussed above, the three-layer HQ:LQ:MM-ME combination is substantially less expensive than the HQ:MM:MM-EE method and slightly more expensive than the HQ:MM:MM-ME method, but the error of the three-layer method is substantially smaller than either of the HQ:MM:MM-EE or -ME methods.

4. Deprotonation Energy in Partition C. We tested another partition scheme C, in which the small model consists of only H_3N^+L, and the middle system is equal to the model system in partition A. As discussed before, this is the minimal model and minimal middle system trying to push the ONIOM method to the extreme. Previously such minimal models worked reasonably well for C−H and C−C bond energies of organic molecules such as RR′R″C−H RR′R′C−CRR′R″,[20,21] and here we examine briefly whether such extreme models work for very polar complexes such as the present system.

The results are shown in Table 4. Comparing Tables 2 and 4, one can see clearly that the errors in deprotonation energy in partition C are significantly larger than in partition B. Only those combinations involving solely HQ and LQ (not MM) performed reasonably well. HQ:HQ:LQ here is

Table 4. One-, Two-, and Three-Layered ONIOM Calculations for the Deprotonation Energy (in kcal/mol) of $NH_3^+-C^nBuH-CO-NH-CH_2-CO-NH-CH^nBu-COO^-$ System with Partition C

combination[a]	*E*(high/ model)	*E*(med/ mid)	*E*(med/ model)	*E*(low/ real)	*E*(low/ mid)	*E*(ONIOM)	*S*(med/ mid-model)	*S*(low/ real-mid)
HQ:HQ:HQ	217.75	241.70	217.75	322.09	241.70	322.09	23.95	80.39
HQ:HQ:LQ	218.14	243.61	218.14	244.03	172.43	315.21	25.46	71.60
HQ:HQ:MM	217.62	236.02	217.62	28.35	28.53	235.83	18.40	−0.18
HQ:LQ:LQ	218.16	158.57	159.13	238.30	158.57	297.32	−0.56	79.73
HQ:LQ:MM	218.19	159.07	159.18	30.75	26.11	222.71	−0.11	4.63
HQ:MM:MM	217.94	28.93	5.58	28.80	28.93	241.16	23.35	−0.13
LQ:LQ:LQ	158.84	158.24	158.84	238.08	158.24	238.08	−0.60	79.84
LQ:LQ:MM	158.95	158.76	158.95	30.85	26.22	163.40	−0.18	4.64
LQ:MM:MM	158.89	28.97	5.64	28.84	28.97	182.10	23.33	−0.13

[a] See footnote *a* of Table 1.

identical to HQ:LQ:LQ in partition B, and HQ:LQ:LQ here has an error of 22 kcal/mol, still noticeably better than any combinations involving MM. It appears that the MM method even at the outermost layer is not able to approximate the HQ method. The *S*-value between the real and middle systems, *S*(low/real-mid), is 80 kcal/mol in the target HQ method and 72−80 kcal/mol in the LQ method but is 0−5 kcal/mol in MM, introducing a huge error in the deprotonation energy. Obviously the minimal models are too small for the MM method to handle this very polar system.

IV. Conclusions

We have systematically tested all possible three- and two-layer ONIOM combinations of high-level QM (HQ=B3LYP/6-31G*), low-level QM (LQ=AM1), and MM (Amber) for the deprotonation energy and structure of a test molecule, an ionic form of a peptide. Depending on the partition, we find the errors introduced in the ONIOM approximation, in comparison with the target HQ (or HQ:HQ:HQ) calculation, generally increase in the following order:

$$\begin{aligned}&\text{HQ:HQ:HQ (target)} < \text{HQ:HQ:LQ} < \text{HQ:LQ:LQ} < \\ &\quad \text{HQ:HQ:MM} < \text{HQ:LQ:MM} \ll \text{HQ:MM:MM} \ll \\ &\qquad \text{LQ:LQ:LQ} < \text{LQ:LQ:MM} < \text{LQ:MM:MM}\end{aligned}$$

For realistic systems, the HQ calculation for the middle system and the LQ calculation for the real system can be expensive. The AM1 as the highest level (semiempirical QM/MM) has too large an error to be useful, while also a QM-MM boundary close to the region of the action, in ONIOM-(QM:MM:MM), produces large errors.

We recommend the three-layer ONIOM(HQ:LQ:MM) method as the best tradeoff between accuracy and computational cost. It treats the innermost active center (small model) with a high-level quantum mechanical (HQ) method. The active center plus nearby environment (middle system) is handled with a low-level quantum mechanical (LQ) method, which provides a proper quantum mechanical description of the exchange as well as charge−charge interaction, can polarize the wave function of the active center, and allows charge-transfer between the active center and the environment. The real system is handled with a molecular mechanics (MM) method. The problematic boundary between the MM layer and the QM layer is sufficiently distant from the active center where the bond breaking and forming takes place, and the intrinsically arbitrary choice of charge assignment and scaling does not affect the outcome of the calculations.

Despite the increasing computer power and further development of theoretical methods, it will not be to fully model very large molecular systems accurately with a single method. Therefore the use of the hybrid methods will remain essential. As we have shown in this case study, the ability of our ONIOM method to partition the system in more than two layers offers a valuable feature to the repertoire of hybrid methods in general.

Acknowledgment. The authors are grateful to Dr. Stephan Irle for very helpful discussions and assistance. This work was supported in part by a grant (CHE-0209660) from the National Science Foundation. Computer resources were provided in part by the Air Force Office of Scientific Research DURIP grant (FA9550-04-1-0321) as well as by the Cherry Emerson Center for Scientific Computation.

Supporting Information Available: XYZ coordinates (in Å) for the optimized geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(2) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.

(3) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(4) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357−19363.

(5) Froese, R. D. J.; Morokuma, K. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Kollman, P. A., Clark, T., Schaefer, H. F., III, Gasteiger, J., Schreiner, P. R., Eds.; Wiley: Chichester, 1998; Vol. 2, pp 1244−1257.

(6) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct. (THEOCHEM)* **1999**, *461−462*, 1−21.

(7) Morokuma, K.; Musaev, D. G.; Vreven, T.; Basch, H.; Torrent, M.; Khoroshun, D. V. *IBM J. Res. Dev.* **2001**, *45*, 367−395.

(8) Vreven, T.; Morokuma, K. *Annual Reports in Computational Chemistry*; in press.

(9) Vreven, T.; Frisch, M. J.; Kudin, K. N.; Schlegel, H. B.; Morokuma, K. *Mol. Phys.* **2006**, *104*, 701−714.

(10) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.

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

(12) Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125−132.

(13) Morokuma, K. *Philos. Trans. R. Soc. London, Ser. A* **2002**, *360*, 1149−1164.

(14) Morokuma, K. *Bull. Kor. Chem. Soc.* **2003**, *24*, 797−801.

(15) Svensson, M.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654−3661.

(16) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.

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

(18) Julian, R. R.; Jarrold, M. F. *J. Phys. Chem. A* **2004**, *108*, 10861−10864.

(19) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(20) Vreven, T.; Morokuma, K. *J. Chem. Phys.* **1999**, *111*, 8799−8803.

(21) Froese, R. D. J.; Morokuma, K. *J. Phys. Chem. A* **1999**, *103*, 4580−4586.

CT600135B

J. Chem. Theory Comput. **2006,** *2,* 1325−1334

JCTC Journal of Chemical Theory and Computation

First Excited State Properties and Static Hyperpolarizability of Ruthenium(II) Ammine Complexes

Talgat M. Inerbaev,* Rodion V. Belosludov, Hiroshi Mizuseki, Masae Takahashi, and Yoshiyuki Kawazoe

Institute for Materials Research, Tohoku University, 2-1-1 Katahira, Aoba-ku, 980-8577 Sendai, Japan

Received August 4, 2005

Abstract: First principles calculations were used to study the electronic excitation energies (E), transition dipole moments (μ), and difference of dipole moments between ground and excited states ($\Delta\mu$) for low-lying singlets of the series of ruthenium(II) ammine complexes. Both cases of the gas phase and the acetonitrile solution were investigated in order to explain the discrepancy between the recent experimental and theoretical results and to develop the optimal way of estimation for the first static hyperpolarizability in the framework of a two-state model introduced by Oudar and Chemla. The present calculations reveal that the effect of solvent on the electronic properties of investigated compounds is not only the change of the excitation energy but also the increasing of ground-state molecular polarization and intensification of metal-to-ligand intramolecular charge transfer for electronic excitations. These effects lead to increasing of the values of $\Delta\mu$ and ground-state dipole moment μ_g in solution as compared with the gas-phase ones. The proposed theoretical approach gives good agreement with experiment and allows one to apply it for designing a new perspective nonlinear optical active organometallics.

Introduction

The design of new molecular materials with large second-order nonlinear optical properties (NLO) is currently the subject of extensive investigations by theoretical and experimental methods, since they are expected to be used as frequency doubler, electro-optic modulator, and photorefractive media.[1−3] Within this field, an increasing amount of attention has recently been paid to organotransition-metal complexes which offer possibilities for combination of NLO effects with many superior characteristics, such as ultrafast response times, lower dielectric constants, better processability as thin-film devices, and enhanced nonresonant NLO responses.[3] Besides organic molecules, organometallic structures are intriguing candidates as second-order NLO chromophores since[4] (1) they show very strong absorption bands (metal-to-ligand charge transfer and ligand-to-metal charge transfer) that are related to high transition dipole moments and low transitions energies; (2) organometallic and coordination compounds are often strong oxidizing or redoxing agents, since the metal centers may be electron rich or poor depending on their oxidation state and ligand environment [Thus, the metal center may be an extremely strong donor or acceptor.]; and (3) metals can be used to fine-tune the electronic properties of organic fragments.

A lot of organometallic and coordination compounds with a variety of metals, ligands, and molecular configurations were designed and synthesized. Their structures and NLO properties of the second and third order were studied.[5−8] Usually, chromophores possessing large molecular hyperpolarizability β contain donor (D) and acceptor (A) groups linked through a π-backbone. The NLO properties of such polarizable dipolar compounds are caused by intense, low-energy $D(\pi) \rightarrow A(\pi^*)$ intramolecular charge-transfer (ICT) transitions. In the case of unidirectional ICT transition the frequency-independent nonresonance first hyperpolarizability β_0 can be described by a simple two-state model (TSM) introduced by Oudar and Chemla[9]

---

* Corresponding author phone: +81-(0)22-215-2057; fax: +81-(0)22-215-2052; e-mail: talgat@imr.edu.

10.1021/ct050193m CCC: $33.50

$$\beta_0 = \frac{3\Delta\mu(\mu)^2}{E^2} \qquad (1)$$

where μ is the transition dipole moment between ground and excited states, $\Delta\mu$ is the dipole moment change, and E is the energy gap between ground and excited states.[10]

A number of ruthenium(II) NLO active complexes were studied in detail using hyper-Rayleigh scattering and Stark spectroscopy techniques.[11−14] There were established that NLO properties of these compounds are caused by strong metal-to-ligand charge transfer (MLCT). Moreover, the strong solvatochromic effect was demonstrated and the validity of TSM was proved.[15] In the present paper we will study ruthenium(II) ammine complexes **1**−**4** presented in Figure 1, to propose the reliable theoretical description of NLO properties of organometallic and coordination compounds.

Theoretical quantum investigations are very helpful to study NLO chromophores. For example, a semiempirical computational study of (4-Y-pyridine)Cr(CO)$_5$ and (4-Y-stilbazole)Cr(CO)$_5$ was carried out by Kanis et al.[16] The computed molecular hyperpolarizabilities were in good agreement with the Electric-Field-Induced-Second-Harmonic-Generation-determined experimental values. For ruthenium complexes such semiempirical calculations demonstrated the poor quantitative agreement with experimental data, and the best accuracy for $\Delta\mu$ and μ calculations was achieved about 25%.[11] The most recent experimental and theoretical study of the effects of polyene chain extension on the NLO properties of ruthenium(II) pyridyl complexes was done.[13] The study based on B3P86[17,18] hybrid functional with the LanL2DZ basis set in the gas phase gave rise to only a qualitative description of experimentally observed trends. Moreover, these calculations predicted that the change in dipole moment caused by the low-energy MLCT decreases with the polyene bridge length growth that is opposite to the experimentally observed tendency. Molecular first hyperpolarizabilities could be calculated by the finite-field (FF) method that involves a double numerical differentiation of the dipole moment with respect to the applied electric field, but the FF approach does not provide a microscopic interpretation of the obtained results.

In the present study, we have examined a time-dependent (TD) B3P86 functional approach to investigate the excited-state properties E, $\Delta\mu$, and μ of ruthenium(II) ammine complexes **1**−**4** (see Figure 1). The solvent effect on these characteristics has also been studied in order to understand the properties of investigated organometallic chromophores in details. Dependences on the choice of both the structure optimization method and the wave function basis set for TD-B3P86 analysis were also studied.

Computation Details

All theoretical calculations were carried out by the Gaussian 03 program.[19] The geometry optimization of investigated structures was performed in the gas phase using HF, B3P86, and MP2 methods with the LanL2DZ basis set and by HF and B3P86 levels using a larger basis set, including LanL2DZ for Ru atom and 6-311++g(d,p) for N, O, C, and H atoms.

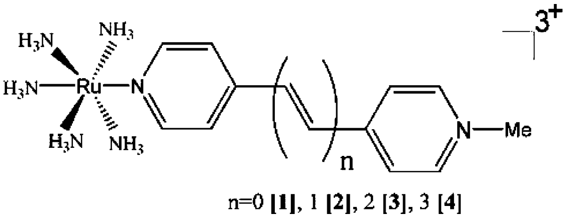


Figure 1. Chemical structures of cations investigated.

The solvent effect was estimated by performing calculations within the framework of the integral equation formalism polarized continuum model (PCM) developed by Tomasi.[20] Solute structure optimization was carried out by the HF/LanL2DZ approach. Electronic transitions were calculated by means of the TD-B3P86 method for all optimized structures using the same basis set that one used for optimization calculations. The HF/LanL2DZ gas-phase optimized structures were also analyzed by the TD-B3P86 theory using the 6-311++g(d,p)/LanL2DZ combined basis set. These data were compared with the data calculated for the HF/6-311++g(d,p)/LanL2DZ optimized structures in order to examine dependence of electronic excitation properties on the choice of the basis set for geometry optimization. To estimate the effect of the Ru(NH$_3$)$_5$-ligand distance variation on the electronic excitation properties, structure **2** was optimized in the gas phase employing the CEP-4G basis set on Ru and LanL2DZ on all other atoms. The choice of the B3P86 functional was based on the more accurate TD-B3P86 results provided by this functional when compared to the more widely used B3LYP functional.[21] The differences of dipole moments between the ground and the *n*th excited states were estimated using the FF calculations (±0.001 au) and the TD-B3P86 excitation energies,[22] and the results were compared with data obtained by one particle RhoCI density. Experimental studies of investigated systems were carried out in acetonitrile (MeCN) at room temperature and butyronitrile (PrCN) at 77 K.[11−14] Since at the present time modeling in the PrCN solution is unavailable in the Gaussian 03 program, we restrict ourselves to study systems in MeCN at zero temperature. Only electronic excitation properties were investigated, and for this reason the effect of zero-point energy correction was not estimated. Net atomic charges were calculated using the natural population analysis (NPA) included in the natural bond orbital algorithm proposed by Weinhold and co-workers.[23−25] NPA analysis was carried out for the HF/LanL2DZ optimized structures by the B3P86/LanL2DZ method. NPA for excited states was performed using one particle RhoCI density. The default Gaussian 03 parameters were used in every case.

Results and Discussion

Geometry Optimization. Since up to the present there were not published crystallographic data of any salt containing the considered Ru(II) ammine complexes, we can perform only an indirect accuracy evaluation of the different methods of molecular geometry optimization by comparison of the theoretical estimation of electronic excitations with experimentally measured ones. It should be mentioned that the most important geometry parameter for push−pull chromophores

Table 1. BLA and $Ru(NH_3)_5$-Ligand Distances for Structures **1**−**4** Calculated by Different Theoretical Methods

		6-311++g(d,p)/LanL2DZ		LanL2DZ			
compd	geometry parameter	HF	B3P86	HF	HF/MeCN	B3P86	MP2
1	$Ru(NH_3)_5$-ligand	2.296	2.109	2.278	2.219	2.083	2.161
2	BLA	0.152	0.114	0.141	0.145	0.105	0.115
	$Ru(NH_3)_5$-ligand	2.275	2.105	2.258	2.316	2.083	2.152
3	BLA	0.137	0.095	0.127	0.129	0.089	0.096
	$Ru(NH_3)_5$-ligand	2.256	2.107	2.245	2.248	2.088	2.147
4	BLA	0.131	0.083	0.120	0.113	0.078	0.089
	$Ru(NH_3)_5$-ligand	2.246	2.107	2.237	2.375	2.091	2.144

responsible for their optical properties is the bond length alternation (BLA) defined as the average difference in the length between single and double bonds in the π-conjugated backbone.[26−27] To study dependence on the geometry optimization method, compounds **1**−**4** were optimized in the gas phase at HF, B3P86, and MP2 levels using the LanL2DZ basis set. As it will be shown below, the TD-B3P86 calculations using the HF optimized molecular structures give the best agreement with experiment. For this reason and since the MP2 approach has a very high computation cost, the optimization using the 6-311++g(d,p)/LanL2DZ basis sets combination was carried out only by the HF and the B3P86 methods. The data of calculated BLA and $Ru(NH_3)_5$-ligand bond length values of compounds **1**−**4** are summarized in Table 1.

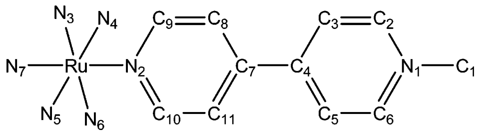
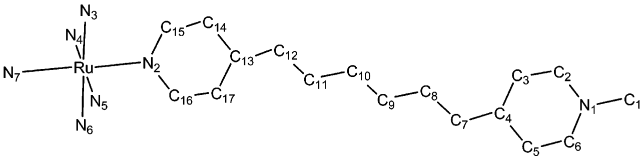
Most of the results presented in this study were performed using the structures optimized in the gas phase. Solution geometry optimization was carried out only by the HF/LanL2DZ approach, and in this case the optimization procedure with used parameters did not converge completely. The values of maximal and residual mean square displacements converged to values less than the corresponding default thresholds, but the same values of forces did not. All nonconverged forces are between Ru and surrounding it are nitrogen atoms and only partial optimization with converged values of displacement and significantly large values of ruthenium−nitrogen interatomic forces. Therefore, the solvent effect on the structural properties of investigated organometallics was not studied in detail such as for the gas-phase optimized geometries. The selected interatomic distances for structures **1** and **4** obtained for all considered cases are presented in Table 2. It can be pointed out that optimization in solution does not significantly affect the geometry of studied molecular complexes. Full geometry optimization of transition-metal complexes in solution using the PCM model needs additional contrivances, and this problem is beyond the boundaries of this study.

Electronic Excitations and NLO Properties. The values of E, μ, and $\Delta\mu$ for the first excited states were calculated by the TD-B3P86 method and compared with experiment.[11−14] Selected calculation results are plotted in Figure 2. Data obtained for all molecular geometries considered in this study are presented in Table 3. To compare the theoretical and experimental results correctly it is necessary to denote that the MLCT bands of investigated chromophores undergo marked red-shifts with decreasing temperature.[14] The measured changes in E on moving from 298 to 77 K are −0.15 and −0.23 eV for complexes **1** and **2** in the PrCN solution, correspondingly. At the same time, the values of oscillator strength that are directly related to the transition dipole moment μ show very little temperature dependence (≤5%).[11]

Calculations demonstrate that results are sensitive to the method of molecular structure optimization. Namely, for B3P86/LanL2DZ and MP2/LanL2DZ optimized structures the values of E and $\Delta\mu$ are in good agreement with data obtained by Stark spectroscopy (SS) measurements, and an error for values μ are found.[12,13] Using the HF/LanL2DZ optimized structures, E and μ values were obtained in good agreement with the SS experiment, but $\Delta\mu$ values are overestimated. In both cases the values of μ and $\Delta\mu$ calculated for all structures demonstrate similar functional behavior. The value of E is a steadily decreasing function of n for B3P86/LanL2DZ and MP2/LanL2DZ optimized structures, while the experimental one is approximately constant for $n > 1$. From this point of view HF/LanL2DZ optimized structures give a better agreement with the experiment.

Calculations using the 6-311++g(d,p)/LanL2DZ combined basis set reveal the same qualitative results but with significant quantitative differences. In this case the values of E and $\Delta\mu$ are larger, and μ is smaller than the same values calculated using the LanL2DZ basis set. For the HF optimized structures the calculated E demonstrates perfect agreement with the room-temperature UV−visible absorption (VA) experiment except for the overestimated value for cation **1**. At the same time for the B3P86 optimized structure **1** E is also overestimated, and for other compounds the calculated excitation energies are between SS and VA measured values. These results are not very sensitive to the variation of basis sets used for molecular structure optimization. Data obtained by 6-311++g(d,p)/LanL2DZ basis set TD-B3P86 analysis for HF/LanL2DZ and HF/6-311++g(d,p)/LanL2DZ optimized structures are qualitatively the same with some small quantitative difference. Employing larger basis set leads to an increasing of E and $\Delta\mu$ with a simultaneous decreasing of μ. We expect the same basis set dependence for structures optimized by the MP2 method that will not give a better agreement between experimental and theoretical results. Besides the reducing of the absorption energy there is also the effect of solvent on the electronic properties of investigated complexes. Gas-phase calculations predict that the change of dipole moment $\Delta\mu$ decreases with the polyene chain length growth. However, the study of these

Table 2. Selected Interatomic Distances (Å) for Structures **1**(**a**) and **4**(**b**) Calculated by Different Theoretical Methods[a]

bond	B3P86/6-311++g(d,p)/ LanL2DZ	HF/6-311++g(d,p)/ LanL2DZ	B3P86/ LanL2DZ	MP2/ LanL2DZ	HF/ LanL2DZ	HF/ LanL2DZ/MeCN
						
$\mathrm{C_1{-}N_1}$	1.480	1.491	1.495	1.530	1.507	1.500
$\mathrm{C_2{-}N_1}$	1.347	1.334	1.362	1.383	1.348	1.350
$\mathrm{C_6{-}N_1}$	1.347	1.334	1.363	1.383	1.348	1.349
$\mathrm{C_2{-}C_3}$	1.379	1.374	1.391	1.415	1.383	1.384
$\mathrm{C_5{-}C_6}$	1.379	1.374	1.390	1.414	1.383	1.384
$\mathrm{C_3{-}C_4}$	1.399	1.391	1.412	1.431	1.401	1.406
$\mathrm{C_4{-}C_5}$	1.399	1.391	1.412	1.431	1.401	1.407
$\mathrm{C_4{-}C_7}$	1.480	1.496	1.484	1.505	1.494	1.517
$\mathrm{C_7{-}C_8}$	1.397	1.389	1.411	1.431	1.399	1.402
$\mathrm{C_7{-}C_{11}}$	1.397	1.389	1.411	1.431	1.399	1.400
$\mathrm{C_8{-}C_9}$	1.384	1.381	1.395	1.419	1.390	1.415
$\mathrm{C_{10}{-}C_{11}}$	1.384	1.381	1.395	1.419	1.390	1.409
$\mathrm{N_2{-}C_9}$	1.350	1.330	1.371	1.390	1.345	1.342
$\mathrm{N_2{-}C_{10}}$	1.350	1.330	1.371	1.390	1.346	1.351
$\mathrm{Ru{-}N_2}$	2.108	2.296	2.083	2.161	2.278	2.219
$\mathrm{Ru{-}N_3}$	2.177	2.269	2.176	2.222	2.270	2.241
$\mathrm{Ru{-}N_4}$	2.177	2.269	2.177	2.222	2.270	2.241
$\mathrm{Ru{-}N_5}$	2.176	2.268	2.176	2.222	2.269	2.241
$\mathrm{Ru{-}N_6}$	2.176	2.268	2.177	2.222	2.269	2.241
$\mathrm{Ru{-}N_7}$	2.195	2.265	2.194	2.233	2.266	2.263
dihedral angle	38.7	47.7	32.7	43.1	42.2	33.7
	(b) 					
$\mathrm{C_1{-}N_1}$	1.473	1.481	1.487	1.521	1.496	1.481
$\mathrm{C_2{-}N_1}$	1.353	1.334	1.368	1.382	1.349	1.358
$\mathrm{C_6{-}N_1}$	1.347	1.332	1.364	1.386	1.354	1.357
$\mathrm{C_2{-}C_3}$	1.370	1.370	1.383	1.411	1.373	1.372
$\mathrm{C_5{-}C_6}$	1.374	1.363	1.386	1.408	1.378	1.371
$\mathrm{C_3{-}C_4}$	1.408	1.396	1.421	1.434	1.406	1.416
$\mathrm{C_4{-}C_5}$	1.405	1.403	1.418	1.438	1.412	1.418
$\mathrm{C_4{-}C_7}$	1.447	1.467	1.454	1.482	1.467	1.451
$\mathrm{C_7{-}C_8}$	1.354	1.333	1.368	1.385	1.345	1.355
$\mathrm{C_8{-}C_9}$	1.432	1.459	1.441	1.469	1.462	1.454
$\mathrm{C_9{-}C_{10}}$	1.358	1.333	1.372	1.389	1.345	1.353
$\mathrm{C_{10}{-}C_{11}}$	1.431	1.459	1.441	1.468	1.462	1.465
$\mathrm{C_{11}{-}C_{12}}$	1.355	1.333	1.369	1.386	1.345	1.348
$\mathrm{C_{12}{-}C_{13}}$	1.446	1.468	1.453	1.481	1.469	1.486
$\mathrm{C_{13}{-}C_{14}}$	1.406	1.399	1.419	1.438	1.408	1.401
$\mathrm{C_{13}{-}C_{17}}$	1.407	1.396	1.420	1.436	1.406	1.405
$\mathrm{C_{14}{-}C_{15}}$	1.379	1.373	1.390	1.415	1.383	1.394
$\mathrm{C_{16}{-}C_{17}}$	1.378	1.375	1.390	1.414	1.384	1.394
$\mathrm{N_2{-}C_{15}}$	1.351	1.336	1.371	1.392	1.351	1.347
$\mathrm{N_2{-}C_{16}}$	1.354	1.332	1.374	1.390	1.348	1.352
$\mathrm{Ru{-}N_2}$	2.106	2.246	2.091	2.144	2.236	2.375
$\mathrm{Ru{-}N_3}$	2.172	2.267	2.173	2.220	2.269	2.248
$\mathrm{Ru{-}N_4}$	2.172	2.267	2.173	2.220	2.269	2.273
$\mathrm{Ru{-}N_5}$	2.171	2.267	2.174	2.220	2.268	2.273
$\mathrm{Ru{-}N_6}$	2.171	2.267	2.174	2.220	2.268	2.273
$\mathrm{Ru{-}N_7}$	2.203	2.278	2.200	2.241	2.276	2.276

[a] Atoms in the compounds are enumerates as it is shown in the structures above the corresponding section. All geometries are calculated in the gas phase excluding those presented in the last column.

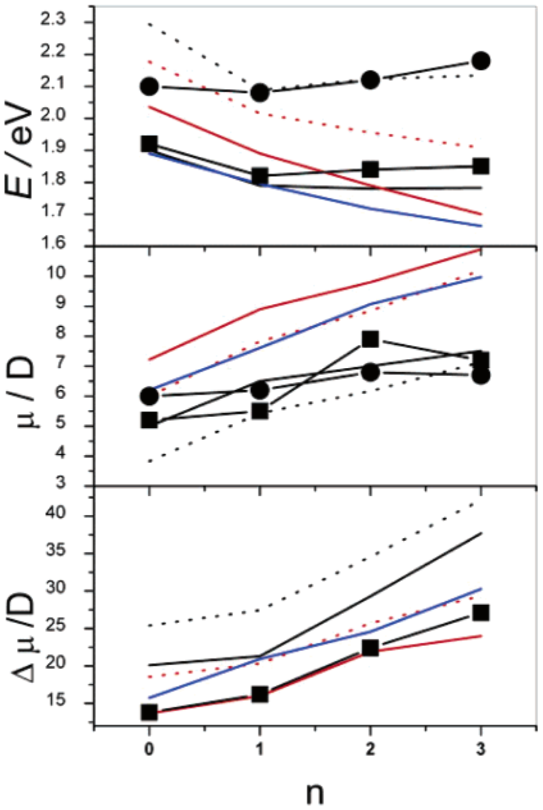


Figure 2. Values of $\Delta\mu$, μ, and E: experimental[11−14] (■ − Stark spectroscopy, ● − UV−visible absorption) and calculated for HF (black lines), B3P86 (red lines), and MP2 (blue line) optimized structures. Solid and dotted lines correspond to molecular geometries obtained by LanL2DZ and 6-311++g-(d,p)/LanL2DZ basis sets, respectively.

compounds in solution gives a steadily increasing of this value that is in good agreement with experiment.[12]

The variation in results obtained using different methods of structure optimization is caused by the distinction between the BLA values and the $Ru(NH_3)_5$-ligand distances. The former effect is well-known,[26,27] and to clarify the influence of the latter one on the excited-state properties structure **2** was optimized by the HF approach using the CEP-4G basis set on the Ru and the LanL2DZ one on all other atoms. Obtained molecular geometry demonstrates that all bond lengths are identical with the HF/LanL2DZ optimized structure but that the $Ru(NH_3)_5$-ligand distance is 0.015 Å longer. At the same time the difference between $Ru(NH_3)_5$-ligand distances for the HF/LanL2DZ/6-311++g(d,p) and HF/CEP-4G/LanL2DZ optimized structures coincide with an accuracy of 0.002 Å and have a different BLA. Comparing the results of TD-B3P86/6-311++g(d,p)/LanL2DZ analysis for these structures (Table 3) it is seen that the effect of the $Ru(NH_3)_5$-ligand distance variation is smaller than the BLA variation but not negligible. In the considered case elongation of the $Ru(NH_3)_5$-ligand bond has the same effect as the BLA increasing. As a result, the discrepancy between the experimental and the theoretical data obtained by employing the different method optimized geometries is caused by a variation of both BLA and $Ru(NH_3)_5$-ligand distances.

From Table 1 it is seen that the BLA values for all compounds are quite the same, but the $Ru(NH_3)_5$-ligand distances are varied for structures **1**, **2**, and **4**. The comparative analysis of TD-B3P86 results obtained at the HF/LanL2DZ method using optimization in the gas phase and the MeCN solution (Table 3) indicates that the calculated values of E, μ, and $\Delta\mu$ are not systematically improved by including the solvent effect in the optimization procedure. Since for solute optimized structures $Ru(NH_3)_5$-ligand distances are not equilibrium, we can say only about the qualitative effect of this geometry parameter on investigated properties. Thus, the accurate estimation of the $Ru(NH_3)_5$-ligand distance could have an affect on the electronic excitation properties of considered chromophores.

Values of ground-state dipole moments μ_g are also strongly influenced by solvent. The absolute values of μ_g of consid-

Table 3. Experimental[11−14] and Theoretical Values of E, μ, and $\Delta\mu$ Obtained by TD-B3P86 Analysis Using Different Basis Sets[a]

salt	E	μ	$\Delta\mu$	optimization method
1	2.1	6.0		MeCN (exp.)
	1.92	5.2	13.8	PrCN (exp.)
	1.90	5.0	20.1	HF/LanL2DZ
	1.94	4.27	23.41	HF/LanL2DZ[b]
	2.17	4.43	23.10	HF[c]
	2.29	3.83	25.40	HF/6-311++g(d,p) LanL2DZ
	2.04	7.22	13.67	B3P86/LanL2DZ
	2.18	5.99	18.55	B3P86 /6-311++g(d,p)/LanL2DZ
	1.89	6.20	15.75	MP2/LanL2DZ
2	2.08	6.2		MeCN (exp.)
	1.82	5.5	16.2	PrCN (exp.)
	1.79	6.5	21.3	HF/LanL2DZ
	1.88	5.43	26.77	HF/LanL2DZ[b]
	2.03	5.89	25.44	HF[c]
	2.04	5.72	26.09	HF[d]
	2.09	5.43	27.43	HF/6-311++g(d,p)/LanL2DZ
	1.89	8.90	16.00	B3P86/LanL2DZ
	2.02	7.83	20.31	B3P86 /6-311++g(d,p)/LanL2DZ
	1.80	7.61	20.94	MP2/LanL2DZ
3	2.12	6.8		MeCN (exp.)
	1.84	7.9	22.4	PrCN (exp.)
	1.78	7.0	29.3	HF/LanL2DZ
	1.79	6.89	29.75	HF/LanL2DZ[b]
	2.04	6.62	32.82	HF[c]
	2.12	6.17	34.57	HF/6-311++g(d,p)/LanL2DZ
	1.79	9.80	21.91	B3P86/LanL2DZ
	1.96	8.84	25.69	B3P86 /6-311++g(d,p)/LanL2DZ
	1.72	9.07	24.57	MP2/LanL2DZ
4	2.18	6.7		MeCN (exp.)
	1.85	7.2	27.1	PrCN (exp.)
	1.78	7.51	37.68	HF/LanL2DZ
	1.85	5.98	34.44	HF/LanL2DZ[b]
	2.06	7.63	43.08	HF[c]
	2.13	7.13	42.15	HF/6-311++g(d,p)/LanL2DZ
	1.70	10.90	24.02	B3P86/LanL2DZ
	1.91	10.21	29.35	B3P86 /6-311++g(d,p)/LanL2DZ
	1.66	9.97	30.27	MP2/LanL2DZ

[a] In all cases the used basis set is the same as that employed for the geometry optimization one, except as otherwise noted. [b] Optimized by the HF/LanL2DZ method in the MeCN solution. [c] Electronic transitions were calculated by the 6-311++g(d,p)/LanL2DZ basis set using the HF/LanL2DZ optimized structures. [d] Same as *c* for the HF/CEP-4G/LanL2DZ optimized structure.

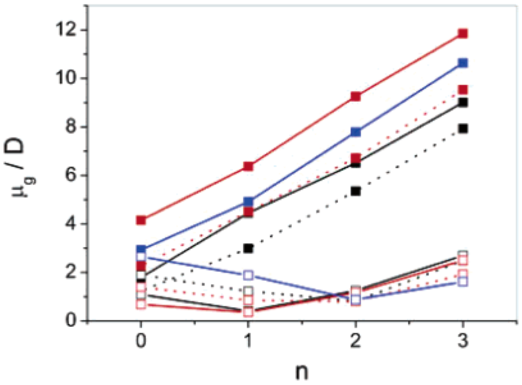


Figure 3. Calculated values of the ground-state dipole moment μ_g for investigated compounds in the gas phase (open symbols) and the MeCN solution (solid symbols). μ_g were estimated by the B3P86 method for the HF (black), B3P86 (red), and MP2 (blue) optimized structures. Solid and dotted lines denote data obtained by LanL2DZ and 6-311++g-(d,p)/LanL2DZ basis sets, respectively.

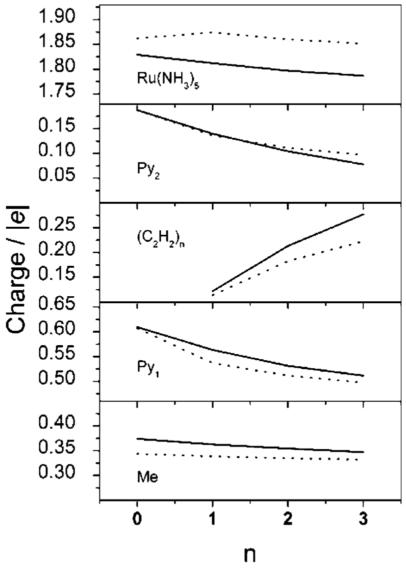


Figure 4. B3P86/LanL2DZ NPA calculated charges in the gas phase (solid lines) and the MeCN solution (dotted lines) for the chemical groups constituting compounds **1**−**4**.

ered complexes in solution are larger that than in the gas phase and also depend on both the way of molecular geometry optimization and used basis set. B3P86 calculations of μ_g for structures optimized by different methods are summarized in Figure 3. NPA results for ground-state charge distribution in the gas phase and the MeCN solution are presented in Figure 4 and illustrate MLCT in considered chromophores. We have also compared results for $\Delta\mu$ calculated by FF and one particle RhoCI density approach. The results of the B3P86 calculations of $\Delta\mu$ by one particle RhoCI density method are presented in Figure 5. For the B3P86 optimized structures in the gas phase the results obtained in the present work and published in a paper (see ref 13) are quite the same. The calculated $\Delta\mu(n)$ in MeCN solution demonstrates the same tendency of change like the experimental one but is approximately two times overestimated, and the FF calculation method is more accurate.

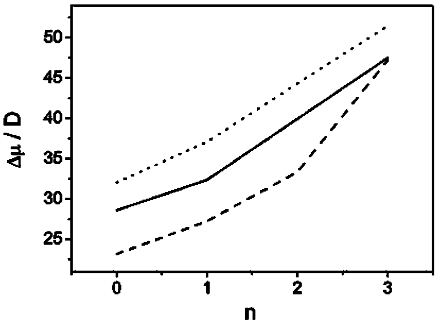


Figure 5. Difference between ground and excited states dipole moments $\Delta\mu$ calculated in MeCN solute by one particle RhoCI density for structures optimized by HF/LanL2DZ (solid), HF/6-311++g(d,p)/LanL2DZ (dotted), and B3P86/LanL2DZ methods (dashed). TD-B3P86 analysis in each case was performed using the same basis set that was used for structure optimization.

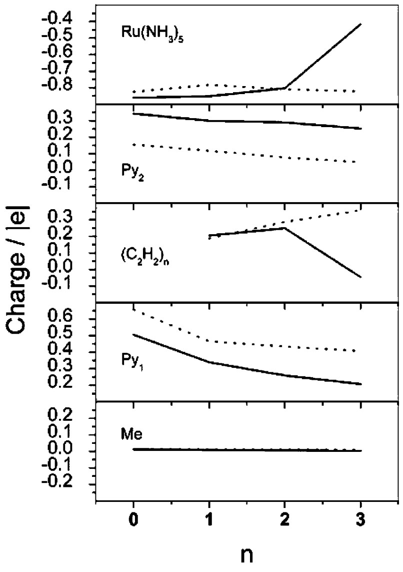


Figure 6. B3P86/LanL2DZ NPA calculated ground-to-excited-state ICT in the gas phase (solid lines) and the MeCN solution (dotted lines) for the chemical groups constituting compounds Me-Py_1-$(C_2H_2)_n$-Py_2-$Ru(NH_3)_5$ **1**−**4**. Py_1 and Py_2 denote the pyridinium rings attached to Me and $Ru(NH_3)_5$ groups, respectively.

The calculations of ground-to-excited-state charge transfer was performed by subtraction of NPA calculated excited-state charge distribution from the same value for the ground state. Obtained results are plotted in Figure 6. It is seen that solvent intensifies the ground-to-excited-state charge transfer on the Py_1 group and reduces it on the Py_2 one for all investigated complexes. For compound **4** this charge transfer is about two times larger in solution for the $Ru(NH_3)_5$ group and significantly larger and has an opposite sign for the polyene chain. To illustrate this, the structures of molecular orbitals for investigated complexes both in the gas phase and the MeCN solution are plotted in Figure 7. For compounds **3** and **4** the MLCT transition is caused by the (HOMO-3, HOMO) → LUMO transition in the gas phase and the HOMO → LUMO one in solution, while for structures **1** and **2** the HOMO → LUMO transition takes place in both media. For structures **1** and **2** HOMO and LUMO are

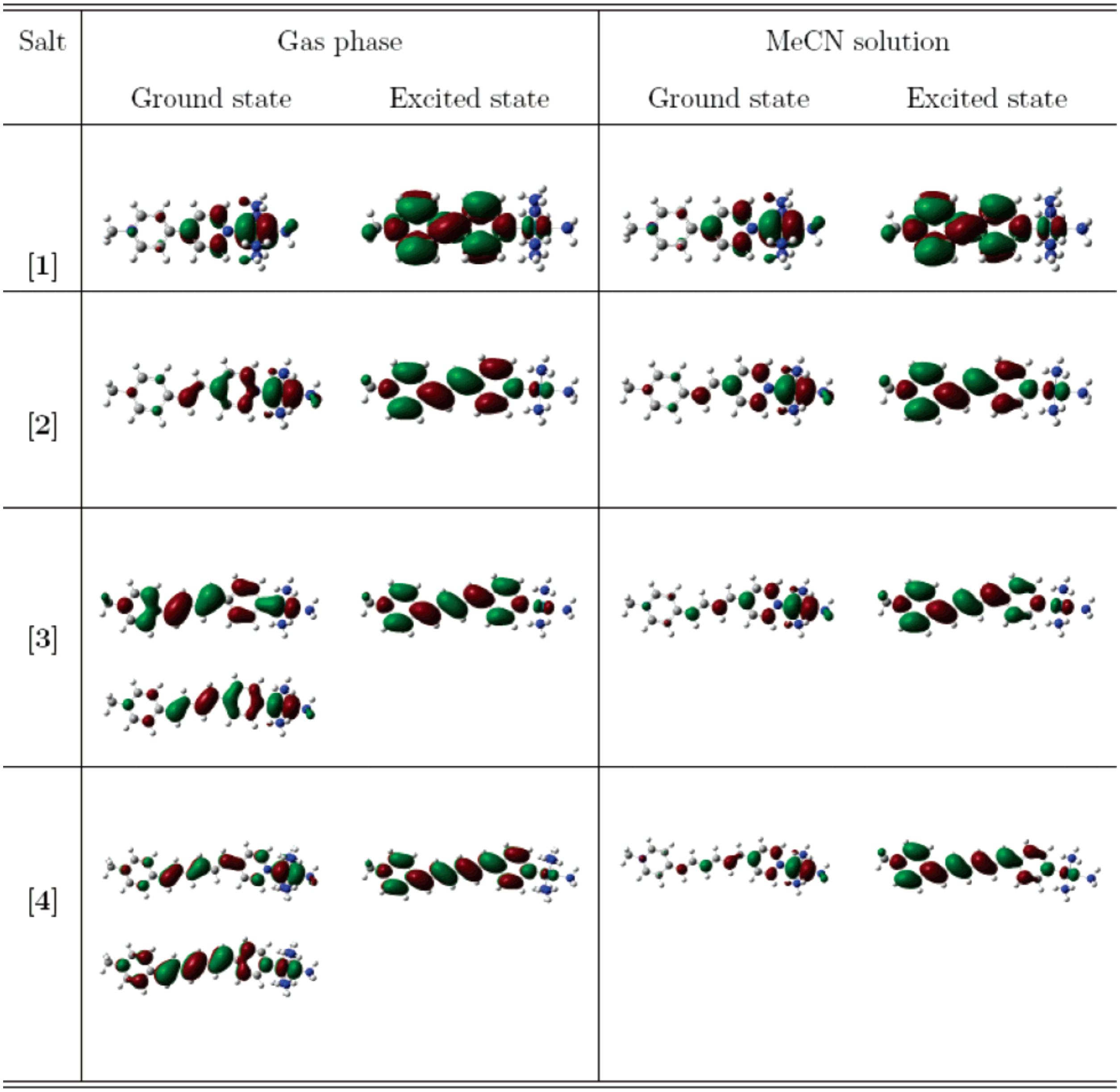


Figure 7. Electron density contours calculated for the MOs involved in the ICT transitions of the cations **1**−**4** in both the gas phase and the MeCN solute.

quantitatively the same and HOMO presents a larger electron density around the metal group, and the LUMO mostly lies on the acceptor part of the molecule in both the gas phase and the MeCN solution (see Figure 4). The same picture is for solute compounds **3** and **4**, but in the gas phase the HOMO-3 and HOMO spread over the whole molecule. This fact explains the large discrepancy between the gas phase and the solution calculated values of $\Delta\mu$. The interaction with the solvent leads to polarization of molecules and molecular dipole moment increases. As it is shown in Figure 6, ground-to-excited ICT also intensifies for structures **3** and **4** that leads to an increasing of excited-state dipole moment too. These two solvent effects on electronic structures of ground and excited states give significant growth of the $\Delta\mu$ in solution compared to the gas phase.

Values of β_0 calculated by eq 1 for considered cases of different optimization methods and basis sets are plotted in Figure 8. The HF/6-311++g(d,p)/LanL2DZ calculated $\beta_0(n)$ is in the best agreement with experiment. The HF/LanL2DZ optimized structures with the following TD-B3P86/6-311++g(d,p)/LanL2DZ analysis also give very good results for all compounds excluding structure **4**, but even in this case agreement with the experiment it is better than for all other methods. These results demonstrate that excitation properties of compounds investigated are significantly dependent on the molecular geometry. As it was shown above, the structures optimized by the HF, B3P86, and MP2 methods give different relationships between bond lengths. The electron−electron correlations taken into consideration by the MP2 approach give the molecular structures that give the values of E, μ, $\Delta\mu$, and β_0 that are not in good agreement with the experiment. The HF and B3P86 optimized structures are sufficiently close to each other, but details of molecular geometry lead to some differences in electronic excitation properties. The HF optimized structures demonstrate good agreement with the experiment for $E(n)$ and $\Delta\mu(n)$, while

Table 4. Calculated Values of E, μ, $\Delta\mu$, and β_0 for the New Palladium-Based Organometallic Chromophore (Structure **6**) in Comparison with the Ru(II) Ammine Complex with the Same Length of the Polyene Chain (Structure **2**) for the Gas Phase (MeCN Solution)

salt	E(eV)	μ(D)	$\Delta\mu$(D)	β_0 (10^{-30} esu)	basis set
2	2.65 (1.82)	4.89 (6.34)	20.59 (22.19)	82 (315)	LanL2DZ
6	2.75 (2.92)	10.36 (10.63)	8.85 (14.07)	147 (218)	
2	2.94 (2.09)	4.60 (5.43)	21.12 (27.43)	60 (216)	LanL2DZ/6-311++g(d,p)
6	2.72 (2.88)	10.98 (11.56)	7.14 (10.34)	136 (195)	

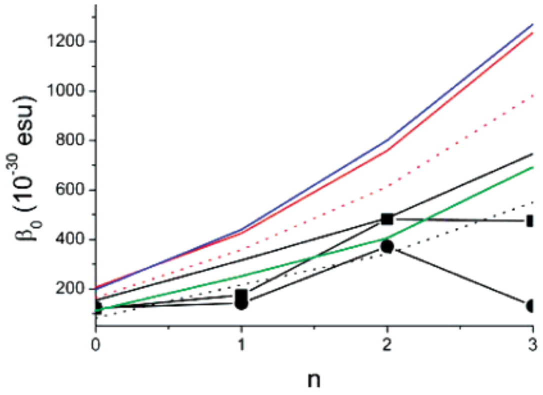


Figure 8. Experimental (■ − Stark spectroscopy, ● − hyper-Rayleigh scattering[12,13]) and estimated by eq 1 values of β_0. Calculated data are presented for HF (black lines), B3P86 (red lines), and MP2 (blue line) optimized structures. Solid and dotted lines denote results obtained by LanL2DZ and 6-311++g(d,p)/LanL2DZ basis sets, correspondingly. Results for HF/LanL2DZ optimized structures that were analyzed by the TD-B3P86/LanL2DZ/6-311++g(d,p) approach are plotted by the green line.

the B3P86 ones give the best agreement for $\mu(n)$. Since in TSM the β_0 is the linear function of $\Delta\mu$ and the quadratic one of μ, the calculation of the value of β_0 by the HF optimized structures is more preferable.

In a short description, the proposed method of a new organometallic NLO active compound design is the HF/6-311++g(d,p)/LanL2DZ geometry optimization with the following TD-B3P86/6-311++g(d,p)/LanL2DZ analysis of electronic excitation properties. The value of $\Delta\mu$ has to be calculated by the FF approach. To keep the computation time the molecular structure optimization may be performed by the HF/LanL2DZ method. As it was shown above, the results obtained by this manner are close to properties predicted for the HF/6-311++g(d,p)/LanL2DZ optimized structures. The proposed approach has a restriction to be taken into account during consideration of the push−pull chromophores with long π-conjugated chains. As it was demonstrated,[28] the density functional theory fails for linear and nonlinear responses in the case of polyene chains with $n > 6$, and some other theoretical approaches have to be used in this situation. In our case use of the B3P86 method is applicable since the maximal length of the investigated chain (n) was smaller than 6. In the next section we use the developed approach to predict a new perspective organometallic chromophore.

Design of a New NLO Active Compound. As an example of using the proposed application for design of an organometallic compound with a high value of β_0, we present a study of a new palladium-based NLO active compound (see Figure 9). This is the stilbazolium cation where the one methyl group was replaced by the $Pd(PH_3)_2CH_3$ ligand (structure **6**). We optimized the design structure by the HF/LanL2DZ approach with the following TD-B3P86 analysis using both of the LanL2DZ and LanL2DZ/6-311++g(d,p) basis sets. The values of E, μ, $\Delta\mu$, and β_0 for the same level optimized structures for both the gas phase and the MeCN solution in comparison with the Ru(II) ammine complex with the same length of the polyene chain (structure **2**) were calculated within the developed method, and the results are presented in Table 4. In opposite to the Ru(II) ammine complexes studied above a newly designed chromophore demonstrates blue-shifted solvatochromic effect. In the case of the soluted complex, the calculated β_0 value of the Pd complex is smaller than that of the Ru one for structure **2**. However, in the gas phase the Pd complex has the larger value of β_0 as compared with the same value of the Ru complex. This is caused by a larger value of transition dipole moment μ. For practical application, it is important to know the characteristics of nonsoluted materials. Therefore, the results obtained for the Pd complex can suggest that metal substitution from Ru to Pd will improve the NLO properties of the organometallic compound.

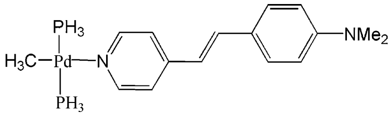


Figure 9. New palladium-based NLO active organometallic chromophore with a predicted large value of the first static hyperpolarizability (structure **6**).

Conclusion

We have presented the B3P86 study of the first singlet excited state of ruthenium(II) ammine complexes with a different number of polyene units. The solvent effect on the electronic structure in the framework of the PCM model has been studied. The TD-B3P86 study for HF, B3P86, and MP2 optimized structures has been carried out using the LanL2DZ and 6-311++g(d,p)/LanL2DZ basis sets, and calculated data were compared with experimental results. It is shown that in solution all considered compounds demonstrate not only significant red-shift of absorption energy maxima but also a sizable polarization of ground state and intensification of MLCT for low-lying excited states. Theoretical results are very sensitive to molecular geometry and basis set. The BLA and Ru(NH3)5-ligand distances are important geometry parameters responsible for the electronic excitation properties. Using the HF/6-311++g(d,p)/LanL2DZ optimized structures with the following TD-B3P86/6-311++g(d,p)/

LanL2DZ treatment is found to be more preferable for the calculation of E, μ, and $\Delta\mu$ values in order to estimate the first static hyperpolarizability in the framework of TSM and is given the best agreement with the experiment. In comparison with the gas phase all investigated compounds demonstrate sizable polarization in the solution, and as a result the values of ground-state dipole moments in the solution are larger than the same values in the gas phase. For electronic excitation the solvent also intensifies the MLCT that leads to the increasing of $\Delta\mu$ in solution and the decreasing of this value in the gas phase. Using the FF method for calculation of $\Delta\mu$ gives better agreement with the experiment than one particle RhoCI density approach. Thus, the developed approach can be used to design a new perspective organometallic NLO active compound.

Acknowledgment. We thank the Information Science Group of the Institute for Materials Research, Tohoku University for their continuous support of the SR8000 supercomputing system. One of the authors (T. M. Inerbaev) was supported by JSPS Research Fellowship Grant No. P 04683.

References

(1) *Nonlinear Optical Properties of Organic Molecules and Crystals*; Chemla, D. S., Zyss, J., Eds.; Academic Press: Orlando, FL, 1987.

(2) *Introduction to Nonlinear Optical Effects in Molecules and Polymers*; Prasad, P. N., Williams, D. J.; Eds.; Wiley: New York, 1991.

(3) Verbiest, T.; Houbrechts, S.; Kauranen, M.; Clays, C.; Persoons, A. Second-Order Nonlinear Optical Materials: Recent Advances in Chromophore Design. *J. Mater. Chem.* **1997**, *7*, 2175−2189.

(4) Calabrese, J. C.; Cheng, L.-T.; Green, J. C.; Marder, S. R.; Tam, W. Molecular Second-Order Optical Nonlinearities of Metallocenes. *J. Am. Chem. Soc*. **1991**, *113*, 7227−7232.

(5) Le Bozec, H.; Renouard, T. Dipolar and Non-Dipolar Pyridine and Bipyridine Metal Complexes for Nonlinear Optics. *Eur. J. Inorg. Chem.* **2000**, 229−239.

(6) Lacroix, P. G. Second-Order Optical Nonlinearities in Coordination Chemistry: the Case of Bis(salicylaldiminato)-metal Schiff Base Complexes. *Eur. J. Inorg. Chem.* **2000**, 229−239.

(7) Qin, J.; Liu, D.; Dai, C.; Chen, C.; Wu, B.; Yang, C.; Zhan, C. Influence of the Molecular Configuration on Second-Order Nonlinear Optical Properties of Coordination Compounds. *Coord. Chem. Rev.* **1999**, *188*, 23−34.

(8) Long, N. J. Organometallic Compounds for Nonlinear Optics − The Search for Enlightenment! *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 21−38.

(9) Oudar, J. L.; Chemla, D. S. Hyperpolarizabilities of the Nitroanilines and Their Relations to the Excited-State Dipole Moment. *J. Chem. Phys.* **1977**, *66*, 2664−2668.

(10) Numerical prefactor in eq 1 can take values 6 or 3/2 instead 3. See: Willets, A.; Rice, J. E.; Burland, D. M. Problems in the Comparison of Theoretical and Experimental Hyperpolarizabilities. *J. Chem. Phys.* **1992**, *97*, 7590−7599.

(11) Coe, B. J.; Harris, J. A.; Brunschwig, B. S. Absorption Spectroscopic Studies of Dipolar Ruthenium (II) Complexes Possessing Large Quadratic Optical Responses. *J. Phys. Chem. A* **2002**, *106*, 897−905.

(12) Coe, B. J.; Jones, L. A.; Harris, J. A.; Brunschwig, B. S.; Asselberghs, I.; Clays, K.; Persoons, A. Highly Unusual Effects of π-Conjugation Extension of the Molecular Linear Quadratic Nonlinear Optical Properties of Ruthenium (II) Ammine Complexes. *J. Am. Chem. Soc.* **2003**, *125*, 862−863.

(13) Coe, B. J.; Jones, L. A.; Harris, J. A.; Brunschwig, B. S.; Asselberghs, I.; Clays, K.; Persoons, A.; Garin, J.; Orduna, J. Syntheses and Spectroscopic and Quadratic Nonlinear Optical Properties of Extended Dipolar Complexes with Ruthenium(II) Ammine Electron Donor and *N*-methylpyridinium Acceptor Groups. *J. Am. Chem. Soc.* **2004**, *126*, 3880−3891.

(14) Shin, Y. K.; Brunschwig, B. S.; Creutz, C.; Sutin, N. Electroabsorption Spectroscopy of Charge-Transfer States of Transition-Metal Complexes. 2. Metal-to-Ligand and Ligand-to-Metal Charge-Transfer Excited States of Pentaammineruthenium Complexes. *J. Phys. Chem.* **1996**, *100*, 8157−8169.

(15) Coe, B. J.; Chamberlain, M. C.; Essex-Lopresti, J. P.; Gaines, S.; Jeffery, J. C.; Houbrechts, S.; Persoons, A. Large Molecular Quadratic Hyperpolarizabilities in Donor/Acceptor-Substituted Trans-tetraammineruthenium(II) Complexes. *Inorg. Chem.* **1997**, *36*, 3284−3292.

(16) Kanis, D. R.; Lacroix, P. G.; Ratner, M. A.; Marks, T. J. Electronic Structure and Quadratic Hyperpolarizabilities in Organotransition-Metal Chromophores Having Weakly Coupled π-Networks. Unusual Mechanism of Second-Order Response. *J. Am. Chem. Soc.* **1994**, *116*, 10089−10102.

(17) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *33*, 8822−8824.D.

(18) Becke A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision B.04*; Gaussian, Inc.: Pittsburgh, PA, 2003.

(20) Menucci, B.; Tomasi, J. Continuum solvation models: A new approach to the problem of solute’s charge distribution and cavity boundaries. *J. Chem. Phys.* **1997**, *106*, 5151−5158.

(21) Wiberg, K. B.; Stratmann, R. E.; Frisch, M. J. A Time-Dependent Density Functional Theory Study of the Electronically Excited States of Formaldehyde, Acetaldehyde and Acetone. *Chem. Phys. Lett*. **1998**, *297*, 60−64.

(22) Cave, R. J.; Burke, K.; Castner, E. W., Jr. Theoretical Investigation of the Ground and Excited States of Courmarins 151 and Courmarin 120. *J. Phys. Chem. A* **2002**, *106*, 9294−9305.

(23) Reed, A. E.; Weinhold, F. Natural Bond Orbital Analysis of Near-Hartree−Fock Water Dimer. *J. Chem. Phys.* **1983**, *78*, 4066−4073.

(24) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735−746.

(25) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular Interactions from a Natural Bond Orbital, Donor−Acceptor Viewpoint. *Chem. Rev.* **1988**, *88*, 899−926.

(26) Bourhill, G.; Brédas, J.-L.; Cheng, L.-T.; Marder, S. R.; Meyers, F.; Perry, J. W.; Tiemann, B. G. Experimental Demonstration of the First Hyperpolarizability of Donor−Acceptor-Substituted Polyenes on the Ground-State Polarization and Bond Length Alternation. *J. Am. Chem. Soc.* **1994**, *116*, 2619−2620.

(27) Meyers, F.; Marder, S. R.; Pierce, B. M.; Brédas, J.-L. Electric-Field Modulated Nonlinear-Optical Properties of Donor−Acceptor Polyenes − Sum-Over-States Investigation of the Relationship between Molecular Polarizabilities (α, β, and γ) and Bond Length Alternation. *J. Am. Chem. Soc.* **1994**, *116*, 10703−10714.

(28) Champagne, B.; Perpète, E. A.; van Gisbergen, S. J. A.; Baerends, E.-J.; Snijders, J. G.; Soubra-Ghaoui, C.; Robins, K. A.; Kirtman, B. Assessment of Conventional Density Functional Schemes for Computing the Polarizabilities and Hyperpolarizabilities of Conjugated Oligomers: An *Ab Initio* Investigation of Polyacetylene Chains. *J. Chem. Phys*. **1998**, *109*, 10489−10498.

CT050193M

J. Chem. Theory Comput. **2006,** *2,* 1335−1341

JCTC Journal of Chemical Theory and Computation

Study of the Topological Properties of Some Pseudohalides

Nora B. Okulik,[†] Alicia H. Jubert,[‡] and Eduardo A. Castro*[,§]

Facultad de Agroindustrias, UNNE, Cte. Fernández 755, 3700 Pcia. R. Sáenz Peña, Chaco, Argentina, and CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas 47 y 115 y Facultad de Ingeniería 1 y 47 and INIFTA, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, 1900 Buenos Aires, Argentina

Received December 3, 2005

Abstract: The pseudohalide principle has been used extensively in nonmetal chemistry to predict the structure and stability of many molecular species. The 1,2,3,4-thiatriazole-5-thiolate anion, $CS_2N_3^-$, is of particular interest. In a short communication we have recently reported the topological study of some $CS_2N_3^-$containing species reported by Crawford et al. Previous reports on these compounds showed that in covalent derivatives not only does the ring remain intact but also the site of attachment of the R group is most likely at the exocyclic sulfur atom in contrast to the previously suggested N−R connectivity. Therefore, the structure and bonding of derivatives of the $CS_2N_3^-$ moiety is clearly an important question. With that in our mind, we undertook a topological analysis, based on the AIM theory, to gain more insight into the bonding in covalent derivatives of the $CS_2N_3^-$ moiety, trying to find an explanation to the origin of the N−H and S−H connectivities. The question is which is the reason that makes all the covalent derivatives prefer the S−R connectivity while the hydracid has an N−H one.

Introduction

The study of pseudohalides has been of interest to nonmetal chemists for many years. Due to the similarities between the halides and the pseudohalides, the pseudohalide family of compounds is of fundamental chemical interest.[1] The pseudohalide concept was introduced in 1925,[2] and since its introduction, the pseudohalide principle has been used extensively in nonmetal chemistry to predict the structure and stability of many species.

The so-called “azidodithiocarbonate” anion, more properly referred to as the 1,2,3,4-thiatriazole-5-thiolate anion, $CS_2N_3^-$, is of particular interest.[3] First described and isolated in 1918 by Sommer,[4] it was not until 1991 that the structure of the $CS_2N_3^-$ anion in the form of its $NaCS_2N_3{\cdot}2H_2O$ salt was determined experimentally using X-ray diffraction techniques to be a five-membered ring,[5] in contrast to the previously postulated chain structure.[4] While early reports suggested the formation of several derivatives of this anion, little or no data were presented to support these claims.[6]

Crawford et al. have reported the characterization of several $CS_2N_3^-$ containing species.[3,7] They also reported the quantum-chemical calculations on the $CS_2N_3^-$ moiety which not only supported the formation of the five-membered ring but also were also in a very good agreement with the experimentally determined structure.[5] In a short communication we have recently reported the topological study of any $CS_2N_3^-$ containing species reported by Crawford et al.[8]

A previous report on these compounds showed that in covalent derivatives[5] not only did the ring remain intact but also the site of attachment of the R group is most likely at the exocyclic sulfur atom in contrast to the previously suggested N−R connectivity.[9] Therefore, the structure and bonding of derivatives of the $CS_2N_3^-$ moiety is clearly an important question. With that in our mind, we undertook a topological analysis to gain more insight into the bonding

---

* Corresponding author e-mail: castro@quimica.unlp.edu.ar.

[†] UNNE.

[‡] CEQUINOR.

[§] INIFTA.

10.1021/ct0503062 CCC: $33.50

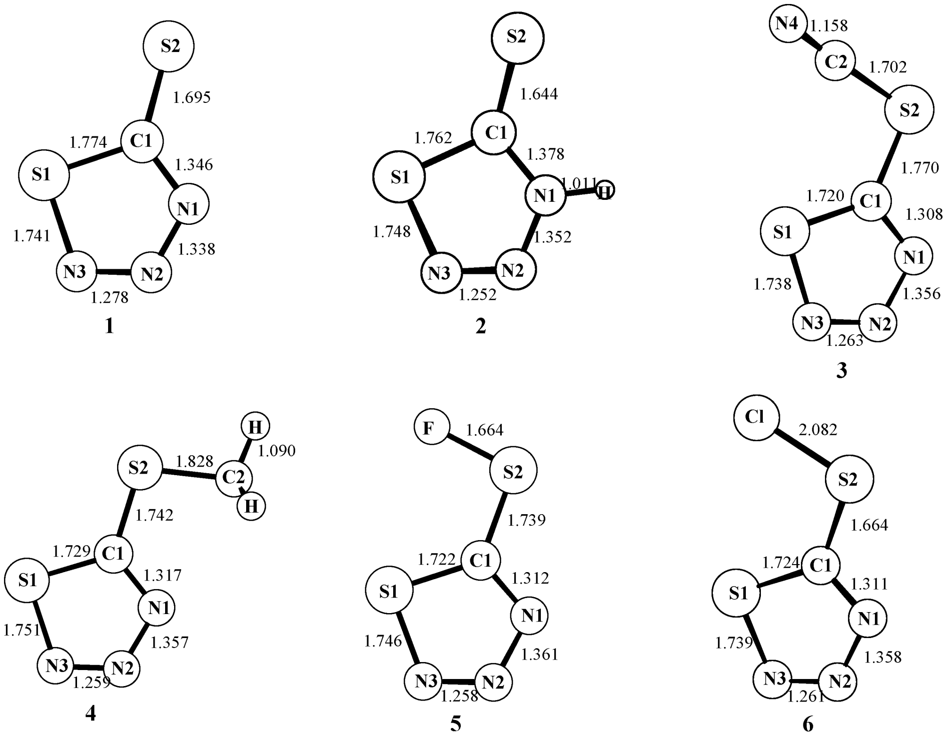


Figure 1. Optimized structures of 1,2,3,4-thiatriazole-5-thiolate anion, $CS_2N_3^-$ (**1**), hydracid, $HN_3SC{=}S$ (**2**), interpseudohalogen, $CS_2N_3{-}CN$ (**3**), methylazidodithiocarbonate, $CH_3CS_2N_3$ (**4**), chloroazidodithiocarbonate, $ClCS_2N_3$ (**5**), and fluoroazidodithiocarbonate, FCS_2N_3 (**6**), calculated at the B3LYP/6-311++G* level. The atomic labeling scheme and the bond lengths are indicated.

in covalent derivatives of the $CS_2N_3^-$ moiety, trying to find an explanation to the origin of the N−H and S−H connectivities. The question is what is the reason that makes all the covalent derivatives prefer the S−R connectivity while the hydracid has an N−H one.

Methods of Calculation and Computational Details

The structures of pseudohalides were calculated using the density functional theory,[10] and their topological properties are analyzed in terms of the atoms-in-molecules (AIM) method.[11] All calculations were carried out using the Gaussian 98 package.[12] The Becke's three-parameter density functional[13] together with the Lee, Yang, and Parr functional, which accounts for both local and gradient-corrected correlation effects,[14] were used to accomplish the calculations. This combination leads to the well-known and widely used B3LYP method. The basis set used for all elements was 6-311++G*.[15] Each optimized structure was tested against imaginary frequencies in order to be sure they are located at an energy minimum.

The topological analysis of all the species under study was accomplished by means of the PROAIM program.[16] The densities used for the topological analysis were obtained through single-point calculations on the above optimized geometries using the B3LYP level of theory and the 6-311++G** basis set provided by the Gaussian 98 package.[12]

Atoms-in-Molecules Theory: An Overview. The AIM theory is a simple, rigorous, and elegant way of defining atoms and bonds within a chemical structure. This theory is based on the critical points (CP) of the electronic density, $\rho(\mathbf{r})$. These are points where the gradient of the electronic density, $\nabla\rho(\mathbf{r})$, vanishes, and they are characterized by the three eigenvalues (λ_1, λ_2, λ_3) of the Hessian matrix of $\rho(\mathbf{r})$. The CPs are labeled as (r, s) according to their rank, r (number of nonzero eigenvalues), and signature, s (the algebraic sum of the signs of the eigenvalues).

Four types of CPs are of interest in molecules: (3, −3), (3,−1), (3, +1), and (3, +3). A (3, −3) point corresponds to a maximum in $\rho(\mathbf{r})$, and it appears generally at the nuclear positions. A (3, +3) point indicates electronic charge depletion and is known as a cage critical point. (3, +1) points, or ring critical points, are merely saddle points. Finally, a (3, −1) point, or bond critical point, is generally found between two neighboring nuclei indicating the existence of a bond between them.

Several properties that can be evaluated at the bond critical point, BCP, constitute very powerful tools to classify the interactions between two fragments.[11] Calculated properties at the BCP of the electronic density are labeled with the subscript "b" throughout the work.

The two negative eigenvalues of the Hessian matrix (λ_1 and λ_2) at the BCP measure the degree of contraction of ρ_b perpendicular to the bond toward the critical point, while the positive eigenvalue (λ_3) measures the degree of contraction parallel to the bond and from the BCP toward each of the neighboring nuclei. When the negative eigenvalues dominate, the electronic charge is locally concentrated within the region of the BCP leading to an interaction characteristic typically found in covalent or polarized bonds and being characterized by large ρ_b values, $\nabla^2\rho_b < 0$, $|\lambda_1|/\lambda_3 > 1$, and $G_b/\rho_b < 1$, with G_b being the local kinetic energy density at the bond critical point. On the other hand, if the positive

Table 1: Comparison of Selected Experimentally Determined (X-ray) and Calculated (B3LYP/6-311++G*) Structural Parameters (Å, deg) for **1**−**6** Structures

	1		**2**		**3**		**4**		**5**		**6**	
parameter[a]	calc	exptl[b]	calc	exptl[b]	calc	exptl[b]	calc	exptl[c]	calc	calc[d]	calc	calc[e]
C_1-S_2	1.695	1.696	1.644	1.661	1.770	1.756	1.742	1.723	1.739	1.739	1.664	1.72
C_1-N_1	1.346	1.340	1.378	1.346	1.308	1.305	1.317	1.314	1.312	1.312	1.311	1.31
C_1-S_1	1.774	1.717	1.762	1.722	1.720	1.699	1.729	1.699	1.722	1.722	1.724	1.75
N_1-N_2	1.338	1.347	1.352	1.351	1.356	1.364	1.357	1.356	1.361	1.361	1.358	1.36
N_2-N_3	1.278	1.286	1.252	1.260	1.263	1.276	1.259	1.274	1.258	1.258	1.261	1.26
S_1-N_3	1.741	1.682	1.748	1.706	1.738	1.670	1.751	1.680	1.746	1.747	1.739	1.74
N_1-H			1.011	0.800								
S_2-C_2					1.702	1.697	1.828	1.799				
C_2-N_4					1.158	1.139						
C_2-H							1.090	0.92				
S_2-X									1.664	1.663	2.082	2.08
$S_1-C_1-N_1$	108.0	109.1	103.2	105.2	113.2	112.7	112.2	111.8	113.5	113.5	112.9	112.9
$S_2-C_1-S_1$	124.3	124.3	130.3	127.65	129.5	124.6	123.4	122.3	127.6	127.5	129.8	129.8
$C_1-S_1-N_3$	90.5	91.7	92.2	92.03	87.9	89.5	88.3	89.9	87.6	87.5	87.9	87.0
$N_1-N_2-N_3$	118.7		113.7		117.0		117.8	116.6	117.2	117.2	117.0	117.1
C_1-N_1-H			122.6	120.0								
$S_2-C_2-N_4$					176.3	172.6						
C_1-S_2-X									98.2	98.2	101.5	101.5

[a] See Figure 1 for labels. [b] Reference 7c. [c] Reference 3. [d] Reference 7a. [e] Reference 7b.

eigenvalue is dominant, the electronic density is locally concentrated at each atomic site. The interaction is now referred to as a closed-shell one, and it is characteristic of highly ionic bonds, hydrogen bonds, and van der Waals interactions. It is characterized by relatively low ρ_b values, $\nabla^2\rho_b > 0$, $|\lambda_1|/\lambda_3 < 1$, and $G_b/\rho_b > 1$. Finally, the ellipticity, ϵ, defined as $\lambda_1/\lambda_2 - 1$ indicates the deviation of the electronic charge density from the axial symmetry of a chemical bond providing a quantitative measure of the π character of the bond or of the delocalization electronic charge.

Other properties are obtained by integrating the corresponding property density over the atomic basin, which is denoted by Ω.[11] The relevant atomic properties for the present work are the average number of electrons, $N(\Omega)$, from which the atomic net charge, $q(\Omega)$, can be calculated as $Z_\Omega - N(\Omega)$, Z_Ω being the nuclear charge of the atom. AIM theory permits the identification of reactive sites by means of the Laplacian of the charge density, $\nabla^2\rho$. AIM defines the valence-shell charge concentration (VSCC) as the outer molecular zone where $\nabla^2\rho < 0$. This zone is the one which, upon chemical combination, is distorted to yield nonbonded critical points (NBCP), which are minima in $\nabla^2\rho$ (maxima of charge concentration), corresponding in number and position to the electron pairs defined by the Lewis and related models.[11,17] NBCP correspond to zones where an electrophilic attack can occur.

Results and Discussion

The Connectivity. The optimized geometries of the 1,2,3,4-thiatriazole-5-thiolate anion, $CS_2N_3^-$ (**1**), the hydracid, $HN_3SC{=}S$ (**2**), the interpseudohalogen, CS_2N_3-CN (**3**), the methylazidodithiocarbonate, $CH_3CS_2N_3$ (**4**), the chloroazidodithiocarbonate, $ClCS_2N_3$ (**5**), and the fluoroazidodithiocarbonate, FCS_2N_3 (**6**), calculated at the B3LYP/6-31G** level of theory, are shown in Figure 1, and the bond lengths are indicated. The calculated parameters are in very good agreement with the available experimental data and theoretical results reported by Crawford et al.[3,7] (Table 1).

The five-membered ring is the standard form for the $CS_2N_3^-$ anion and its derivatives. The planarity of the $CS_2N_3^-$ anion is an indication for both the aromatic and pseudohalide character of this species.[3] Table 2 lists relevant topological properties obtained from this study. The topological analysis of BCPs in ρ of $CS_2N_3^-$ anion, **1**, and the covalent derivatives, **2**−**6**, reveal that all bonds forming the five-member ring correspond to covalent interactions, namely, a relatively large value for ρ_b and a negative value for $\nabla^2\rho_b$. The $|\lambda_1|/\lambda_3$ quotient is appreciably greater than 1, except for the N_1-N_2 bond for which these values are near 1. The ellipticities of bonds forming the ring have considerably large numerical values, and it reveals their partial double bond character due to the electronic charge delocalization over the ring surface. The G_b/ρ_b relationship is lower than 1 as usual in covalent bonds.

The exocyclic C−S bond is described by lower values of ρ_b, and $\nabla^2\rho_b$ shows small although negative values. The $|\lambda_1|/\lambda_3$ quotient is always more than 1 for the C_1-S_2 bonds. A remarkable situation lies in the HCS_2N_3 compound, **2**, in which the Laplacian of the electronic density, $\nabla^2\rho_b$, shows a small, positive value at the BCP. In the same way, the $|\lambda_1|/\lambda_3$ quotient is less than 1, and the G_b/ρ_b relationship is > 1. These topological properties reflect the bond polarity. Figures 2 and 3 show the $\nabla^2(\mathbf{r})$ contour maps for structures **1**−**3** and **4**−**6**, respectively, which clearly show the concepts previously mentioned. It can be seen that the interaction pattern of the ring atoms is nearly the same and clearly suggests the presence of shared interactions.

The same situation is found in the exocyclic bonds which exhibit a region of charge concentration along each bond, a characteristic feature associated with shared interactions. Nevertheless, it is necessary to highlight the interaction

Table 2: Topological Analysis of BCP in ρ of 1,2,3,4-thiatriazole-5-thiolate Anion, $CS_2N_3^-$ (**1**), Hydracid, $HN_3SC{=}S$ (**2**), Interpseudohalogen, $CS_2N_3{-}CN$ (**3**), Methylazidodithiocarbonate, $CH_3CS_2N_3$ (**4**), Chloroazidodithiocarbonate, $ClCS_2N_3$ (**5**), and Fluoroazidodithiocarbonate, FCS_2N_3 (**6**)[a]

	$C_1{-}S_2$	$C_1{-}N_1$	$C_1{-}S_1$	$S_1{-}N_3$	$N_1{-}N_2$	$N_2{-}N_3$	$N_1{-}H$	$S_2{-}C_2$	$C_2{-}N_4$	$C_2{-}H$	$S_2{-}X$
						1					
ρ_b	0.2065	0.3385	0.1943	0.1979	0.3804	0.4340					
$\nabla^2\rho_b$	−0.2846	−0.9986	−0.2941	−0.3149	−0.7416	−0.9551					
$\lvert\lambda_1\rvert/\lambda_3$	1.6724	1.9937	1.1137	1.1896	0.9122	1.0352					
ϵ	0.0693	0.1366	0.1834	0.1783	0.0914	0.1630					
G_b/ρ_b	0.7956	0.6015	0.3258	0.4502	0.5315	0.6062					
						2					
ρ_b	0.2166	0.3040	0.2014	0.1956	0.3680	0.4626	0.3387				
$\nabla^2\rho_b$	0.0295	−0.7912	−0.3452	−0.3129	−0.7017	−1.1329	−1.7923				
$\lvert\lambda_1\rvert/\lambda_3$	0.4718	1.6004	1.2901	1.2041	0.9131	1.1507	1.5992				
ϵ	0.0313	0.0775	0.2111	0.1893	0.1406	0.1910	0.0493				
G_b/ρ_b	1.2175	0.7006	0.3173	0.4484	0.5084	0.6215	0.1291				
						3					
ρ_b	0.1948	0.3670	0.2115	0.2002	0.3656	0.4510		0.2096	0.4688		
$\nabla^2\rho_b$	−0.3473	−1.0768	−0.3928	−0.3111	−0.6834	−1.0553		−0.4705	−1.0728		
$\lvert\lambda_1\rvert/\lambda_3$	1.3505	2.0007	1.6079	1.1400	0.8734	1.0848		4.0575	0.5338		
ϵ	0.2198	0.2107	0.2315	0.1841	0.0621	0.1586		0.3118	0.0156		
G_b/ρ_b	0.2849	0.7185	0.3679	0.4500	0.5141	0.6159		0.4575	1.7325		
						4					
ρ_b	0.2027	0.3589	0.2096	0.1945	0.3647	0.4551		0.1738		0.2774	
$\nabla^2\rho_b$	−0.3953	−1.0517	−0.3816	−0.2777	−0.6762	−1.0758		−0.2525		−0.9351	
$\lvert\lambda_1\rvert/\lambda_3$	1.7716	2.000	1.5023	1.0445	0.8712	1.0966		1.0133		1.4081	
ϵ	0.2524	0.1761	0.2139	0.1803	0.0653	0.1610		0.0739		0.0127	
G_b/ρ_b	0.3360	0.6859	0.3530	0.4427	0.5155	0.6206		0.2636		0.1432	
						5					
ρ_b	0.2093	0.3636	0.2099	0.1967	0.3604	0.4559					0.1677
$\nabla^2\rho_b$	−0.4140	−1.0712	−0.3829	−0.2895	−0.6527	−1.0808					−0.0327
$\lvert\lambda_1\rvert/\lambda_3$	1.6161	2.0165	1.5487	1.0765	0.8611	1.0992					0.6225
ϵ	0.3068	0.1884	0.2216	0.1800	0.0673	0.1605					0.3410
G_b/ρ_b	0.3000	0.6525	0.3654	0.4474	0.5172	0.6210					0.8777
						6					
ρ_b	0.2024	0.3642	0.2099	0.1995	0.3633	0.4529					0.1302
$\nabla^2\rho_b$	−0.3795	−1.0715	−0.3826	−0.3083	−0.6674	−1.0651					−0.0379
$\lvert\lambda_1\rvert/\lambda_3$	1.5006	2.0060	1.5377	1.1339	0.8679	1.0908					0.5957
ϵ	0.2647	0.1917	0.2238	0.1829	0.0674	0.1605					0.1064
G_b/ρ_b	0.3033	0.7024	0.3628	0.4493	0.5172	0.6180					0.3771

[a] ρ_b and $\nabla^2\rho_b$ in au.

pattern in the $S_2{-}X$ bonds in structures **5** and **6**, because of the electronegativity difference between the S and F and Cl atoms, respectively.

Regarding the situation related with the site of protonation in the HCS_2N_3 compound, that is to say, whether the site of protonation was at the exocyclic sulfur atom or at one of the ring nitrogen atoms (see Chart 1), the isomer with $N_1{-}H$ connectivity (**2**) was found to be the lowest energy isomer for HCS_2N_3 in the gas phase and the isomer with an exocyclic S−H bond (**2′**) was calculated to be the next higher energy isomer.[3] However, there is still no answer to the question about why the N−H isomer is the favored structure, in opposition to the higher stability of those covalent derivatives with an exocyclic S−R bond ($R{=}CH_3$, X, CN). To look for an answer to this question we also performed topological calculations on both structures (**2** and **2′**).

The results show that the higher stability of the covalent derivatives with S−R connectivites instead of N−R connectivity is due to a steric reason. It seems that the substituent bond to the exocycle sulfur has “more space” than the bond to the nitrogen of the ring. Obviously, we cannot use the same reasoning for the hydrogen atom. Thus, which would be the reason that makes the hydracid more stable with N−H connectivity?

Up to now, our calculations have allowed us to find two alternative explanations regarding the following aspects:

1. The proton affinity calculated by Crawford et al.[3] of 327.0 kcal mol^{-1} is similar to the proton affinity calculated[3] for the HCl of 339.3 kcal mol^{-1} (experimental[18] 333.2 kcal mol^{-1}). These values indicate that hydracid is a strong acid. If this is true, its anion (conjugate base) is weak and, therefore, more stable when it is dissociated.

2. The contour map of the Laplacian of the distribution of the charge density of the hydracid with N−H connectivity is much more similar to the contour map of the anion (at least more similar than the map of the hydracid with the

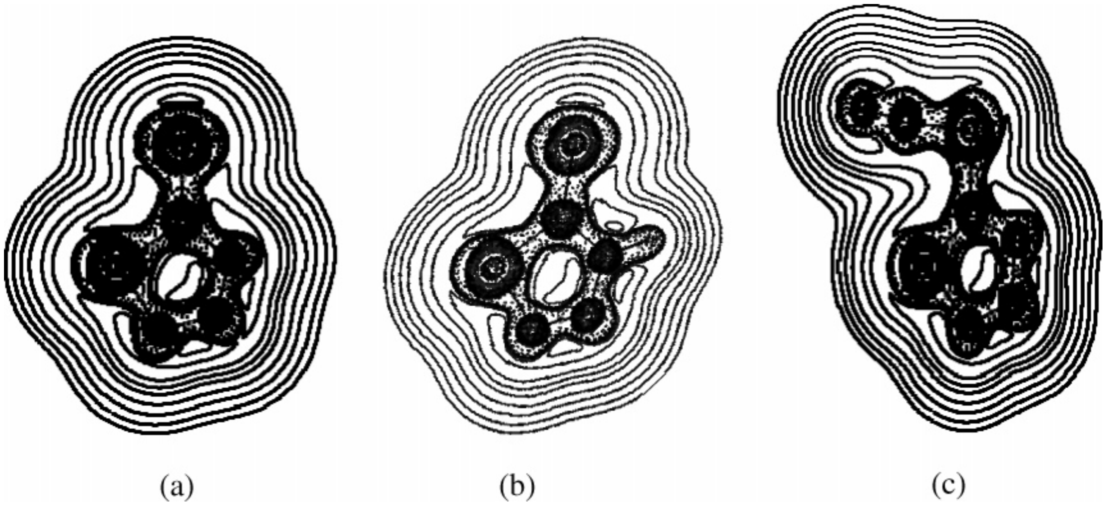


Figure 2. Laplacian of the electronic density of (a) 1,2,3,4-thiatriazole-5-thiolate anion, $CS_2N_3^-$ (**1**), (b) hydracid, $HN_3SC{=}S$ (**2**), and (c) interpseudohalogen, $CS_2N_3{-}CN$ (**3**), in the plane of rings. Broken lines denote regions of electronic charge concentration, and solid lines represent regions of electronic charge depletion. BCP are indicated with black circles. The molecular graphs are also indicated. The contours of the Laplacian of the electronic density increase and decrease from a zero contour in steps of ±2, ±4, ±8 × 10^n, with *n* beginning at −3 and increasing by unity. The same set of contours is used in all the figures of the present work.

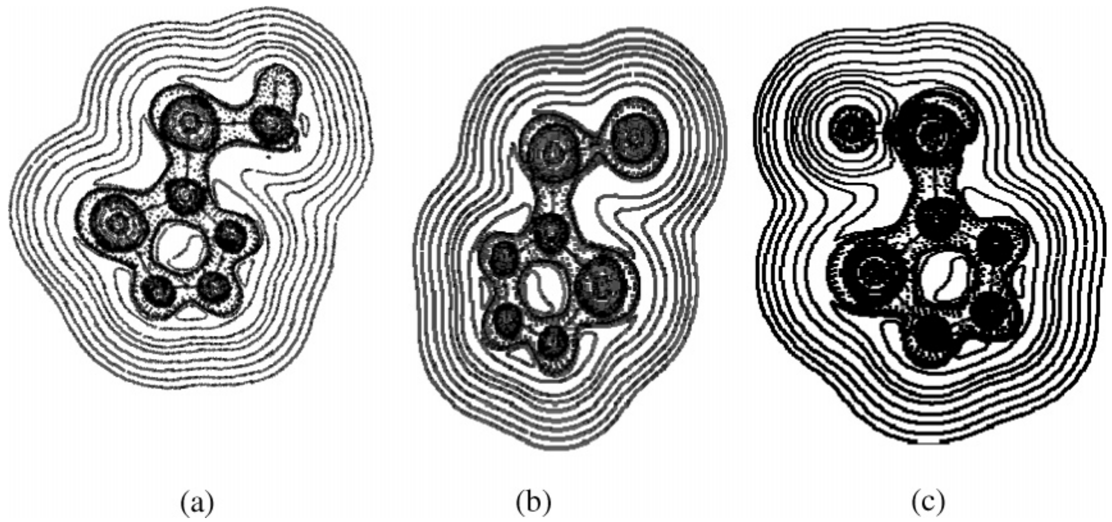
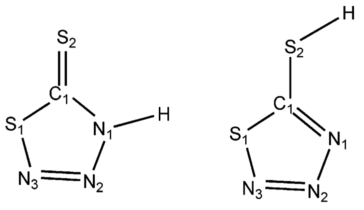


Figure 3. Laplacian of the electronic density of (a) methylazidodithiocarbonate, $CH_3CS_2N_3$ (**4**), (b) chloroazidodithiocarbonate, $ClCS_2N_3$ (**5**), and (c) fluoroazidodithiocarbonate, FCS_2N_3 (**6**), in the plane of rings. Broken lines denote regions of electronic charge concentration, and solid lines represent regions of electronic charge depletion. BCP are indicated with black circles. The molecular graphs are also indicated.

Chart 1: HCS_2N_3 Isomers with N−H Connectivity (**2**) and with an Exocyclic S−H Bond (**2′**)



2 **2′**

S−H connectivity) which could indicate that after the loss of the proton an anion is formed practically without changes in the electronic distribution. This feature can be seen when confronting the map of structure **2** (Figure 2b) with the map of structure **2′** (Figure 4).

Table 3: Net Charges, $q(\Omega)$, on Five-Member Rings and Exocyclic Sulfur Atoms for All Studied Structures[a]

	1	2	2′	3	4	5	6
q(C1)	+0.074	−0.055	+0.232	+0.264	+0.209	+0.238	+0.246
q(N1)	−0.608	−0.707	−0.601	−0.598	−0.608	−0.604	−0.605
q(N2)	−0.094	−0.007	−0.017	−0.003	−0.020	−0.003	−0.007
q(N3)	−0.456	−0.329	−0.326	−0.336	−0.335	−0.334	−0.344
q(S1)	+0.286	+0.476	+0.461	+0.539	+0.187	+0.520	+0.509
q(S2)	−0.199	+0.166	+0.214	+0.524	+0.457	+0.737	+0.380

[a] The units are atomic units.

Charge Distribution. The atomic charge polarizations which take place with chemical bonding are of fundamental interest to chemists. Therefore, to study the charge distribution that takes place with the intensity and direction of the

Table 4: Values of Laplacian of the Charge Density, $\nabla^2\rho$ [au], at the Nonbonded Critical Points (NBCP) of the Selected Atoms in Studied Structures[d]

atom	1	2	2′	3	4	5	6
N1	−2.7012 (1)[b]	−1.6284 (2)[a]	−2.8106 (1)[b]	−2.8427 (1)[b]	−2.8063 (1)[b]	−2.8078 (1)[b]	−2.8101 (1)[b]
N2	−3.0714 (1)[b]	−3.1253 (1)[b]	−3.1009 (1)[b]	−3.1459 (1)[b]	−3.0975 (1)[b]	−3.1279 (1)[b]	−3.1427 (1)[b]
N3	−2.7039 (1)[b]	−2.9092 (1)[b]	−2.9011 (1)[b]	−2.9203 (1)[b]	−2.8916 (1)[b]	−2.8651 (1)[b]	−2.8771 (1)[b]
S1	−0.4786 (2)[c]	−0.5026 (2)[c]	−0.4641 (2)[c]	−0.4619 (2)[c]	−0.4667 (2)[c]	−0.4689 (2)[c]	−0.4668 (2)[c]
S2	−0.4619 (2)[b]	−0.4883 (2)[b]	−0.5464 (2)[c]	−0.5627 (2)[c]	−0.5556 (2)[c]	−0.6728 (2)[c]	−0.6117 (2)[c]

[a] NBCP position: top and down of plane. [b] NBCP position: in plane. [c] NBCP position: top and down plane in sp^3 hybridization. [d] Numbers in parentheses indicate the NBCP encountered.

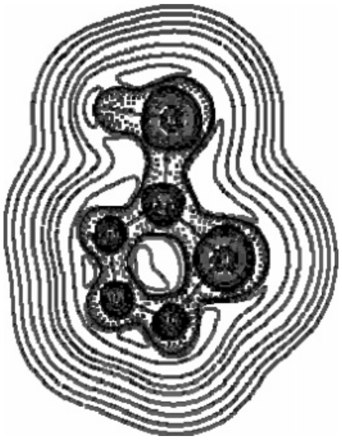


Figure 4. Laplacian of the electronic density of HCS_2N_3 isomers with an exocyclic S−H bond (**2′**), in the plane of ring. Broken lines denote regions of electronic charge concentration, and solid lines represent regions of electronic charge depletion. BCP are indicated with black circles. The molecular graphs are also indicated.

charge transfer, the net charges, $q(\Omega)$, were computed. Table 3 summarizes the atomic charges on five-member rings and on exocyclic sulfur atom for the structures studied (**1**−**6**). These results show that exocyclic S atom possesses a negative charge only in the anion, structure **1**, although practically the whole negative charge is located at the N1 atom. In structures **3**−**6**, the exocyclic S atom carries a considerable positive charge taking into account structure **2** due to the charge transfer of the S atom toward the R groups (R=CN, CH_3, X).

It is interesting to note that the charge transfer is important in structure **5**, where the S atom is bonded to the F atom. This situation is clearly evidenced in the Laplacian of the electronic density distribution (Figure 3c). On the other hand, in all cases the cyclic S atom carries a positive charge, which is bigger in structure **3**. The considerable values of the positive charges in structures **3**, **5**, and **6** on the cyclic S atom reveal a decreased electronic population on this atom in derivatives where the R groups contain electronegative atoms. In all derivatives studied, the charge of the cyclic C atom is positive (about +0.2 au).

Sites of Electrophilic Attack. To gain more insight into the site of attachment of the R group is most likely at the exocyclic sulfur atom in contrast to the N−R connectivity in the HCS_2N_3 compound, we undertook a topological analysis of the Laplacian of the electronic density. Therefore, NBCPs have been determined on the sulfur and nitrogen atoms value for all the studied structures.

Results are collected in Table 4. A single NBCP is found at the pyridine nitrogen N1, N2, and N3 in **1** and **3**−**6**. Those NBCP are coplanar with the ring. On the other hand, two NBCP are found for the amine nitrogen N1 in 2. The first of these NBCP is located at the apex of the pyramidic nitrogen (the place where the lone electron pair is usually represented). The second NBCP appears pointing toward the base of the pyramid. The values of $\nabla^2\rho$ in both NBCP are similar. The existence of these two NBCP can be explained in light of the conjugation between the amine group and the pseudoaromatic ring. In the absence of conjugation we can expect a single local maximum of charge concentration, a NBCP, corresponding to the lone electron pair located in a hybrid orbital pointing to the apex of the pyramid. However, when the amine group is conjugated with the aromatic ring, the electron pair is more similar to a two-lobed p orbital than to a hybrid orbital with a single lobe.

On the cyclic sulfur atom two NBCP are found, compatible with two lone electron pairs localized at the upper and lower positions of the plane corresponding to a sulfur atom with sp^3 hybridization. At the same time, on the exocyclic sulfur atom bound to the C atom in the ring are found two NBCP located in the plane, compatible with two lone electron pairs corresponding to a sulfur atom with sp^2 hybridization in structures **1** and **2**. On the other hand, in the other ones the two lone electron pairs are placed in two sp^3 hybrid orbitals.

In all studied cases the nitrogen atoms have $\nabla^2\rho$ values of the NBCP larger than those of the sulfur atoms. When the values of $\nabla^2\rho$ are compared at the different nonbonded critical points on N atoms of the studied species, the NBCP at the N2 exhibits the highest concentration of charge (see Table 4). Provided that $\nabla^2\rho$ can be considered as an indicator of the site electrophilic attack, the center with the highest $\nabla^2\rho$ value seems to correspond to the preferred attack site.

The data analysis of Table 2 shows clearly that the hydracid N−H has in the exocyclic S atom the same array of lone electron pairs corresponding to a hybridization approximately sp^2 (two lone electron pairs in the plane and with an angle of 155.5°). This point can reinforce what was pointed out with regard to the connectivity since in the hydracid S−H: the exocycle S atom has a hybridization of approximately sp^3 (two lone electron pairs up and down the molecular plane and forming between them an angle of 134.4°). Then, with the loss of the proton it would be possible to undergo a rehybridization to sp^2 when forming the anion. Looking at what happens with the nitrogen atom of the ring which is bonded to the hydrogen atom in the hydracid N−H (**2**), this atom has two NBCP: one up and another down the plane and forming an angle of only 8° with the N atom, resembling a pure p orbital, while the N atom maintains the sp^2 hybridization. We believe that this situation allows maintaining the aromatic character of the species, and, therefore, with a single change in the electronic distribution, the N atom reorders a single NBCP.

Conclusions

We have undertaken a topological analysis of the so-called "azidodithiocarbonate" anion, 1,2,3,4-thiatriazole-5-thiolate anion, $CS_2N_3^-$ (**1**), and its covalent derivatives: the hydracid, $HN_3SC{=}S$ (**2**), the interpseudohalogen, $CS_2N_3{-}CN$ (**3**), the methylazidodithiocarbonate, $CH_3CS_2N_3$ (**4**), the chloroazidodithiocarbonate, $ClCS_2N_3$ (**5**), and the fluoroazidodithiocarbonate, FCS_2N_3 (**6**).

The results reveal that all bonds forming the five-member ring correspond to covalent interactions. The higher stability of the covalent derivatives with S−R connectivites instead of the N−R connectivity is explained on the basis of an existing steric reason. Additionally, the distribution of the charge density of the hydracid with N−H connectivity and the anion seem to indicate that after the loss of the proton the anion is formed practically without changes in the electronic distribution. In this way, with a single change in the electronic distribution the N atom reorders two at a single NBCP.

A remarkable result of this work, merging from this analysis, is that the nitrogen atoms are sites that exhibit the highest concentration of electronic charge. The topological studies have shown that they are useful and effective tools for elucidating the structures of many CS_2N_3-containing species.

Acknowledgment. The authors acknowledge the Supercomputer Center of the Secretary for the Technology, Science, and Productive Innovation, Argentina, for computational time. N.B.O. thanks the Facultad de Agroindustrias and SECYT UNNE for financial support. A.H.J. is a member of the Carrera del Investigador CIC, Buenos Aires, and E.A.C. and N.B.O. are members of the career researcher of CONICET, Argentina.

References

(1) (a) Greenwood, N. N.; Earnshaw, A. A. In *Chemistry of the Elements*; Pergamon Press: Oxford, U.K., 1984. (b) Klapötke, T. M.; Janiak, C.; Meyer, H.-J. In *Modern Inorganic Chemistry;* Walter de Gruyter: Berlin, 1998.

(2) Birkenbach, L.; Kellermann, K. *Ber. Dtsch. Chem. Ges.* **1925**, *58B*, 786.

(3) Crawford, M.-J.; Klapotke, T. M.; Mayer, P.; Vogt, M. *Inorg. Chem.* **2004**, *43*, 1370−1378.

(4) Sommer, F. *Ber. Dtsch. Chem. Ges.* **1915**, *48*, 1833.

(5) Perman, C. A.; Gleason, W. B. *Acta Crystallogr.* **1991**, *C47*, 1018.

(6) (a) Browne, A. W.; Hoel, A. B. *J. Am Chem. Soc.* **1922**, *44*, 2315. (b) Howlett-Gardner, W.; Browne, A. W. *J. Am Chem. Soc.* **1927**, *49*, 2759. (c) Smith, G. B. L.; Wrattman, P.; Browne, A. W. *J. Am Chem. Soc.* **1930**, *52*, 2806.

(7) (a) Crawford, M.-J.; Klapötke, T. M. *J. Fluorine Chem.* **1998**, *92*, 153−156. (b) Crawford, M.-J.; Klapötke, T. M. *Inorg. Chim. Acta* **1999**, 294, 68−72. (c) Crawford, M. J.; Klapötke, T. M.; Klüfers, P.; Mayer, P.; White, P. S. *J. Am. Chem. Soc.* **2000**, *122*, 9052.

(8) Okulik, N.; Jubert, A. H.; Castro, E. A. *J. Mol. Struct. (THEOCHEM)* **2002**, *589-590*, 79−87.

(9) Lieber, E.; Offentdahl, E.; Rao, C. N. R. *J. Org. Chem.* **1963**, *28*, 194.

(10) (a) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136B*, 864. (b) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140A*, 1133. (c) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules;* Oxford University Press: Oxford, 1989.

(11) (a) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Clarendon: Oxford, 1990. (b) Popelier, P. L. A. *Atoms in Molecules. An Introduction;* Pearson Education: Harlow, U.K., 1999.

(12) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, revision A.6*; Gaussian, Inc.: Pittsburgh, PA, 1998.

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

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

(15) Huzinaga, S.; Andzelm, J.; Klobukowski, M.; Radzio-Andzelm, E. Sakai, T.; Tatewaki, H. *Gaussian Basis Set for Molecular Calculations*; Elsevier: Amsterdam, 1984.

(16) Biegler-König, F. W.; Bader, R. F. W.; Tang, T. H. *J. Comput. Chem.* **1982**, *13*, 317.

(17) Bader, R. F. W. *Chem. Rev.* **1990**, *91*, 893.

(18) NIST Service. Web-book. http://www.nist.gov (accessed Nov 10, 2005).

CT0503062

1342 *J. Chem. Theory Comput.* **2006,** *2,* 1342−1359

JCTC Journal of Chemical Theory and Computation

Density-Functional Theory Investigation of the Geometric, Energetic, and Optical Properties of the Cobalt(II)tris(2,2′-bipyridine) Complex in the High-Spin and the Jahn−Teller Active Low-Spin States

Alfredo Vargas,[†] Mohamed Zerara,[†] Elmars Krausz,[‡] Andreas Hauser,[†,*] and
Latévi Max Lawson Daku[†,*]

*Département de Chimie Physique, Université de Genève, 30 quai Ernest-Ansermet,
CH-1211 Genève 4, Switzerland, and Research School of Chemistry, Australian
National University, Building 31 Science Road, Canberra ACT 0200, Australia*

Received April 13, 2006

Abstract: State-of-the-art generalized gradient approximation (GGA) (PBE, OPBE, RPBE, OLYP, and HCTH), meta-GGA (VSXC and TPSS), and hybrid (B3LYP, B3LYP*, O3LYP, and PBE0) functionals are compared for the determination of the structure and the energetics of the D_3 $[Co(bpy)_3]^{2+}$ complex in the $^4A_2$ and ^4E trigonal components of the high-spin $^4T_{1g}(t_{2g}^5\,e_g^2)$ state and in the low-spin ^2E state of octahedral $^2E_g(t_{2g}^6\,e_g^1)$ parentage. Their comparison extends also to the investigation of the Jahn−Teller instability of the ^2E state through the characterization of the extrema of C_2 symmetry of this spin state's potential energy surface. The results obtained for $[Co(bpy)_3]^{2+}$ in either spin manifold are very consistent among the functionals used and are in good agreement with available experimental data. The functionals, however, perform very differently with respect to the spin-state energetics because the calculated values of the high-spin/low-spin energy difference ΔE_{HL}^{el} vary between −3212 and 3919 cm^{-1}. Semilocal functionals tend to give too large ΔE_{HL}^{el} values and thus fail to correctly predict the high-spin state as the ground state of the isolated complex, while hybrid functionals tend to overestimate the stability of the high-spin state with respect to the low-spin state. Reliable results are, however, obtained with the OLYP, HCTH, B3LYP*, and O3LYP functionals which perform best for the description of the isolated complex. The optical properties of $[Co(bpy)_3]^{2+}$ in the two spin states are also analyzed on the basis of electronic excitation calculations performed within time-dependent density functional response theory. The calculated absorption and circular dichroism spectra agree well with experimental results.

1. Introduction

The electronic ground state of octahedral d^{7}complexes is either the high-spin (HS) $^4T_{1g}(t_{2g}^5e_g^2)$ state or the low-spin (LS) $^2E_g(t_{2g}^6e_g^1)$ state depending on the strength of the ligand field. The d^7 $[Co(bpy)_3]^{2+}$ (bpy = 2,2′-bipyridine) complex is an intermediate-field complex with normally a HS ground state and a low-lying excited LS state. However, when $[Co(bpy)_3]^{2+}$ is incorporated into zeolite-Y supercages[1−3] or into the cavities of three-dimensional oxalate networks,[4,5] the LS state becomes the ground state and the complex exhibits spin crossover. The X-ray structures of the complex in the HS and LS states[4] show a large difference of the metal−ligand bond length, $\Delta r_{HL} = r_{HS} - r_{LS}$, of as much as ~0.1 Å, which is due to the fact that the change of spin state involves an electron transfer between the metallic nonbond-

---

* Corresponding author e-mail: max.lawson@unige.ch (L.M.L.D.), andreas.hauser@unige.ch (A.H.).

[†] Universitée de Genève.

[‡] Australian National University.

10.1021/ct6001384 CCC: $33.50

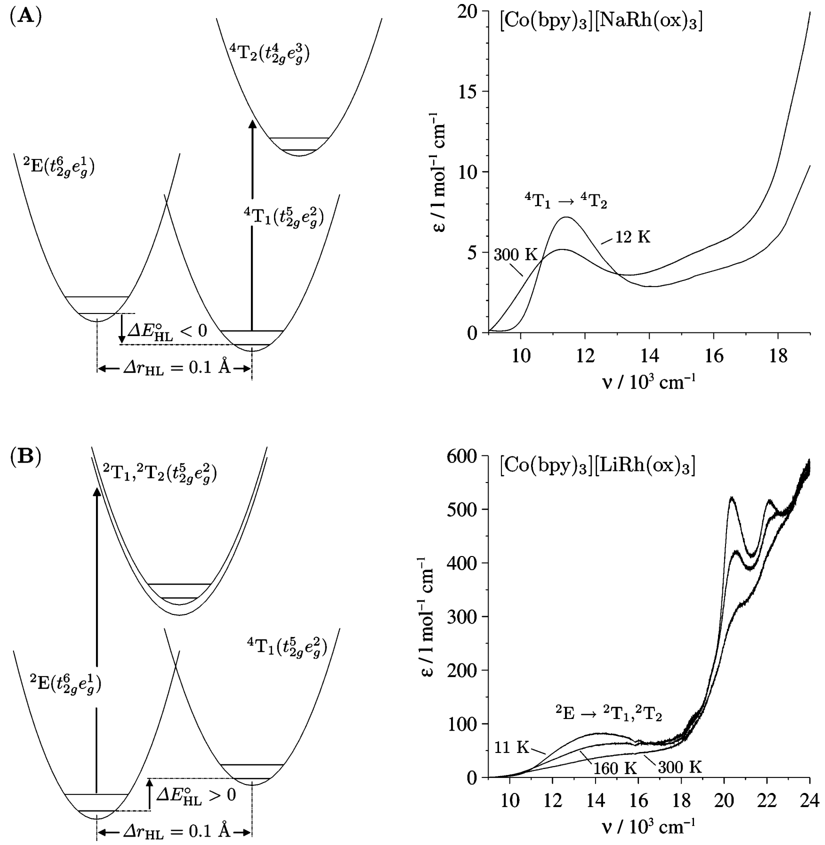


Figure 1. (A) Configurational coordinate diagram for the HS $[Co(bpy)_3]^{2+}$ complex along the Co−N totally symmetric normal mode and temperature-dependent single-crystal absorption spectra of the HS compound $[Co(bpy)_3][NaRh(ox)_3]$. (B) Configurational coordinate diagram for the LS $[Co(bpy)_3]^{2+}$ complex and temperature-dependent single-crystal absorption spectra of the spin-crossover compound $[Co(bpy)_3][LiRh(ox)_3]$.

ing t_{2g} and antibonding e_g orbitals. The change of spin states is therefore accompanied by a variation of the molecular volume, $\Delta V_{HL} = V_{HS} - V_{LS}$, estimated to ~17 Å^3.[5] The molecular volume of the complex being larger in the HS state than in the LS state, the observed inversion of the energetic order of the two spin states can be rationalized in terms of an internal or chemical pressure exerted by the confining environment of the hosts on the $[Co(bpy)_3]^{2+}$ guest which destabilizes the HS state with respect to the LS state.[4,5] Figure 1 helps illustrate this phenomenon: in $[Co(bpy)_3]$-$[NaRh(ox)_3]$ (Figure 1A), the complex has a classic HS electronic ground state, as shown in the configurational diagram along the Co−N distance coordinate. The zero-point energy E°_{LS} of the LS state is only slightly higher than the zero-point energy E°_{HS} of the HS state ($\Delta E^\circ_{HL} = E^\circ_{HS} - E^\circ_{LS} < 0$). The population of the HS state being favored by the entropy, this state remains the only populated state at all temperatures, in accordance with the temperature-dependent single-crystal absorption spectra which are typical for the HS species. On passing to $[Co(bpy)_3][LiRh(ox)_3]$ (Figure 1B), the size of the cavity occupied by $[Co(bpy)_3]^{2+}$ is reduced because of the substitution of the Na^+ by the smaller Li^+ cations on the oxalate backbone. The resulting increase in the chemical pressure exerted onto the complex destabilizes the HS state with respect to the LS state sufficiently for $[Co(bpy)_3]^{2+}$ to become a spin-crossover complex ($\Delta E^\circ_{HL} > 0$). The low-temperature (12 K) absorption spectrum is typical for the LS species; the spectra recorded at higher temperatures allow one to follow the entropy-driven thermal transition to the HS state.

Although the concept of chemical pressure provides insight into the manner in which the electronic properties of transition-metal complexes can be tuned by the guest−host interactions,[6] it does not provide a detailed picture of the involved interactions. The use of quantum chemical methods should help improve our understanding of such guest−host interactions. However, prior to the investigation of these interactions for $[Co(bpy)_3]^{2+}$ incorporated in the cavities of zeolite Y or oxalate networks, the ability of such methods to provide an accurate description of the isolated complex has to be examined.

For the present study of $[Co(bpy)_3]^{2+}$, methods based on density functional theory (DFT) only were considered, because the large number of atoms in the complex precludes the use of computationally demanding high-level ab initio methods. The calculations were performed within the Kohn−Sham (KS) formulation of spin-density functional theory,[7−12] wherein the ground-state energy E_0 of an N-electron system in an external local potential $v(\mathbf{r})$ is expressed as a functional of the spin-up $\rho_\uparrow(\mathbf{r})$ and spin-down $\rho_\downarrow(\mathbf{r})$ densities as in eq 1. $\rho(\mathbf{r}) = \rho_\uparrow(\mathbf{r}) + \rho_\downarrow(\mathbf{r})$ is the charge density; $T_s[\rho_\uparrow,\rho_\downarrow]$ is the

$$E_0[\rho_\uparrow,\rho_\downarrow] = T_s[\rho_\uparrow,\rho_\downarrow] + \frac{1}{2}\int \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'\,\frac{\rho(\mathbf{r})\,\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r})\,v(\mathbf{r}) + E_{xc}[\rho_\uparrow,\rho_\downarrow] \quad (1)$$

kinetic energy of the KS noninteracting N-electron system with the same spin densities as those of the interacting system and is given by eqs 2 and 3

$$T_s[\rho_\uparrow,\rho_\downarrow] = \int d\mathbf{r}\, [\tau_\uparrow(\mathbf{r}) + \tau_\downarrow(\mathbf{r})] \tag{2}$$

$$\tau_\sigma(\mathbf{r}) = \frac{1}{2}\sum_i f_{i\sigma}|\nabla\psi_{i\sigma}(\mathbf{r})|^2 \tag{3}$$

with $\sigma = \uparrow, \downarrow$. $f_{i\sigma}$ and $\psi_{i\sigma}(\mathbf{r})$ are the occupation number and the spin orbital of the ith σ-type single-particle level of the KS noninteracting system.

The exchange-correlation (XC) energy, $E_{xc}[\rho_\uparrow,\rho_\downarrow]$, is the sum of the exchange and the correlation energy, $E_x[\rho_\uparrow,\rho_\downarrow]$ and $E_c[\rho_\uparrow,\rho_\downarrow]$, respectively, for which approximate functionals are devised, requiring that either the largest number of the mathematical properties of the exact exchange and correlation functionals or, at least, a few conditions known to be limiting for the evaluation of selected properties are satisfied. The approximate XC functionals of widespread use either belong to the family of local and semilocal approximations or are hybrid functionals. Local and semilocal functionals admit the general form given in eq 4.

$$E_{xc}[\rho_\uparrow,\rho_\downarrow] = \int d\mathbf{r}\, \rho(\mathbf{r})\, \epsilon_{xc}\rho_\uparrow(\mathbf{r}),\rho_\downarrow(\mathbf{r}),\nabla\rho_\uparrow(\mathbf{r}),\nabla\rho_\downarrow(\mathbf{r}),\nabla^2\rho_\uparrow(\mathbf{r}),\nabla^2\rho_\downarrow(\mathbf{r}),\tau_\uparrow(\mathbf{r}),\tau_\downarrow(\mathbf{r}) \tag{4}$$

ϵ_{xc} is the XC energy *per* particle. In the early local density approximation (LDA),[8,13−16] it only depends on the spin densities. Despite being valid only for slowly varying densities, the LDA performs reasonably well for the description of atomic and molecular systems, but the accuracy of energy parameters is not always sufficient.

Semilocal approximations improve on the LDA by including corrections which reflect the strongly inhomogeneous densities of real systems. Thus, within the generalized gradient approximation (GGA), ϵ_{xc} depends on the spin densities but also on their gradients $\nabla\rho_\sigma(\mathbf{r})$. The GGA outperforms the LDA,[17−20] and this nowadays more than compensates for the increase in the computation time tied to the evaluation of the gradients. The meta-generalized gradient approximation (meta-GGA) goes beyond the GGA in that ϵ_{xc} becomes also a function of the second derivatives of the spin densities $\nabla^2\rho_\sigma(\mathbf{r})$ or the kinetic energy densities $\tau_\sigma(\mathbf{r})$, or of both. Although the evaluation of the semilocal variables incurs a further increase in the computational time as compared to the GGA, their inclusion allows the enrichment, in a straightforward manner, of the built-in physics of the approximate functionals (see refs 21−24). Hybrid functionals[25,26] are characterized by the fact that they include a contribution of the exact-exchange or Hartree−Fock (HF) exchange energy $E_x^{HF}[\rho_\uparrow,\rho_\downarrow]$ given by eq 5.

$$E_x^{HF}[\rho_\uparrow,\rho_\downarrow] = -\frac{1}{2}\sum_{\sigma=\uparrow,\downarrow}\int d\mathbf{r}\, d\mathbf{r}' \frac{|\sum_{i\sigma} f_{i\sigma}\psi_{i\sigma}(\mathbf{r})\, \psi^*_{i\sigma}(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|} \tag{5}$$

The evaluation of $E_x^{HF}[\rho_\uparrow,\rho_\downarrow]$ and the presence of the nonlocal exchange operator in the KS equations put the computational demand of the hybrid methods on the same level as that for the HF method. Note that the correct contribution of exact exchange in the hybrids is the subject of ongoing studies.[27−29] The commonly admitted value is around 25% as in the B3LYP[26,30] and PBE0 (also known as PBE1PBE)[27,29,31,32] functionals. Dropping the explicit reference to the spin densities in the expression of the functionals, the B3LYP hybrid is given by eq 6

$$E_{xc}^{B3LYP} = (1 - a_0)E_x^{LDA} + a_0E_x^{HF} + a_x\Delta E_x^B + (1 - a_c)E_c^{VWN} + a_cE_c^{LYP} \tag{6}$$

E_c^{VWN} is the Vosko−Wilk−Nusair[14] parametrization of the random phase approximation correlation energy; ΔE_x^B stands for Becke's 1988 gradient corrections to the LDA exchange energy,[20] and E_c^{LYP} is the Lee−Yang−Parr (LYP) correlation GGA.[33] The coefficients were determined by fitting to a data set of atomization energies (a_0, a_x, a_c) = (0.20, 0.72, 0.81).[26] The PBE0 hybrid is parameter-free and is given by eq 7

$$E_{xc}^{PBE0} = E_{xc}^{PBE} + a_0(E_x^{HF} - E_x^{PBE}) \tag{7}$$

with $a_0 = {}^1/_4$; PBE refers to the recent XC GGA of Perdew et al.[34,35] Although semilocal and hybrid functionals are routinely used in DFT applications, the quality that can be achieved for the description of a given property is not guaranteed to be as satisfactory as one may expect nor to increase with the degree of refinement of the functionals used.

We aim to achieve an accurate description of the structure and the energetics of $[Co(bpy)_3]^{2+}$ in the LS and the HS states. In that respect, most post-LDA functionals give a very satisfactory description of the geometry of transition-metal complexes. However, they fail to reliably predict the electronic energy difference between states of different spin multiplicities, as shown in recent DFT studies performed on d^6 iron(II) complexes in the quintet HS and the singlet LS states.[36−43] The LDA strongly overestimates the electronic HS−LS energy difference given by eq 8

$$\Delta E_{HL}^{el} = E_{HS}^{el} - E_{LS}^{el} \tag{8}$$

The situation improves with the GGA functionals, which, however, still tend to overestimate ΔE_{HL}^{el}. In contrast, hybrid functionals tend to underestimate the HS−LS energy difference. The difficulties met by the approximate functionals for the evaluation of ΔE_{HL}^{el} are due to the fact that they do not properly account for the variation of exchange when the spin polarization and the metal−ligand bond length significantly vary, as is the case upon the LS ↔ HS conversion in octahedral d^4 to d^7 transition-metal complexes. Promisingly, recent functionals were reported to give satisfactory results for iron(II) complexes.[37−43] These are the revised PBE (RPBE) XC GGA functional of Hammer et al.,[44] the OLYP functional which combines the OPTX GGA exchange of Handy and Cohen[45] with the LYP GGA correlation, and the B3LYP* reparametrized hybrid of Reiher and co-workers,[37,38] which differs from the B3LYP by the fact that the amount of exact exchange is reduced to 15% to allow a better description of the spin-state energetics in transition-metal complexes. Note that the study of the dependence of the calculated HS−LS energy difference in the $[Fe(bpy)_3]^{2+}$ complex on the exact-exchange contribution a_0 in the PBE0 and B3LYP hybrids (eqs 6 and 7) supports the idea that this contribution should effectively be reduced to some 10%.[42]

The fact that with regard to energy calculations the situation is very delicate requires a careful assessment of the performance of the different XC functionals for correctly predicting the HS state as the ground state of the isolated $[Co(bpy)_3]^{2+}$ complex. Thus, in addition to the above-mentioned PBE, RPBE, OLYP, PBE0, B3LYP, and B3LYP* functionals, we have characterized $[Co(bpy)_3]^{2+}$ in the HS and LS states, using the OPBE GGA made of the OPTX exchange and the PBE correlation functional, the HCTH (also known as HCTH/407) GGA;[46] the O3LYP hybrid, similar in form to the B3LYP, but with ΔE_x^B replaced by the OPTX gradient correction to the LDA exchange ΔE_x^{OPTX} and with reoptimized parameter values;[47] and also the VSXC[48] and TPSS[23] meta-GGAs.

The photophysical properties of $[Co(bpy)_3]^{2+}$ provide clear-cut evidence for its spin state, as shown by the absorption spectra in Figure 1. The low-energy absorption bands present in the spectra of the complex in either spin state could be assigned to d−d transitions on the basis of ligand field theory (LFT).[3,4] The higher energy bands observed for the complex in the LS state were likewise attributed to d−d transitions,[3] but this assignment was later questioned.[4] To assign these electronic transitions, we performed electronic excitation calculations within linear response theory in time-dependent density functional theory (TDDFT).[49−52]

Our paper is organized as follows. Section 2 gives the technical details. In section 3, we present the results obtained for the geometry and the energetics of the complex in the two spin states. This includes the results obtained for the analysis of the trigonal splitting in the HS state and of the Jahn−Teller (JT) instability in the LS state. In section 4, we present the results obtained for the optical and chiroptical properties of the complex.

2. Computational and Experimental Details

The Amsterdam Density Functional (ADF)[53,54] program package was used to perform calculations with the PBE, OPBE, RPBE, and OLYP GGAs. Calculations with the B3LYP, B3LYP*, O3LYP, and PBE0 hybrids and with the VSXC and TPSS meta-GGAs were carried out with the Gaussian program package.[55] The two packages use different types of basis sets, namely, Slater-type orbital (STO) basis sets for ADF and Gaussian-type orbital (GTO) basis sets for Gaussian. Therefore, to be able to compare the results obtained with the two package, we used for all atoms the STO TZP basis set from the ADF basis set database and the GTO TZVP basis set,[56] respectively, which both are large basis sets of valence triple-ζ polarized quality and which were shown to perform similarly for the description of $[Fe(bpy)_3]^{2+}$ in the LS and HS states.[42] To assert the similar performance of the TZP and TZVP basis sets, to which we will refer as the $\mathcal{S}$ and the $\mathcal{G}$ basis sets, respectively, the Gaussian package was also used to perform calculations with the PBE GGA. Fractional occupation numbers were used in the ADF calculations but are not allowed in the present implementation of Gaussian (except as an intermediate step in the self-consistent field convergence strategy). For the two spin states, spin-unrestricted calculations were performed constraining M_S, the projection of the total electronic spin, to $M_S = +^1/_2$ and $M_S = +^3/_2$, for the complex in the LS and HS states, respectively. For the geometry optimization calculations, except for those carried out for the study of the JT instability in the LS state, the symmetry of the complex was constrained to D_3. For the study of the JT effect in the LS state, the symmetry was lowered to C_2.

Single crystals of $[Co(bpy)_3][NaRh(ox)_3]$ and $[Co(bpy)_3][LiRh(ox)_3]$ were grown as described in ref 5 and prepared for the optical measurements as explained in ref 4. The absorption spectra were recorded on a Fourier transform spectrometer equipped with the appropriate light sources, beam splitters, and detectors (Bruker IFS66). Temperatures between 11 and 300 K were achieved with a closed-cycle cryosystem with the sample sitting in He exchange gas (Oxford Instruments CCC 1204). The circular dichroism (CD) spectrum of $[Co(bpy)_3][LiRh(ox)_3]$ was recorded at 15 K using the apparatus described in ref 57. For analyzing the spectra of the complex in either spin state, the energy E, the oscillator strength f, and the rotatory strength R of its 80 lowest-lying dipole-allowed electronic excitations were calculated within the adiabatic local density approximation of TDDFT as implemented in the Gaussian package.[49−52] The calculations were performed at the B3LYP/$\mathcal{G}$ level on optimized HS and LS geometries obtained at the same theoretical level. The individual absorption and CD bands were simulated by convoluting the oscillator and rotatory strengths with Gaussian functions having a full width at half-maximum of 2000 cm^{-1}. For the CD bands, the normalization of the Gaussian functions is such that[58]

$$R = 22.97 \times 10^{-40} \int \mathrm{d}E \frac{\Delta\epsilon(E)}{E} \tag{9}$$

where R is in cgs units; the difference $\Delta\epsilon$ in the extinction coefficients for left- and right-hand circular polarized light is given in M^{-1} cm^{-1}.

The oscillator strength f_{ij} and the rotatory strength R_{ij} of an $i \rightarrow j$ transition of energy E_{ij} actually read, in atomic units ($m_e = 1$, $e = 1$, $\hbar = 1$, $4\pi\epsilon_0 = 1$, $c \approx 137.036$),

$$f_{ij} = \frac{2}{3} E_{ij} |\mu_{ij}|^2 \tag{10}$$

where $\mu_{ij} = \langle i|\mu|j\rangle$ and $\mathbf{m}_{ij} = \langle i|\mathbf{m}|j\rangle$ are the electric and

$$R_{ij} = \mathrm{Im}(\mu_{ij} \cdot \mathbf{m}_{ji})$$

magnetic transition dipole moment vectors, respectively. The magnetic dipole moment operator is given by $\mathbf{m} = (-i/2c)\mathbf{r} \times \nabla$. The electric dipole moment operator is given in the dipole-length form by $\mu = -\mathbf{r}$, and in the dipole-velocity form, its matrix elements read $\mu_{ij} = -(1/E_{ij})\nabla_{ij}$. The two representations are equivalent in a complete basis set, but

Table 1. Comparison between Experimental and Optimized HS $^4A_2$ $[Co(bpy)_3]^{2+}$ Geometries

		ADF			
	exp.[a]	PBE/$\mathcal{S}$	OPBE/$\mathcal{S}$	RPBE/$\mathcal{S}$	OLYP/$\mathcal{S}$
		Bond Lengths [Å]			
Co−N	2.123	2.141	2.149	2.180	2.155
N−C_2	1.350	1.359	1.350	1.365	1.355
N−C_6	1.338	1.347	1.339	1.352	1.343
C_2−C_3	1.386	1.401	1.397	1.408	1.400
C_3−C_4	1.381	1.393	1.387	1.398	1.390
C_4−C_5	1.364	1.395	1.390	1.401	1.393
C_5−C_6	1.369	1.392	1.387	1.397	1.389
C_2−C_2'	1.488	1.480	1.475	1.489	1.480
		Angles [deg]			
β	77.4	76.7	76.0	76.6	75.9
γ	0.5	4.4	4.3	4.5	4.5
τ	47.7	50.7	50.5	50.8	50.1
θ	58.6	60.2	60.6	60.3	60.5

	Gaussian							
	PBE/$\mathcal{G}$	PBE0/$\mathcal{G}$	B3LYP/$\mathcal{G}$	B3LYP*/$\mathcal{G}$	O3LYP/$\mathcal{G}$	HCTH/$\mathcal{G}$	VSXC/$\mathcal{G}$	TPSS/$\mathcal{G}$
				Bond Lengths [Å]				
Co−N	2.145	2.156	2.182	2.173	2.139	2.187	2.143	2.148
N−C_2	1.360	1.344	1.350	1.352	1.343	1.351	1.352	1.360
N−C_6	1.347	1.334	1.339	1.340	1.332	1.338	1.346	1.347
C_2−C_3	1.402	1.391	1.395	1.397	1.389	1.397	1.396	1.400
C_3−C_4	1.393	1.385	1.388	1.389	1.382	1.386	1.395	1.392
C_4−C_5	1.396	1.386	1.389	1.390	1.383	1.388	1.402	1.395
C_5−C_6	1.393	1.384	1.387	1.388	1.381	1.386	1.392	1.391
C_2−C_2'	1.481	1.480	1.485	1.484	1.472	1.479	1.481	1.480
				Angles [deg]				
β	76.7	76.1	75.8	76.0	76.4	75.4	78.0	76.5
γ	4.5	5.2	4.8	4.8	5.0	4.2	33.3	4.1
τ	50.9	50.5	49.9	50.1	50.8	50.1	58.6	51.1
θ	60.3	60.5	60.5	60.452	60.4	60.8	63.0	60.5

[a] Experimental values from ref 4.

there are issues tied to their use in finite basis sets.[52,59−61] Indeed, the dipole-length representation gives origin-dependent rotatory strengths, which is not the case for the dipole-velocity form. This latter form, however, gives rotatory strengths which are sensitive to the quality of the wave functions. In the present case, the two representations have been employed. We only present the results obtained with the dipole-length representation, which generally gives more reliable results.[61]

3. Geometry and Energetics of $[Co(bpy)_3]^{2+}$

3.1. $[Co(bpy)_3]^{2+}$ in the HS State. The calculations carried out for the study of the D_3 complex in the HS state led to its characterization in the $^4A_2$ and ^4E states, which result from the trigonal splitting of the octahedral HS $^4T_{1g}$ state. The $^4A_2$ state proves to be the lower-lying trigonal component of the HS state. This was also evidenced experimentally.[3] Because of the difference in the way they handle orbitally degenerate states, ADF actually allowed the characterization of the complex in both states whereas Gaussian was restricted to the $^4A_2$ state.

3.1.1. Optimized HS Geometries. 3.1.1.1. The $^4A_2$ State. Table 1 gives selected bond lengths and the values for the angles β, γ, τ, and θ (see Figure 2) for the optimized $^4A_2$ geometries. The experimental values included in Table 1 are taken from the X-ray crystal structure reported for $[Co(bpy)_3][NaCr(ox)_3]$.[4] In this oxalate network, $[Co(bpy)_3]^{2+}$ has a C_3 site symmetry. The experimental values are averaged such that they can be compared to those calculated in D_3 symmetry.

Table 1 shows that, with the exception of the VSXC/$\mathcal{G}$ results, the optimized parameter values are consistent with one another and compare well with the experimental values. This is, in particular, the case for the Co−N bond length. As experimentally observed, the N−C_2 bond remains longer than the N−C_6 bond. Similarly, the optimized values of the angular parameters are close to the experimental values. This indicates that the arrangement of the ligands around the metal is well-described in the optimized geometries. It is noteworthy that the geometries obtained with the PBE GGA and the $\mathcal{S}$ and $\mathcal{G}$ basis sets are very close to each other, emphasizing the previously reported[42] equivalence of the STO and GTO basis sets. Generally, the optimized HS geometries are slightly more expanded than experimentally observed. The minor differences can be ascribed to the absence of counterion and packing effects in gas-phase calculations.

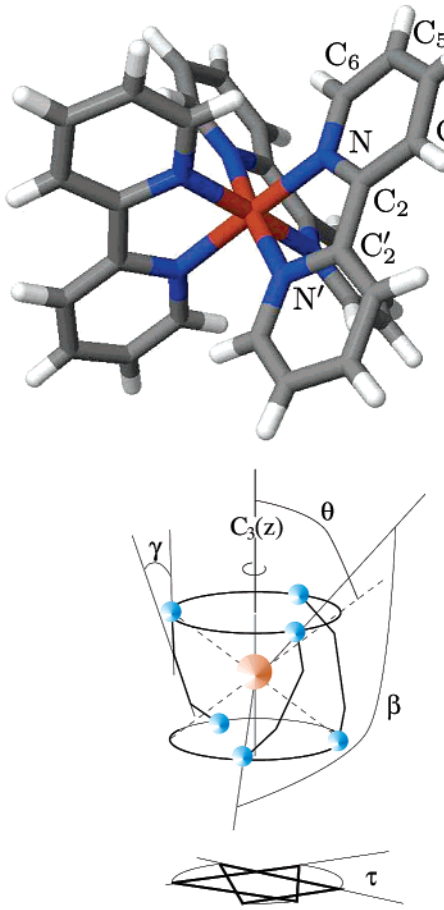


Figure 2. Atom labeling used for the D_3 $[Co(bpy)_3]^{2+}$ complex (top) and angles characterizing the arrangement of the ligands around the cobalt center (bottom): β is the bite angle, γ the dihedral angle between the ligand moieties, τ the twist angle, and θ the angle between the z axis and the generator of the cone on which the nitrogen atoms are located.

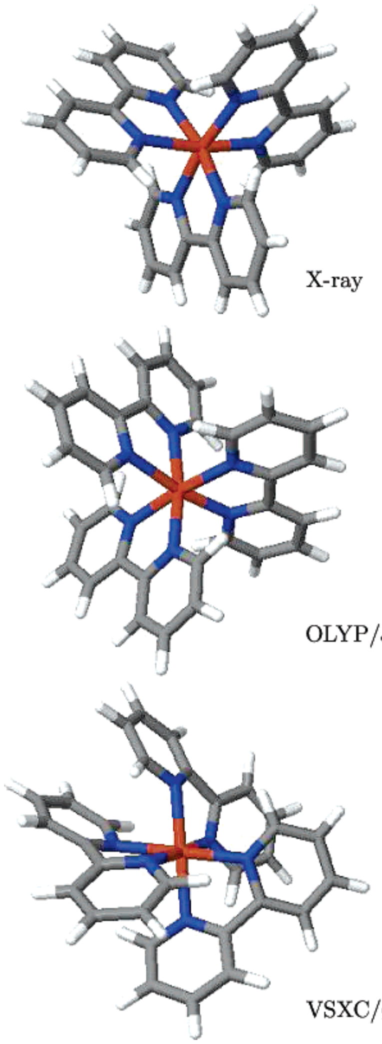


Figure 3. X-ray crystal structure[4] and optimized OLYP/$\mathcal{S}$ and VSXC/$\mathcal{G}$ geometries of the HS $[Co(bpy)_3]^{2+}$ complex.

Table 2. Optimized HS ^4E $[Co(bpy)_3]^{2+}$ Geometries (ADF Results)

	PBE/$\mathcal{S}$	OPBE/$\mathcal{S}$	RPBE/$\mathcal{S}$	OLYP/$\mathcal{S}$
		Bond Lengths [Å]		
Co−N	2.141	2.149	2.180	2.159
N−C_2	1.359	1.352	1.365	1.356
N−C_6	1.347	1.340	1.353	1.344
C_2−C_3	1.401	1.398	1.407	1.401
C_3−C_4	1.393	1.388	1.399	1.391
C_4−C_5	1.395	1.390	1.400	1.393
C_5−C_6	1.392	1.388	1.397	1.390
C_2−C_2'	1.479	1.476	1.490	1.482
		Angles [deg]		
β	76.7	75.9	75.8	76.0
γ	4.9	4.7	5.0	4.8
τ	50.3	50.1	49.7	50.1
θ	60.0	60.5	60.4	60.5

Frequency calculations performed with the Gaussian package show that the geometries obtained with this package are true minima because no imaginary frequency was found. Hence, because of the consistency between the geometries obtained with both Gaussian and ADF, it can be concluded that the above results indeed describe the most stable geometry of $[Co(bpy)_3]^{2+}$ in the $^4A_2$ state. The VSXC geometry markedly differs from the experimental and other calculated HS geometries by its angular parameter values (see Table 1 and Figure 3). The ligands are very much twisted and this in a way that favors π stacking of the neighboring pyridinyl rings. This is a direct consequence of the known tendency of the VSXC functional to overestimate long-range interactions such as $\pi-\pi$ interactions.[43,62,63] Therefore, the VSXC functional is not adequate for the description of the geometric structure for the type of transition-metal complex considered here.

3.1.1.2. The ^4E State. Table 2 gives the geometrical parameters of the ^4E HS state calculated using ADF. The comparison of these data with those obtained at the same theoretical levels for the $^4A_2$ state (see Table 1) shows that the geometries of the complex in the two components of the HS are very similar.

The lack of significant geometrical change in passing from one quartet component to the other stems from the fact that, upon inspection of the KS orbital levels, passing from the $^4A_2$ state to the ^4E state involves the transfer of an of electron from the metallic $e_{\downarrow}$ spin-down orbital level of Co(t_{2g}) parentage to the metallic $a_{1\downarrow}$ spin-down orbital level of the

Table 3. Calculated $E(^4E) - E(^4A_2)$ Energy Differences (cm^{-1}) Obtained with ADF

	$E(^4E) - E(^4A_2)$
PBE/$\mathcal{S}$	1968
OPBE/$\mathcal{S}$	2000
RPBE/$\mathcal{S}$	2194
OLYP/$\mathcal{S}$	1848

same Co(t_{2g}) parentage. Indeed, upon the $O_h \rightarrow D_3$ symmetry lowering, the octahedral metallic t_{2g} level splits into a_1 and e levels, and the occupation of the spin-down orbital levels for the $^4A_2$ state is $(e_\downarrow)^2(a_{1\downarrow})^0$, and that for the ^4E state is $(e_\downarrow)^1$-$(a_{1\downarrow})^1$. These levels are nonbonding, which explains why no significant variation in the geometry of the complex is expected.

3.1.2. The Trigonal Splitting in the HS State. Table 3 gives the values found for the electronic energy difference between the ^4E and $^4A_2$ HS components at the respective optimized geometries. All four functionals predict that the $^4A_2$ component is lower in energy than the ^4E component. The calculated energy differences range from 1848 to 2194 cm^{-1}. However, this spread is comparatively small, all of the functionals predicting a trigonal splitting of about 2000 cm^{-1}. This value is in reasonable agreement with the experimental estimate of 1000 cm^{-1}obtained from the fitting of the magnetization data of the HS $[Co(bpy)_3](ClI_4)_2$ compound.[3]

The ^4E state is JT unstable.[64] However, given that this instability is due to the partial occupation of the degenerate metallic nonbonding e orbital level of Co(t_{2g}) parentage, this is not expected to have a significant influence on the results obtained for the energetics of the HS state. As will be discussed below, this is not the case for the LS ^2E state, for which the JT instability results from the single occupation of the doubly degenerate metallic antibonding orbital level of octahedral Co(e_g) parentage.

3.2. $[Co(bpy)_3]^{2+}$ in the LS State. This section first presents the results obtained for the characterization of the LS ^2E state of $[Co(bpy)_3]^{2+}$ in the high-symmetry D_3 point group. This is followed by a discussion of the JT instability in this state.

3.2.1. LS ^2E Geometries in D_3. Table 4 shows selected bond lengths and the values of the angles β, γ, τ, and θ found for the optimized LS ^2E geometries. The experimental data are taken from the X-ray crystal structure[4] of the spin-crossover compound $[Co(bpy)_3][LiCr(ox)_3]$ at 10 K, where $[Co(bpy)_3]^{2+}$ is effectively in the LS state. This compound is isostructural with $[Co(bpy)_3][NaCr(ox)_3]$, with a C_3 site symmetry for $[Co(bpy)_3]^{2+}$. The optimized values are again very consistent among the functionals used, and they are in good agreement with the experimental values. We may conclude that the LS structure is described well enough by the different functionals and may thus be used to discuss the structural changes upon spin-state transition. The most important feature to be noticed is that the LS → HS transition is characterized by a lengthening of the Co−N bonds by ~0.10 Å, which is due to the promotion of an electron from the nonbonding metallic levels of t_{2g} parentage to the antibonding metallic level of e_g parentage. Concomitantly, there is a marked rearrangement of the ligands around the metal ion, manifesting itself by quite large changes in the angular parameters (see Tables 1 and 4).

Table 4. Comparison between Experimental and Optimized LS ^2E Geometries (ADF Results)

	exp.[a]	PBE/$\mathcal{S}$	OPBE/$\mathcal{S}$	RPBE/$\mathcal{S}$	OLYP/$\mathcal{S}$
			Bond Lengths [Å]		
Co−N	2.024	2.039	2.031	2.079	2.063
N−C_2	1.343	1.362	1.355	1.368	1.359
N′−C_6	1.338	1.346	1.342	1.354	1.346
C_2−C_3	1.399	1.400	1.397	1.407	1.400
C_3−C_4	1.388	1.393	1.387	1.398	1.391
C_4−C_5	1.373	1.396	1.391	1.401	1.393
C_5−C_6	1.389	1.392	1.388	1.398	1.390
C_2−C_2'	1.481	1.480	1.466	1.480	1.474
			Angles [deg]		
β	80.1	79.9	79.6	78.9	79.0
γ	0.1	6.0	6.1	5.6	6.4
τ	51.2	52.4	52.7	51.9	53.2
θ	58.1	58.7	59.0	59.2	59.2

[a] Experimental results taken from ref 4.

3.2.2. The Jahn−Teller Instability in the LS State. A complete study of $[Co(bpy)_3]^{2+}$ in the LS state requires that one also takes into account the JT instability[64] due to the orbital degeneracy of the LS ^2E state. Even though the X-ray structure shows no evidence for a symmetry lowering for the LS state, electron paramagnetic resonance (EPR) spectra indicate that a dynamic or statically disordered JT effect must be operative.[65]

For such a large molecular system as $[Co(bpy)_3]^{2+}$ with numerous (177) vibrational degrees of freedom, the JT effect[64] in the LS ^2E state is probably a multimode one. That is, it may involve the coupling of the 2-fold degenerate electronic state to several vibrational modes of a_1, a_2, or e symmetries, according to the decomposition $E \otimes E = A_1 + A_2 + E$, which gives the symmetries of the vibrational modes that could couple linearly to the ^2E state. For displacements along an a_1 vibrational mode, the full D_3 symmetry of the complex is preserved, and for displacements along an a_2 mode, the C_3 symmetry is maintained. Thus, only the e modes can be JT-active and lift the orbital degeneracy. For a full discussion of the JT effect and the many aspects of vibronic interactions, see refs 66−70. In the following, we restrict ourselves to the characterization of the extrema of low symmetry on the LS adiabatic potential energy surface (APES). Therefore, the geometry of the complex was optimized while constraining the symmetry of the molecular framework to C_2. This led to the identification of two extrema, which correspond to electronic states of ^2A and ^2B symmetry. The vibrational analysis performed for the ^2A and ^2B geometries revealed zero and one imaginary frequency, respectively. The vibrational mode of imaginary frequency found for the ^2B geometry is of b symmetry and transforms the ^2B geometry into the ^2A geometry. Consequently, the ^2A geometry corresponds to a minimum and the ^2B geometry to a saddle point between two such minima. The trigonal symmetry of the JT problem implies that there actually exist three equivalent minima and three associated equivalent saddle points. Additionally, the presence of the saddle points

Table 5. Selected Distances (Å) and Angles (deg) Characterizing the JT-Distorted Geometry of $[Co(bpy)_3]^{2+}$ at the Minima (A) and Saddle Points (B) of the LS APES

Part A

	L1				L2				
	Co−N = Co−N′	C_2−C_2′	β	γ	Co−N	Co−N′	C_2−C_2′	β	γ
	ADF Results								
PBE/$\mathscr{S}$	1.956	1.464	82.2	0.5	1.982	2.191	1.475	78.7	11.5
OPBE/$\mathscr{S}$	1.950	1.457	81.8	3.4	1.972	2.209	1.472	78.0	5.3
RPBE/$\mathscr{S}$	1.990	1.471	81.5	3.6	2.015	2.252	1.486	77.5	4.3
OLYP/$\mathscr{S}$	1.980	1.464	81.4	3.5	2.005	2.243	1.479	77.4	5.3
	Gaussian Results								
PBE/$\mathscr{G}$	1.959	1.464	82.1	0.4	1.985	2.190	1.476	78.7	11.9
PBE0/$\mathscr{G}$	1.982	1.468	81.6	0.0	2.000	2.218	1.477	77.9	10.1
B3LYP/$\mathscr{G}$	2.004	1.472	81.2	0.0	2.027	2.256	1.483	77.3	10.1
B3LYP*/$\mathscr{G}$	1.993	1.470	81.5	0.1	2.016	2.244	1.482	77.7	10.1
O3LYP/$\mathscr{G}$	1.953	1.457	82.1	0.0	1.974	2.199	1.469	78.4	10.8
HCTH/$\mathscr{G}$	1.974	1.460	81.6	0.1	2.011	2.261	1.475	77.3	12.4
TPSS/$\mathscr{G}$	1.961	1.464	82.1	0.4	1.987	2.193	1.476	78.6	11.5

Part B

	L1				L2				
	Co−N = Co−N′	C_2−C_2′	β	γ	Co−N	Co−N′	C_2−C_2′	β	γ
	ADF Results								
PBE/$\mathscr{S}$	2.129	1.477	77.3	13.9	1.935	2.057	1.467	81.2	3.7
OPBE/$\mathscr{S}$	2.172	1.473	75.6	17.3	1.925	2.063	1.463	80.6	0.1
RPBE/$\mathscr{S}$	2.202	1.486	75.8	18.5	1.959	2.094	1.476	80.5	1.8
OLYP/$\mathscr{S}$	2.188	1.480	75.7	17.7	1.947	2.083	1.469	80.4	0.2
	Gaussian Results								
PBE/$\mathscr{G}$	2.131	1.478	77.3	14.3	1.938	2.059	1.468	81.2	3.9
PBE0/$\mathscr{G}$	2.142	1.479	76.7	11.2	1.956	2.091	1.472	80.4	3.0
B3LYP/$\mathscr{G}$	2.178	1.484	76.0	11.1	1.988	2.122	1.477	79.8	3.1
B3LYP*/$\mathscr{G}$	2.167	1.483	76.3	11.4	1.969	2.105	1.475	80.3	3.0
O3LYP/$\mathscr{G}$	2.123	1.471	77.1	12.1	1.929	2.062	1.461	80.9	3.5
HCTH/$\mathscr{G}$	2.193	1.478	75.5	14.8	1.952	2.091	1.465	80.4	3.9
TPSS/$\mathscr{G}$	2.128	1.478	77.2	13.8	1.941	2.061	1.468	81.0	4.1

indicates that quadratic vibronic coupling is effective. Thus, assuming that the JT instability is well-described by an ideal single-mode E ⊗ e problem, the section of the LS APES along the effective e mode resembles the warped Mexican hat potential.[71]

3.2.2.1. JT-Distorted Geometries of $[Co(bpy)_3]^{2+}$ at the Minima and the Saddle Points. In these geometries, one ligand (L1) lies on the C_2 axis and is left unchanged by the C_2 rotation, contrarily to the other two ligands (L2), which are interchanged by this operation. In the L1 ligand moiety, the two Co−N bond lengths are the same, while for L2, they differ. Parts A and B of Table 5 give the Co−N and C_2−C_2′ bond lengths and the values of the angles β and γ calculated with all of the previously used functionals (with the exception of the VSXC functional) for the ^{2}A and ^{2}B geometries, respectively. All functionals predict very similar geometries for the complex at the minima and at the saddle points. There is also a nice match between the values obtained at the PBE/$\mathscr{S}$ and PBE/$\mathscr{G}$ levels for the structural parameters, which illustrates the equivalence of these GTO and STO basis sets.

From the comparison of the structural data given in Tables 4 and 5, the JT distortion to which the geometry of the complex is subjected in going from the D_3 apex to one of the minima can be depicted as follows. Starting from the D_3 geometry, one ligand moves closer to the cobalt atom. Its displacement takes place along the C_2 axis in such a way that its local C_2 symmetry is preserved. For this ligand, which will be designated L1, the shortening of the Co−N bonds amounts to ∼0.08 Å. The other two ligands undergo rearrangements while remaining equivalent to the C_2 rotation that leaves L1 unchanged. For each of these two ligands, designated L2, one Co−N bond shortens by ∼0.07 Å while the other lengthens by ∼0.17 Å. Compared to L1, the Co−N bonds in L2 are longer by ∼0.02−0.03 Å for one of the pyridinyl moieties and by ∼0.24−0.27 Å for the other. The larger distance of the L2 ligands to the cobalt is accompanied by a lengthening of the C_2−C_2′ bonds by ∼0.01 Å, a decrease of the bite angle β by about 4°, and a generalized increase of the dihedral angle γ, which helps maintain optimal metal−ligand orbital interactions.

The distortion that brings $[Co(bpy)_3]^{2+}$ from its high-symmetry configuration to its geometry at one of the saddle points can be described in the same manner. In this case, however, the L1 ligand moves away from the metal atom,

Table 6. Low-Spin $[Co(bpy)_3]^{2+}$ Complex: Values of the Co−N Bond Length in D_3 Symmetry and Averaged Values of the Co−N Bond Lengths in C_2 Symmetry (ADF Results)

	PBE/$\mathcal{S}$	OPBE/$\mathcal{S}$	RPBE/$\mathcal{S}$	OLYP/$\mathcal{S}$
D_3	2.039	2.031	2.079	2.063
C_2 (av.)	2.042	2.049	2.085	2.074

Table 7. Calculated Values of the JT Stabilization Energy E_{JT} and Barrier Height Δ[a]

	E_{JT} [cm^{-1}]	Δ [cm^{-1}]
	ADF Results	
PBE/$\mathcal{S}$	1542	308
OPBE/$\mathcal{S}$	1443	391
RPBE/$\mathcal{S}$	1466	419
OLYP/$\mathcal{S}$	1483	334
	Gaussian Results	
PBE/$\mathcal{G}$		304
PBE0/$\mathcal{G}$		342
B3LYP/$\mathcal{G}$		282
B3LYP*/$\mathcal{G}$		320
O3LYP/$\mathcal{G}$		383
HCTH/$\mathcal{G}$		354
TPSS/$\mathcal{G}$		−45 (311)

[a] The value given in parentheses for the TPSS functional was obtained using the “ultrafine” integration grid (see text).

which leads to a ∼0.10 Å increase of its Co−N bond lengths. As for the L2 ligands, their rearrangement translates into the shortening of one Co−N bond by about 0.11 Å and into a lengthening of the other by ∼0.02 Å. The Co−N bond length in L1 is about 2.13−2.20 Å; it is larger than the two different Co−N bond lengths of L2 by ∼0.07 Å and ∼0.20 Å. L1 being the most distant ligand to the cobalt, it exhibits a smaller bite angle, a slightly larger C_2-C_2' distance, and a larger dihedral angle γ than those of L2.

In the solid state, the JT distortion can be less than predicted because of the steric constraints imposed by the environment of the complex. If the JT effect is static, then there is a disorder of the JT distortion over the three C_2 axes, which explains the high symmetry of the X-ray structure of LS $[Co(bpy)_3]^{2+}$. If the JT effect is dynamic, then, on the time scale of the X-ray measurements, the quantum-mechanical zero-point delocalization of the complex around the moat of the LS APES averages out the instantaneous JT distortions. From the discussion above, the average value of the Co−N bond lengths in C_2 symmetry is expected to be quite close to the value of the single Co−N bond length in D_3 symmetry. That this is indeed the case is confirmed in Table 6.

3.2.2.2. The Energetics of the JT Effect. The energetic terms in the JT effect are the JT stabilization energy E_{JT}, which is the energy difference between the D_3 apex and the global minimum of the LS APES, and the height Δ of the energy barrier between the minima, which is the energy difference between the minima and the saddle points. The values found for these two energetic parameters are summarized in Table 7. All functionals give for either parameter very consistent values, with deviations that fall within the chemical accuracy of 350 cm^{-1} (≈1 kcal/mol).

The negative Δ value obtained with the TPSS meta-GGA corresponds to an unphysical description of the LS APES wherein the global minima are higher in energy than the saddle points. It is, however, well-known that the meta-GGAs are sensitive to the integration grid.[62] For the TPSS functional, the results obtained for the JT energetics can be significantly improved by using the “ultrafine” integration grid in Gaussian in place of the smaller default grid. This leaves the optimized geometries unchanged, but the calculated energy barrier now amounts to $\Delta = 311$ cm^{-1}, in very good agreement with the values obtained with the other functionals. In summary:

$$E_{JT} \approx 1500 \text{ cm}^{-1};\ \Delta \approx 340 \text{ cm}^{-1} \tag{12}$$

Further insight into the JT effect requires the determination of the JT-active modes and the characterization of the LS APES along each of the JT modes in order to determine their respective contributions to E_{JT} and Δ. This is a work in progress.

3.2.2.3. The Nature of the JT Effect. A static JT effect requires high barriers along each of the JT modes; that is, one must verify[69,71] $\delta_i \gg \hbar\omega_i$ $(i = 1, ..., p)$, where ω_i is the angular frequency and δ_i the contribution to Δ of the ith JT-active mode. Given that the JT distortions in the LS state mainly involve drastic changes in the Co−N bond lengths, the JT-active modes have to be among the vibrational modes which involve the Co−N bonds. For the complex in the HS state, the results of our vibrational analysis show that the frequencies of such modes are in the 110−270 cm^{-1} range, in agreement with the proposed identification of Co−N stretching modes in the far-infrared region.[72] The vibrational analysis similarly performed for the geometries found for the C_2 minima of the LS APES gives the frequencies of such modes in the 110−360 cm^{-1} range (for calculated vibrational frequencies for $[Co(bpy)_3]^{2+}$, see the Supporting Information). It is therefore reasonable to assume that the frequencies $\hbar\omega_i$ $(i = 1, ..., p)$ of the JT-active modes are in the same 110−360 cm^{-1} interval. Because the value found for Δ provides an upper bound for the δ_i's, it turns out that, for any JT-active mode “i”, δ_i is on the same order of magnitude as the vibrational quantum $\hbar\omega_i$: $\delta_i \lesssim \hbar\omega_i$, $(i = 1, ..., p)$. The condition for a static JT effect clearly is not obeyed.

3.3. The High-Spin/Low-Spin Energy Difference. Because the isolated $[Co(bpy)_3]^{2+}$ complex has a HS ground state, the zero-point energy difference between the HS and the LS state is negative:

$$\Delta E^{\circ}_{HL} = E^{\circ}_{HS} - E^{\circ}_{LS} < 0. \tag{13}$$

This inequality can therefore be used as a criterion to assess the functionals with regard to their ability to correctly predict the ground state. The zero-point energies E° are calculated as the sum of the electronic and vibrational contributions, E^{el} and E^{vib}, respectively. E^{el} is obtained from the geometry optimization, while E^{vib} can be accessed from the vibrational analysis performed on the optimized geometries within the harmonic approximation.

Given that the HS state is split by the trigonal field, the relevant zero-point energy in this spin state must equal that

Table 8. Calculated HS−LS Electronic Energy Differences[a]

	ΔE_{HL}^{el} [cm^{-1}]
ADF Results	
PBE/$\mathcal{S}$	3919
OPBE/$\mathcal{S}$	893
RPBE/$\mathcal{S}$	1526
OLYP/$\mathcal{S}$	334
Gaussian Results	
PBE/$\mathcal{G}$	3836
PBE0/$\mathcal{G}$	−3212
B3LYP/$\mathcal{G}$	−2076
B3LYP*/$\mathcal{G}$	−543
O3LYP/$\mathcal{G}$	168
HCTH/$\mathcal{G}$	−596
TPSS/$\mathcal{G}$	2781(2847)

[a] The value given in parentheses was obtained by using the “ultrafine” integration grid.

of the lowest-lying trigonal component $^4A_2$: $E_{HS}^{\circ} = E^{\circ}(^4A_2)$. As for the JT-unstable LS state, the zero-point energy is the one evaluated at the JT distorted minima of C_2 symmetry: $E_{LS}^{\circ} = E^{\circ}(^2A)$. An additional correction to the calculated LS zero-point energy must be considered when one is to treat the dynamic nature of the JT instability.[71] However, this can be expected to be a minor correction to the ground vibronic level of the LS state as given by $E^{\circ}(^2A)$. Hence, this correction can be neglected. The HS−LS zero-point energy difference can, in turn, be separated into its electronic ΔE_{HL}^{el} and vibrational ΔE_{HL}^{vib} contributions according to $\Delta E_{HL}^{\circ} = \Delta E_{HL}^{el} + \Delta E_{HL}^{vib}$. The vibrational analyses performed using the Gaussian package give a ΔE_{HL}^{vib} value of between −330 and −110 cm^{-1} depending on the functional used. Its minus sign reflects the weakening of the cobalt−ligand bond upon the LS → HS change of spin states. Using for ΔE_{HL}^{vib} the average value of −200 cm^{-1}, the criterion given by eq 13 now reads

$$\Delta E_{HL}^{el} < -\Delta E_{HL}^{vib} \approx 200 \text{ cm}^{-1} \qquad (14)$$

The calculated values of ΔE_{HL}^{el} given in Table 8 show that most functionals do not fulfill this condition. Indeed, the different functionals give very different values for ΔE_{HL}^{el}, ranging from −3212 cm^{-1} (PBE0) to 3919 cm^{-1} (PBE). This large spread contrasts with the consistency observed when comparing the results obtained for the HS and LS geometries of the complex (with the exception of the discarded VSXC functional) and for the energetics of the JT effect in the LS state as well as for the trigonal splitting in the HS state. Thus, consistency among the functionals is achieved for the description of the geometry of the complex in either spin state. However, when it comes to energetics, consistency is retained only when comparing results for a given spin multiplicity, and large deviations are observed when dealing with different spin states. The values obtained with the PBE functional show a remarkable consistency between the results obtained with the GTO $\mathcal{G}$ basis set in Gaussian and the STO $\mathcal{S}$ basis set in ADF. The equivalence of these two basis sets was already established when treating geometrical parameters. The consistency observed for the spin-state energetics fully validates the comparison of results from these two basis sets notably when using the Gaussian and ADF packages.

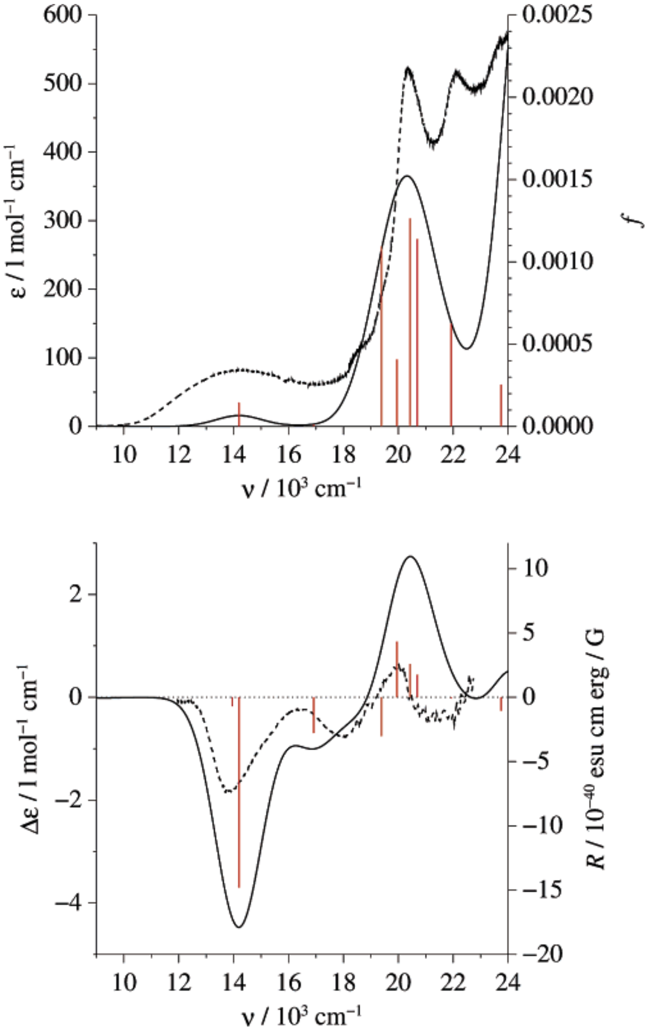


Figure 4. Calculated absorption (top) and CD (bottom) spectra of LS Λ-[Co(bpy)$_3$]$^{2+}$. The experimental single-crystal absorption and CD spectra of [Co(bpy)$_3$][LiRh(ox)$_3$], recorded at 11 and 15 K, respectively, are shown (dashed lines) for comparison.

In line with previous observations,[40−42] the spin-state energetics provide a very stringent criterion for assessing the performance of the approximate functionals. The PBE, RPBE, and OPBE GGAs and the TPSS meta-GGA functionals fail to correctly predict the HS state as the ground state of the isolated [Co(bpy)$_3$]$^{2+}$ complex. With ΔE_{HL}^{el} values of between 893 and 3919 cm^{-1}, these semilocal functionals overestimate the stability of the LS state with regard to the HS state. The OLYP and HCTH GGAs whose ΔE_{HL}^{el} values are 334 and −596 cm^{-1}, respectively, correctly predict the HS state to be the ground state or at least to be almost equienergetic with the LS state. Compared to the other semilocal functionals, their built-in nonlocality overcomes the tendency of their LDA component to destabilize states of higher multiplicities and thus to overestimate the HS−LS energy difference. As for the hybrids, they all predict the correct ground state. The PBE0, B3LYP, B3LYP*, and O3LYP functionals give ΔE_{HL}^{el} values of −3212, −2076, −543, and 168 cm^{-1}, respectively. Their exact

Table 9. Excitation Energies, Oscillator Strengths, Rotational Strengths for the Λ-Enantiomer, and Assignments Calculated for $[Co(bpy)_3]^{2+}$ in the LS ^2A State

excited state	E [cm^{-1}]	f [$\times$ 10^{-5}]	R [10^{-40} cgs]	major MO → MO transitions	main character
^2B	3790	0.18	−0.076	$69a_{\uparrow} \rightarrow 69b_{\uparrow}$ (33%) $69a_{\downarrow} \rightarrow 71b_{\downarrow}$ (29%) $69a_{\downarrow} \rightarrow 72b_{\downarrow}$ (18%)	d−d[a]
^2A	5203	0.09	−0.052	$65b_{\uparrow} \rightarrow 69b_{\uparrow}$ (40%) $66b_{\downarrow} \rightarrow 71b_{\downarrow}$ (29%) $66b_{\downarrow} \rightarrow 72b_{\downarrow}$ (18%)	d−d ($^2T_1$, $^2T_2$)
^2B	6823	20.90	1.139	$70a_{\uparrow} \rightarrow 69b_{\uparrow}$ (89%)	d−d ($^2T_1$, $^2T_2$)
^2B	8738	0.18	−0.095	$66a_{\uparrow} \rightarrow 69b_{\uparrow}$ (38%) $67a_{\downarrow} \rightarrow 71b_{\downarrow}$ (11%) $68a_{\downarrow} \rightarrow 71b_{\downarrow}$ (20%) $68a_{\downarrow} \rightarrow 72b_{\downarrow}$ (12%)	d−d ($^2T_1$, $^2T_2$)
^2A	13951	0.01	−0.680	$65b_{\uparrow} \rightarrow 69b_{\uparrow}$ (11%) $69a_{\downarrow} \rightarrow 74a_{\downarrow}$ (70%)	d−d ($^2T_1$, $^2T_2$)
^2B	14196	14.51	−14.795	$69a_{\uparrow} \rightarrow 69b_{\uparrow}$ (11%) $66b_{\downarrow} \rightarrow 74a_{\downarrow}$ (64%) $69a_{\downarrow} \rightarrow 71b_{\downarrow}$ (10%)	d−d ($^2T_1$, $^2T_2$)
^2B	16917	1.25	−2.789	$66a_{\uparrow} \rightarrow 69b_{\uparrow}$ (42%) $67a_{\downarrow} \rightarrow 71b_{\downarrow}$ (11%) $68a_{\downarrow} \rightarrow 71b_{\downarrow}$ (19%) $68a_{\downarrow} \rightarrow 72b_{\downarrow}$ (12%)	d−d + MLCT
^2B	19383	108.96	−2.973	$70a_{\uparrow} \rightarrow 67b_{\uparrow}$ (61%) $66b_{\downarrow} \rightarrow 74a_{\downarrow}$ (16%) $69a_{\downarrow} \rightarrow 71b_{\downarrow}$ (9%)	MLCT
^2A	19953	40.75	4.333	$65b_{\uparrow} \rightarrow 69b_{\uparrow}$ (17%) $66b_{\downarrow} \rightarrow 71b_{\downarrow}$ (25%) $66b_{\downarrow} \rightarrow 72b_{\downarrow}$ (15%) $69a_{\downarrow} \rightarrow 74a_{\downarrow}$ (31%)	d−d + MLCT
^2B	20426	126.37	2.586	$70a_{\uparrow} \rightarrow 67b_{\uparrow}$ (51%) $66b_{\downarrow} \rightarrow 74a_{\downarrow}$ (27%) $69a_{\downarrow} \rightarrow 71b_{\downarrow}$ (10%)	MLCT
^2A	20690	113.86	1.762	$70a_{\uparrow} \rightarrow 71a_{\uparrow}$ (97%)	MLCT
^2B	21924	62.59	−0.080	$70a_{\uparrow} \rightarrow 68b_{\uparrow}$ (94%)	MLCT
^2B	23746	25.36	−1.059	$67a_{\uparrow} \rightarrow 67b_{\uparrow}$ (9%) $69a_{\uparrow} \rightarrow 67b_{\uparrow}$ (10%) $66a_{\downarrow} \rightarrow 67b_{\downarrow}$ (16%) $69a_{\downarrow} \rightarrow 67b_{\downarrow}$ (51%)	MLCT + intraligand
^2A	24541	464.05	3.846	$67a_{\downarrow} \rightarrow 74a_{\downarrow}$ (18%) $68a_{\downarrow} \rightarrow 74a_{\downarrow}$ (33%) $66b_{\downarrow} \rightarrow 67b_{\downarrow}$ (40%)	MLCT

[a] Transition toward the highest sheet of the LS adiabatic potential energy surface (see JT effect).

exchange contributions amount to 25, 20, 15, and 11.61%, respectively. That is, the HS−LS energy difference calculated within the hybrid formalism decreases with an increasing amount of exact exchange. However, hybrids with large exact-exchange contributions such as the B3LYP and PBE0 clearly overestimate the stability of the HS state with regard to the LS state in transition-metal complexes, and this can be remedied by reducing this contribution to ~10%.[37,38,42] In conclusion, the B3LYP* and O3LYP hybrids as well as the OLYP and HCTH GGAs give the best results regarding the calculation of the HS−LS energy difference in the isolated $[Co(bpy)_3]^{2+}$ complex.

4. Photophysical Properties

In the characterization of $[Co(bpy)_3]^{2+}$, the absorption and circular dichroism spectra are significant signatures for the population of the different states of the complex. In the following, they are analyzed on the basis of TDDFT calculations carried out at the B3LYP/$\mathscr{G}$ level.

4.1 $[Co(bpy)_3]^{2+}$ in the LS State. Figure 4 shows the single-crystal absorption spectrum of the spin-crossover compound $[Co(bpy)_3][LiRh(ox)_3]$, which has been recorded at 10 K where $[Co(bpy)_3]^{2+}$ is in the LS state. The slightly asymmetric band centered at 14 000 cm^{-1} was previously attributed to the spin-allowed and overlapping d−d transitions of the low-spin species, namely, the $^2E(t_{2g}^6e_g^1) \rightarrow {}^2T_1(t_{2g}^5e_g^2)$, $^2T_2(t_{2g}^5e_g^2)$ transitions according to the Tanabe−Sugano diagram for octahedral d^7 complexes. The higher-energy bands were likewise attributed to d−d transitions,[3] but this assignment was later questioned.[4] The electronic excitation calculations were performed using the JT-distorted C_2 geometry of the complex in the ^2A state. Figure 4 shows

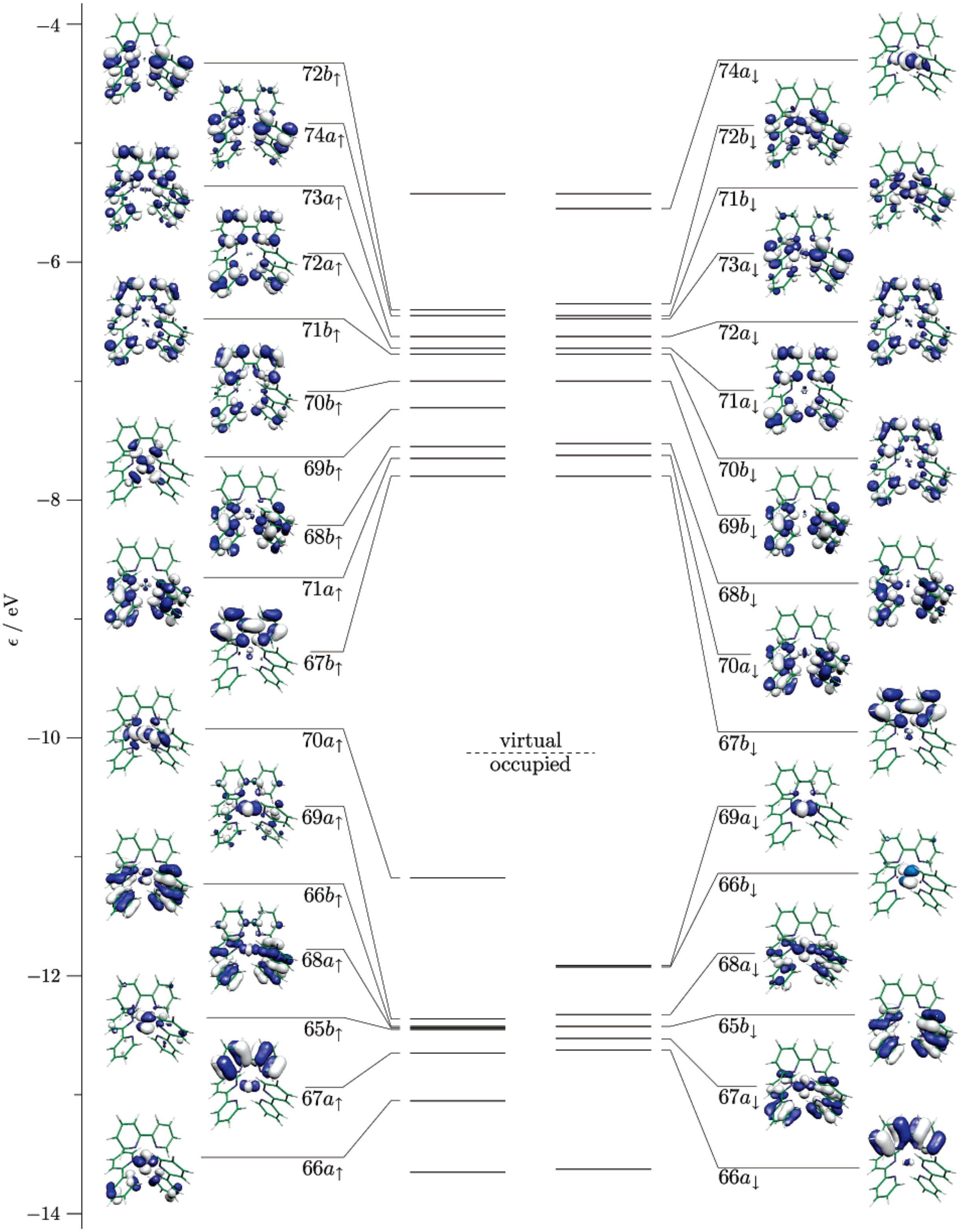


Figure 5. Frontier KS MOs of the C_2 complex [Co(bpy)$_3$]$^{2+}$ in the LS ^{2}A state (B3LYP/$\mathcal{G}$ results).

the stick plot of the calculated oscillator strengths along with the plot of the simulated absorption spectrum. Although there are some differences in the intensities, the relative intensities of the experimental absorption bands are well-reproduced and the maxima of the measured and calculated absorption bands coincide to within ~1000 cm^{-1}, which leads to a very good overall agreement between the experimental and the predicted spectra.

In going from the O_h to the C_2 symmetry, the d−d transitions predicted for [Co(bpy)$_3$]$^{2+}$ on the basis of LFT considerations split as follows. The ^{2}E → ^{2}T$_1$ transition decomposes into a ^{2}A → ^{2}A and two ^{2}A → ^{2}B transitions, and the ^{2}E → ^{2}T$_2$ transition decomposes into two ^{2}A → ^{2}A and a ^{2}A → ^{2}B transition. We thus expect to identify six corresponding d−d transitions in the calculated spectra. Additionally, with the LS ^{2}E state being split into the ^{2}A and ^{2}B states due to the Jahn−Teller effect, we also expect a lower-energy intraconfigurational transition corresponding to the ^{2}A → ^{2}B excitation. The characteristics of all the calculated spin-allowed transitions of interest are given in Table 9. This includes the excitation energies, the oscillator strengths, and the rotational strengths for the Λ-enantiomer as well as the assignments of these transitions. The assignments were made on the basis of the diagram of the frontier KS molecular orbitals (MOs) in Figure 5 in order to identify the major monoelectronic MO → MO excitations contributing to a given transition.

In Table 9, the lowest-lying transition at 3790 cm^{-1} corresponds to the intraconfigurational ^{2}A → ^{2}B transition. However, it has comparatively low oscillator strength and has not been identified experimentally so far. It is likewise difficult to identify all six of the transitions of ^{2}E → ^{2}T$_1$, ^{2}T$_2$ parentage. Three of them are expected in the range between 5000 and 9000 cm^{-1}, but only one of them carries significant oscillator strength. It has not yet been observed experimentally either. Two such transitions are predicted at about 14 000 cm^{-1}, one of them carrying significant oscillator strength. Thus, we may indeed attribute the experimental absorption band at this energy to a d−d transition. At higher energy, the transitions are strongly mixed, carrying both d−d and metal-to-ligand charge-transfer (MLCT) characters, but the most intense transitions in the region around 20 000 cm^{-1} are predominantly MLCT transitions.

CD spectroscopy constitutes a versatile tool to elucidate the nature of electronic transitions in chiral complexes. [Co(bpy)$_3$][LiRh(ox)$_3$] crystallizes in the chiral and cubic space group $P2_13$ with spontaneous resolution of the racemate.[4] Figure 4 shows the single-crystal CD spectrum of this compound at 15 K. It also includes the results of the TDDFT calculations performed for the Λ-enantiomer, namely, the stick plot of the rotatory strengths and the plot of the simulated CD spectrum. The calculated and experimental CD spectra present a very good agreement from which one can infer the configuration of the complex in the crystal; that is, in the crystal under investigation, the complexes are present as the Λ-[Co(bpy)$_3$]$^{2+}$ enantiomer. d−d transitions have usually comparatively large CD intensities, whereas MLCT transitions have smaller intensities with respect to their molar extinction coefficients. The fact that above 18 000 cm^{-1} the values of the relative CD signals expressed as $\Delta\epsilon/\epsilon_{\max}$ are lower than 0.002 as compared to the value of 0.02 for the signals at 14 000 cm^{-1} corroborates the assignments of the high-energy bands to MLCT transitions. It is noteworthy that the electronic transition predicted at 16 917 cm^{-1} has a very weak oscillator strength but carries a relatively large CD intensity. Although the analysis based on the MOs led to the conclusion that it exhibits both an MLCT character and a d−d character, the last observations suggest that this transition may predominantly be viewed as a higher-lying d−d transition.

4.2. [Co(bpy)$_3$]$^{2+}$ in the HS State. Figure 6 shows the single-crystal absorption spectrum of the HS compound [Co(bpy)$_3$] [LiRh(ox)$_3$] measured at 12 K. The experimental absorption band at 11 500 cm^{-1} was assigned to the spin-allowed d−d transition ^{4}T$_1$(t$_{2g}^5$e$_g^2$) → ^{4}T$_2$(t$_{2g}^4$e$_g^3$).[4]

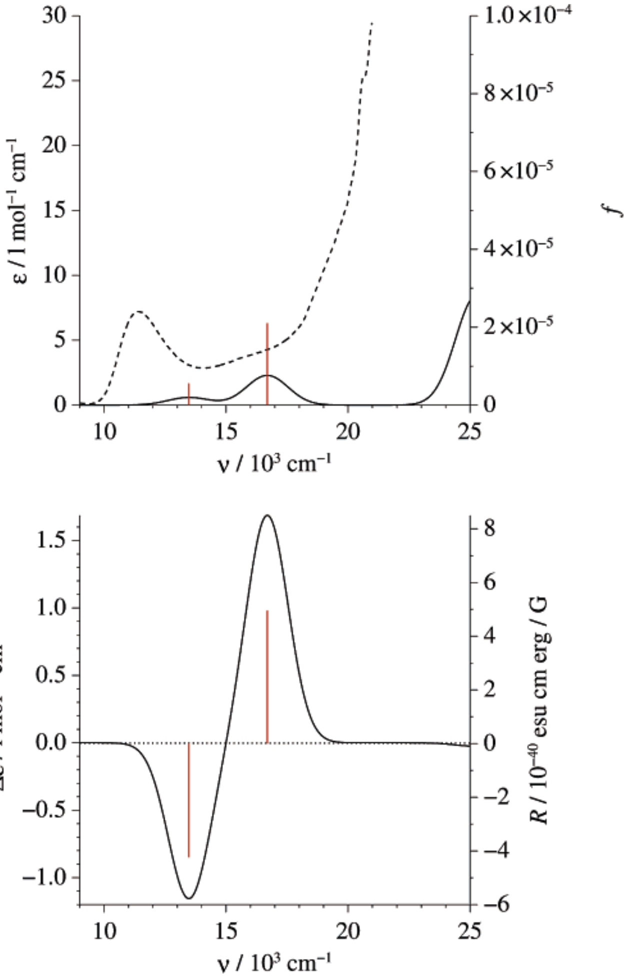


Figure 6. Calculated absorption (top) and CD (bottom) spectra of HS Λ-[Co(bpy)$_3$]$^{2+}$. The 12 K single-crystal absorption spectrum of [Co(bpy)$_3$][NaRh(ox)$_3$] is shown (dashed line) for comparison purposes (see text).

The TDDFT calculations were carried out using the D_3 geometry of [Co(bpy)$_3$]$^{2+}$ in the HS ^{4}A$_2$ state. In the effective D_3 symmetry of the complex, because of the trigonal splitting of the HS ^{4}T$_1$ and ^{4}T$_2$ states, one expects to identify a ^{4}A$_2$ → ^{4}A$_1$ and a ^{4}A$_2$ → ^{4}E transition in addition to a ^{4}A$_2$ → ^{4}E intraconfigurational transition at lower energy. Table 10 gives for the spin-allowed transitions of interest the excitation energies, the oscillator strengths, and the rotational strengths for the Λ-enantiomer. It also gives their assignments, which are based on the MO diagram of Figure 7. The intraconfigurational transition is calculated at 3203 cm^{-1} and has not been observed so far. The two components ^{4}A$_2$ → ^{4}A$_1$, ^{4}E of the d−d ^{4}T$_1$ → ^{4}T$_2$ transition are predicted at 13 473 and 16 687 cm^{-1}, respectively. The dipole-forbidden d−d transition at 20 403 cm^{-1} corresponds to an excitation toward the ^{4}A$_2$ trigonal component of the LFT ^{4}T$_1$(t$_{2g}^4$e$_g^3$) state.

The absorption and CD spectra simulated using these results are plotted in Figure 6. The calculated absorption spectrum does not agree with the experimental spectrum on an absolute energy scale. However, it very nicely predicts the region with no absorption band before the rise of a relatively intense band corresponding to an intraligand $\pi-\pi^*$ transition (Table 10). The shift in energy of ~4000 cm^{-1} between the experimental and calculated spectra indicates

Table 10. Excitation Energies, Oscillator Strengths, Rotational Strengths for the Λ-Enantiomer, and Assignments Calculated for $[Co(bpy)_3]^{2+}$ in the HS $^4A_2$ State

state	*E* [cm^{-1}]	*f* [× 10^{-5}]	*R* [10^{-40} cgs]	major MO → MO transitions	main character
^4E	3203	0.40	4.900	$43e_{\downarrow} \rightarrow 25a_{1\downarrow}$ (15%)	d−d (HS $^4T_1$ state)[a]
				$44e_{\downarrow} \rightarrow 25a_{1\downarrow}$ (66%)	
$^4A_1$	13473	0.51	−4.215	$43e_{\downarrow} \rightarrow 48e_{\downarrow}$ (18%)	d−d ($^4T_2$)
				$44e_{\downarrow} \rightarrow 48e_{\downarrow}$ (74%)	
^4E	16687	1.05	2.476	$43e_{\downarrow} \rightarrow 48e_{\downarrow}$ (16%)	d−d ($^4T_2$)
				$44e_{\downarrow} \rightarrow 48e_{\downarrow}$ (74%)	
$^4A_2$	20403	0.00	0.000	$43e_{\downarrow} \rightarrow 48e_{\downarrow}$ (16%)	d−d ($^4T_1$)
				$44e_{\downarrow} \rightarrow 48e_{\downarrow}$ (74%)	
^4E	25255	3.16	0.047	$25a_{1\uparrow} \rightarrow 46e_{\uparrow}$ (15%)	intraligand ($\pi-\pi^*$)
				$44e_{\uparrow} \rightarrow 46e_{\uparrow}$ (15%)	
				$44e_{\uparrow} \rightarrow 23a_{2\uparrow}$ (15%)	
				$24a_{1\downarrow} \rightarrow 45e_{\downarrow}$ (14%)	
				$43e_{\downarrow} \rightarrow 45e_{\downarrow}$ (10%)	
				$43e_{\downarrow} \rightarrow 23a_{2\downarrow}$ (11%)	

[a] Trigonal component of the HS state.

an overestimation of the excitation energies of the allowed and forbidden transitions in the near-IR/visible domain. This also implies for the two transitions $^4A_2 \rightarrow {}^4A_1$, ^4E that only the latter has been observed experimentally.

The discrepancy between the calculated and experimental excitation energies could be due to the fact that, in carrying out the calculations in D_3, we do not consider the possibility of a pseudo-JT effect between the close-lying $^4A_2$ and JT-unstable ^4E components of the HS state. This is being investigated.

5. Concluding Remarks

Density-functional theory has been applied to the study of the structural, energetic, and photophysical properties of $[Co(bpy)_3]^{2+}$ in the HS and LS states. The structural and energetic features of the complex in each spin state were thus determined by carrying out geometry optimizations using several GGA (PBE, OPBE, RPBE, OLYP, and HCTH), meta-GGA (VSXC and TPSS), and hybrid (B3LYP, B3LYP*, O3LYP, and PBE0) functionals in combination with the STO $\mathscr{S}$ and GTO $\mathscr{G}$ basis sets. Its electronic absorption and CD spectra were likewise analyzed on the basis of the results of TDDFT calculations performed at the B3LYP/$\mathscr{G}$ level. The main results and conclusions obtained can be summarized as follows.

Calculations performed with the PBE functional and the two basis sets gave almost identical results, thus validating the direct comparison of results from both basis sets.[42] With the exception of the VSXC functional, all of the functionals perform equally well with regard to the description of the geometry of the complex and with regard to its energetics in a given spin multiplicity. The VSXC functional gives unphysical geometries because of its well-known tendency toward the overestimation of long-range bonding interactions,[43,62,63] and it therefore cannot be expected to give reliable results for the energetics of the complex even within the same spin multiplicity. The TPSS meta-GGA does give correct geometries, but an integration grid higher than usual must be used for it to give reliable results for the energetics.

Density-functional calculations carried out in D_3 symmetry allowed the characterization of the complex in the LS ^2E state of the octahedral $^2E(t_{2g}^6e_g^1)$ parentage and in the HS $^4A_2$ and ^4E states of the octahedral $^4T_1(t_{2g}^5e_g^2)$ parentage. The $^4A_2$ state is found to be the lower-lying trigonal component of the HS state, as observed experimentally, and the trigonal splitting is estimated to be ~2000 cm^{-1}, in reasonable agreement with the experimental value of 1000 cm^{-1}.[3] There is no significant structural change in passing from one quartet component to the other because this principally involves an electronic rearrangement within nonbonding metallic levels of octahedral $Co(t_{2g})$ parentage. The calculated ^2E and $^4A_2$ geometries agree well with the X-ray crystal structures[4] of the complex in either spin state. The lengthening of the Co−N bonds by ~0.10 Å observed upon the LS → HS transition can be ascribed to the fact that the change of spin states involves the promotion of an electron from a nonbonding metallic level of octahedral $Co(t_{2g})$ parentage into an antibonding metallic level of octahedral $Co(e_g)$ parentage.

The JT instability of the LS ^2E state was investigated by determining the extrema of low symmetry (C_2) of the LS APES. Two extrema could be identified, which correspond to the ^2A and ^2B states and which are a true minimum and a saddle point, respectively. Because of the trigonal symmetry of the JT problem, there are three such minima and three such saddle points. In addition, the presence of the saddle points shows that the JT effect involves quadratic vibronic coupling. The geometries of $[Co(bpy)_3]^{2+}$ in the ^2A and ^2B states are different, each being characterized by distinct variations in the Co−N bond lengths and ligand arrangement with respect to the high-symmetry configuration. The JT stabilization energy and the height of the barrier to pseudorotation amount to $E_{JT} \approx 1500$ cm $^{-1}$ and $\Delta \approx 340$ cm^{-1}, respectively. The theoretical results support the idea of a dynamic JT instability in the LS state, which can be inferred from the high symmetry of the X-ray crystal structure of LS $[Co(bpy)_3]^{2+}$ [4] and its EPR spectra.[65]

The spin-state energetics prove to be the most stringent criterion for assessing the performance of the functionals. To correctly give the HS state as the ground state of the isolated $[Co(bpy)_3]^{2+}$ complex, the calculated values of the HS−LS energy difference ΔE_{HL}^{el} must verify $\Delta E_{HL}^{el} \lesssim$

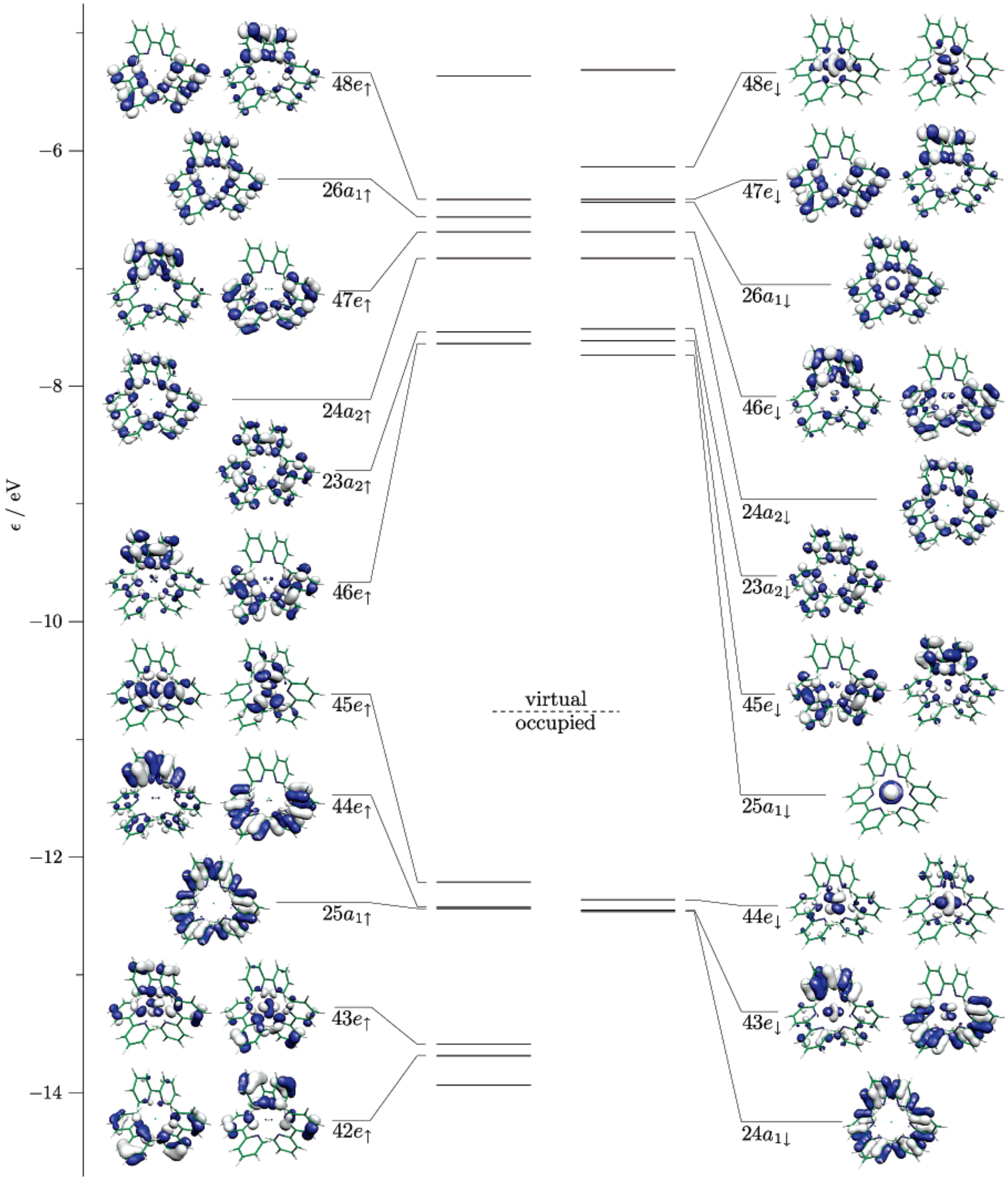


Figure 7. Frontier KS MOs of the D_3 complex $[Co(bpy)_3]^{2+}$ in the HS $^4A_2$ state (B3LYP/$\mathscr{G}$ results).

$-\Delta E_{HL}^{vib} \approx 200$ cm $^{-1}$. However, this criterion is not met by all of the functionals despite the fact that they all perform equally well for the geometry of the complex and for its energetics within a given spin multiplicity. Of all the functionals tested, the OLYP and HCTH GGAs and the B3LYP* and O3LYP hybrids give the best performance with ΔE_{HL}^{el} values of 334, −596, −543, and 168 cm $^{-1}$, respectively, that is, around zero as expected for a complex close to the spin-crossover point.

The absorption and CD spectra deduced from the calculated oscillator and rotatory strengths based on TDDFT are in very good agreement with the experimental spectra. This allows a clear-cut assignment of the observed electronic transitions and of the chirality of the complex in the sample studied by CD spectroscopy. For the study of the photophysical properties of the complex in the HS state, the TDDFT calculations were performed using the geometry of D_3 symmetry found for $[Co(bpy)_3]^{2+}$ in the $^4A_2$ state. Although the calculated absorption spectrum for the near-IR/visible region proves to be very similar in shape to the experimental one, it presents a shift toward higher energies by some ∼4000 cm^{-1}. This discrepancy is tentatively ascribed to the neglect of the pseudo-JT effect between the two trigonal components of the HS state.

In summary, the present theoretical study has allowed us to gain new insight into the structural, energetic, and photophysical properties of the $[Co(bpy)_3]^{2+}$ complex in the LS and HS states. Among the 11 state-of-the-art functionals that have been used, only four prove to perform well for the HS−LS energy difference in the isolated complex. These functionals, namely, the OLYP, HCTH, B3LYP*, and O3LYP functionals, are therefore good candidates for the study of the guest−host interactions that are responsible for turning the HS complex into a spin-crossover system.

Acknowledgment. This work has benefitted from the financial support of the Swiss National Science Foundation and the MAGMANet NoE of the European Union (contract NMP3-CT-2005-515767-2). We acknowledge supercomputer time at the Centro Svizzero di Calcolo Scientifico (CSCS) in the framework of the CSCS project entitled “Photophysics and Photochemistry of Transition Metal Compounds: Theoretical Approaches”. We thank Claudio Redaelli, Angelo Mangili, Claudio Bareato, and Djordje Maric of the CSCS for valuable technical support. We also thank Pierre-Yves Morgantini and Mihail Atanasov for helpful discussions.

Supporting Information Available: Tables S1 and S2 list calculated vibrational frequencies for $[Co(bpy)_3]^{2+}$ in the HS $^4A_2$ state and in the LS ^2A state, respectively (PBE/$\mathscr{G}$ and B3LYP*/$\mathscr{G}$ results). For comparison purposes, Table S1 also gives the IR frequencies experimentally determined for the HS species by Saito et al.[72] The calculated IR spectra of the complex in the two spin states are shown in Figure S1. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Mizuno, K.; Lunsford, J. H. Electron Paramagnetic Resonance Study of Tris(2,2′-bipyridine)cobalt(II) Complexes in Zeolite Y: Evidence for Spin Equilibrium. *Inorg. Chem.* **1983**, *22*, 3484−3486.

(2) Tiwary, S. K.; Vasudevan, S. Spin Crossover in the Ship-in-a-Bottle Compound: Cobalt(II)tris(bipyridyl) Encapsulated in Zeolite-Y. *Chem. Phys. Lett.* **1997**, *277*, 84−88.

(3) Tiwary, S. K.; Vasudevan, S. Void Geometry Driven Spin Crossover in Zeolite-Encapsulated Cobalt Tris(bipyridyl) Complex Ion. *Inorg. Chem.* **1998**, *37*, 5239−5246.

(4) Sieber, R.; Decurtins, S.; Stoeckli-Evans, H.; Wilson, C.; Yufit, D.; Howard, J. A. K.; Capelli, S. C.; Hauser, A. A Thermal Spin Transition in $[Co(bpy)_3][LiCr(ox)_3]$ (ox = $C_2O_4^{2-}$); bpy = 2, 2′-bipyridine. *Chem.*—*Eur. J.* **2000**, 361−368.

(5) Zerara, M.; Hauser, A. Cobalt(II)-tris-2,2′-bipyridine as a Spin-Crossover Complex: Evidence for Cooperative Effects in Three-Dimensional Oxalate Networks. *ChemPhysChem* **2004**, *5*, 395−399.

(6) Hauser, A.; Amstutz, N.; Delahaye, S.; Sadki, A.; Schenker, S.; Sieber, R.; Zerara, M. Fine Tuning the Electronic Properties of $[M(bpy)_3]^{2+}$ Complexes by Chemical Pressure (M = Fe^{2+}, Ru^{2+}, Co^{2+}; bpy = 2,2′-Bypridine). *Struct. Bonding* **2004**, *106*, 81−96.

(7) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−B871.

(8) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(9) Gunnarsson, O.; Lundqvist, B. I. Exchange and Correlation in Atoms, Molecules, and Solids by the Spin-Density-Functional Formalism. *Phys. Rev. B: Solid State* **1976**, *13*, 4274−4298.

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

(11) Dreizler, R. M.; Gross, E. K. U. *Density Functional Theory, An Approach to the Quantum Many-Body Problem;* Springer-Verlag: New York, 1990.

(12) Koch, W.; Holthausen, M. C. *A Chemist’s Guide to Density Functional Theory*; Wiley-VCH: New York, 2000.

(13) Ceperley, D. M.; Adler, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **1980**, *45*, 566−569.

(14) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.

(15) Wang, Y.; Perdew, J. P. Correlation Hole of the Spin-Polarized Gas, with Exact Small-Wave-Vector and High-Density Scaling. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *44*, 13298−13307.

(16) Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13244−13249.

(17) Ma, S.-K.; Brueckner, K. E. Correlation Energy of an Electron Gas with a Slowly Varying High Density. *Phys. Rev.* **1968**, *165*, 18−31.

(18) Langreth, D. C.; Mehl, M. J. Beyond the Local-Density Approximation in Calculations of Ground-State Electronic Properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1983**, *28*, 1809−1834.

(19) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822−8824.

(20) Becke, A. D. Density-Functional Exchange Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(21) Becke, A. D. Simulation of Delocalized Exchange by Local Density Functionals. *J. Chem. Phys.* **2000**, *112*, 4020−4026.

(22) Nekovee, M.; Foulkes, W. M. C.; Needs, R. J. Quantum Monte Carlo Analysis of Exchange and Correlation in the Strongly Inhomogeneous Electron Gas. *Phys. Rev. Lett.* **2001**, *87*, 036401.

(23) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(24) Cancio, A. C.; Chou, M. Y. Beyond the Local Approximation to Exchange and Correlation: The Role of the Laplacian of the Density in the Energy Density of Si. http://xxx.lanl.gov/pdf/cond-mat/0506462 (accessed Dec 2005).

(25) Becke, A. D. A New Mixing of Hartree−Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.

(26) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(27) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for Mixing Exact Exchange with Density Functional Approximations. *J. Chem. Phys.* **1996**, *105*, 9982−9985.

(28) Ernzerhof, M. Construction of the Adiabatic Connection. *Chem. Phys. Lett.* **1996**, *263*, 499−506.

(29) Burke, K.; Ernzerhof, M.; Perdew, J. P. The Adiabatic Connection Method: A Non-empirical Hybrid. *Chem. Phys. Lett.* **1997**, *265*, 115−120.

(30) Becke3LYP Method References and General Citation Guidelines. *Gaussian NEWS*; Gaussian, Inc.: Wallingford, CT; Vol 5, no. 2, summer 1994, p 2.

(31) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.

(32) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew−Burke−Ernzerhof Exchange-Correlation Functional. *J. Chem. Phys.* **1999**, *110*, 5029−5036.

(33) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle−Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

(34) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(35) Perdew, J. P.; Burke, K.; Ernzerhof, M. Erratum: Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865; *Phys. Rev. Lett.* **1997**, *78*, 1396 Errata.

(36) Paulsen, H.; Duelund, L.; Winkler, H.; Toftlund, H.; Trautwein, A. X. Free Energy of Spin-Crossover Complexes Calculated with Density Functional Methods. *Inorg. Chem.* **2001**, *40*, 2201−2204.

(37) Reiher, M.; Salomon, O.; Hess, B. A. Reparameterization of Hybrid Functionals Based on Energy Differences of State of Different Multiplicity. *Theor. Chem. Acc.* **2001**, *107*, 48−55.

(38) Salomon, O.; Reiher, M.; Hess, B. A. Assertion and Validation of the Performance of the B3LYP* Functional for the First Transition Metal Row and the G2 Test Set. *J. Chem. Phys.* **2002**, *117*, 4729−4737.

(39) Reiher, M. Theoretical Study of the Fe(phen)$_2$(NCS)$_2$ Spin-Crossover Complex with Reparametrized Density Functionals. *Inorg. Chem.* **2002**, *41*, 6928−6935.

(40) Fouqueau, A.; Mer, S.; Casida, M. E.; Lawson Daku, L. M.; Hauser, A.; Mineva, T. Comparison of Density Functionals for Energy and Structural Differences between the High-[$^5T_{2g}$: $(t_{2g})^4(e_g)^2$] and Low-[$^1T_1$: $(t_{2g})^6(e_g)^0$] Spin States of the Hexaquoferrous Cation [Fe(H$_2$O)$_6$]$^{2+}$. *J. Chem. Phys.* **2004**, *120*, 9473−9486.

(41) Fouqueau, A.; Casida, M. E.; Lawson Daku, L. M.; Hauser, A.; Neese, F. Comparison of Density Functionals for Energy and Structural Differences between the High-[$^5T_{2g}$: $(t_{2g})^4(e_g)^2$] and Low-[$^1T_1$: $(t_{2g})^6(e_g)^0$] Spin States of Iron(II) Coordination Compounds: II. Comparison of Results for More than Ten Modern Functionals with Ligand Field Theory and ab Initio Results for the Hexaquoferrous Cation [Fe(H$_2$O)$_6$]$^{2+}$, and the Hexaminoferrous Cation [Fe(NH$_3$)$_6$]$^{2+}$. *J. Chem. Phys.* **2005**, *122*, 044110.

(42) Lawson Daku, L. M.; Vargas, A.; Hauser, A.; Fouqueau, A.; Casida, M. E. Assessment of Density Functionals for the High-Spin/Low-Spin Energy Difference in the Low-Spin Iron(II)tris(2,2′-bipyridine) Complex. *ChemPhysChem* **2005**, *6*, 1393−1410.

(43) Ganzenmüller, G.; Berkaïne, N.; Fouqueau, A.; Casida, M. E.; Reiher, M. Comparison of Density Functionals for Differences between the High-(^{5}T$_{2g}$) and Low-(^{1}A$_{1g}$) Spin States of Iron(II) Compounds. IV. Results for the Ferrous Complexes [Fe(*L*)('NHS$_4$')]. *J. Chem. Phys.* **2005**, *122*, 234321.

(44) Hammer, B.; Hansen, L. B.; Nørskov, J. K. Improved Adsorption Energetics within Density-Functional Theory Using Revised Perdew−Burke−Ernzerhof Functionals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 7413−7421.

(45) Handy, N. C.; Cohen, A. J. Left−Right Correlation Energy. *Mol. Phys.* **2001**, *99*, 403−412.

(46) Boese, A. D.; Handy, N. C. A New Parametrization of Exchange-Correlation Generalized Gradient Approximation Functionals. *J. Chem. Phys.* **2001**, *114*, 5497−5503.

(47) Cohen, A. J.; Handy, N. C. Dynamic Correlation. *Mol. Phys.* **2001**, *99*, 607−615.

(48) Van Voorhis, T.; Scuseria, G. E. A Novel Form for the Exchange-Correlation Energy Functional. *J. Chem. Phys.* **1998**, *109*, 400−410.

(49) Casida, M. E. Time-Dependent Density Functional Response Theory for Molecules. In *Recent Advances in Density Functional Methods*; Chong, D. P., Ed.; World Scientific: Singapore, 1995; Vol. 1.

(50) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. An Efficient Implementation of Time-Dependent Density Functional Theory for the Calculation of Excitation Energies of Large Molecules. *J. Chem. Phys.* **1998**, *109*, 8218−8224.

(51) Bauernschmitt, R.; Ahlrichs, R. Treatment of Electronic Excitations within the Adiabatic Approximation of Time Dependent Density Functional Theory. *Chem. Phys. Lett.* **1996**, *256*, 454−464.

(52) Autschbach, J.; Ziegler, T.; van Gisbergen, S. J. A.; Baerends, E. J. Chiroptical Properties from Time-Dependent Density Functional Theory. I. Circular Dichroism Spectra of Organic Molecules. *J. Chem. Phys.* **2002**, *116*, 6930−6940.

(53) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931−967.

(54) *Amsterdam Density Functional Program*, Release ADF2004.01; Theoretical Chemistry, Vrije Universiteit: Amsterdam, The Netherlands. http://www.scm.com (accessed Mar 2006).

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

T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.03/B.04; Gaussian, Inc.: Pittsburgh, PA, 2003.

(56) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets of Triple-ζ Valence Quality for Atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829−5835.

(57) Årsköld, S. P.; Masters, V. N.; Prince, B. J.; Smith, P. J.; Pace, R. J.; Krausz, E. Optical Spectra of Synechocystis and Spinach Photosystem II Preparations at 1.7 K: Identification of the D1-Pheophytin Energies and Stark Shifts. *J. Am. Chem. Soc.* **2003**, *125*, 13063−13074.

(58) Schellman, J. A. Circular Dichroism and Optical Rotation. *Chem. Rev.* **1975**, *75*, 323−331.

(59) Furche, F.; Ahlrichs, R.; Wachsmann, C.; Weber, E.; Sobanski, A.; Vögtle, F.; Grimme, S. Circular Dichroism of Helicenes Investigated by Time-Dependent Density Functional Theory. *J. Am. Chem. Soc.* **2000**, *122*, 1717−1724.

(60) Liang, W. Z.; Yokojima, S.; Chen, G. H. Localized-Density-Matrix Calculation of Circular Dichroism Spectrum of Optically Active Molecule. *Chem. Phys.* **2003**, *289*, 175−189.

(61) Diedrich, C.; Grimme, S. Systematic Investigation of Modern Quantum Chemical Methods to Predict Electronic Circular Dichroism Spectra. *J. Phys. Chem. A* **2003**, *107*, 2524−2539.

(62) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. Application of 25 Density Functionals to Dispersion-Bound Homomolecular Dimers. *Chem. Phys. Lett.* **2004**, *394*, 334−338.

(63) Sancho-García, J. C.; Cornil, J. Assessment of Recently Developed Exchange-Correlation Functionals for the Description of Torsion Potential in π-Conjugated Molecules. *J. Chem. Phys.* **2004**, *121*, 3096−3101.

(64) Jahn, H. A.; Teller, E. Stability of Polyatomic Molecules in Degenerate Electronic States. *Proc. R. Soc. London, Ser. A* **1937**, *161*, 220−235.

(65) Zerara, M. Transition de Spin Dans le Complexe Cobalt(II)tris(2,2′-bipyridine). Ph.D. Thesis, Université de Genève, Genève, Switzerland, 2003; Thesis no. 3445.

(66) O'Brien, M. C. M.; Chancey, C. C. The Jahn−Teller Effect: An Introduction and Current Review. *Am. J. Phys.* **1993**, *61*, 688−697.

(67) Ham, F. S. The Jahn−Teller Effect: A Retrospective View. *J. Lumin.* **2000**, *85*, 193−197.

(68) Bersuker, I. B. Modern Aspects of the Jahn−Teller Effect Theory and Applications To Molecular Problems. *Chem. Rev.* **2001**, *101*, 1067−1114.

(69) Applegate, B. E.; Barckholtz, T. A.; Miller, T. A. Exploration of Conical Intersections and Their Ramifications for Chemistry through the Jahn−Teller Effect. *Chem. Soc. Rev.* **2003**, *32*, 38−49.

(70) Sabin, J. R.; Brändas, E.; *Manifestations of Vibronic Coupling in Chemistry and Physics*; Elsevier: Amsterdam, 2003; volume 44 of Advances in Quantum Chemistry.

(71) Bersuker, I. B. *The Jahn−Teller Effect and Vibronic Interactions in Modern Chemistry*; Fackler, J. P., Jr., Ed.; Plenum Press: New York, 1984; Modern Inorganic Chemistry Series.

(72) Saito, Y.; Takemoto, J.; Hutchinson, B.; Nakamoto, K. Infrared Studies of Coordination Compounds Containing Low-Oxidation-State Metals. I. Tris(2,2′-bibyridine) and Tris(1,10-phenanthroline) Complexes. *Inorg. Chem.* **1972**, *11*, 2003−2011.

CT6001384

JCTC Journal of Chemical Theory and Computation

The Basis Set Convergence of Spin−Spin Coupling Constants Calculated by Density Functional Methods

Frank Jensen*

Department of Chemistry, University of Southern Denmark,
DK-5230 Odense, Denmark

Received May 11, 2006

Abstract: The previously proposed polarization-consistent basis sets, optimized for density functional calculations, are evaluated for calculating indirect nuclear spin−spin coupling constants. The basis set limiting values can be obtained by performing a series of calculations with increasingly larger basis sets, but the convergence can be significantly improved by adding functions with large exponents. An accurate calculation of the Fermi-contact contribution requires the addition of tight *s* functions, while the paramagnetic spin−orbit contribution is sensitive to the presence of tight *p* functions. The spin-dipolar contribution requires the addition of *p*, *d*, and *f* functions. The optimal exponents for the tight functions can be obtained by optimizing the absolute sum of all contributions to the spin−spin coupling constant. On the basis of a series of test cases, we propose a standard set of tight *s*, *p*, *d*, and *f* functions to be added to the polarization-consistent basis sets. The resulting pcJ-*n* basis sets should be suitable for calculating spin−spin coupling constants with density functional methods.

I. Introduction

The use of theoretical predicted molecular properties for correlating with experimental information is becoming an increasingly popular method for obtaining information regarding molecular structure. While some properties, like vibrational frequencies and intensities, have been used for many years, spin-dependent properties have only come into focus more recently.[1] The theory for calculating nuclear magnetic resonance (NMR) spin−spin coupling constants is well-known[2] and was implemented in a semiempirical framework already in 1970.[3] In recent years, the methodology for calculating such properties with a variety of sophisticated methods has become generally available and a number of studies have appeared where the accuracy of various methods has been tested. Of prime concern from a practical point of view is the use of density functional theory,[4] because these methods are applicable to quite large systems at a favorable computational cost.[5]

NMR spin−spin coupling constants pose a challenging problem, as the results depend strongly on the quality of the wave function and the basis set. Hartree−Fock wave functions are in general unsuitable for calculating spin−spin coupling constants because of the presence of triplet instabilities, while correlated methods perform much better. Density functional methods are significantly better than Hartree−Fock and often provide an acceptable accuracy with only modest computational requirements. Recently, Keal and Tozer have proposed a family of functionals designed specifically for calculating nuclear shielding constants,[6] and one could imagine similar functionals optimized for spin-dependent properties.

The second component for performing calculations is the use of a basis set for expanding the molecular orbitals. While many standard basis sets are available for predicting, for example, geometries and relative energies, the calculation of molecular properties often requires more specialized basis sets. The basis set convergence for properties depending on the region of the wave function far from the nucleus can be improved by the addition of basis functions with small exponents,[7] while properties depending on the core region require the addition of functions with large exponents.[8−10] The calculation of indirect spin−spin coupling constants belongs to the latter category, and a nonrelativistic perturbational approach involves the four operators shown in eqs 1−4.[2,11]

---

* Corresponding author fax: +45-6615-8780; e-mail: frj@dou.dk.

10.1021/ct600166u CCC: $33.50

$$\mathbf{H}^{\mathrm{DSO}} = \frac{g_A g_B {\mu_N}^2}{2c^4} \frac{(\mathbf{r}_{iA}^t \mathbf{r}_{iB} - \mathbf{r}_{iB} \mathbf{r}_{iA}^t)}{{\mathbf{r}_{iA}}^3 {\mathbf{r}_{iB}}^3} \tag{1}$$

$$\mathbf{H}^{\mathrm{PSO}} = \frac{g_A \mu_N}{c^2} \frac{\mathbf{r}_{iA} \times \mathbf{p}_i}{{r_{iA}}^3} \tag{2}$$

$$\mathbf{H}^{\mathrm{SD}} = \frac{g_e \mu_B g_A \mu_N}{c^2} \frac{(\mathbf{r}_{iA}^t \mathbf{r}_{iA} - 3\mathbf{r}_{iA} \mathbf{r}_{iA}^t)}{{r_{iA}}^5} \mathbf{s}_i \tag{3}$$

$$\mathbf{H}^{\mathrm{FC}} = \frac{8\pi g_e \mu_B g_A \mu_N}{3c^2} \delta(\mathbf{r}_{iA}) \mathbf{s}_i \tag{4}$$

Here, $\mathbf{r}_{iA}$ denoted the position vector between electron i and nucleus A; $\mu_{B/N}$ is the Bohr and nuclear magneton; $g_{e/A}$ is the electron and nuclear g factor, and $\mathbf{s}_i$ is the electron spin operator. The diamagnetic spin−orbit operator $\mathbf{H}^{\mathrm{DSO}}$ gives a contribution which can be evaluated as an expectation value over the wave function, while the other three must be evaluated by response methods, formally given as a sum over all excited states in perturbation theory.

$$\mathbf{J} = \langle \Psi_0 | \mathbf{H}^{\mathrm{DSO}} | \Psi_0 \rangle + \sum_{k}^{\mathrm{PSO,SD,FC}} \sum_{n \neq 0} \frac{\langle \Psi_0 | \mathbf{H}^k | \Psi_n \rangle \langle \Psi_n | \mathbf{H}^k | \Psi_0 \rangle}{E_0 - E_n} \tag{5}$$

The spin−spin coupling is a 3 × 3 tensor, but only the average isotropic component corresponding to $^1/_3$ of the trace of $\mathbf{J}$ is observed in solution, and we will consequently focus on this. It should be noted that the magnitude of coupling constants varies from essentially zero to several thousand hertz depending on the specific system and pairs of nuclei, and only part of the large variation is due to differences in the g factor for different elements and isotopes (e.g., 5.58 for ^{1}H and 0.40 for ^{14}N). Each of the four contributions may be either positive or negative, and the total coupling constant may consequently be either positive or negative, or close to zero because of the cancellation of terms.

From the form of the operators it is clear that the diamagnetic and paramagnetic spin−orbit ($\mathbf{H}^{\mathrm{DSO}}$ and $\mathbf{H}^{\mathrm{PSO}}$) and the spin-dipolar operators ($\mathbf{H}^{\mathrm{SD}}$) primarily sample the inner region of the wave function, while the Fermi-contact term ($\mathbf{H}^{\mathrm{FC}}$) evaluates the quality of the wave function *at* the nuclear positions. The representation of the excited states in the response part, on the other hand, is expected to be sensitive to the quality of the wave function in the region far from the nucleus. It is therefore clear that calculation of spin−spin coupling constants requires a careful consideration of the basis set, as the results are expected to be sensitive to functions with both large and small exponents. Indeed, previous calculations for correlated wave functions have shown that basis sets of at least triple-ζ-type quality augmented with both diffuse and tight functions are required to obtain accuracy results.[8,9,12] A similar study using density functional methods has also stressed the importance of adding tight *s* functions in order to evaluate the Fermi-contact contribution.[5,13] These studies have used the correlation consistent basis sets[14] where additional tight *s* functions were added by extending the regular basis set inward by a geometrical progression and decontraction of the basis sets. During the course of this work, Manninen and Vaara proposed to use basis sets complete to within a given threshold in a given exponent range for calculating magnetic properties.[15] After this paper was submitted, Deng et al. proposed to use the aug-cc-pVXZ basis sets augmented by tight *s* and *d* functions.[16]

We have previously proposed a series of basis sets optimized specifically for density functional methods, called polarization-consistent, which have been shown to improve the basis set convergence for several properties relative to other standard basis sets.[17] The notation is pc-*n*, where *n* indicates the level of polarization beyond the atomic system; that is, pc-0 is unpolarized, pc-1 is of double-ζ quality with a single polarization function, pc-2 is of triple-ζ quality with *d*- and *f*-type polarization functions, and so forth. In the present case, we will investigate the basis set convergence for spin−spin coupling constants with the polarization-consistent basis sets and suggest that the convergence can be improved by the addition of tight *s*, *p*, *d*, and *f* functions.

II. Computational Details

All calculations have been performed with the Dalton program package[18] and the B3LYP functional,[19] as this has been one of the more successful for predicting NMR properties.[11,20] Coupling constants refer to the most common isotope in all cases. Molecular geometries have been optimized at the B3LYP/6-31G(d,p) level. We emphasize that only the convergence with respect to basis set is investigated in the present case, and no attempt is made for comparing with experimental results. The latter requires attention to, for example, the molecular geometry and the quality of the exchange-correlation functional. It is expected that a very similar basis set convergence is observed for other functionals, and the present results should thus be valuable for probing the accuracy of various exchange-correlation functionals for predicting NMR spin−spin coupling constants.

III. Basis Set Convergence for Polarization-Consistent Basis Sets

The spin−spin coupling is a potential observable between all nuclei with a spin in a molecule, although only one-, two- and three-bond couplings are usually sufficiently large to be observed. The number of systems investigated by theoretical methods is relatively limited, and only a few guidelines are available for determining the importance of the above four contributions. One- and two-bond coupling constants (^1J and ^2J) are often dominated by the FC term, especially if one of the nuclei is hydrogen. Three-bond coupling constants (^3J), on the other hand, are often found to have substantial PSO and SD contributions. The DSO contribution is small in absolute magnitude (a few hertz) but may represent a sizable fraction for small coupling constants. These generalizations depend strongly on the system; the coupling constant involving fluorine and chlorine, for example, has large PSO and SD contributions for one-, two-, and three-bond coupling.[10,21] Furthermore, most of the investigations so far have concentrated on systems containing only elements from the first row in the periodic table. In the

Table 1. Molecular Systems Used in the Present Study

HF, H_2O, NH_3, CH_4
CO, N_2, F_2, SiS, P_2, Cl_2, F_2O, Cl_2S
HCN, C_2H_2, C_2H_4, C_2F_2, H_2CO, H_2CS

present case, we have used the systems in Table 1 for probing the performance. The basis set convergence for each of the four terms is found to be very similar for all of the systems, but only a few of the present systems have coupling constants where all four terms make a significant contribution. One of the exceptions is the F_2 system, which will be used as a representative case. We note that current exchange-correlation functionals have problems calculating accurate values for especially the FC term for fluorine-containing systems,[10,11,22] often being in error by a factor of 2. This problem is less relevant for the present case, where only the basis set convergence is investigated.

An initial exploration showed that the spin−spin coupling constant displays an erratic behavior when contracted basis sets are used, and all basis set optimizations have consequently been done using completely uncontracted basis sets. Furthermore, for some systems, the results obtained with the aug-pc-*n* basis sets, which include diffuse functions, are significantly different from those obtained with the regular pc-*n* basis sets. The difference diminishes rapidly as the size of the basis set increases. The results shown in Figure 1 indicate that both types of basis sets converge to a limiting value of ∼15 500 Hz for F_2, but the convergence is slow. It is therefore of interest to investigate to what extent the basis set can be improved in order to speed up the convergence.

The optimization of basis sets has typically employed a variational criterion, that is, minimizing the energy as a function of the basis function exponents. For properties other than the energy, there is no analogous variational principle, although a basis set converged value can be defined by the requirement that the addition of further functions does not cause changes within a specified threshold. In our initial exploration of extending basis sets by adding tight functions with exponents generated as an even-tempered sequence, we noted that the individual components of the coupling constant display a monotonic variation as more tight functions are added. Except for a few cases where the contribution is near zero and changes sign upon the addition of extra functions, this translates into a maximization of the absolute value of each of the contributions. We may consequently use the absolute value of each contribution as a parameter for optimizing the exponents of the added functions and to assess the importance of each type of function. For probing the sensitivity of the total coupling constant, which may contain both positive and negative contributions, we have used the quantity obtained by summation of the absolute values of each of the four contributions and used this in a variational sense to optimize the exponents of added functions.

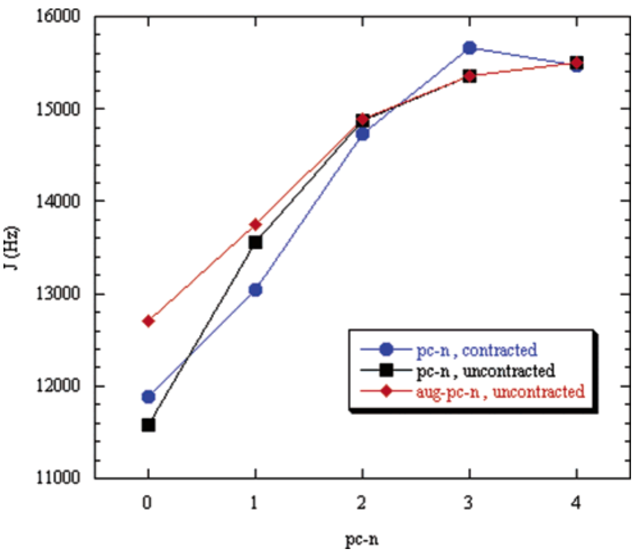


Figure 1. Coupling constant for F_2 calculated by the pc-*n* and aug-pc-*n* basis sets.

Table 2. DSO Contributions (Hz) to the Spin−Spin Coupling Constant in F_2[a]

basis	none	+s	+p	+d	+f
pc-0	3.41	3.41	3.04		
pc-1	2.53	2.53	2.39	2.36	
pc-2	2.20	2.20	2.17	2.14	2.13
pc-3	2.12	2.12	2.12	2.11	2.11
pc-4	2.11	2.11	2.11	2.11	2.11
aug-pc-0	3.30	3.30	2.99		
aug-pc-1	2.49	2.49	2.36	2.33	
aug-pc-2	2.20	2.20	2.17	2.14	2.13
aug-pc-3	2.12	2.12	2.12	2.11	2.11

[a] Notation +*x* indicates addition of a tight *x* function.

The effect of the basis set on the coupling constant may conceptually be divided into three categories: (1) improving the representation of the wave function (density), (2) improving the representation of the operators, that is, improving the matrix elements in the nominator of the response part in eq 5, and (3) improving the excitation energies in the denominator of the response part in eq 5.

The pc-*n* basis sets have been designed to systematically improve category 1, which is the primary contributor to energetic quantities. On the basis of other work, it is expected that the addition of diffuse functions (aug-pc-*n* basis sets) will improve the performance for category 3. Our primary concern in the present case is therefore to add functions tailored to improve the representation of the operators in eqs 1−4. The basis set limiting value of the spin−spin coupling constant in F_2 is 15 556 Hz with the B3LYP functional, and we will in the following establish the requirements for converging the value to within ∼1 Hz of the limiting value.

A. The Diamagnetic Spin−Orbit Contribution. The DSO contribution is calculated as an expectation value over the unperturbed wave function and is in the employed implementation calculated by a grid approach.[10] It is therefore expected that this contribution is insensitive to the addition of tight functions, as their influence is only indirect by changes in the density. The calculated values for F_2 are given in Table 2, and the convergence is illustrated in Figure 2. The basis set limiting value is 2.1 Hz, and both the pc-*n* and aug-pc-*n* basis sets display a smooth convergence toward the limiting value. The convergence can be slightly improved by adding a tight *p* function, while the augmentation with diffuse functions has only a marginal effect (Table 2). The

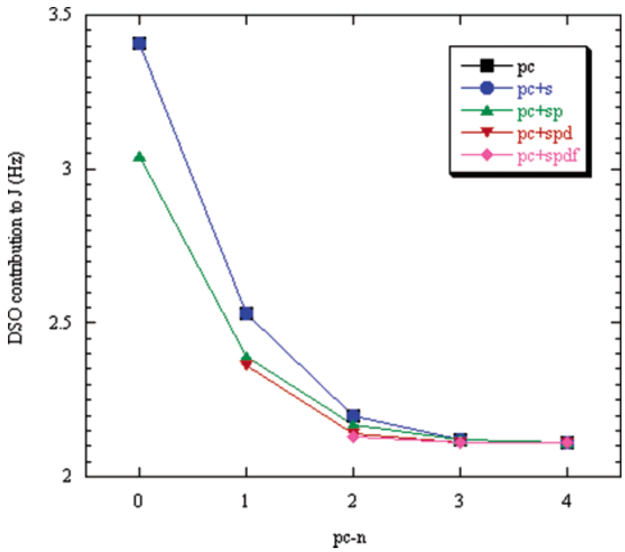


Figure 2. DSO contribution to the coupling constant for F_2 calculated by the pc-*n* basis sets augmented with tight functions.

Table 3. PSO Contributions (Hz) to the Spin−Spin Coupling Constant in F_2[a]

basis	none	+s	+p	+d	+f
pc-0	7032	7033	9497		
pc-1	8542	8542	9569	9565	
pc-2	9245	9245	9418	9416	9409
pc-3	9399	9399	9417	9417	9416
pc-4	9409	9409	9414	9414	9414
aug-pc-0	7745	7746	10 483		
aug-pc-1	8623	8623	9661	9658	
aug-pc-2	9245	9245	9419	9416	9409
aug-pc-3	9395	9395	9414	9413	9413

[a] Notation +*x* indicates addition of a tight *x* function.

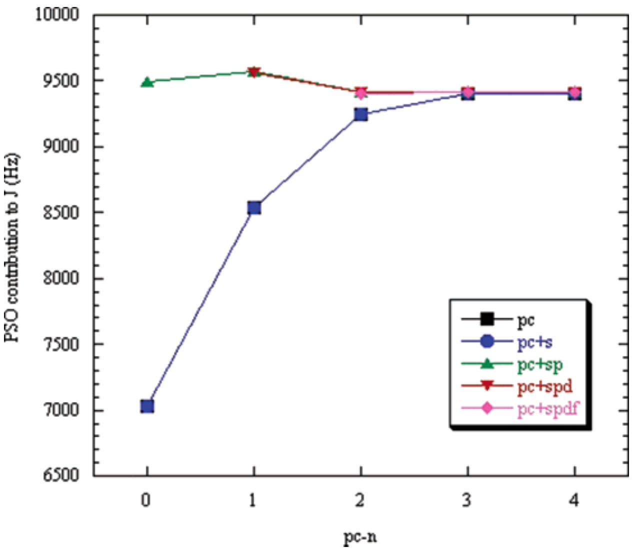


Figure 3. PSO contribution to the coupling constant for F_2 calculated by the pc-*n* basis sets augmented with tight functions.

Table 4. SD Contributions (Hz) to the Spin−Spin Coupling Constant in F_2[a]

basis	none	+s	+p	+d	+f
pc-0	3773	3774	5172		
pc-1	4805	4805	5406	5691	
pc-2	5426	5426	5527	5800	5926
pc-3	5738	5738	5749	5873	5927
pc-4	5865	5865	5867	5873	5911
aug-pc-0	4341	4341	5969		
aug-pc-1	4870	4870	5481	5770	
aug-pc-2	5428	5428	5530	5803	5929
aug-pc-3	5737	5737	5748	5872	5926

[a] Notation +*x* indicates addition of a tight *x* function.

effect by addition of a tight *p* function is largest for the pc-0 and pc-1 basis sets, but because the DSO term is calculated as an expectation value, the change as a function of the basis set quality is due to improvements in representing the electronic density. The effect of the tight *p* function for the two smaller basis sets is therefore mainly due to inadequacies in the underlying pc-*n* basis set.

B. The Paramagnetic Spin−orbit Contribution. The PSO operator provides the largest contribution to the coupling constant in F_2, and the calculated values are given in Table 3 with the convergence illustrated in Figure 3. The addition of a tight *s* function has no effect, while the addition of a tight *p* function drastically improves the performance. For the pc-0 basis set, the change by adding a single tight *p* function is ∼2500 Hz, and the corresponding value for the pc-1 basis set is ∼1000 Hz. The addition of more tight *p* functions leads to further increases by ∼10%, that is, ∼250 and ∼100 Hz, respectively, but for the larger basis sets, the changes due to a second *p* function become negligible. Because the addition of just one tight *p* function slightly overestimates the basis set limiting value (Figure 3), this indicates that the addition of multiple tight *p* functions will be counterproductive, as it would cause the calculated value to move away from the limiting value, at the expense of increasing the computational time. Higher angular momentum functions (*d* and *f* functions) have very marginal effects (i.e., a few hertz at most). Augmentation with diffuse functions has a large effect for the pc-0 basis set, but as seen in Table 3, the effect rapidly diminishes as the quality of the underlying basis improves.

C. The Spin Dipole Contribution. The SD operator provides the second largest contribution to the coupling constant in F_2, and the calculated values by adding tight functions are given in Table 4 and shown in Figure 4. In contrast to the PSO contribution, the SD contribution is sensitive to the addition of *p*, *d*, and *f* functions. Again, the majority of the effect (90+%) is obtained by adding a single tight function. The effect of adding a tight *g* function to the pc-3 basis set is only a fraction of a hertz. It should be noted that *removal* of the *g* function inherent in the pc-3 basis set causes changes of ∼10 Hz for the total coupling constant; that is, *g* functions are somewhat important, but the function already present in the pc-3 basis set is sufficient for converging the calculated value. The corresponding effect for the *h* function present in the pc-4 basis set is ∼1 Hz. Augmentation with diffuse functions (Table 4) again has a significant effect for the pc-0 and pc-1 basis sets but not for the larger basis sets.

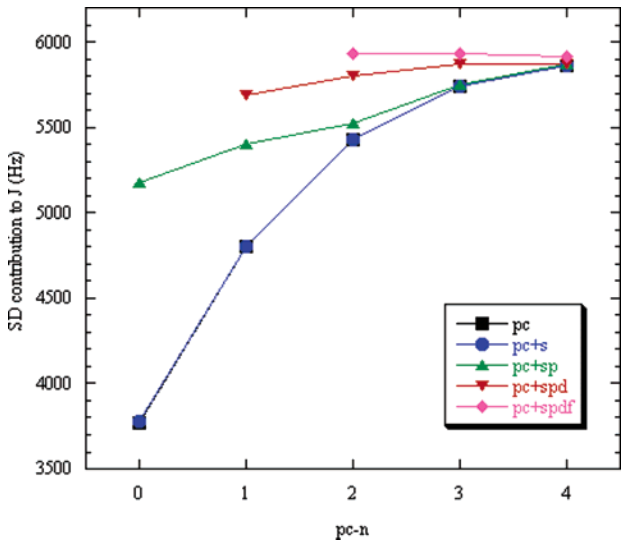


Figure 4. SD contribution to the coupling constant for F_2 calculated by the pc-*n* basis sets augmented with tight functions.

Table 5. FC Contributions (Hz) to the Spin−Spin Coupling Constant in F_2[a]

basis	none	+1s	+2s	+3s	+4s	+5s	+p	+d	+f
pc-0	774	986	1038	1053	1058	1059	1119		
pc-1	201	237	246	248	249	249	219	219	
pc-2	203	216	219	220	220		220	220	224
pc-3	226	230	231	231			231	231	232
pc-4	227	229	229				229	229	229
aug-pc-0	610	781	823	834	838	840	810		
aug-pc-1	256	302	313	316	317	317	288	288	
aug-pc-2	214	228	231	231			231	231	235
aug-pc-3	223	227	228	228			228	228	229

[a] Notation +*nx* indicates addition of *n* tight *x* function.

D. The Fermi-Contact Contribution. The FC operator contains a δ operator at the nuclear positions, and it is known that calculation of this term requires the addition of *s*-type functions with large exponents. In other work, it has been suggested that the addition of four to six functions with exponents generated by an even-tempered extension should be used for generating accurate results.[8,12,13,16] Higher-angular-momentum functions have nodes at the nuclear position and can only affect the FC term indirectly; that is, the density at a given nucleus can have contributions from functions centered at neighboring nuclei. Table 5 and Figure 5 show how the FC contribution converges upon the addition of *s* functions with optimized exponents. From Figure 5, it is clear that the first tight *s* function provides ∼70% of the total effect possible by saturating the function space with tight *s* functions. It is furthermore clear that the number of *s* functions required to reach the converged value decreases as the size of the underlying basis is increased. This is understandable as the larger basis sets already contain *s* functions with large exponents and, thus, have fewer requirements for additional tight functions. Indeed, when examining the absolute value of the exponents, it is clear that the FC contribution is converged once the largest exponent reaches a value of $\sim 60 \times 10^6$ for F_2. For an optimum set of exponents, this requires five *s* functions for the pc-0 basis set, four for pc-1, three for pc-2, two for pc-3, and one for the pc-4 basis set.

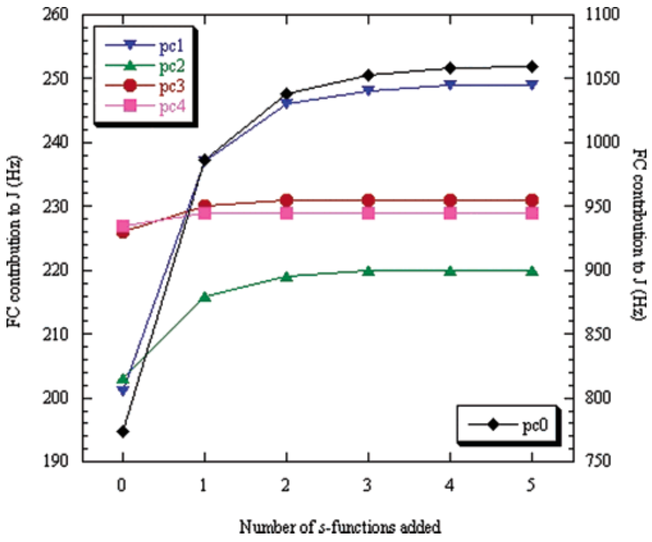


Figure 5. FC contribution to the coupling constant for F_2 as a function of added *s* functions with large exponents. The values on the left-hand axis refer to the pc-1, -2, -3, and -4 results; the values on the right-hand axis refer to the pc-0 results.

While the addition of one tight *s* function clearly is desirable, the addition of multiple functions requires some considerations. Adding, for example, five tight *s* functions to the pc-0 basis set in order to converge the FC contribution will double the number of *s* functions and, thereby, make the number of *s* functions similar to that of the pc-2 basis set. The other contributions to the coupling constant, however, are relatively poorly represented with the pc-0 basis set. Adding many tight *s* functions in order to converge the FC contribution will thus not necessarily improve the total coupling constant, as the error in the other contributions may be dominating. Even for cases where the FC contribution dominates the total coupling constant, the representation of the density and excitation part of the response calculation is not significantly improved by the addition of tight *s* functions, and achieving the limiting FC contribution for a given quality of the underlying basis set is not necessarily a worthy goal. For the pc-1 basis set, for example, the changes by adding more than two *s* functions are only 3 Hz (Table 5), which should be compared to the error of 20 Hz relative to the basis set limit. Furthermore, the addition of several tight *s* functions increases the computational time, and from Figure 5, it is clear that this is a case of rapidly diminishing return. We suggest as a compromise between balance in the accuracy of all four terms, computational efficiency, and inherent accuracy of the underlying basis set, that two tight *s* functions should be added to the pc-0, pc-1, pc-2, and pc-3 basis sets but only one additional *s* function for the pc-4, as this essentially saturates the *s*-function space.

Table 5 and Figure 6 (pc-0 results are not show, as they clearly are of much lower accuracy) show that the FC contribution in addition to *s* functions is sensitive to the addition of a tight *p* function for the pc-0 and pc-1 basis sets. Analogous to the DSO contribution, this is most likely due to inadequacies in the underlying basis sets for repre-

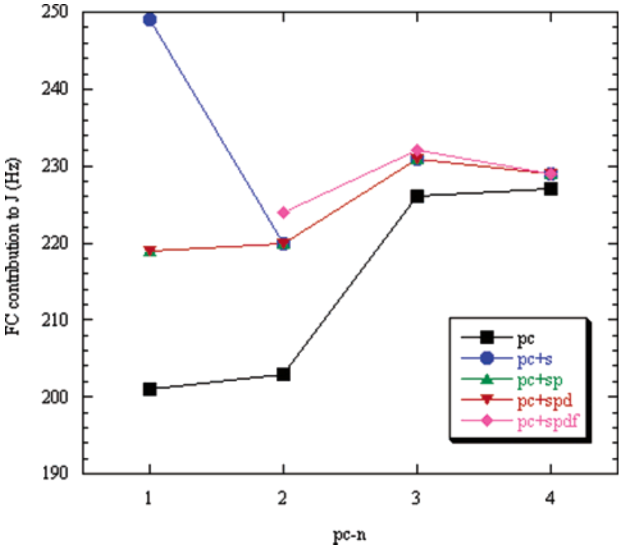


Figure 6. FC contribution to the coupling constant for F_2 calculated by the pc-*n* basis sets augmented with tight functions.

senting the density. Table 5 also shows that the FC contribution is sensitive to augmentation with diffuse functions, although the effect rapidly diminishes as the basis set quality is improved. It can also be noted that the presence of diffuse functions to some extent is counterproductive for this system, as the value moves further away from the limiting value.

IV. Designing Basis Sets for Calculating Spin−Spin Coupling Constants

The analysis in the previous section strongly indicates that a substantial improvement in the basis set convergence can be obtained by the addition of functions up to *f*-type with larger exponents than present in the standard basis sets. An explicit optimization of exponents for each specific system is impractical, and a procedure for selecting exponents for the additional tight functions is therefore required.

Previous work has recommended the addition of four tight *s* functions to basis sets of triple-ζ quality (e.g., pc-2) to converge the FC contribution.[8,12] The exponents of these functions were generated by multiplying the largest exponent with the ratio between the two innermost functions, that is, an even-tempered sequence. From our previous experience with basis set optimization, it is clear that this is unlikely to be optimum, as the ratio between exponents in a fully optimized basis set increases with the exponent values in the core region; that is, the spacing between exponents is small in the valence region and becomes larger as the functions extend into the core region. Furthermore, it is noticeable that the ratio between the two innermost *s* functions is a near constant value of 6.6 for all elements and basis sets ranging from pc-0 to pc-4. The same regularity is observed for the optimum exponents determined by optimizing the FC contribution for a variety of systems and basis sets, and suitable average values are given in Table 6. The exponent for a single tight *s* function should thus be generated by multiplying the innermost *s* function by a factor of 20; the exponents for two additional functions are obtained by sequentially multiplying with factors of 12.5 and 25 and so forth. The exponent ratios in Table 6 allow a systematic and efficient convergence toward the basis set limit for the FC contribution, but as argued above, we recommend only the addition of two tight *s* functions, obtained by the sequential scaling of the innermost exponent by ratios of 12.5 and 25.0. We note that these two functions effectively cover the same exponent space as three functions generated by an even-tempered sequence.

Table 6. Recommended Exponent Ratios for Generating Tight Basis Functions

							hydrogen		1. row		2. row	
all elements						basis	p	d	d	f	d	f
1s	20.0					pc-1	9.0		20.0		180.0	
2s	12.5	25.0				pc-2	6.0	9.0	10.0	5.0	45.0	45.0
3s	9.0	12.5	25.0			pc-3	6.0	6.0	5.5	3.5	15	20.0
4s	8.0	9.0	12.5	25.0		pc-4	6.0	6.0	15.0	3.5	5.5	9.0
5s	8.0	8.0	9.0	12.5	25.0							
p	6.5											

Both the PSO and SD contributions are sensitive to the addition of tight *p*-type functions. An optimization of either the PSO or SD contribution separately indicated that the optimum exponent in both cases is close to that obtained by multiplying the innermost *p* function by a constant of 6.5. This validates the procedure used in section III where the exponents are optimized on the basis of the absolute sum of all four contributions. The ratio of 6.5 has been found to be remarkably constant for all of the basis sets and for a variety of systems. For the addition of two tight *p* functions, the corresponding ratios are 5.0 and 6.5, but the contribution from the second *p* function is only a few percent of the effect of the first function and is only significant for the smaller basis sets (pc-0 and pc-1). Arguing again for a balance in the accuracy of the four contributions, the quality of the underlying basis set, and the computational cost, the recommendation is to only extend the pc-*n* basis sets by one tight *p* function.

The SD contribution in addition requires augmentation with both tight *d*- and *f*-type functions. A determination of the optimum exponents for a selection of systems indicated that the optimum value is nearly constant for a given element, until the underlying pc-*n* basis set acquires exponents in the same range. This suggests that the SD contribution is sensitive to basis functions covering a rather specific distance from the nucleus. As expected the optimum value increases with the atomic number. The optimum exponents are again fairly close to those obtained by multiplying the innermost exponents by a fixed ratio, with the recommended values given in Table 6.

Our final recommendation is thus to add two tight *s* functions and one tight *p*, *d*, and *f* function to the pc-*n* basis sets, except that only one tight *s* function is added to the pc-4 basis set. The addition is of course conditioned on the presence of the corresponding type of function in the underlying basis set (e.g., *f* functions are not added to the pc-1 basis set). We will denote these pc-*n* basis sets augmented with tight functions as pcJ-*n*, where the J indicates that the additional functions have been optimized for

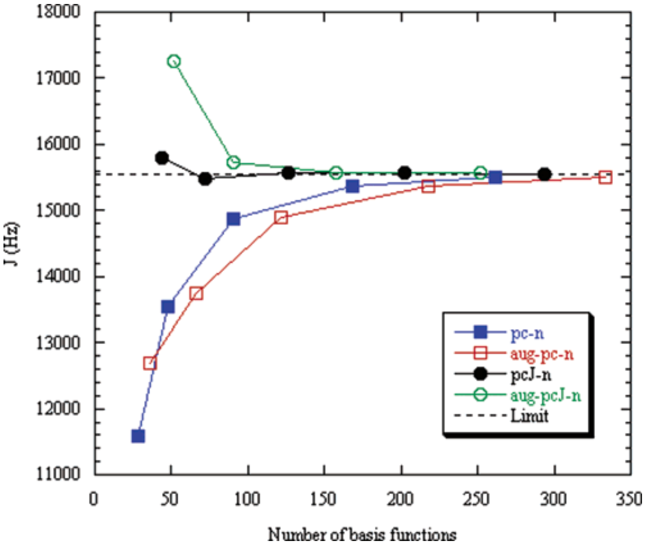


Figure 7. Coupling constant for F_2 calculated by the (uncontracted) pc-*n*, aug-pc-*n*, pcJ-*n*, and aug-pcJ-*n* basis sets.

calculation of the spin−spin coupling constants. Figure 7 shows the basis set convergence of the coupling constant in F_2 for the (uncontracted) pc-*n* and aug-pc-*n* basis sets and the two corresponding J-augmented ones, pcJ-*n* and aug-pcJ-*n*. The plot in Figure 7 has been done using the total number of basis functions as the variable, rather than the maximum angular momentum (Figure 1) to provide a more fair comparison. It is clear that the pcJ-*n* basis sets display a much improved performance, and even the double-ζ-quality basis pcJ-1 provides results close to the limiting value. It is also clear that the inclusion of diffuse functions is counterproductive for this specific system. The performance dependence on diffuse functions for a larger variety of systems will be quantified in section VI.

It is possible that the pcJ-*n* basis sets may also be useful for improving the basis set convergence for other properties depending on the atomic core region, but this clearly will require a careful calibration for each property.

V. Basis Set Contraction

The results in Figure 1 show that the standard contraction scheme is unsuitable for the calculation of spin−spin coupling constants. It should be noted that the F_2 system in this respect is relatively well-behaved, as significantly larger deviations between contracted and uncontracted basis sets have been observed for other systems. The sensitivity to contraction is not unexpected, as an increased flexibility in the core region is required for the calculation of spin−spin coupling constants. Although standard basis sets such as 6-311G** and cc-pVTZ have been tested for the calculation of spin−spin coupling constants,[16,21,23] the lack of tight functions and heavy contraction of the core region clearly makes such standard basis sets unsuitable.

The contraction of a basis set is always a balance between computational efficiency and a loss of accuracy, that is, what is the acceptable degradation of the results for a given gain in computational time. The acceptable loss of accuracy should be graded against the inherent error in the basis set relative to the basis set limiting result; that is, a relatively large contraction error is acceptable if the underlying basis set has a large error relative to the limiting value anyway.

Previous work using the aug-cc-pVTZ basis set augmented with tight *s* functions recommended a partly uncontraction of the outer *s* function, while the contraction coefficients for the additional tight *s* functions were determined from molecular calculations.[8] It was shown that molecular contraction coefficients for one system could be used for others with little loss of accuracy, although the conclusion was based on only a few systems with similar bonding. The use of molecular information has in our opinion an inherent risk of biasing the results. The pc-*n* basis sets employ a general contraction using coefficients from atomic calculations, and we have shown that the contraction coefficient from one exchange-correlation functional can be used for other functionals with little loss of accuracy.[24] The maximum contraction is determined from the condition that the contraction error should be smaller than the inherent error of the uncontracted basis set relative to the basis set limit, and we have used the same approach in the present case.

Contraction of the inner *s* functions is expected to influence primarily the FC contribution. As the very tight functions describe the inner part of the 1s orbital, which is insensitive to the molecular environment, one might expect that this could be contracted with little loss of accuracy. On the other hand, the FC term is a δ function at a specific position, and even minor variations with the molecular species could potentially give substantial changes in the FC contribution. The results in Table 7 show that it is difficult to contract the *s* functions to any significant extent without destroying the accuracy of the uncontracted pcJ-*n* basis sets.

The PSO and SD contributions are insensitive to contraction of the *s*-function space but are sensitive to contraction of the *p*-function space. The contraction errors for the PSO and SD contributions are shown in Tables 8 and 9. It is again seen that only a small contraction of the inner functions is possible without compromising the accuracy. It is also seen that the PSO term is the most sensitive to contraction.

The results in Tables 7−9 illustrate that even the smaller of the pcJ-*n* basis sets (e.g., pcJ-1) are capable of providing results close to the basis set limiting values, and contraction errors must consequently be kept small. Given that the coupling constant depends primarily on the region near the nucleus, this means that it is difficult to contract the basis sets to a significant extent. It should be noted that the results in Tables 7−9 are specific for the F_2 system, but we have performed an analogous contraction analysis on the H_2, CO, C_2H_2, P_2, and SiS systems, with similar results. On the basis of these analyses, we performed calculations for all of the systems in Table 1 with a selection of contraction schemes, and the final recommended contraction schemes are given in Table 10.[25] The basis sets for the second row elements can be contracted somewhat harder than for the first row elements, presumably reflecting that the core region for second row elements is less affected by the molecular environment than for first row elements. The recommended contraction schemes are biased toward computational efficiency; that is, the contraction errors are comparable to or sometimes exceed the error of the uncontracted basis set

Table 7. Contraction Errors (Hz) for the FC Term in the F_2 Molecule[a]

basis	uncontracted	3s	4s	5s	6s	7s	8s	9s	10s	11s	12s	13s	14s	15s	16s	17s	18s	19s
pcJ-0	7s4p (866)	−243	92	46	1	0												
pcJ-1	9s5p (−13)		−515	71	−31	4	−1	0										
pcJ-2	12s7p (−6)				31	−37	20	−11	4	1	0							
pcJ-3	16s10p (3)					16	14	5	−5	3	−2	1	0	0	0			
pcJ-4	19s12p (0)								16	−13	9	−7	4	−4	1	−2	0	0

[a] Value in parentheses is the error of the uncontracted basis set relative to the pcJ-4 result.

Table 8. Contraction Errors (Hz) for the PSO Term in the F_2 Molecule[a]

basis	uncontracted	2p	3p	4p	5p	6p	7p	8p	9p	10p	11p	12p
pcJ-0	7s4p (76)	−201	6	0								
pcJ-1	9s5p (149)	−146	−45	−2	0							
pcJ-2	12s7p (−5)		−65	−20	−3	1	0					
pcJ-3	16s10p (2)			−70	−16	−8	−1	0	0	0		
pcJ-4	19s12p (0)					−9	−6	0	1	0	0	0

[a] Value in parentheses is the error of the uncontracted basis set relative to the pcJ-4 result.

Table 9. Contraction Errors (Hz) for the SD Term in the F_2 Molecule[a]

basis	uncontracted	2p	3p	4p	5p	6p	7p	8p	9p	10p	11p	12p
pcJ-0	7s4p (−743)	−134	2	0								
pcJ-1	9s5p (−227)	−149	−35	3	0							
pcJ-2	12s7p (16)		−83	−19	2	0	0					
pcJ-3	16s10p (16)			−76	−18	−1	1	0	0	0		
pcJ-4	19s12p (0)					−10	1	0	0	0	0	0

[a] Value in parentheses is the error of the uncontracted basis set relative to the pcJ-4 result.

Table 10. Recommended Contraction Schemes for the pcJ-*n* Basis Sets

	hydrogen		1. row		2. row	
basis	uncontracted	contracted	uncontracted	contracted	uncontracted	contracted
pcJ-0	5s	3s	7s4p	4s3p	10s7p	5s4p
pcJ-1	6s2p	4s	9s5p2d	5s4p	13s9p2d	6s5p
pcJ-2	8s3p2d	5s	12s7p3d2f	7s5p	15s11p3d2f	8s6p
pcJ-3	11s5p3d1f	8s	16s10p5d3f1g	10s8p	19s14p5d3f1g	11s10p
pcJ-4	12s7p4d2f1g	10s	19s12p7d4f2g1h	15s10p	22s17p7d4f2g1h	16s12p

relative to the basis set limit. It is, however, much easier to decontract a given basis set than to increase the contraction level, and the pcJ-*n* basis sets can easily be decontracted to reduce the contraction error if required by the specific system. It may be argued that the computational savings by contraction of the pcJ-3 and pcJ-4 basis sets are small, as the computational time tends to be dominated by the many polarization functions, and for benchmark calculations, one should consider using these basis sets in their uncontracted forms.

Although the requirement of using only a low contraction level will significantly increase the computational requirements, there are a few redeeming quantities to be considered. First, the pcJ-*n* basis sets already provide quite reasonable values for pcJ-1 and pcJ-2, which are applicable for a variety of systems. Furthermore, because the coupling constant to a large extent is determined by the atomic region, it is possible that the pcJ-*n* basis sets can be used as local dense basis sets for the nuclei between which the spin−spin coupling is desired, and regular (contracted) pc-*n* basis sets can be used for spectator atoms.[26] How well this will work in practice will require a calibration study and is outside the scope of the present work. For specific classes of coupling constants, one may also consider pruning the pcJ-*n* basis sets selectively; for example, tight *p*, *d*, and *f* functions are not required for coupling constants that are dominated by the FC contribution, and tight *d* and *f* functions are only required if the SD contribution is a large component.

VI. Basis Set Convergence for a Larger Test Set

We have examined the performance of the pcJ-*n* and aug-pcJ-*n* families of basis sets for the systems in Table 1. Because coupling constants vary in magnitude by several thousand hertz, the error at a given level is reported as a percent deviation from the basis set limiting value for coupling constants larger than 10 Hz, as the percent deviation for small absolute values displays large nonsignificant fluctuations. The limiting value has in all cases been taken as the (uncontracted) aug-pcJ-4 result, and the convergence behavior indicates that this value is converged to at least 0.1%.

Table 11 shows the percent-wise mean and maximum absolute deviations (MAD and MaxAD) for the different basis sets over 33 coupling constants for the systems in Table 1. The effect of diffuse functions is system-dependent; that

Table 11. Percent-wise Mean and Maximum Absolute Deviations (MAD and MaxAD) Relative to the aug-pcJ-4 Results for Coupling Constants Larger than 10 Hz for the species in Table 1 (33 Data Points)

	MAD						MaxAD					
	uncontracted				contracted		uncontracted				contracted	
	pc-n	aug-pc-n	pcJ-n	aug-pcJ-n	pcJ-n	aug-pcJ-n	pc-n	aug-pc-n	pcJ-n	aug-pcJ-n	pcJ-n	aug-pcJ-n
pc-0	54.5	41.2	46.6	29.4	47.7	29.9	168.9	98.6	248.0	92.3	250.1	88.2
pc-1	18.5	18.6	4.5	3.2	6.5	4.8	34.1	32.4	19.9	11.7	20.8	25.8
pc-2	7.5	7.6	1.1	0.7	1.6	1.4	13.6	13.4	5.4	6.4	6.2	9.5
pc-3	2.2	2.1	0.3	0.1	0.4	0.8	4.2	3.8	3.1	0.7	3.4	2.8
pc-4	0.9	0.9	0.0	(0)	0.1	0.2	1.9	1.9	0.3	(0)	0.7	0.6

is, the HF, H_2O, F_2, CO, F_2O, C_2F_2, and P_2 systems are sensitive to augmentation with diffuse functions, while the effect for the other systems is very marginal. Furthermore, the effect of diffuse functions is largest for the smaller basis sets, that is, pcJ-0 and pcJ-1, while it rapidly diminishes as the underlying basis set is improved. The pcJ-0 basis set provides rather erratic results and is not recommended for general use. Given the relatively small improvement by adding diffuse functions to the pcJ-1 and pcJ-2 basis sets and the resulting increase in computational time, we do not generally recommend the inclusion of diffuse functions, which is in contrast to previous work using a wave function approach.[7]

The improved convergence indicated in Figure 7 for the F_2 system is also displayed by the MAD and MaxAD values in Table 11. It is encouraging that even the pcJ-1 basis set provides results within ~5% of the limiting value, and it will probably only in special cases be necessary to go beyond the pcJ-2 basis set in practical calculations. For very accurate work, the pcJ-3 and pcJ-4 basis sets should be able to provide the required accuracy. For the slightly smaller test set corresponding to removal of the systems containing Si, P, and Cl, the performance can be compared to the aug-cc-pVTZ-J basis set[7] which is comparable in size to pcJ-2. The MAD and MaxAD values for the aug-cc-pVTZ-J basis set are 1.8% and 8.0%, compared to the pcJ-2 values of 1.4% and 6.2%. Although this improvement is rather marginal, the present work has the advantage of defining a sequence of basis sets which allow a systematic improvement of the accuracy. It should also be noted that it is straightforward to define pcJ-*n* basis sets from the regular pc-*n* ones, when the latter are developed for more elements.

It should be noted that the basis set error is only one possible error component in a comparison with experimental values, as the reference geometry, vibrational averaging, solvent effects, and inadequacies in the exchange-correlation functional will need to be addressed in order to provide a direct comparison with experiments. The vibrational and environmental effects will typically change the coupling constant by ~5%,[5,27] while a typical error arising from the exchange-correlation functional is ~10%.[16] The present pcJ-*n* basis sets are optimized for density functional methods but may also be suitable for wave-function-based methods. It is likely, however, that more optimum basis sets for calculating spin−spin coupling constants with wave-function-based methods could be constructed from, for example, the cc-pVXZ basis sets by augmenting them with tight functions determined along the principles used in the present work.

VII. Conclusions

The previously proposed polarization-consistent basis sets have been augmented with tight *s*, *p*, *d*, and *f* functions for improving the basis set convergence for calculating spin−spin coupling constants with density functional methods. The analysis shows that tight *s* functions are required for the Fermi-contact term, tight *p* functions are necessary for the paramagnetic spin−orbital contribution, while tight *p*, *d*, and *f* functions are required to converge the spin-dipolar contribution. The necessity of adding tight *s* functions has been noted by several other groups, but the requirement of tight *p*, *d*, and *f* functions for the noncontact terms is novel. The optimum exponents can be determined by a variational procedure based on the sum of the absolute value of all contributions. The optimum exponents display sufficient regularity from which general rules for assigning standard exponents for a given element can be derived. Using these rules, we propose a sequence of pcJ-*n* basis sets which is capable of converging the spin−spin coupling constant to within ~0.1% of the basis set limiting value. The basis set convergence of the pcJ-*n* basis sets is significantly improved relative to the regular basis sets, and the pcJ-1 and pcJ-2 basis sets should be efficient basis sets for calculating spin−spin coupling constants. While these basis sets by default contain tight *s*, *p*, *d*, and *f* functions, some of these functions may be omitted for specific cases. For coupling constants dominated by the Fermi-contact term, only the tight *s* functions are required, while a tight *p* function is only necessary if either the paramagnetic spin−orbital or spin-dipolar contribution is significant. The latter is often small, and in such cases, the tight *d* and *f* functions can be removed. Such basis set pruning, however, must be decided on a case-by-case basis.

Acknowledgment. This work was supported by grants from the Danish Center for Scientific Computation and the Danish Natural Science Research Council. The author thanks Stephan Sauer for helpful discussions during this work.

Supporting Information Available: Exponents and contraction coefficients for the pcJ-*n* and aug-pcJ-*n* basis sets for the elements H, C, N, O, F, Si, P, S, and Cl. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Provasi, P. F.; Aucar, G. A.; Sanchez, M.; Alkorta, I.; Elguero, J.; Sauer, S. P. A. *J. Phys. Chem. A* **2005**, *109*, 6555. Gräfenstein, J.; Tuttle, T.; Cremer, D. *J. Phys. Chem. A* **2005**, *109*, 2325. Wu, A.; Cremer, D.; Auer, A. A.; Gauss, J. *J. Phys. Chem. A* **2002**, *106*, 657. Del Bene, J. E.; Perera, S. A.; Bartlett, R. J. *J. Am. Chem. Soc*. **2000**, *122*, 3560.

(2) Kirpekar, S.; Jensen, H. J. A.; Oddershede, J. *Chem. Phys.* **1994**, *188*, 171. Perera, S. A.; Nooijen, M.; Bartlett, R. J. *J. Chem. Phys.* **1996**, *104*, 3290. Helgaker, T.; Jaszunski, M.; Ruud, K. *Chem. Rev*. **1999**, *99*, 293. Melo, J. I.; Ruiz de Azua, M. C.; Peralta, J. E.; Scuseria, G. E. *J. Chem. Phys*. **2005**, *123*, 204112. Jensen, F. *Introduction to Computational Chemistry*; Wiley: New York, 1999. Kaupp, M.; Bühl, M.; Vladimir, M. G. *Calculation of NMR and ESR Parameters*; Wiley-VCH: New York, 2004. Sauer, S. P. A.; Packer, M. J. In *Computational Molecular Spectroscopy*; Jensen, P., Bunker, P. R., Eds.; Wiley: New York, 2000.

(3) Maciel, G. E.; McIver, J. W., Jr.; Ostlund, N. S.; Pople, J. A. *J. Am. Chem. Soc*. **1970**, *92*, 4151.

(4) Parr, R. G.; Yang, W. *Density Functional Theory*; Oxford University Press: New York, 1989. Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: New York, 2000.

(5) Lutnaes, O. B.; Ruden, T. A.; Helgaker, T. *Magn. Reson. Chem*. **2004**, *42*, S117.

(6) Keal, T. W.; Tozer, D. *J. Chem. Phys*. **2005**, *121*, 5654.

(7) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys*. **1992**, *96*, 6796. Jensen, F. *J. Chem. Phys.* **2002**, *117*, 9234.

(8) Provasi, P. F.; Aucar, G. A.; Sauer, S. P. A. *J. Chem. Phys*. **2001**, *115*, 1324.

(9) Enevoldsen, T.; Oddershede, J.; Sauer, S. P. A. *Theor. Chem. Acc*. **1988**, *100*, 275. Provasi, P. F.; Aucar, G. A.; Sauer, S. P. A. *Int. J. Mol. Sci*. **2003**, *4*, 231. Provasi, P. F.; Aucar, G. A.; Sauer, S. P. A. *J. Phys. Chem. A* **2004**, *108*, 5393.

(10) Barone, V.; Provasi, P. F.; Peralta, J. E.; Snyder, J. P.; Sauer, S. P. A.; Contreras, R. H. *J. Phys. Chem. A* **2003**, *107*, 4748.

(11) Helgaker, T.; Watson, M.; Handy, N. C. *J. Chem. Phys*. **2000**, *113*, 9402.

(12) Helgaker, T.; Jaszunski, M.; Ruud, K.; Gorska, A. *Theor. Chem. Acc*. **1998**, *99*, 175.

(13) Peralta, J. E.; Scuseria, G. E.; Cheeseman, J. R.; Frisch, M. J. *Chem. Phys. Lett*. **2003**, *375*, 452.

(14) Dunning, T. H., Jr. *J. Chem. Phys*. **1989**, *90*, 1007. Wilson, A. K.; van Mourik, T.; Dunning, T. H., Jr. *J. Mol. Struct*. **1996**, *388*, 339.

(15) Manninen, P.; Vaara, J. *J. Comput. Chem*. **2006**, *27*, 434.

(16) Deng, W.; Chesseman, J. R.; Frisch, M. J. *J. Chem. Theory Comput*. **2006**, *2*, 1028.

(17) Jensen, F. *J. Chem. Phys.* **2001**, *115*, 9113. Jensen, F. *J. Chem. Phys.* **2002**, *116*, 3502. Jensen, F. *J. Chem. Phys.* **2002**, *116*, 7372. Jensen, F. *J. Chem. Phys.* **2002**, *117*, 9234. Jensen, F. *J. Chem. Phys.* **2003**, *118*, 2459. Jensen, F.; Helgaker, T. *J. Chem. Phys.* **2004**, *121*, 3462.

(18) Helgaker, T.; Jensen, H. J. A.; Jørgensen, P.; Olsen, J.; Ruud, K.; Ågren, H.; Auer, A. A.; Bak, K. L.; Bakken, V.; Christiansen, O.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Hättig, C.; Hald, K.; Halkier, A.; Heiberg, H.; Hettema, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Pedersen, T. B.; Ruden, T. A.; Sanchez, A.; Saue, T.; Sauer, S. P. A.; Schimmelpfenning, B.; Sylvester-Hvid, K. O.; Taylor, P. R.; Vahtras, O. *DALTON*, release 2.0; 2005.

(19) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem*. **1994**, *98*, 11623.

(20) Maximoff, S. N.; Peralta, J. E.; Barone, V.; Scuseria, G. E. *J. Chem. Theory Comput*. **2005**, *1*, 541. Keal, T. W.; Helgaker, T.; Salek, P.; Tozer, D. J. *Chem. Phys. Lett*. **2006**, *425*, 163.

(21) Barone, V.; Peralta, J. E.; Contreras, R. H.; Snyder, J. P. *J. Phys. Chem. A* **2002**, *106*, 5607.

(22) Malkin, V. G.; Malkina, O. L.; Salahub, D. R. *Chem. Phys. Lett.* **1994**, *221*, 91. Malkin, O. L.; Salahub, D. R.; Malkin, V. G. *J. Chem. Phys*. **1996**, *105*, 8793. Peralta, J. E.; Barone, V.; Contreras, R. H.; Zaccari, D. G.; Snyder, J. P. *J. Am. Chem. Soc.* **2001**, *123*, 9162. Lantoo, P.; Vaara, J.; Helgaker, T. *J. Chem. Phys*. **2002**, *117*, 5998.

(23) Hansen, M. J.; Wendt, M. A.; Weinhold, F.; Farrar, T. C. *Mol. Phys*. **2002**, *100*, 2807.

(24) Jensen, F. *Chem. Phys. Lett.* **2005**, *402*, 510.

(25) All of the pcJ-*n* basis sets are available as Supporting Information.

(26) Sanchez, M.; Provasi, P. F.; Aucar, G. A.; Sauer, S. P. A. *Adv. Quantum Chem*. **2005**, *48*, 161.

(27) Mikkelsen, K. V.; Ruud, K.; Helgaker, T. *J. Comput. Chem*. **1999**, *20*, 1281. Ruden, T. A.; Helgaker, T.; Jaszunski, M. *Chem. Phys*. **2004**, *296*, 53.

CT600166U

JCTC Journal of Chemical Theory and Computation

An Efficient Linear-Scaling Electrostatic Coupling for Treating Periodic Boundary Conditions in QM/MM Simulations

Teodoro Laino,*,† Fawzi Mohamed,‡ Alessandro Laio,‡ and Michele Parrinello‡

Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, I-56125 Pisa, Italy, and
Computational Science, Department of Chemistry and Applied Biosciences,
ETH Zürich, USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

Received March 29, 2006

Abstract: A new linear-scaling method based on a multigrid approach to treat long-range electrostatic interactions in hybrid quantum mechanics/molecular mechanics (QM/MM) simulations is described. The scheme has been implemented in the context of a QM calculation based on density functional theory (DFT). The method is tested on an analytical model to validate the new algorithm. Two realistic problems in α-quartz crystals and a zwitterionic dipeptide (GLY-ALA) in water have been chosen as further tests. Results from QM/MM calculations with periodic boundary conditions (PBC) show that the use of PBC is essential when studying highly ordered crystal structures, unless a carefully designed MM crystal is used for the calculation. With a general shaped MM subsystem, the absence of PBC leads to an incorrect description of Kohn−Sham band gaps and charge density. The present method allows periodic boundary conditions to be used in molecular simulations of biological and material science systems.

1. Introduction

Understanding and predicting the properties of condensed systems requires, among other things, the reliable treatment of the long-range Coulomb interactions.[1] Many schemes have been proposed, and a consensus has emerged regarding the use of periodic boundary conditions (PBC) to treat properly the interactions between the periodic replicas. In fact, it is usually believed that PBC have to be used together with an exact treatment of the long-range interactions between the periodically repeated images in order to obtain reliable MD simulations.[2−7]

Although numerous methods have been developed to avoid the truncation of the electrostatic interactions beyond a given cutoff, they have all been applied in the context of classical simulations.[8−13] These methods are all based on different techniques to compute the lattice sum involved in the evaluation of the long-range electrostatic interactions within PBC.

The treatment of long-range forces in conjunction with PBC is much less well established for hybrid quantum mechanics/molecular mechanics (QM/MM) simulations. So far, most of the QM/MM implementations have relied on a spherical truncation scheme, in which the solute(QM)−solvent(MM) electrostatic interactions are neglected beyond a certain cutoff distance R_c. There are only a few exceptions to this implementation. A very popular and inexpensive approach is the reaction field method, which couples the spherical truncation scheme with a polarizable continuum medium that extends beyond a cutoff distance R_c.[14−19] Ewald's lattice summation techniques were also investigated to treat the long-range QM/MM electrostatic interactions. Within a semiempirical framework, the first implementation is due to Gao and Alhambra.[20] In their scheme only the long-range QM/MM interactions are evaluated, while the QM/QM ones are omitted. For the particular set of applications tested by these authors, namely solvation phenomena, the solute−solvent (QM/MM) interactions were considered as the determining ones. Recent implementations of Ewald

---

* Corresponding author fax: +41919138817; e-mail: teodoro.laino@sns.it.

† Scuola Normale Superiore di Pisa.

‡ ETH Zürich.

10.1021/ct6001169 CCC: $33.50

techniques extended to the full QM/MM long-range interactions[21,22] show indeed that even for solvation cases long-range QM/QM electrostatic interactions play a significant role. Within a self-consistent DFT scheme, to the best of our knowledge there is only one QM/MM scheme that allows PBC[23] to be used. This approach is conceptually similar to that of the present work, and it relies on the use of splines in reciprocal space (k-space), optimally designed for use within plane wave (PW) codes. The present work[24] is, on the other hand, based on real space techniques and is designed to be used with Gaussian basis codes, as is CP2K.[25]

In a recent paper we proposed a new computational scheme based on the Gaussian expansion of the electrostatic potential (GEEP).[24] This technique can be used efficiently to map functions on a grid, and we applied it to the development of a new QM/MM electrostatic coupling for isolated systems which exhibits linear scaling.

The new scheme, implemented into the CP2K package,[25] extends the recently implemented QM/MM method[24] to applications where the use of PBC is required. It relies on the most efficient methods for calculating long-range electrostatic interactions of point charges within PBC and scales linearly with respect to the number of MM atoms. Moreover the evaluation of the MM electrostatic potential using PBC is independent of the number of QM atoms, depending only on the dimension of the coarsest grid used in the multigrid approach.

An accuracy test was first performed on an analytically solvable model in order to provide a clear and unambiguous validation of the new approach. As a more realistic test case we apply the new method to the study of SiO_2 and its charged oxygen vacancy defect and on a zwitterionic dipeptide (GLY-ALA) in water. The tests show clearly that to achieve a correct description of the QM/MM system there are only two possibilities: the first is based on the optimization of the shape and charges of the MM crystal, to reproduce correctly the long-range stabilizing effects within a truncation scheme,[26] and the other is to treat long-range interactions with a proper PBC scheme. Although results are similar for both approaches the use of QM/MM-PBC schemes avoids the preparation steps related to the fine-tuning of the MM subsystem. For the solvated zwitterionic dipeptide we find, as expected, that the use of QM/MM-PBC can be avoided due to the high dielectric shield and to the lack of long-range stabilizing effects.

2. Methodology

Assuming the overall charge neutrality condition, the total energy of a QM/MM simulation within PBC can be easily evaluated

$$E^{\mathrm{TOT}} = \frac{1}{2}\int\int \mathrm{d}r\mathrm{d}r' \frac{\rho(r)\rho(r')}{|r-r'|} \tag{1}$$

with $\rho = \rho^{\mathrm{QM}} + \rho^{\mathrm{MM}}$ being the total charge density of the system (see Figure 1a). Once the total density is split into a QM and a MM part both subsystems could in principle possess an overall net charge different from zero. Therefore the use of a neutralizing background charge (ρ^{B}) is necessary to avoid divergence in treating electrostatic within PBC. The total energy term can be split into three separate terms (see Figure 1a):

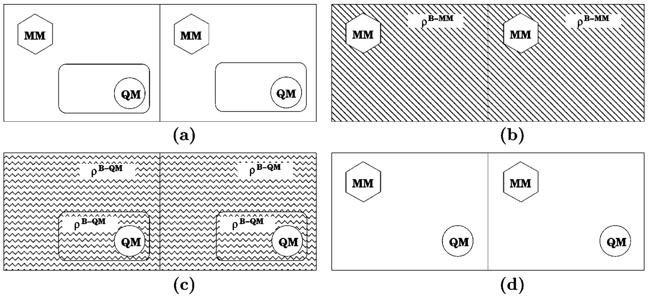


Figure 1. These frames show the decomposition of the total QM/MM energy. In each frame two of the many periodic replica have been shown. Frame (a) shows the total system. Frame (b) shows the energy of the MM subsystem embedded in the neutralizing background charge (deriving from the division of the QM and MM subsystems). Frame (c) shows the energy of the QM subsystem with the neutralizing background charge of the QM cell and that relating to the MM cell. The last frame (see frame (d)) depicts the QM/MM pure electrostatic mutual interaction term.

$$E^{\mathrm{MM}} = \frac{1}{2}\int\int \mathrm{d}r\mathrm{d}r' \frac{(\rho^{\mathrm{MM}}(r) + \rho^{\mathrm{B,MM}})(\rho^{\mathrm{MM}}(r') + \rho^{\mathrm{B,MM}})}{|r-r'|} \tag{2}$$

$$E^{\mathrm{QM}} = \frac{1}{2}\int\int \mathrm{d}r\mathrm{d}r' \frac{(\rho^{\mathrm{QM}}(r) + \rho^{\mathrm{B,QM}})(\rho^{\mathrm{QM}}(r') + \rho^{\mathrm{B,QM}})}{|r-r'|} \tag{3}$$

$$E^{\mathrm{QM/MM}} = \int\int \mathrm{d}r\mathrm{d}r' \frac{(\rho^{\mathrm{QM}}(r) + \rho^{\mathrm{B,QM}})(\rho^{\mathrm{MM}}(r') + \rho^{\mathrm{B,MM}})}{|r-r'|} \tag{4}$$

The physical nature of these terms is illustrated pictorially in Figure 1. Assuming the total charge of the system is zero (although this assumption can be relaxed with no modifications to the formalism) the mixed terms involving the neutralizing background charge of the $E^{\mathrm{QM/MM}}$ cancel the interaction terms of the QM and MM density with their own background charges. The total expression for the three terms is

$$E^{\mathrm{MM}} = \frac{1}{2}\int\int \mathrm{d}r\mathrm{d}r' \frac{(\rho^{\mathrm{MM}}(r))(\rho^{\mathrm{MM}}(r'))}{|r-r'|} \tag{5}$$

$$E^{\mathrm{QM}} = \frac{1}{2}\int\int \mathrm{d}r\mathrm{d}r' \frac{(\rho^{\mathrm{QM}}(r))(\rho^{\mathrm{QM}}(r'))}{|r-r'|} \tag{6}$$

$$E^{\mathrm{QM/MM}} = \int\int \mathrm{d}r\mathrm{d}r' \frac{(\rho^{\mathrm{QM}}(r))(\rho^{\mathrm{MM}}(r'))}{|r-r'|} \tag{7}$$

The first term (Figure 1b) is evaluated using standard techniques such as particle−particle or particle−mesh schemes. The second term (Figure 1c) is the evaluation of the energy of the QM subsystem. Since the total energy of the QM subsystem is usually evaluated exploiting a smaller cell, care needs to be taken to include the correct electrostatic interactions of the periodic QM replicas. The last term (Figure 1d) is the evaluation of the periodic MM electrostatic potential, partitioned into a real space contribution and a periodic correction. The real space term contains the interaction due to the short-range part of the electrostatic potential of the MM charges with the total quantum charge distribution (electrons plus nuclei). Only MM atoms close to the QM region will contribute to this term. The periodic term contains instead the long-range effects of the MM subsystem.

In the next section, the standard Ewald method is briefly revised for a N-point charge particle system interacting in an orthorhombic box of edge L_x, L_y, L_z. Particular care is then devoted to the discussion of the use of the Ewald lattice summation with the GEEP scheme.[24] Finally we discuss the algorithm to decouple/recouple multiple QM images. In the following Latin letters a and b will be used to index the MM atoms, while Greek letters α and β will be used for QM atoms.

2.1. Ewald Lattice Summation for Electrostatic Interactions. Given an N-point charge particle system, the electrostatic potential $\Phi_{\text{tot}}(\mathbf{r})$ at position $\mathbf{r}$ is evaluated using the Ewald lattice sum technique.[8] In this approach, $\Phi_{\text{tot}}(\mathbf{r})$ is split into the sum of two potentials, using a Gaussian screening charge of width κ:

$$\Phi_{\text{tot}}(\mathbf{r}) = \Phi_{\text{rec}}(\mathbf{r}) + \Phi_{\text{real}}(\mathbf{r}) \tag{8}$$

The reciprocal space potential term $\Phi_{\text{rec}}(\mathbf{r})$ can be determined using the Fourier series

$$\Phi_{\text{rec}}(\mathbf{r}) = \frac{4\pi}{V}\sum_{\mathbf{k}\neq 0}\frac{e^{-k^2/4\kappa}}{k^2}\sum_{a}^{\text{MM}} q_a e^{-i\mathbf{k}\cdot|\mathbf{r}-\mathbf{r}_a|} \tag{9}$$

where $\mathbf{k} = [2\pi n_x/L_x^2, 2\pi n_y/L_y^2, 2\pi n_z/L_z^2]$, and V is the volume of the primary unit cell. The real space part of the Ewald potential is given by

$$\Phi_{\text{real}}(\mathbf{r}) = \sum_{a}^{\text{MM}}\sum_{|\mathbf{L}|\leq L_{\text{cut}}} q_a \frac{\text{Erfc}(\kappa|\mathbf{r}-\mathbf{r}_a+\mathbf{L}|)}{|\mathbf{r}-\mathbf{r}_a+\mathbf{L}|} \tag{10}$$

where $\mathbf{L} = [n_xL_x, n_yL_y, n_zL_z]$ counts the periodic images with n_k integers. As the Erfc has a real space short-range property, only the $|\mathbf{L}| \leq L_{\text{cut}}$ periodic images will contribute to the real space term of the electrostatic potential.

2.2. GEEP. The evaluation of a scalar field on grids can be efficiently achieved by exploiting the recently proposed technique of the Gaussian expansion of the electrostatic potential, namely GEEP,[24] that we applied to the evaluation of the QM/MM electrostatic potential. Following ref 24, the charge on MM atoms is represented with a Gaussian distribution

$$\rho(|\mathbf{r}-\mathbf{r}_a|) = \left(\frac{1}{\sqrt{\pi}r_{c,a}}\right)^3 \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_a|^2}{r_{c,a}^{\ 2}}\right) \tag{11}$$

where $r_{c,a}$ is the width of the Gaussian charge of the classical atom a. An accurate analysis of the use of nonpointlike atomic potentials in QM/MM simulations has recently been published.[27] The electrostatic potential originated by the Gaussian charge distribution can be evaluated analytically:

$$v_a(\mathbf{r},\mathbf{r}_a) = \frac{\text{Erf}(|\mathbf{r}-\mathbf{r}_a|/r_{c,a})}{|\mathbf{r}-\mathbf{r}_a|} \tag{12}$$

Using GEEP we can write this potential energy function as a series of short-range functions plus a residual function with a very low cutoff:

$$v_a(\mathbf{r},\mathbf{r}_a) = \frac{\text{Erf}(|\mathbf{r}-\mathbf{r}_a|/r_{c,a})}{|\mathbf{r}-\mathbf{r}_a|} = \sum_{N_g} A_g \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_a|^2}{G_g^{\ 2}}\right) + R_{\text{low}}(|\mathbf{r}-\mathbf{r}_a|) \tag{13}$$

The smoothed Coulomb potential is then expressed as a sum of N_g Gaussian functions and of a residual function R_{low}. The A_g are the amplitudes of the Gaussian functions, G_g their width. The Gaussian functions will be mapped on grids and their contribution neglected if their value is less than a preassigned threshold. This is equivalent to assuming that Gaussians are compact support functions, namely that they are numerically zero beyond a certain distance from the MM atom on which the Gaussian is centered. If the parameters A_g and G_g are properly chosen, the residual function R_{low} is smooth, i.e., its Fourier transform is approximately zero for $|\mathbf{k}| \gg k_{\text{cut}}$. The k_{cut} parameter is related to the spacing of the grid on which the R_{low} function will be mapped. Within a multigrid framework this implies that a good representation of the function is obtained on a coarse grid, chosen in order to have the same grid cutoff as the cutoff of the residual function.

It can be easily shown that the evaluation of the electrostatic potential scales linearly for systems as small as hundreds of atoms, with a prefactor which depends only on the number of grid points of the coarsest grid level.[24]

2.3. QM/MM Periodic Potential. The QM/MM periodic potential (see Figure 1d) on a generic point i of the finest grid level can be computed using the real space lattice sum

$$V^{\text{fine}}(\mathbf{r}_a)_i = \sum_{a}^{\text{MM}}\sum_{\mathbf{L}}^{\infty}{}' q_a v_a(\mathbf{r}_i,\mathbf{r}_a+\mathbf{L}) \tag{14}$$

where $\mathbf{r}_i$ is the coordinate of the point i of the finest grid level and $\mathbf{r}_a$ indexes the functional dependence from the set of MM atomic coordinates. The summation over $\mathbf{L}$ involves all integer translations of the real space lattice vectors $\mathbf{L} = [n_xL_x, n_yL_y, n_zL_z]$ for integers n_k, and the prime symbol indicates that when $\mathbf{L} = 0$ the term $|\mathbf{r}_i - \mathbf{r}_a| = 0$ is neglected. $v_a(\mathbf{r}_i,\mathbf{r}_a + \mathbf{L})$ represents the functional form of the electrostatic potential. The summation in eq 14 has the same convergency properties as the standard Ewald summation schemes.[8]

The total QM/MM electrostatic energy can be split into two rapidly convergent terms,[8,28] one over real space and the other over reciprocal space lattice vectors

$$E^{\text{QM/MM}}(\mathbf{r}_\alpha,\mathbf{r}_a) = E^{\text{QM/MM}}_{\text{real}}(\mathbf{r}_\alpha,\mathbf{r}_a) + E^{\text{QM/MM}}_{\text{recip}}(\mathbf{r}_\alpha,\mathbf{r}_a) \tag{15}$$

where

$$E_{\mathrm{real}}^{\mathrm{QM/MM}}(\mathbf{r}_\alpha,\mathbf{r}_a) = \int d\mathbf{r}\rho(\mathbf{r},\mathbf{r}_\alpha)V_{\mathrm{real}}^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a) \tag{16}$$

and

$$E_{\mathrm{recip}}^{\mathrm{QM/MM}}(\mathbf{r}_\alpha,\mathbf{r}_a) = \int d\mathbf{r}\rho(\mathbf{r},\mathbf{r}_\alpha)V_{\mathrm{recip}}^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a) \tag{17}$$

The definition of the two terms is strictly connected to the type of functional form used to describe the Coulomb interactions. In our case, since we decided to treat the MM classical charges through a Gaussian charge distribution, the electrostatic potential function has the analytical form

$$v_a(\mathbf{r},\mathbf{r}_a) = \frac{\mathrm{Erf}(|\mathbf{r}-\mathbf{r}_a|/r_{c,a})}{|\mathbf{r}-\mathbf{r}_a|} \tag{18}$$

easily represented as a sum of two terms:[24]

$$v_a(\mathbf{r},\mathbf{r}_a) = \frac{\mathrm{Erf}(|\mathbf{r}-\mathbf{r}_a|/r_{c,a})}{|\mathbf{r}-\mathbf{r}_a|} = \sum_{N_g} A_g \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_a|^2}{G_g^{\,2}}\right) + R_{\mathrm{low}}(|\mathbf{r}-\mathbf{r}_a|) \tag{19}$$

The best choice is to use the mathematical properties of the two functional forms (short-range term and long-range term) to define the division into real and reciprocal space contributions:

$$v_a(\mathbf{r},\mathbf{r}_a) = \frac{\mathrm{Erf}(|\mathbf{r}-\mathbf{r}_a|/r_{c,a})}{|\mathbf{r}-\mathbf{r}_a|} = \sum_{N_g} A_g \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_a|^2}{G_g^{\,2}}\right) + R_{\mathrm{low}}(|\mathbf{r}-\mathbf{r}_a|) \tag{20}$$

$$= v_a^{rs}(\mathbf{r},\mathbf{r}_a) + v_a^{\mathrm{recip}}(\mathbf{r},\mathbf{r}_a) \tag{21}$$

All short-range interactions will be evaluated in the real space, while all long-range interactions will be taken into account in the reciprocal space formalism.

The real space term $V_{\mathrm{real}}^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a)$ is defined as

$$V_{\mathrm{real}}^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a) = \sum_{|\mathbf{L}| \le L_{\mathrm{cut}}} \sum_a q_a v_a^{rs}(\mathbf{r},\mathbf{r}_a+\mathbf{L}) = \sum_{|\mathbf{L}| \le L_{\mathrm{cut}}} \sum_a q_a \left[\sum_{N_g} A_g \exp\left(-\frac{|\mathbf{r}-\mathbf{r}_a+\mathbf{L}|^2}{G_g^{\,2}}\right)\right] \tag{22}$$

where a labels the MM atoms. The radii of the Gaussians are such that only a few periodic images ($|\mathbf{L}| \le L_{\mathrm{cut}}$) are needed to achieve convergence of the real space term, while others give zero contribution. As in ref 24, each Gaussian of eq 22 is mapped on the appropriate grid level.

The effect of the periodic replicas of the MM subsystem is only in the long-range term, and it comes entirely from the residual function $R_{\mathrm{low}}(\mathbf{r},\mathbf{r}_a)$ of eq 20:

$$V_{\mathrm{recip}}^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a) = \sum_{\mathbf{L}}^{\infty}{}' \sum_a q_a v_a^{\mathrm{recip}} = \sum_{\mathbf{L}}^{\infty}{}' \sum_a q_a R_{\mathrm{low}}(|\mathbf{r}-\mathbf{r}_a+\mathbf{L}|) \tag{23}$$

Performing the same manipulation used in Ewald summation[8] (see Appendix B) the previous equation can be computed more efficiently in the reciprocal space:

$$V_{\mathrm{recip}}^{\mathrm{QM/MM}}(\mathbf{r}_i,\mathbf{r}_a) = L^{-3}\sum_{\mathbf{k}}^{k_{\mathrm{cut}}}{}' \sum_a \tilde{R}_{\mathrm{low}}(\mathbf{k}) q_a \cos[2\pi\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_a)] \tag{24}$$

The term $\tilde{R}_{\mathrm{low}}(\mathbf{k})$, representing the Fourier transform of the smooth electrostatic potential, can be evaluated analytically:

$$\tilde{R}_{\mathrm{low}}(\mathbf{k}) = \left[\frac{4\pi}{|\mathbf{k}|^2}\right]\exp\left(-\frac{|\mathbf{k}|^2 r_{c,a}^{\,2}}{4}\right) - \sum_{N_g} A_g (\pi)^{3/2} G_g^{\,3} \exp\left(-\frac{G_g^{\,2}|\mathbf{k}|^2}{4}\right) \tag{25}$$

The potential in eq 24 can be mapped on the coarsest grid. In fact, the long-range contribution is physically very smooth, and a good representation can be achieved with large grid spacings. Furthermore, since the R_{low} function is a low cutoff function, $\tilde{R}_{\mathrm{low}}(\mathbf{k})$ is zero for all $\mathbf{k}$-vectors larger than a well defined k_{cut}. The k_{cut} parameter depends strongly on the number of Gaussian functions used in the GEEP scheme (as described in ref 24).

Once the electrostatic potential of a single MM charge within periodic boundary conditions is derived, the evaluation of the electrostatic potential due to the MM subsystem is easily computed employing the same multigrid operators (interpolation and restriction) described in ref 24.

2.4. Periodic Coupling with QM Images. In the present section we complete the description of the electrostatic coupling, discussing the interaction between the periodic images of the QM replicas (see Figure 1c). The Quickstep[29,30] algorithm uses a mixed plane wave/Gaussian basis set to solve the DFT equations for the quantum subsystem. The plane waves are used to compute efficiently the Hartree potential. Therefore, unless the quantum box and the MM box have the same dimensions, the QM images, interacting by PBC implicit in the evaluation of the Hartree potential, have the wrong periodicity.

To avoid this error, the QM problem is usually solved using standard decoupling techniques.[31,32] This approximation is legitimate when the evaluation of the QM/MM potential is performed using spherical truncation schemes for Coulomb interactions.

Since we want to describe the long-range QM/MM interaction with periodic boundary conditions, we may not neglect the QM/QM periodic interactions, which play a significant role if the QM subsystem has a net charge different from zero or a significant dipole moment. Therefore we exploit a technique recently proposed by Blöchl,[32] which decouples the periodic images in order to restore the correct periodicity also for the QM part. A full and comprehensive description of the methods to evaluate energy corrections and derivatives is given in ref 32. Here we summarize Blöchl's decoupling scheme. Given a QM total density charge $\rho(\mathbf{r},\mathbf{r}_\alpha)$, the electrostatic energy of this isolated density is

$$E = \frac{1}{2}\int_V \mathrm{d}\mathbf{r} \int \mathrm{d}\mathbf{r}' \frac{\rho(\mathbf{r},\mathbf{r}_\alpha)\rho(\mathbf{r}',\mathbf{r}_\alpha)}{|\mathbf{r} - \mathbf{r}'|} \tag{26}$$

Let us introduce a new model charge density $\hat{\rho}(\mathbf{r},\mathbf{r}_\alpha)$, which is localized within the same volume V as $\rho(\mathbf{r},\mathbf{r}_\alpha)$ and which reproduces the multipole moments of the correct charge distribution. The representation adopted in ref 32 is given as a sum

$$\hat{\rho}(\mathbf{r},\mathbf{r}_\alpha) = \sum_\alpha q_\alpha g_\alpha(\mathbf{r},\mathbf{r}_\alpha) \tag{27}$$

of atom-centered spherical Gaussians, which are normalized such that they possess a charge of one

$$g_i(\mathbf{r},\mathbf{r}_\alpha) = \frac{1}{(\sqrt{\pi} r_{\mathrm{c},\alpha})^3} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}_\alpha|^2}{{r_{\mathrm{c},\alpha}}^2}\right) \tag{28}$$

where $\mathbf{r}_\alpha$ denotes a particular atomic site. Every atomic site may be the center of various Gaussians with different decay lengths $r_{\mathrm{c},\alpha}$. By construction, the multipole moments of the model charge density agree with those of the original charge distribution. Since the electrostatic interaction of separated charge distribution (the array of periodic QM charge densities) depends only on its multipole moments, the model charge density is used to modify the Hartree potential and to cancel the electrostatic interactions between the periodic images. In Appendix A, we briefly summarize with a matrix formalism the charge fit scheme as derived in ref 32. In the same way that the Blöchl scheme cancels the electrostatic interactions between periodic images, it is possible to use it to include the electrostatic interactions between periodic images with the periodicity of the MM box.

2.5. QM/MM Forces. The derivatives on MM atoms can be easily evaluated taking the derivative of both terms in real space and in reciprocal space and summing the contribution of the different grid levels. The derivatives of the real space term are the same as the one presented in ref 24. The derivatives of the reciprocal space term need to be evaluated by deriving the MM nuclei potential energy contribution and integrating this derivative with the quantum charge distribution

$$\frac{\partial E^{\mathrm{QM/MM}}_{\mathrm{recip}}(\mathbf{r}_\alpha,\mathbf{r}_a)}{\partial \mathbf{r}_a} = \int d\mathbf{r}\rho(\mathbf{r},\mathbf{r}_\alpha)\frac{\partial V^{\mathrm{QM/MM}}_{\mathrm{recip}}(\mathbf{r},\mathbf{r}_a,\mathbf{r}_\alpha)}{\partial \mathbf{r}_a} = \tag{29}$$

$$\Delta\omega \sum_{\mathbf{r}_i} \rho(\mathbf{r}_i,\mathbf{r}_\alpha) L^{-3} \sum_{\mathbf{k}}^{k_{\mathrm{cut}}}{}' \sum_{a}^{\mathrm{MM}} \tilde{R}_{\mathrm{low}}(\mathbf{k}) q_a \frac{\partial \cos[2\pi\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_a)]}{\partial \mathbf{r}_a} \tag{30}$$

where $\Delta\omega$ is the volume element of the coarsest grid level. This contribution is summed with the terms in real space to obtain the total derivatives on MM atoms. The derivatives on QM atoms are computed in the same way as we described in ref 24, the only difference being that the QM derivatives are modified by the coupling/decoupling terms. These corrections have been derived and extensively discussed in ref 32.

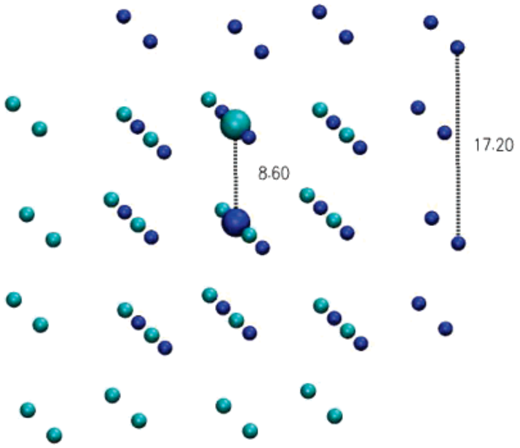


Figure 2. Orthorhombic cell of face centered cubic lattice of Gaussian charges. The two big spheres represent the QM atoms. Lattice parameter 17.2 Å. The Gaussian charges have a width of $0.5\sqrt{2}$ Å.

3. Tests and Applications

Four systems were selected to test the new method. The first one, an infinite array of Gaussian alternating opposite charges, can be solved analytically and therefore provides a clear and unambiguous test of the accuracy of our new approach.

The second system is a periodic model of α-quartz (α-SiO_2) where a bulk fragment, described at the DFT level, is embedded in the environment of classical atoms described with MM force fields. The third system analyzes a charged oxygen vacancy defect in α-quartz, in the same periodic model. These two systems do not possess an analytical solution but both have been extensively studied experimentally[33−40] and theoretically.[41−49]

The last system is a zwitterionic dipeptide (GLY-ALA) in water. It was chosen since it represents an extreme test where the use of PBC is expected to have minor effects on the electronic structure properties.

3.1. Analytical Test. To validate this new algorithm, let us consider a simple Gaussian charge distribution

$$\rho(\mathbf{r}_\alpha) = (\kappa/\pi)^{3/2}\exp(-\kappa^2|\mathbf{r}_\alpha|^2) \tag{31}$$

with κ being the width of the Gaussian charge density. Using this system we test and show that for a suitable choice of parameters the results can be made arbitrarily close to the analytical results.

We consider an ordered array of 64 Gaussian charges of alternating signs (32 positively charged (+1) and 32 negatively charged (−1)) arranged to form a NaCl lattice. The potential generated by such a set of charges can be calculated exactly by noting that the electrostatic potential of this charge density at an arbitrary distance $\mathbf{r}$ can be determined analytically, $V_{\mathrm{ext}}(\mathbf{r}) = \mathrm{Erf}(\kappa\mathbf{r})/\mathbf{r}$. We select two neighboring charges (see Figure 2) and calculate the Hartree potential in a smaller orthorhombic cell centered around the two chosen charges. This calculation would have been a necessary step had we treated the two selected centers quantum mechanically instead of with a fixed nuclear charge distribution. The

Table 1: Interaction of a Gaussian Charge Distribution in a 3-Dimensional Lattice as Shown in Figure 2 as a Function of the Number of Gaussians Used in GEEP and as a Function of the QM Cell

QM cell (x,y,z) (Å)	no. of Gaussians	k_{cut} (bohr^{-1})	E_{tot} (Hartree)	ΔE (mHartree)
34.4 34.4 34.4	analytical calculation		3.441010	
34.4 34.4 34.4	6	0.5	3.440520	0.49
34.4 34.4 34.4	6	0.7	3.441176	−0.17
34.4 34.4 34.4	6	1.0	3.441119	−0.11
34.4 34.4 34.4	6	2.0	3.441070	−0.06
34.4 34.4 34.4	6	0.5	3.440520	0.49
34.4 34.4 34.4	9	0.5	3.440687	0.33
34.4 34.4 34.4	12	0.5	3.440885	0.12
34.4 34.4 34.4	15	0.5	3.440895	0.11
34.4 34.4 34.4	15	0.5	3.440895	0.11
27.0 27.0 27.0	15	0.5	3.440978	0.03
34.4 27.0 27.0	15	0.5	3.440951	0.06
22.0 22.0 12.0	15	0.5	3.440865	0.14
12.0 12.0 12.0	15	0.5	3.441356	−0.35
34.4 34.4 34.4	QM/MM nonperiodic[a]		3.443106	2.10

[a] The QM/MM nonperiodic calculation was performed with 64 000 MM atoms arranged in a cube cell of 344.0 Å.

calculation was performed using a plane wave cutoff of 25 Ry, and 3 Gaussians were used for each selected atom to build the model density used to decouple/recouple the periodic images.

In Table 1 we show how this pseudo QM/MM calculation depends on parameters such as the QM cell dimension (affecting the coupling/decoupling between QM periodic images), the k_{cut} parameter of eq 24, and the number of Gaussians used in the GEEP scheme. In particular we note that the number of Gaussians is strictly correlated to the k_{cut} value. In fact, the more Gaussians that are used in the GEEP scheme, the more the R_{low} will be a low cutoff function. This permits a smaller k_{cut} parameter to be used in order to reach the same accuracy (see Table 1).

The choice of the dimension of the QM box is almost irrelevant for the accuracy of the results (see Table 1). In fact even using a box of 12.0 Å, which is the smallest possible box size usable with this QM subsystem, we find accurate results. We remark that other decoupling techniques[23,31] require boxes twice the size of the minimum box, leading to a substantial computational overhead.

Moreover we computed the pseudo QM/MM interaction energy for the nonperiodic pseudo QM/MM calculation, using an MM environment of 64 000 atoms (MM cell side of 344.0 Å). The result shows that for ordered structures surface effects are very important, and the only way to include correctly the electrostatic interactions is by using PBC. Overall this test indicates that the new proposed scheme is both valid and efficient. In terms of computational time no additional overhead was noted when performing pseudo QM/MM calculation with or without PBC.

3.2. SiO_2. We now consider a realistic problem, a crystal of α-SiO_2 (α-quartz) in an orthorhombic cell, subject to periodic boundary conditions. Several QM/MM schemes have been proposed in the literature for silica-based systems,[26,50−56] differing in the description of the quantum-classical interface and of the classical region. All of them treat the QM/MM long-range interaction with a truncation scheme, properly optimizing the charges of the H-atoms terminating the MM cluster or its shape in order to recover the correct long-range effects.

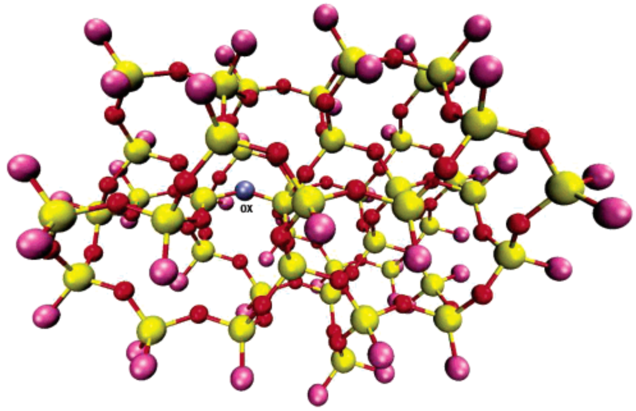


Figure 3. The picture shows the QM cluster. Silicon atoms in yellow, oxygen in red, boundary oxygen atoms (treated increasing the core charge by 0.4) in purple, and in blue the oxygen atom (OX) removed to create the oxygen vacancy defect.

The MM crystal we used for this test is made up of 15 552 atoms (5184 SiO_2 units) in an orthorhombic cell of 49.94, 57.66, and 63.49 Å. The system was optimized using the empirical pair potential of van Beest[57] which is known to provide a reliable description of bulk α-SiO_2.[58] A fragment of 160 atoms was described at the QM level Figure 3, describing the oxygen boundary atoms with a core charge increased by 0.4 in order to maintain the neutrality of the overall system. This boundary scheme has been extensively tested in a recent publication.[26] DFT calculations with Gödecker-Tetter-Hutter (GTH)[59] pseudopotentials[59] using local density approximation to describe the exchange-correlation functional were performed on the QM site using a cutoff of 200 Ry. We optimized the wave function with and without the use of periodic boundary conditions. The results show that the use of periodicity is essential to treat highly ordered crystal structures. Without periodic boundary conditions we find the Kohn−Sham gap to be 0.12 eV which is much lower than the experimental band gap of about 9 eV[60,61] and even than the computed Kohn−Sham gap of 5.8 eV.[42] Also the population analysis gives us an indication that the lack of PBC leads to an incorrect description of the system. In fact by population analysis[32] we find that many oxygen atoms have a positive charge, while some silicon atoms have a negative charge. If we use periodic boundary conditions, on the other hand, we find results that agree with those previously published. In particular, using PBC, we find for the Kohn−Sham band gap a value of 6.23 eV using the same computational parameters as in the case of non-PBC. The population analysis shows the proper charge distribution with charges close to + 2.0 and −1.0 for silicon and oxygen, respectively.

After removing the atom depicted in Figure 3 from the same crystal structure, we studied the charged oxygen vacancy defect in SiO_2 with the same computational setup used for stoichiometric SiO_2. As for quartz the lack of PBC leads to an incorrect description for both the electronic structure and the population analysis. The use of the present scheme gives a Kohn–Sham band gap of 3.18 eV, as against the theoretical result[42] of 3.30 eV. The value obtained without PBC is 0.0089 eV. Unlike the other QM/MM schemes used for silica we do not use any additional charge to terminate the MM cluster, and no particular attention was paid to the choice of its shape. The computational cost for the evaluation of the QM/MM-PBC electrostatic potential on this system accounts for 5% of the total CPU time of a single MD step. In particular 1 MD step (energy and forces), on Cray-XT3 using 32 processors, requires 7 min. Twenty-three seconds out of 7 min are used to evaluate the QM/MM electrostatic potential. Without using the long-range PBC option the computational time used for the construction of the QM/MM electrostatic potential is roughly 12 s. Thus, the use of PBC does not represent a significant overhead in the overall computation scheme, making feasible the study of crystal structures within a QM/MM-PBC framework.

4. GLY-ALA

The results obtained for SiO_2 both with and without PBC could in principle be attributed to the peculiar electronic structure of this material (with a partial ionic/covalent structure). Therefore we further tested both QM/MM schemes, with and without PBC, on a zwitterionic dipeptide (GLY-ALA) in water, where the long-range stabilizing effects due to the Madelung potential are not present. We expect small effects from using PBC with this particular system.

Using the same description for the QM system as the one previously published,[24] we find that PBC do not affect the value of the Kohn–Sham gap or the charge population analysis with respect to the common implementations where a truncation scheme was used. This is to be expected due to the large dielectric constant of water and to the lack of long-range stabilizing effects such as the Madelung potential in an ionic crystal.

5. Conclusions

A new scheme has been designed to include the effects of periodic boundary conditions into hybrid QM/MM descriptions of chemical/biological systems. The present scheme uses the recently proposed Gaussian expansion of the electrostatic potential (GEEP)[24] and is implemented in the CP2K package.[25] Through the use of a modified Ewald lattice summation it is possible to include the effects of the periodic boundary conditions in the evaluation of the MM electrostatic potential. The scheme preserves the linear-scaling property of the GEEP technique and is computationally efficient. The method has no additional overhead with respect to the evaluation of the QM/MM electrostatic potential with a truncation scheme using a spherical cutoff. The new scheme is validated with an analytical model and with three real test cases: the α-quartz crystal and its charged oxygen vacancy defect and a zwitterionic dipeptide (GLY-ALA) in water. It is clear from these tests that the use of periodic boundary conditions together with proper treatment of the long-range interactions is required for ordered systems, unless a careful truncation scheme optimizing the shape and dipole of the MM environment is used. Therefore, it is now possible to perform routinely ab initio molecular dynamics and electronic structure calculations in crystal systems. The scheme has been developed describing the electrons with DFT, but the extension to other quantum chemical schemes (Hartree–Fock and post Hartree–Fock methods) is straightforward.

Acknowledgment. The authors wish to thank F. Zipoli and D. Donadio for useful discussions regarding the treatment of the silica system.

Appendix A. Construction of the Model Charge Density

The model density $\hat{\rho}(\mathbf{r},\mathbf{r}_\alpha)$ can be derived by minimizing the multipole moments and the net charge of the system:

$$\Delta Q_L = |\int d\mathbf{r}\mathbf{r}^l Y_l(\mathbf{r})(\rho(\mathbf{r},\mathbf{r}_\alpha) - \hat{\rho}(\mathbf{r},\mathbf{r}_\alpha))| \tag{32}$$

$$\Delta W = |\int d\mathbf{r}\mathbf{r}^2(\rho(\mathbf{r},\mathbf{r}_\alpha) - \hat{\rho}(\mathbf{r},\mathbf{r}_\alpha))| \tag{33}$$

The parameters of the model density are obtained from a fit to the original charge density, which is biased by a weight function. In the reciprocal space, both requirements eqs 32 and 33 can be translated into expressions that are sensitive only to the intermediate neighborhood of the origin. Thus the fit uses a weighting function of the form

$$w(\mathbf{k}) = 4\pi\frac{(|\mathbf{k}|^2 - |\mathbf{k}_{\mathrm{cut}}|^2)^2}{|\mathbf{k}|^2|\mathbf{k}_{\mathrm{cut}}|^2} \tag{34}$$

for $|\mathbf{k}| < |\mathbf{k}_{\mathrm{cut}}|$ and zero elsewhere. The weight function enhances the importance of the low **k**-vectors while ignoring the high **k**-vectors of the density.

Using the method of Lagrange multipliers, the parameters of the model density q_α are obtained from the extremal condition of

$$\mathcal{L}(q_\alpha,\lambda) = \frac{V}{2}\sum_{\mathbf{k}\neq 0} w(\mathbf{k})|\rho(\mathbf{k}) - \sum_\alpha q_\alpha g_\alpha(\mathbf{k})|^2 - \lambda V[\rho(\mathbf{k}=0) - \sum_\alpha q_\alpha g_\alpha(\mathbf{k}=0)] \tag{35}$$

In matrix form the equation can be written in

$$\mathbf{Aq} + \lambda\mathbf{C} = \mathbf{BCq} = N \tag{36}$$

where the matrix elements of **A**, **C**, and **B** are given by

$$A_{i,j} = V\sum_{\mathbf{k}\neq 0} w(\mathbf{k})[g_i^\dagger(\mathbf{k})g_j(\mathbf{k})] \tag{37}$$

$$C_i = Vg_i(\mathbf{k}=0) = 1 \tag{38}$$

$$B_i = V\sum_{\mathbf{k}\neq 0} w(\mathbf{k})Re[\rho^\dagger(\mathbf{k})g_i(\mathbf{k})] \tag{39}$$

and **q** is the array of parameters of the model charge density. The solution to this linear equation system is given by

$$\mathbf{q} = \mathbf{A}^{-1}\left[\mathbf{B} - \mathbf{C}\frac{\mathbf{CA}^{-1}\mathbf{B} - N}{\mathbf{CA}^{-1}\mathbf{C}}\right] \tag{40}$$

Appendix B. Derivation of the Long-Range QM/MM Potential

The effect of the periodic copies of the MM subsystem is only in the long-range term, and it comes entirely from the residual function $R_{\mathrm{low}}(\mathbf{r},\mathbf{r}_a)$ of eq 20:

$$V_{\mathrm{recip}}^{\mathrm{QM/MM}}(\mathbf{r},\mathbf{r}_a) = \sum_{\mathbf{L}}^{\infty}{}'\sum_a v_a^{\mathrm{recip}} = \sum_{\mathbf{L}}^{\infty}{}'\sum_a R_{\mathrm{low}}(|\mathbf{r} - \mathbf{r}_a + \mathbf{L}|) \tag{41}$$

This summation has the same convergence properties as the Ewald series and can be efficiently computed in the reciprocal space. To derive the expression of this modified Ewald sum, let us assume we know the analytical expression of the density $\sigma(\mathbf{r},\mathbf{r}_a)$ originating from the atomic potential R_{low}. The potential at point $\mathbf{r}_i$ due to the charge distribution $\sigma(\mathbf{r},\mathbf{r}_a)$ is

$$V_{\mathrm{recip}}^{\mathrm{QM/MM}}(\mathbf{r}_i,\mathbf{r}_a) = \int d\mathbf{r}\frac{\sigma(\mathbf{r} + \mathbf{r}_i,\mathbf{r}_a)}{\mathbf{r}} = L^{-3}\int \mathrm{d}\mathbf{r}\sum_{\mathbf{k}}^{k_{\mathrm{cut}}}{}'\frac{\tilde{\sigma}(\mathbf{k})\exp[-\Delta 2\pi\mathbf{k}(\mathbf{r} + \mathbf{r}_i - \mathbf{r}_a)]}{\mathbf{r}} \tag{42}$$

The use of the identity[62]

$$\int \mathrm{d}\mathbf{r}\frac{\exp[-\Delta 2\pi\mathbf{k}(\mathbf{r} + \mathbf{r}_i - \mathbf{r}_a)]}{\mathbf{r}} = \int_0^{\infty} r\mathrm{d}r \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta \mathrm{d}\theta \exp[-\Delta 2\pi|\mathbf{k}||\mathbf{r} + \mathbf{r}_i - \mathbf{r}_a|\cos\theta] \tag{43}$$

$$= \frac{4\pi}{k^2}\cos[2\pi\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_a)] \tag{44}$$

in eq 42 leads to

$$V_{\mathrm{recip}}^{\mathrm{QM/MM}}(\mathbf{r}_i,\mathbf{r}_a) = 4\pi L^{-3}\sum_{\mathbf{k}}^{k_{\mathrm{cut}}}{}'\frac{\tilde{\sigma}(\mathbf{k})}{k^2}\cos[2\pi\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_a)] \tag{45}$$

Using the Maxwell equation $\nabla^2 V = 4\pi\rho$ and its representation in Fourier space, the term in eq 45

$$4\pi\frac{\tilde{\sigma}(\mathbf{k})}{k^2} = \tilde{R}_{\mathrm{low}}(\mathbf{k}) \tag{46}$$

is the Fourier transform of the potential originated by the density of charge $\sigma(\mathbf{r},\mathbf{r}_a)$. Then the previous equation can be written as

$$V_{\mathrm{recip}}^{\mathrm{QM/MM}}(\mathbf{r}_i,\mathbf{r}_a) = L^{-3}\sum_{\mathbf{k}}^{k_{\mathrm{cut}}}{}'\sum_a \tilde{R}_{\mathrm{low}}(\mathbf{k})q_a\cos[2\pi\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_a)] \tag{47}$$

References

(1) Sagui, C.; Darden, A. *Ann. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155−179.

(2) Resat, H.; McCammon, J. A. *J. Chem. Phys.* **1996**, *104*, 7645−7651.

(3) Buono, G. S. D.; Cohen, T. S.; Rossky, P. J. *J. Mol. Liq.* **1994**, *60*, 221−236.

(4) Resat, H. *J. Chem. Phys.* **1999**, *110*, 6887−6889.

(5) Rozanska, X.; Chipot, C. *J. Chem. Phys.* **2000**, *112*, 9691−9694.

(6) Brunsteiner, M.; Boresch, S. *J. Chem. Phys.* **2000**, *112*, 6953−6955.

(7) York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345−8349.

(8) Ewald, P. P. *Ann. Phys.* **1921**, *64*, 253−268.

(9) Toukmaji, A. Y.; Jr., J. A. B. *Comput. Phys. Commun.* **1995**, *95*, 73−92.

(10) Deserno, M.; Holm, C. *J. Chem. Phys.* **1998**, *109*, 7678−7693.

(11) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys* **1995**, *103*, 8577−8593.

(12) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(13) Shan, Y.; Klepeis, J. L.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *J. Chem. Phys.* **2005**, *122*, No. 054101.

(14) Tongraar, A.; Liedl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1998**, *102*, 10340−10347.

(15) Tongraar, A.; Rode, B. M. *J. Phys. Chem. A* **2001**, *105*, 506−510.

(16) Schwenk, C. F.; Loeffer, H. H.; Rode, B. M. *J. Am. Chem. Soc.* **2003**, *126*, 1618−1624.

(17) Chalmet, S.; Ruiz-Lopez, M. F. *J. Chem. Phys.* **2001**, *115*, 5220−5227.

(18) Chalmet, S.; Rinaldi, D.; Ruiz-Lopez, M. F. *Int. J. Quantum Chem.* **2001**, *84*, 559−564.

(19) Bandyopadhyay, P.; Gordon, M. S. *J. Chem. Phys.* **2000**, *113*, 1104−1109.

(20) Gao, J.; Alhambra, C. *J. Chem. Phys.* **1997**, *107*, 1212−1217.

(21) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2−13.

(22) Dehez, F.; Martins-Costa, M. T. C.; Rinaldi, D.; Millot, C. *J. Chem. Phys.* **2005**, *122*, No. 234503.

(23) Yarne, D. A.; Tuckerman, M. E.; Martyna, G. J. *J. Chem. Phys.* **2001**, *115*, 3531−3539.

(24) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2005**, *1*, 1176−1184.

(25) Freely available at the URL http://cp2k.berlios.de, released under GPL license.

(26) Zipoli, F.; Laino, T.; Laio, A.; Bernasconi, M.; Parrinello, M. *J. Chem. Phys.* **2006**, *124*, 154707.

(27) Biswas, P.; Gogonea, V. *J. Chem. Phys.* **2005**, *123*, No. 164114.

(28) *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987.

(29) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103−128.

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

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

(32) Blöchl, P. E. *J. Chem. Phys.* **1995**, *103*, 7422−7428.

(33) Levien, L.; Prewitt, C. T.; Weidner, D. J. *Am. Mineral.* **1980**, *65*, 920−930.

(34) Nelson, C. M.; Weeks, R. A. *J. Am. Ceram. Soc.* **1960**, *43*, 396−399.

(35) Weeks, R. A. *J. Appl. Phys.* **1956**, *27*, 1376−1381.

(36) Weeks, R. A.; Nelson, C. M. *J. Am. Ceram. Soc.* **1960**, *43*, 399−404.

(37) Silsbee, R. H. *J. Appl. Phys.* **1961**, *32*, 1459−1461.

(38) Jani, M. G.; Bossoli, R. B.; Halliburton, L. E. *Phys. Rev. B* **1983**, *27*, 2285−2293.

(39) Warren, W. L.; Poindexter, E. H.; Offenberg, M.; Müller-Warmuth, W. *J. Electrochem. Soc.* **1992**, *139*, 872−880.

(40) Poindexter, E. H.; Warren, W. L. *J. Electrochem. Soc.* **1995**, *142*, 2508−2516.

(41) Hamann, D. R. *Phys. Rev. Lett.* **1996**, *76*, 660−663.

(42) Blöchl, P. E. *Phys. Rev. B* **2000**, *62*, 6158−6179.

(43) Snyder, K. C.; Fowler, W. B. *Phys. Rev. B* **1993**, *48*, 13238−13243.

(44) Allan, D. C.; Teter, M. P. *J. Am. Ceram. Soc.* **1990**, *73*, 3247−3250.

(45) Boero, M.; Pasquarello, A.; Sarnthein, J.; Car, R. *Phys. Rev. Lett.* **1997**, *78*, 887−890.

(46) Pacchioni, G.; Ierano, G. *Phys. Rev. Lett.* **1998**, *81*, 377−380.

(47) Edwards, A. H.; Fowler, W. B. *J. Phys. Chem. Solids* **1985**, *46*, 841−857.

(48) Rudra, J. K.; Fowler, W. B. *Phys. Rev. B* **1987**, *35*, 8223−8230.

(49) Pacchioni, G.; Ferrari, A. M.; Ierano, G. *Faraday Discuss.* **1997**, *107*, 155−172.

(50) Sauer, J.; Sierka, M. *J. Comput. Chem.* **2000**, *21*, 1470−1493.

(51) Sulimov, V.; Sushko, P. A. H.; Edwards, A. S.; Stoneham, A. *Phys. Rev. B* **2002**, *66*, 24108−24114.

(52) Mysovsky, A.; Sushko, P.; S. Mukhopadhyay, A. E.; Shluger, A. *Phys. Rev. B* **2004**, *69*, No. 085202.

(53) Sulimov, V.; Casassa, S.; Pisani, C.; Garapon, J.; Poumellac, B. *Modell. Simul. Mater. Sci. Eng.* **2000**, *8*, 763−773.

(54) Pisani, C.; Busso, M.; Lopez-Gejo, F.; Casassa, S.; Maschio, L. *Theor. Chem. Acc.* **2004**, *111*, 246−254.

(55) Erbetta, D.; Ricci, D.; Pacchioni, G. *J. Chem. Phys.* **2000**, *113*, 10744−10752.

(56) Nasluzov, V.; Ivanova, E.; Shor, A.; Vayssilov, G.; Birkenheuer, U.; Rosch, N. *J. Phys. Chem. B* **2003**, *107*, 2228−2241.

(57) van Beest, B. W. H.; Kramer, G. J.; van Santen, R. A. *Phys. Rev. Lett.* **1990**, *64*, 1955−1958.

(58) Tse, J.; Klug, D. D.; Page, Y. L. *Phys. Rev. B* **1992**, *46*, 5933−5938.

(59) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703−1710.

(60) Phillipp, H. R. *Solid State Commun.* **1966**, *4*, 73−75.

(61) Miyazaki, S.; Nishimura, H.; Fukuda, M.; Ley, L.; Ristein, J. *Appl. Surf. Sci.* **1997**, *113/114*, 585−589.

(62) *The Feynman Lectures on Physics*; Addison-Wesley: 1963; Vol. 1, pp 30−11.

CT6001169

J. Chem. Theory Comput. **2006,** *2,* 1379−1387

JCTC Journal of Chemical Theory and Computation

Active Role of Hydrogen Bonds in Rupe and Meyer−Schuster Rearrangements

Shinichi Yamabe, Noriko Tsuchida, and Shoko Yamazaki*

Department of Chemistry, Nara University of Education, Takabatake-cho, Nara 630-8528, Japan

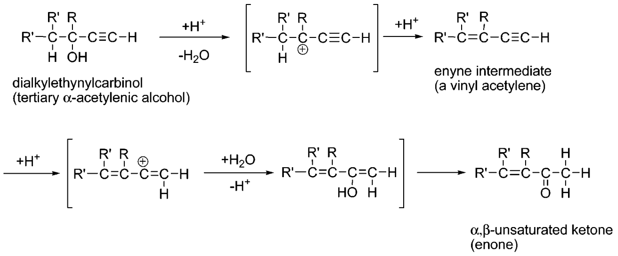
Received April 10, 2006

Abstract: Rupe and Meyer−Schuster rearrangements for the $R_2C(OH)-C\equiv C-H + H_3O^+$ and $(H_2O)_9$ model (R = methyl and phenyl groups) have been investigated by the use of density functional theory calculations. In the substrate $R_2C(OH)-C\equiv CH$ catalyzed by $H_3O^+(H_2O)$, three reaction channels, the two rearrangements and S_N (nucleophilic substitution), were predicted by the frontier molecular orbital theory. The S_N (the OH-group exchange) path was found to have a large activation energy. For 2-methylbut-3-yn-2-ol (R = Me), the Rupe rearrangement has been found to be much more favorable than the Meyer−Schuster rearrangement. For 1,1-diphenylprop-2-yn-1-ol (R = Ph), the occurrence of Meyer−Schuster rearrangement is very likely with the small activation energy. Both rearrangements do not involve the carbonium ion intermediates. However, the calculated geometries of the first transition state are carbonium-ion-like. Dehydration and hydration may occur via the intermolecular proton relay along the hydrogen-bond chains. Minimal models were proposed to represent reaction mechanisms of both rearrangements.

I. Introduction

The Rupe rearrangement is the acid-catalyzed rearrangement of tertiary α-acetylenic alcohols (propargylic alcohols) leading to the formation of α,*β*-unsaturated ketones.[1] The Meyer−Schuster rearrangement is the 1,3-shift isomerization of secondary and tertiary α-acetylenic alcohols to ketones or aldehydes.[2,3] The two rearrangements are shown in Schemes 1 and 2, respectively.

Scheme 1. Mechanism of the Rupe Rearrangement Generally Considered So Far



It is now agreed that the Rupe rearrangement proceeds through a dehydration−hydration sequence with enyne intermediates. These intermediates were sometimes isolated.[4] When a C−H bond adjacent to the C−OH bond is present, the Rupe rearrangement takes place. When a C−H bond adjacent to the C−OH is absent, the Meyer−Schuster rearrangement occurs. In both rearrangements, carbonium ions are thought to be involved. This is because the dehydration in the first step gives the tertiary carbon atom with the acetylene substituent for the enhanced delocalization of the cation charge. The mechanism of the Meyer−Schuster rearrangement of tertiary arylpropargyl alcohols to α,*β*-carbonyl compounds was studied kinetically.[5] Two transition-state models were suggested (Scheme 3).

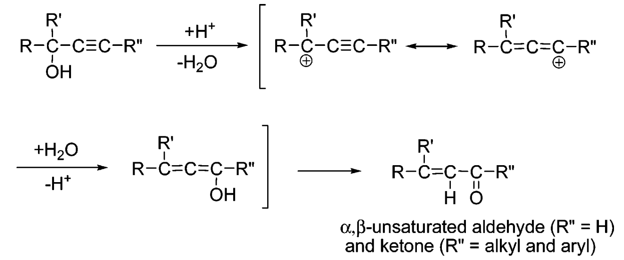
The synthetic utility of the two classic rearrangements is well-established. The Rupe rearrangement was utilized for the synthesis of tricyclic skeletons such as (±)-capnellene[6] and cyathins.[7] The rearrangement was also used for the large-scale synthesis of unsaturated ketones from steroidal propargylic alcohols.[8] The Meyer−Schuster rearrangement was used as key steps for syntheses of a histamine H_3-receptor antagonist[9] and an antifungal mold metabolite.[10] The re-

---

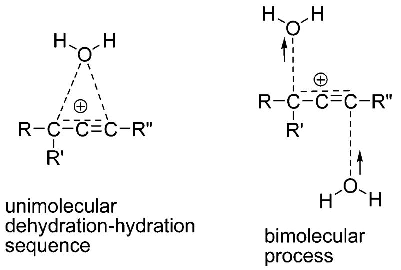
* Author to whom correspondence should be addressed. Fax: +81-742-27-9289. E-mail: yamazaks@nara-edu.ac.jp.

10.1021/ct600132k CCC: $33.50

Scheme 2. Mechanism of the Meyer−Schuster Rearrangement Generally Considered So Far



Scheme 3. Transition-State Models for the Meyer−Schuster Rearrangement



arrangement was also applied to the synthesis of α,β-unsaturated thioesters.[11] Both rearrangements were investigated in water at high temperatures (200−290 °C), and it was shown that an elevated temperature without acids is effective.[12] Furthermore, recently, various metal-catalyzed Rupe and Meyer−Schuster rearrangements were reported.[13] In addition, to suppress Rupe and Meyer−Schuster rearrangements is an important task for some C−C bond formation reactions involving propargyl alcohols.[14]

On the other hand, their precise mechanisms are still open to question. Tapia et al. carried out computational studies of the Meyer−Schuster rearrangement.[15] However, their models were too small or inappropriate to describe the dehydration−hydration sequence. For instance, a propargylic alcohol substrate, $Me_2C(OH)-C\equiv CH$, was employed in a combined quantum-chemical/Monte Carlo study.[15a] However, the substrate with the C−H bond should undergo not the Meyer−Schuster but the Rupe rearrangement. Except for those of Tapia et al., there have been no computational studies. To elucidate the entire process and the driving force of both rearrangements, we conducted a computational study on the reacting system, $R_2C(OH)-C\equiv C-H + H_3O^+ + (H_2O)_9$ (R = Me and Ph). Particular interest lies both in the role of hydrogen bonds to control the (de)hydration and in the intermediacy of the carbonium ions, exhibited in brackets in Schemes 1 and 2. The reacting systems of both rearrangements are composed of only C, H, and O atoms, which are fundamental elements of organic compounds. Elucidation of their mechanisms is expected to give deep insight into other acid-catalyzed solvolytic reactions.

II. Method of Calculations

The geometries of the reacting system starting from $R_2C(OH)-C\equiv C-H + H_3O^+(H_2O)_9$ were determined by density functional theory calculations. The B3LYP method[16] was used. B3LYP seems to be a suitable method, because it

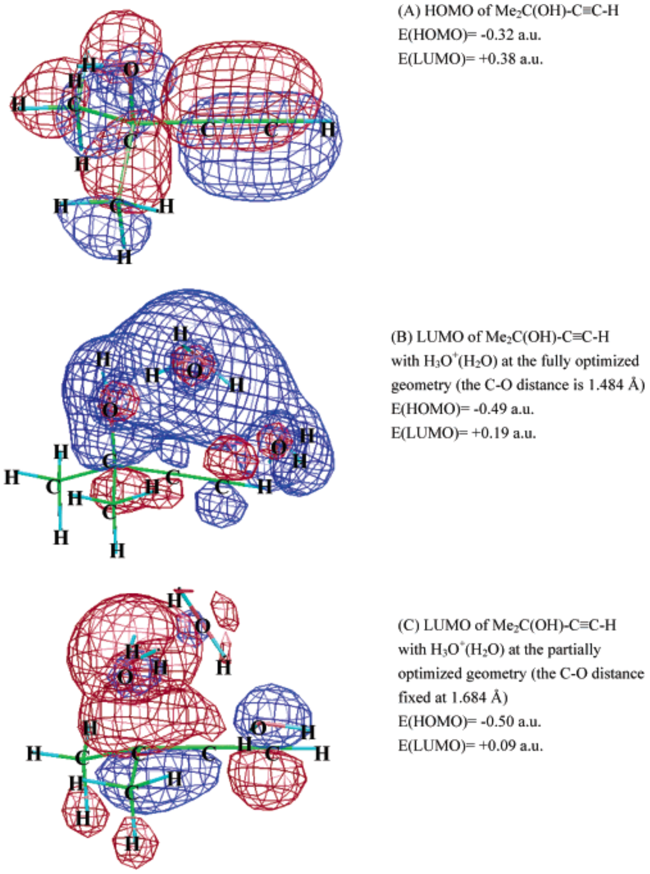


Figure 1. Shapes of frontier orbitals; HOMO and LUMO of the $Me_2C(OH)-C\equiv C-H$ substrate without (A) and with the $H_3O^+(H_2O)$ catalyst (B) and (C). *E*(HOMO) and *E*(LUMO) are orbital energies computed by RHF/STO-3G.

includes the electron correlation effect to some extent. The basis set employed is 6-31G*. Then, the geometry optimizations were carried out by RB3LYP/ 6-31G*.

Transition states (TSs) were characterized by vibrational analysis, which checked whether the obtained geometries have single imaginary frequencies ($\nu^\ddagger$). From TSs, reaction paths were traced by the intrinsic reaction coordinate method[17] to obtain the energy-minimum geometries. Relative electronic (and Gibbs free) energies were refined by single-point calculations of RB3LYP/6-311+G(d,p) [self-consistent reaction field (SCRF) = dipole, solvent = water][18] on the RB3LYP/6-31G* geometries and its zero-point vibrational [and thermal correction (T = 300 K, P = 1 atm)] energies (ZPEs).

All of the calculations were carried out using the Gaussian 03[19] program package installed on a Compaq ES 40 at the Information Processing Center (Nara University of Education).

III. Results and Discussions

Frontier Molecular Orbital Analysis of 2-Methylbut-3-yn-2-ol, $Me_2C(OH)-C\equiv C-H$. The substrate, $Me_2C(OH)-C\equiv C-H$, has two nucleophilic sites, the O−H oxygen lone pair and the triple bond. Figure 1A shows its highest occupied molecular orbital (HOMO) shape. An acid catalyst, $H_3O^+(H_2O)$, may be bound to those sites. That is, the HOMO of the substrate may overlap effectively with the lowest vacant

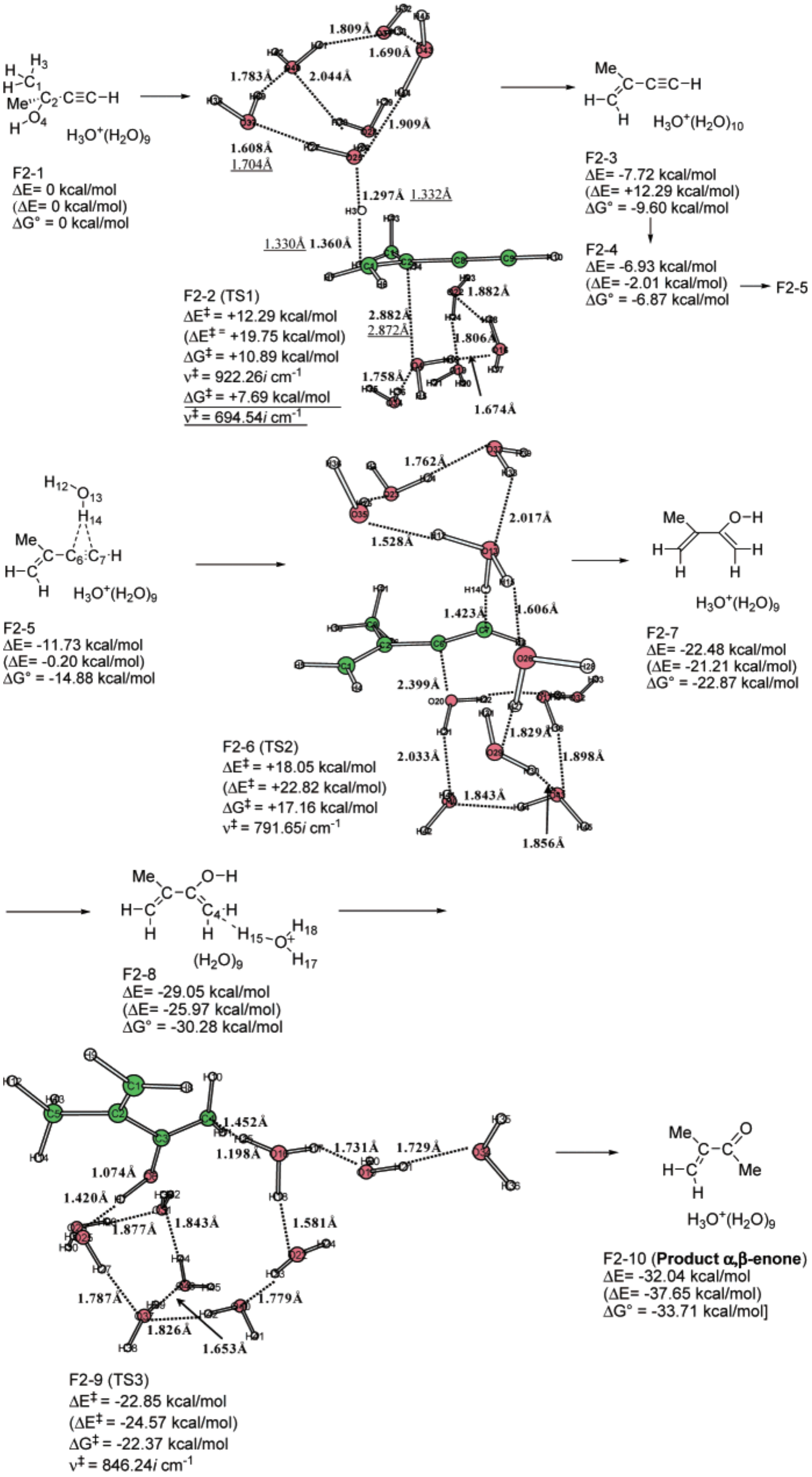
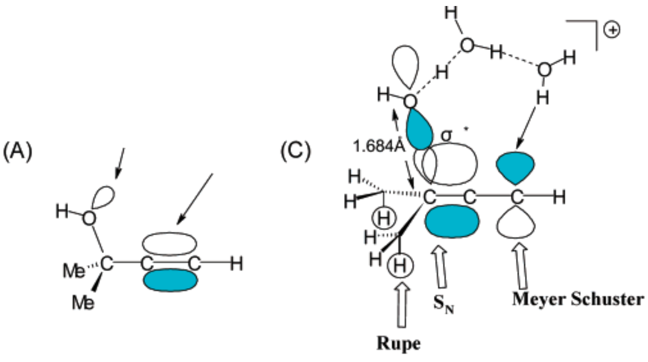


Figure 2. Reaction paths of the Rupe rearrangement by the model in Scheme 5. TS geometries are shown. The reactant, intermediates, and product geometries and reaction-coordinate vectors corresponding to the respective sole imaginary frequencies for TSs are shown in Figure S2 in the Supporting Information. Relative energies (Δ*E*'s) are by RBLYP/6-31G* ZPEs and RB3LYP/6-311+G(d,p) SCRF = dipole electronic energies (without parentheses) and RB3LYP/6-31G* ZPEs and electronic energies (with parentheses). The Gibbs free energy changes are composed of the RB3LYP/6-311+G(d,p) SCRF = dipole electronic energies and RB3LYP/6-31G* thermal corrections to the free energies. The underlined numbers for F-2 (TS1) were obtained by a model excluding three water molecules [7], [8], and [9] in Scheme 5, i.e., $Me_2C(OH)-C\equiv C-H + H_3O^+(H_2O)_6$. F2-1 and F2-2 of the model are shown in Figure S6 (Supporting Information).

Scheme 4. (A) Electrophile $H_3O^+(H_2O)$ Coordinated to the HOMO of the Substrate along the Direction of the Two Arrows and (C) Slight C−O Elongation (1.484 Å in B → 1.684 Å in C) in the Acid-Catalyzed Substrate Providing the LUMO Shape for the Three Reaction Channels



HOMO of (A) in Figure 1 **LUMO of (C) in Figure 1**

MO of $H_3O^+(H_2O)$, leading to the geometry of Figure 1B. By the coordination, the HOMO and lowest unoccupied molecular orbital (LUMO) energies are lowered considerably, *E*(HOMO) = −0.32 → −0.49 au and *E*(LUMO) = +0.38 → +0.19 au. However, LUMO(B) of the complex $Me_2C(OH)-C\equiv CH(H_3O)^+H_2O$ is localized at the catalyst part, $-OH\cdots H_3O^+\cdots OH_2$, and does not apparently show the reactivity of the substrate toward the nucleophilic attack by other water molecules. Some distortion from the equilibrium structure (B) would cause the reactivity. To examine the distortion effect, we made a partial geometry optimization of the $H_3O^+(H_2O)$ complexed substrate with the C−O length elongated (1.684 Å) and fixed. At that geometry, LUMO(C) is depicted and is shown also in Figure 1C. It is noteworthy that, while *E*[HOMO(C)] = −0.50 au is almost the same as *E*[HOMO(B)] = −0.49 au, *E*[LUMO(C)] = +0.09 au is further lowered relative to *E*[LUMO(B)] = +0.19 au. Besides the enhanced electrophilic character, LUMO(C) has large spatial extensions on three sites of the substrate (Scheme 4).

A nucleophile may attack the sites, leading to three different reaction channels. When a methyl C−H bond is attacked, the Rupe rearrangement is initiated. When the backside of the C−O bond is attacked, a solvolytic S_N (nucleophilic substitution) type displacement occurs. When the acetylenic terminal carbon is attacked, the Meyer−Schuster rearrangement is brought about. According to the

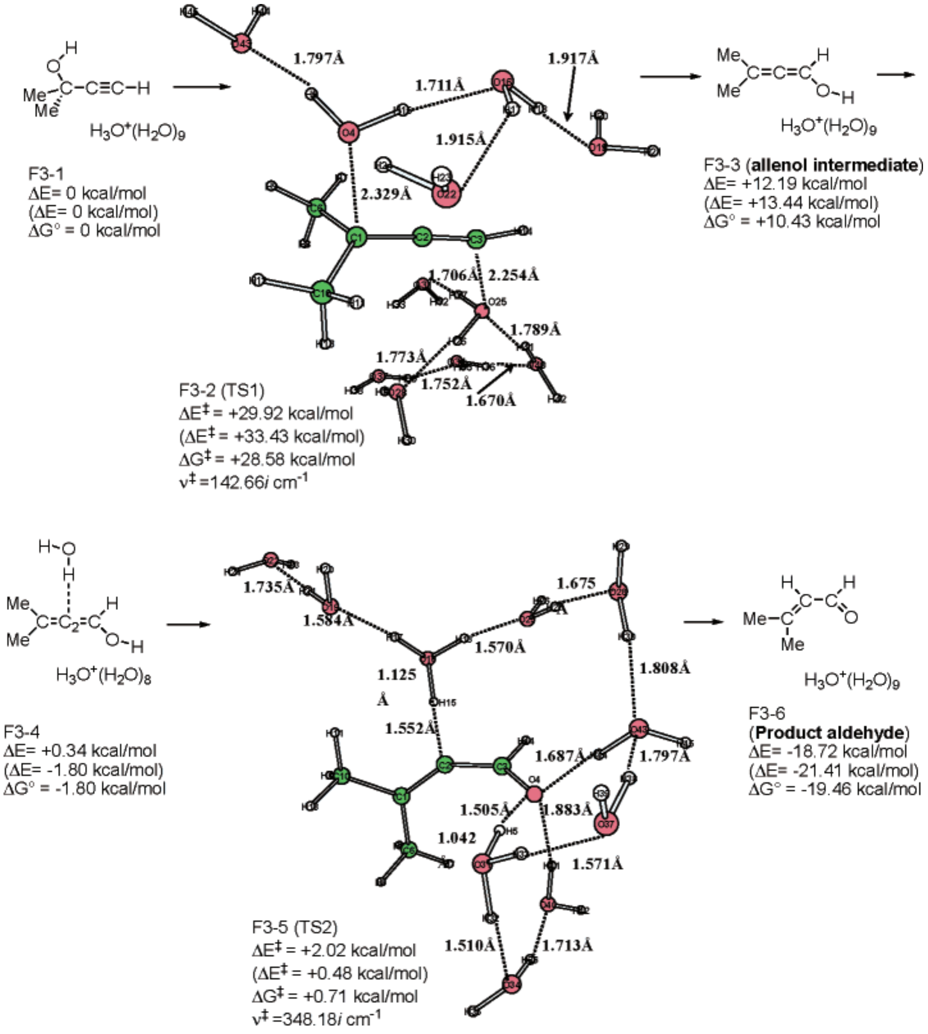
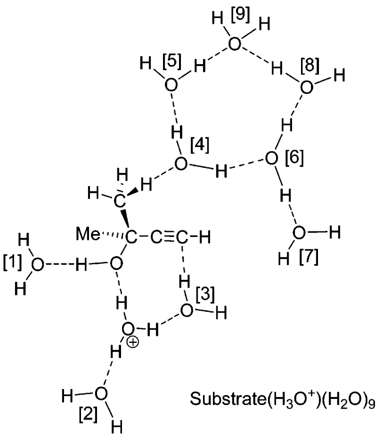


Figure 3. Reaction paths of the Meyer−Schuster rearrangement. TS geometries are shown. The reactant, intermediate, and product geometries and reaction-coordinate vectors corresponding to the respective sole imaginary frequencies for TSs are shown in Figure S3 in the Supporting Information.

Scheme 5. Model of the Rupe Rearrangement to Trace the Reaction Path[a]



[a] To the nine water molecules the numbers [1], [2], ... [9] are attached for explanation.

LUMO(C) shape, the S_N route needs to also be examined. The S_N product is the same as the substrate with the stereochemical scrambling.

Figure 1 should be reviewed. The acid-catalyzed substrate itself appears not to have a clear reactivity of those reaction channels. But, when the geometry is distorted along the 28th normal-mode vibration, the reactivity is successfully presented (Figure S1 in the Supporting Information). In this regard, the driving force of the two rearrangements is the C−O bond elongation.

Reaction Paths of the Rupe and Meyer−Schuster Rearrangements of $\mathbf{Me_2C(OH)-C{\equiv}C-H}$. To investigate the role of the proton relay along hydrogen bonds on the Rupe rearrangement, a model, $\mathrm{Me_2C(OH)-C{\equiv}C-H + H_3O^+ + (H_2O)_9}$, was made and is shown in Scheme 5. In the scheme, the first water molecule [1] is linked with the hydroxyl group. [2] and [3] are needed to cover the hydronium ion. Also, [3] is for protonation to the terminal acetylene carbon. Water molecule [4] is the nucleophile (base) to remove the methyl proton. The resultant H_3O^+ must be surrounded by water molecules, which are [5], [6], [8], and [9]. Molecule [7] is added to the model as a potential linkage between H_3O^+[1][2][3] and [4][5][6][8][9].

If more water molecules were added to the model, the calculated results would give more reliable results. However, the model expansion is extremely difficult, because we need to treat the $\mathrm{Ph_2C(OH)-C{\equiv}C-H}$ substrate for the Meyer−Schuster rearrangement with the same solvation pattern. Figure 2 exhibits the reaction path of $\mathrm{Me_2C(OH)-C{\equiv}C-H + H_3O^+ + (H_2O)_9 \rightarrow H_2C{=}C(Me)-CO-Me + H_3O^+ + (H_2O)_9}$. The geometry of F2-1 is shown in Figure S2 in the Supporting Information and corresponds to that in Scheme 5. The O(25) atom in the $(H_2O)_6$ cluster is a nucleophile (base) to cause the trans elimination (TS1, F2-2).[20] The trans orientation of the E2 reaction is confirmed by the dihedral angle, ∠H3−C1−C2−O4 = −179.04°. At TS1, the C1−H3 and C2−O4 bonds are cleaved at the same time. That is, the carbonium ion does not intervene. The nucleophilicity (basicity) of the $(H_2O)_6$ cluster is strong enough to preclude the carbonium ion and to give a small activation energy, $\Delta E^{\ddagger} = +12.29$ kcal/mol, in comparison with the average C−H bond energy, 98 kcal/mol. After TS1, the enyne intermediate mediated by the $H_3O^+(H_2O)_{10}$ is formed and is shown in F2-3 (Figure S2 of the Supporting Information). Through the proton transfer along the hydrogen bond, the cation center (i.e., H_3O^+) is moved readily and the isomers of the hydrogen-bond network around the substrate are generated (F2-3, F2-4, and F2-5 shown in Figure S2 of the Supporting Information). In the isomer F2-5, a water molecule (H14O13H12) is bound weakly to the C6≡C7 acetylenic

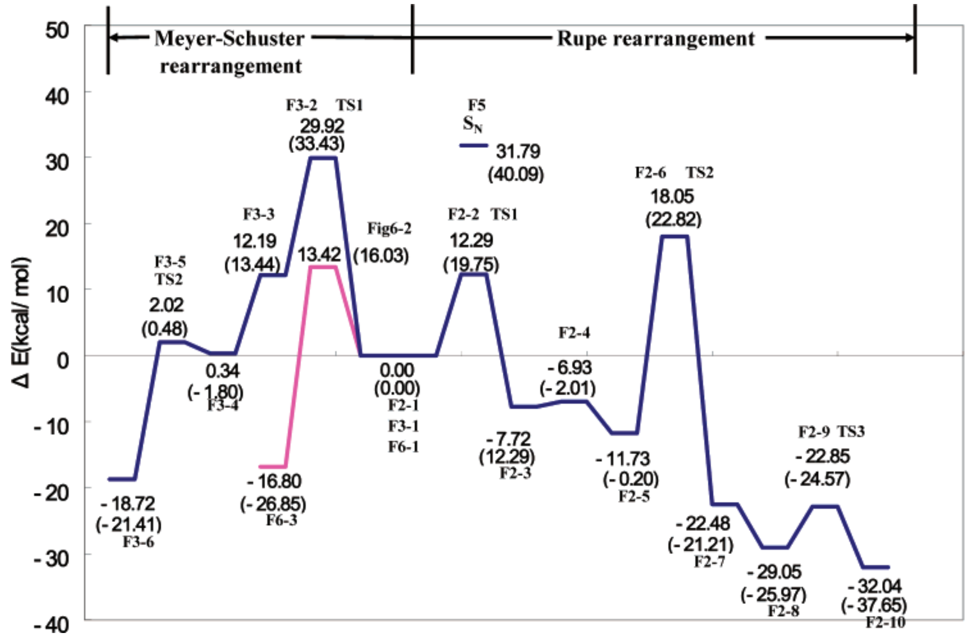
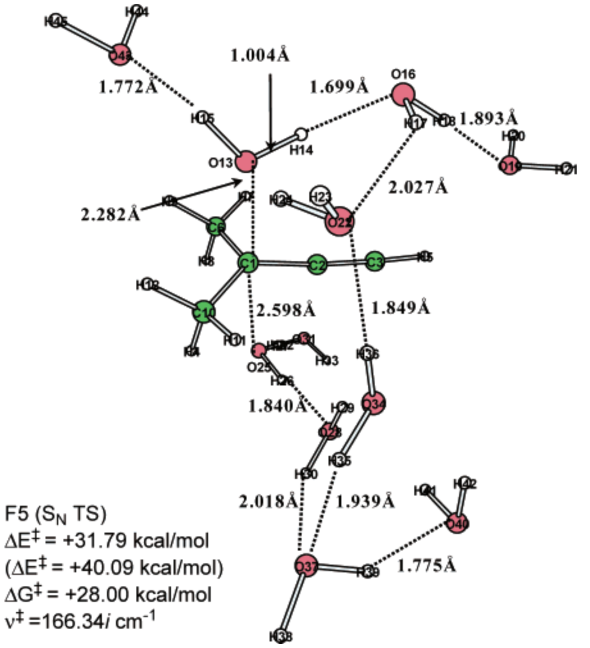


Figure 4. Energy changes for Rupe and Meyer−Schuster rearrangements. Values without parentheses were obtained by the sum of the RB3LYP/6-311+G(d,p) SCRF electronic energy and RB3LYP/6-31G(d) ZPEs. Those with parentheses were obtained by the sum of RB3LYP/6-31G(d) electronic energies and ZPEs. The experimental activation energy of the Meyer−Schuster rearrangement of $\mathrm{Ph_2C(OH)-C{\equiv}C-H}$ is 18.5 kcal/mol.[5]



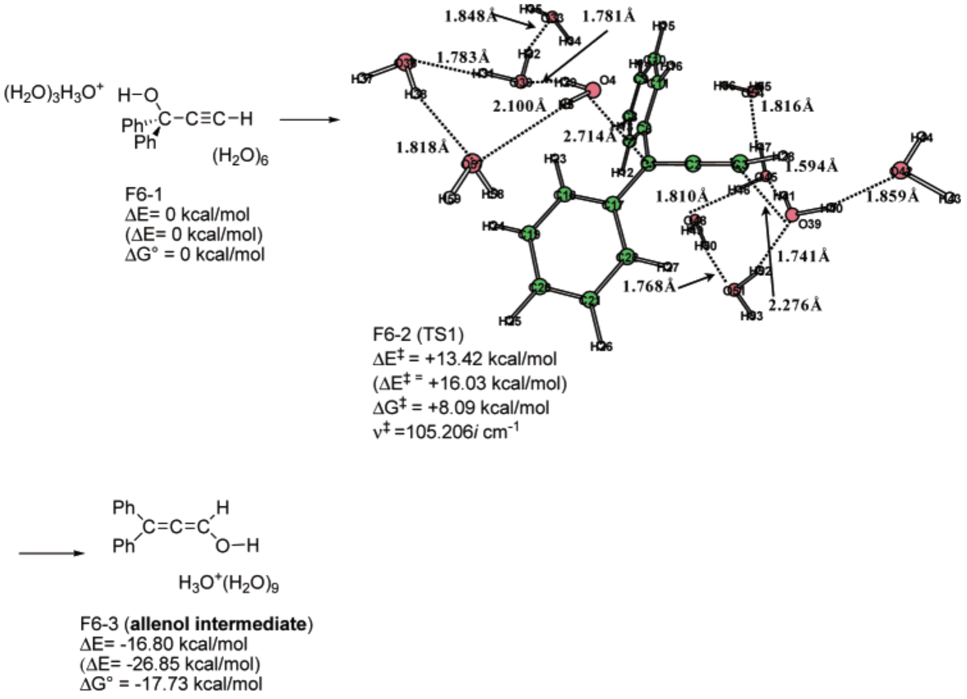
F5 (S_N TS)
$\Delta E^{\ddagger}$ = +31.79 kcal/mol
($\Delta E^{\ddagger}$ = +40.09 kcal/mol)
$\Delta G^{\ddagger}$ = +28.00 kcal/mol
$\nu^{\ddagger}$ =166.34*i* cm^{-1}

Figure 5. TS geometry of the nucleophilic displacement reaction (S_N) at the C1 center. Reaction-coordinate vectors corresponding to the sole imaginary frequency are shown in Figure S4 in the Supporting Information.

bond (2.344 and 2.211 Å). When a proton (H14) approaches the terminal carbon atom, TS2 (F2-6) is obtained. The C7–H14 and C6–O20 bonds are formed synchronously, leading to a dienal, $H_2C{=}CMe{-}C(OH){=}CH_2$, mediated by H_3O^+-$(H_2O)_9$ (F2-7 in Figure S2 of the Supporting Information). Its isomer is shown in F2-8 (Figure S2), where the hydronium ion H15H18H17(O16)$^+$ is in contact with the terminal vinyl carbon, C4 (1.871 Å). The protonation H15 → C4 and deprotonation O6 → H7 occur simultaneously at TS3 (F2-9). After TS3, the product α,*β*-enone solvated by H_3O^+-$(H_2O)_9$ (F2-10 in Figure S2 of the Supporting Information) is formed.

Thus, the proton attachment and detachment at the reaction center occur readily in the combination of the dimethylpropargyl alcohol, H_3O^+ and $(H_2O)_9$ cluster. In Scheme 3, the two transition-state models have been shown. The results in Figure 2 basically support the bimolecular dehydration–hydration sequence in the criterion of the hydrogen-bond directionality. There are three transition states: TS1 is the elimination, TS2 is the addition, and TS3 is the isomerization.

Experimentally, the substrate undergoes not the Meyer–Schuster but the Rupe rearrangement. However, the former rearrangement is examined here both to seek its elementary processes and to check whether it is unfavorable relative to the Rupe rearrangement in the present model. Figure 3 exhibits the reaction paths of the Meyer–Schuster rearrangement. The hydrogen-bond chain of the precursor (F3-1 shown in Figure S3, Supporting Information) is slightly different from but practically similar to that in F2-1. At TS1 (F3-2), C1–O4 bond scission and C3–O25 bond formation occur at the same time. Again, no carbonium-ion intermediate is formed. After TS1, an allenol [$Me_2C{=}C{=}CH(OH)$] mediated by $H_3O^+(H_2O)_9$ is obtained, which is shown in F3-3 (Figure S3 of the Supporting Information). Next, the hydrogen-bond chain is reorganized (F3-4 in Figure S3) so that the protonation to the central carbon (C2) of the allenol intermediate is ready. At TS2 (F3-5), protonation (H15 → C2) and deprotonation (O4 → H5) occur simultaneously. After TS2, the product, the α,*β*-unsaturated aldehyde ($Me_2C{=}CH{-}CHO$) mediated by $H_3O^+(H_2O)_9$, is generated (F3-6 in Figure S3). The dehydration–hydration sequence has been



F6-1
ΔE= 0 kcal/mol
(ΔE= 0 kcal/mol)
$\Delta G°$ = 0 kcal/mol

F6-2 (TS1)
$\Delta E^{\ddagger}$ = +13.42 kcal/mol
($\Delta E^{\ddagger}$ = +16.03 kcal/mol)
$\Delta G^{\ddagger}$ = +8.09 kcal/mol
$\nu^{\ddagger}$ =105.206*i* cm^{-1}

F6-3 (**allenol intermediate**)
ΔE= -16.80 kcal/mol
(ΔE= -26.85 kcal/mol)
$\Delta G°$ = -17.73 kcal/mol

Figure 6. First and rate-determining steps of the Meyer–Schuster rearrangement for the system of $Ph_2C(OH){-}C{\equiv}C{-}H$ + H_3O^+ + $(H_2O)_9$. TS geometry is shown. The reactant and intermediate geometries and reaction-coordinate vectors corresponding to the sole imaginary frequencies for TS1 are shown in Figure S5 in the Supporting Information.

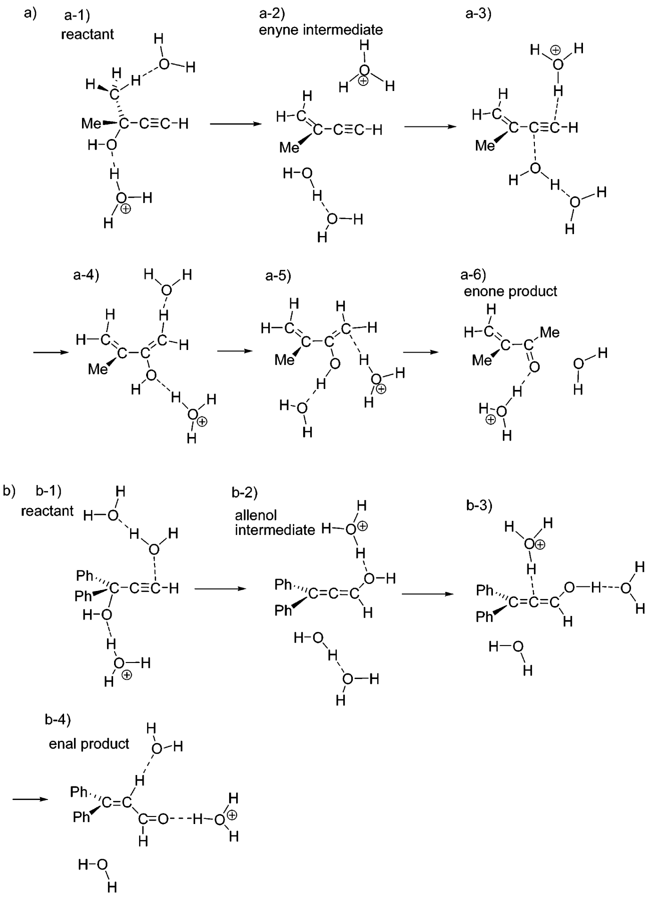
attained by the proton relay through the hydrogen-bond chain, that is, the bimolecular model in Scheme 3.

Figure 4 shows the energy changes (blue color) along both rearrangements. In the Rupe rearrangement, the rate-determining step is TS2, which is consistent with the experimental evidence that the enyne intermediate could be isolated.[4] If TS1 were the rate-determining step, the isolation would be infeasible. On the other hand, the rate-determining step of the Meyer−Schuster rearrangement is TS1. Its activation energy is significantly larger than that of TS2 (Rupe). Therefore, it is found that the reaction channel of the Meyer−Schuster rearrangement is present for the $Me_2C(OH){-}C{\equiv}C{-}H$ substrate but has a large activation energy and is unlikely.

According to the LUMO(C) extension in Figure 1 and Scheme 4, the S_N path was also investigated. Its TS structure (F5) is shown in Figure 5. It is a carbonium-ion-like structure, which is similar to the TS1s of both rearrangements (with large C−O distances). The TS (S_N) energy is also shown in Figure 4 and is found to be larger even than that of TS1 of the Meyer−Schuster rearrangement. Consequently, although the S_N path is present as predicted in Scheme 4, it has an significantly large activation energy and is unlikely. The stereochemistry of the substrate should be conserved.

Reaction Paths of the Meyer−Schuster Rearrangement of $Ph_2C(OH){-}C{\equiv}C{-}H$. The energy diagram in Figure 4 has demonstrated that TS1 is obviously the rate-determining step in the Meyer−Schuster rearrangement. For 1,1-diphenylprop-2-yn-1-ol (R = Ph), accordingly, only the first step was investigated and is shown in Figure 6. The precursor geometry of F6-1 in Figure S5 (Supporting Information) corresponds to that of F3-1 for 2-methylbut-3-yn-2-ol (R = Me). The two hydrogen-bonded water clusters are separated in F6-1, which is in contrast to the linked water clusters in F3-1. The π electron densities on two phenyl groups block the linkage. F6-2 shows TS1 where the C1···O4 bond is to be cleaved (dehydration) and the C3···O39 bond is to be formed (hydration). Noteworthy are their distances, C1···O4 = 2.714 Å and C3···O39 = 2.276 Å. While the latter is close to the corresponding distance, C3···O25 = 2.254 Å (TS1 in F3-2), the former is appreciably larger than C1···O4 = 2.329 Å (TS1 in F3-2). Owing to the extension of the delocalization of the cation charge to the two phenyl rings, TS1 in F6-2 is more carbonium-ion-like than that in F3-2. After TS1, a diphenyl allenol intermediate solvated by $H_3O^+(H_2O)_9$ is formed (F6-3 in Figure S5 of the Supporting Information). In Figure 4, the energy change for the rearrangement of the diphenyl substrate is shown. The calculated activation energy, 13.42 (16.03) kcal/mol, is somewhat smaller than but comparable to the experimental value, 18.5 kcal/mol.[5] The substitution of the methyl group to the phenyl group in the $R_2C(OH){-}C{\equiv}C{-}H$ substrate lowers the activation energy remarkably, 29.92 (33.43) kcal/mol for R = Me → 13.42 (16.03) for R = Ph. Thus, the condition for the Meyer−Schuster rearrangement is not only the absence of the Rupe-shift C−H bond but also the substituent for the cation charge delocalization.

Scheme 6. Minimal Models to Represent the Reaction Mechanisms[a]



[a] For the Rupe rearrangement, (a) $Me_2C(OH){-}C{\equiv}CH + H_3O^+(H_2O)_1$. For the Meyer−Schuster rearrangement, (b) $Ph_2C(OH){-}C{\equiv}CH + H_3O^+(H_2O)_2$.

IV. Concluding Remarks

In this work, Rupe and Meyer−Schuster rearrangements for the $R_2C(OH){-}C{\equiv}C{-}H + H_3O^+$ and $(H_2O)_9$ model have been investigated by the use of RB3LYP calculations. In the substrate $R_2C(OH){-}C{\equiv}CH$ catalyzed by $H_3O^+(H_2O)$, the electrophilic reactivities of the three reaction channels, the two rearrangements and S_N, have been unveiled by the slight C−O bond elongation. The S_N path is unfavorable because the LUMO(C) in Scheme 4 is bonding along the C−C bond (a vague reaction center at the backside of the C−O bond). For the substrate (R = Me), the Rupe rearrangement has been found to be much more favorable than the Meyer−Schuster rearrangement. For the substrate (R = Ph), this shift is very likely with the small activation energy. Both rearrangements do not involve the carbonium-ion intermediates. However, the calculated geometries of TS1 are carbonium-ion-like. Dehydration and hydration may occur via the intermolecular proton relay along the hydrogen-bond chains.

In view of the present calculated results, the reaction mechanism of the two rearrangements may be described by the use of $R_2C(OH){-}C{\equiv}C{-}H$ and $H_3O^+(H_2O)_n$ ($n = 1$ in Scheme 6a and $n = 2$ in Scheme 6b).

In a-1, the substrate is coordinated by H_3O^+ and H_2O. An E2 elimination leads to the enyne intermediate in a-2. The acetylenic terminal carbon is coordinated by H_3O^+ and the other carbon by $(H_2O)_2$ in a-3. The formation of C−H and C−O bonds gives the enol intermediate in a-4. Its isomerization leads to a-5 and results in the enone product of a-6.

In b-1, the substrate is coordinated by H_3O^+ and $(H_2O)_2$. The C−O bond cleavage and formation leads to an allenol intermediate in b-2. The allenic center carbon atom and the hydroxy proton of the intermediate are coordinated by H_3O^+ and H_2O, respectively, in b-3. Protonation and deprotonation give the enal product in b-4. Thus, the Rupe and Meyer−Schuster rearrangements are composed of protonation, deprotonation, C−O bond cleavage, and its formation. The $H_3O^+(H_2O)_n$, $n = 1$ for a or $n = 2$ for b, species may describe the bond interchanges in the minimal sizes. Bond interchanges along hydrogen bonds work effectively to cause both rearrangements.

Supporting Information Available: The 28th normal-mode vibration of $Me_2C(OH)C{\equiv}CH$ with $H_3O^+(H_2O)$ (Figure S1) and reactant, intermediate, and product geometries and reaction-coordinate vectors corresponding to respective sole imaginary frequencies for TSs (Figures S2−S6). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Rupe, H.; Glenz, K. *Liebigs Ann. Chem*. **1924**, *436*, 184−204. (b) Rupe, H.; Kambli, E. *Helv. Chim. Acta* **1926**, *9*, 672. (c) Rupe, H.; Kambli, E. *Liebigs Ann. Chem*. **1927**, *459*, 195−217. (d) Rupe, H.; Giesler, L. *Helv. Chim. Acta* **1928**, *11*, 656−669. (e) Rupe, H.; Messner, W.; Kambli, E. *Helv. Chim. Acta* **1928**, *11*, 449−462. (f) Rupe, H.; Wirz, A.; Lotter, P. *Helv. Chim. Acta* **1928**, *11*, 965−971. (g) Rupe, H.; Gassmann, A. *Helv. Chim. Acta* **1929**, *12*, 193−204. (h) Rupe, H.; Hirschmann, H. *Helv. Chim. Acta* **1931**, *14*, 687−701. (i) Rupe, H.; Kuenzy, F. *Helv. Chim. Acta* **1931**, *14*, 701−708. (j) Rupe, H.; Kuenzy, F. *Helv. Chim. Acta* **1931**, *14*, 708−718. (k) Rupe, H.; Haecker, R.; Kambli, E.; Wassieleff, N. *Helv. Chim. Acta* **1933**, *16*, 685−700. (l) Rupe, H.; Gassmann, A. *Helv. Chim. Acta* **1934**, *17*, 283−285. (m) Rupe, H.; Werdenberg, H. *Helv. Chim. Acta* **1935**, *18*, 542−546.

(2) Meyer, K. H.; Schuster, K. *Chem. Ber*. **1922**, *55*, 819−823.

(3) Swaminathan, S.; Narayanan, K. V. *Chem. Rev*. **1971**, *71*, 429−438.

(4) (a) Hurd, C. D.; Jones, R. N. *J. Am. Chem. Soc*. **1934**, *56*, 1924−1926. (b) Ansell, M. F.; Hancock, J. W.; Hickinbottom, W. J. *J. Chem. Soc*. **1956**, 911−917. (c) Coles, L. E.; Linnell, W. H.; Mathieson, D. W.; Shoukri, A. S. *J. Chem. Soc*. **1954**, 2617−2622. (d) Hamlet, J. C.; Henbest, H. B.; Jones, E. R. H. *J. Chem. Soc.* **1951**, 2652−2659. (e) Inhoffen, H. H.; Longemann, W.; Hohlweg, W.; Serini, A. *Chem. Ber*. **1938**, *71*, 1024−1032. (f) Smissmann, E. E.; Johnson, R. H.; Carlson, A. W.; Aycock, B. F. *J. Am. Chem. Soc*. **1956**, *78*, 3395−3400.

(5) Edens, M.; Boerner, D.; Chase, C. R.; Nase, D.; Schiavelli, M. D. *J. Org. Chem*. **1977**, *42*, 3403−3408.

(6) Stevens, K. E.; Paquette, L. A. *Tetrahedron Lett*. **1981**, *22*, 4393−4396.

(7) Takeda, K.; Nakane, D.; Takeda, M. *Org. Lett*. **2000**, *2*, 1903−1905.

(8) Weinmann, H.; Harre, M.; Neh, H.; Nickisch, K.; Skötsch, C.; Tilstam, U. *Org. Process Res. Dev*. **2002**, *6*, 216−219.

(9) Stark, H.; Sadek, B.; Krause, M.; Hüls, A.; Ligneau, X.; Ganellin, C. R.; Arrang, J.-M.; Schwartz, J.-C.; Schunack. W. *J. Med. Chem*. **2000**, *43*, 3987−3994.

(10) Welch, S. C.; Hagan, C. P.; White, D. H.; Fleming, W. P.; Trotter, J. W. *J. Am. Chem. Soc*. **1977**, *99*, 549−556.

(11) Yoshimatsu, M.; Naito, M.; Kawahigashi, M.; Shimizu, H.; Kataoka, T. *J. Org. Chem*. **1995**, *60*, 4798−4802.

(12) An, J.; Bagnell, L.; Cablewski, T.; Strauss, C. R.; Trainor, R. W. *J. Org. Chem*. **1997**, *62*, 2505−2511.

(13) (a) Imagawa, H.; Asai, Y.; Takano, H.; Hamagaki, H.; Nishizawa, M. *Org. Lett*. **2006**, *8*, 447−450. (b) Narasaka, K.; Kusama, H.; Hayashi, Y. *Chem. Lett.* **1991**, 1413−1416. (c) Narasaka, K.; Kusama, H.; Hayashi, Y. *Tetrahedron* **1992**, *48*, 2059−2068. (d) Lorber, C. Y.; Osborn, J. A. *Tetrahedron Lett.* **1996**, *37*, 853−856. (e) Chabardes, P. *Tetrahedron Lett.* **1988**, *29*, 6253−6256. (f) Suzuki, T.; Tokunaga, M.; Wakatsuki, Y. *Tetrahedron Lett.* **2002**, *43*, 7531−7533. (g) Mercier, C.; Chabardes, P. *Pure Appl. Chem.* **1994**, *66*, 1509−1518. (i) Cadierno, V.; Diez, J.; Garcia-Garrido, S. E.; Gimeno, J. *Chem. Commun.* **2004**, 2716−2717.

(14) (a) Bigi, F.; Carloni, S.; Maggi, R.; Muchetti, C.; Sartori, G. *J. Org. Chem*. **1997**, *62*, 7024−7027. (b) Luzung, M. R.; Tosle, F. D. *J. Am. Chem. Soc*. **2003**, *125*, 15760−15761. (c) Ishikawa, T.; Manabe, S.; Aikawa, T.; Kudo, T.; Saito, S. *Org. Lett*. **2004**, *6*, 2361−2364. (d) Zhao, W.; Carreira, E. M. *Org. Lett*. **2003**, *5*, 4153−4154.

(15) (a) Tapia, O.; Lluch, J. M.; Cardenas, R.; Andres, J. *J. Am. Chem. Soc*. **1989**, *111*, 829−835. (b) Andres, J.; Cardenas, R.; Silla, E.; Tapia, O. *J. Am. Chem. Soc*. **1988**, *110*, 666−674. (c) Andres, J.; Silla, E.; Tapia, O. *J. Mol. Struct.* **1986**, *138*, 171−177. (d) Tapia, O.; Andres, J. *Chem. Phys. Lett.* **1984**, *109*, 471−477. (e) Andres, J.; Silla, E.; Tapia, O. *Chem. Phys. Lett*. **1983**, *94*, 193−197. (f) Andres, J.; Silla, E.; Tapia, O. *J. Mol. Struct.* **1983**, *105*, 307−314.

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

(17) (a) Fukui, K. *J. Phys. Chem*. **1970**, *74*, 4161−4163. (b) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem*. **1990**, *94*, 5523−5527.

(18) (a) Onsager, L. *J. Am. Chem. Soc*. **1936**, *58*, 1486-1493. (b) Tapia, O.; Goscinski, O. *Mol. Phys*. **1975**, *29*, 1653−1661.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; 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.: Pittsburgh, PA, 2003.

(20) The TS1 geometry of the substrate$-H_3O^+(H_2O)_6$ system was also calculated. In the system, three water molecules [7], [8], and [9] in Scheme 5 were excluded. The underlined numbers for the system are similar to those for the substrate$-H_3O^+(H_2O)_9$. Data of TS1 (F2-2) and the precursor (F2-1) of the $H_3O^+(H_2O)_6$ system are shown in Figure S6 (Supporting Information).

CT600132K

Journal of Chemical Theory and Computation

Oxygen Reduction on $Pd_{0.75}Co_{0.25}$ (111) and $Pt_{0.75}Co_{0.25}$ (111) Surfaces: An ab Initio Comparative Study

Eduardo J. Lamas and Perla B. Balbuena*

Department of Chemical Engineering, Texas A&M University, College Station, Texas 77843

Received June 17, 2006

Abstract: Density functional theory studies of adsorption of oxygen electroreduction intermediates and free energy profiles are used to discuss possible reaction mechanisms: one leading directly to H_2O production and another having H_2O_2 as an intermediate, on (111) surfaces of pure Pt, pure Pd, and bimetallic systems $Pd_{0.75}Co_{0.25}$ and $Pt_{0.75}Co_{0.25}$. It is found that the calculated affinities toward the different ORR intermediates in the studied surfaces follow the Hammer-Norskov d-band model predictions. The calculated free energy profiles and the magnitude of the barriers in both mechanisms seem to favor the hypothesis that both the direct and series O_2 reduction mechanisms might be operating in parallel. The highest thermodynamic barriers at $^1/_4$ of a monolayer atomic oxygen coverage and without solvent are located in the first hydrogenation reaction for both mechanisms.

1. Introduction

In the past few years the quest toward a hydrogen-based economy has intensified the interest for effective and less expensive catalysts for fuel cell applications. Due to its slow kinetics alternative catalysts for the oxygen electroreduction reaction (OERR) are actively researched. Platinum alloys with different transition metals (for example: Ni, Co, and Fe) have shown improved activity over pure Pt.[1−3] The design of a Pt-free catalyst is also highly desirable, and different alternatives including metalloporphyrins[4] and Pd-based catalysts are being researched.[5−7] Pd-based catalysts constitute an attractive alternative to Pt alloys in fuel cell applications, not only because of lower costs but also because of the lower reactivity of Pt alloys toward methanol, which is important for improved methanol crossover tolerance on direct methanol fuel cells.[8]

The oxygen electroreduction reaction and its intermediates have been the object of intensive theoretical[9−14] and experimental studies.[1,2,15,16] The important role of OH adsorption on the catalyst surface has been identified; OH bonds strongly to the surface and has a tendency to accumulate on the surfaces that slows down the OERR kinetics by occupying active sites on the catalyst.[17]

The Hammer-Norskov d-band model[18] that successfully explains reactivity trends on transition-metal surfaces is a key contribution to the understanding of the main electronic structure characteristics of the surface atoms and their reactivity with adsorbates. This concept, together with Sabatier's principle,[19] can help to perform a rapid screening of candidate catalysts. Sabatier's principle states that the reactivity of a surface should reach a compromise between having enough strength to break bonds and generate the intermediates but low enough interaction not to stabilize them on the surface in order for the catalyst to be efficient on improving the reaction kinetics.

Recently, Norskov[20,21] proposed an approach for including the cell potential where the Gibbs free energy for the half-cell reactions are derived from density functional theory (DFT) calculations and thermodynamic data; this method is applied in this work to obtain free energy profiles at the cell equilibrium potential. We report the analyses of two possible oxygen electroreduction mechanisms on Pt- and Pd-based alloys and the interaction of the OERR intermediates with selected transition-metal surfaces.

In this work, the hydration effect of intermediates on adsorption energies and free energy profiles was not included. The effect of water in the overall reaction, its similarities and differences among the different catalysts, changes in

---

* Corresponding author e-mail: balbuena@tamu.edu.

10.1021/ct6002059 CCC: $33.50

Table 1. Transferability Tests for the Pd Pseudopotential[a]

state	*E* all electron (Ry)	*E* pseudo (Ry)	diff (Ry)
4d10	−0.297504	−0.297505	0.000001
5s0	−0.24414	−0.24414	0.000000
5p0	−0.023	−0.023	0.000000
4d9.75	−0.342907	−0.343394	0.000487
5s0.25	−0.262584	−0.262315	−0.000269
5p0	−0.032494	−0.032367	−0.000127
4d9	−0.564201	−0.568206	0.004005
5s0.5	−0.370635	−0.370208	−0.000427
5p0.5	−0.095426	−0.095157	−0.000269
4d9	−0.491517	−0.495178	0.003661
5s1	−0.320069	−0.319349	−0.00072
5p0	−0.058215	−0.057809	−0.000406

[a] The last column provides the difference between the energies from the pseudopotential and those of a full electron calculation for the various states in column 1.

adsorption energies, and effects in the overall reaction will be addressed in future work where the results will also be contrasted with the ones here presented.

2. Methodology

DFT calculations of the oxygen reduction reaction intermediates were performed on Pt, Pd, $Pd_{0.75}Co_{0.25}$, and $Pt_{0.75}Co_{0.25}$ using the plane wave framework. The unit cell is modeled as a three-layer − 4 atoms per layer (111) surfaces. In the bimetallic case, we studied two skin systems with the same Co distributions on Pt-based and Pd-based systems; in both cases the proportions of Co are 0% in the first layer, 50% in the second layer, and 25% in the third layer. Periodic boundary conditions are applied in the three spatial directions, and 10 Å of vacuum space is left between periodic images in the direction perpendicular to the surface.

We adopted Vanderbilt ultrasoft pseudopotentials[22] to decrease the computational requirements associated with the description of inert core electrons, together with the PBE exchange correlation functional. All pseudopotentials are either obtained from Vanderbilt's library[23] or derived from that reference. The Pd pseudopotential is based on a Rh pseudopotential available in the library and developed by Hansen designed with 2 s, 2 p, and 2 d nonlocal projectors and nonlinear core correction (rcloc = 2.2 au); the cut off radii are set to rc = 2.64 au for the s and p channels, and rc = 1.55 au for the d channel. In all cases transferability is verified; the results for Pd are shown in Table 1, and this pseudopotential gives a cell constant expanded nearly 1% with respect to experimental results with $a_0 = 3.926$ Å. The plane wave cut off is set to 50 Ry, and the first Brillouin zone is sampled with a Monkhorst Pack mesh with 7 × 7 × 1 k-points. To facilitate convergence, fractional occupancies are allowed applying the Mazzari-Vanderbilt cold smearing scheme.[24] The electronic temperature is set to 1000 K, with a smearing parameter of $kT \sim 0.0325$ eV.

The geometries of the different species in this study are relaxed using the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[25−28] method ensuring that residual forces are below 10^{-3} Ry/au and with a change in total energy under 10^{-4} Ry between BFGS steps. The wave functions are converged in such a way that energy changes are under 10^{-6} Ry during self-consistent iterations; in all cases the calculations include spin polarization, and the nonzero magnetization states are taken into account in the Co alloys. All these convergence criteria are verified by running a few calculations with tighter convergence parameters and checking geometries and energies against the data obtained with the above-mentioned parameters.

In all cases the adsorption energy is calculated according to

$$\Delta E = \frac{1}{n}(E_{S+nA} - (E_s - n\cdot E_A)) \qquad (1)$$

where *S* is the clean surface, *A* is the adsorbate, and *n* is the number of adsorbate molecules in the unit cell.

The changes in electron density originated by the surface−adsorbate interactions have been calculated according to

$$\Delta\rho(\vec{x}) = \rho(\vec{x})_{\mathrm{s+ads}} - (\rho(\vec{x})_{\mathrm{s}} + \rho(\vec{x})_{\mathrm{ads}}) \qquad (2)$$

To gain understanding into the reactivity potential of the different studied surfaces, the local density of states is calculated by projecting the plane wave expansion into a linear combination of atomic orbitals. All simulations and data analysis were performed with the pw.x program and postprocessing utilities that are part of the quantum ESPRESSO package.[29] The visualizations were obtained with VMD.[30]

3. Surface Characterization

After alloying, surface properties can be generally affected in many different ways. Succinctly these changes can be classified as electronic (for example, band shifting due to the presence of a second component) and structural such as one component island formation, skin formation, and/or lattice constant changes. These changes are also entangled; for example, a change in the lattice constant will also generate a change in the system's band structure. In this section, we discuss some of the studied alloy characteristics focusing mainly on the lattice constant and d-band structure changes that, as mentioned in the introductory section, have proven of capital importance in the characterization of transition-metal catalysts.

The surfaces studied in this work have their lattice constants for pure Pt and Pd fixed to their experimental values (3.92 Å and 3.89 Å, respectively), while their Co-alloys lattice constants are estimated from variable cell calculations using the Wentzcovitch damped dynamics algorithm.[31] In all cases where Co is present we found a lattice contraction with final lattice constant $a_0 = 3.85$ Å for the PdCo alloy and $a_0 = 3.89$ Å in the PtCo alloy in good agreement with previously published data (Xu et al.[14] calculation yielded 3.85 Å for the same PtCo alloy). The PdCo and PtCo alloys are found to be magnetic in their ground states, whereas the Pt and Pd surfaces are restricted to the nonmagnetic state during the calculations (see ref 32 for an extended discussion on magnetization issues associated with modeling Pd systems).

Figure 1 shows the localized density of states vs the energy (referred to the Fermi level of the metal or alloy calculated

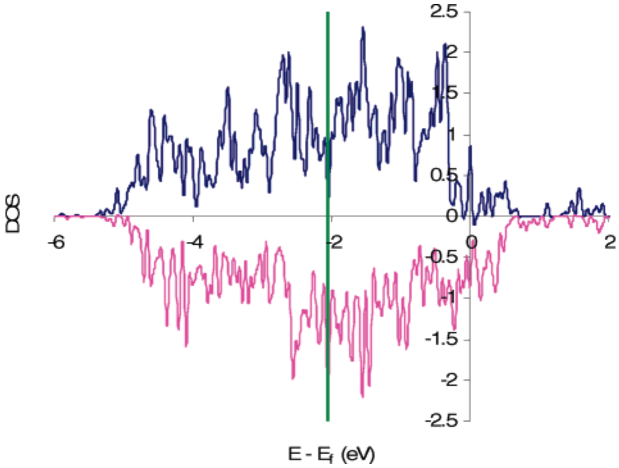


Figure 1. d-Band density of states for surface Pd atoms in the $Pd_{0.75}Co_{0.25}$ alloy. The d-band center (averaged for the spin up and down states) is also shown in the figure (green line). The positive intensities correspond to states of spin-up, and the negatives to spin-down.

by the model) of the d-band in the surface Pd atoms of the skin $Pd_{0.75}Co_{0.25}$ (111) alloy. The location of the first moment of the d-band structure (the d-band center) is also indicated in the picture. The d-band center is located 2.06 eV below the Fermi level in the $Pd_{0.75}Co_{0.25}$ surface, 1.88 eV for the Pd surface, 2.36 eV for Pt, and 2.49 eV for $Pt_{0.75}Co_{0.25}$. If the d-band model holds for these metals the strengths of the adsorbate−surface interactions should be in increasing order: PtCo < Pt < PdCo < Pd; an analysis of adsorption energy values in Tables 2−6 shows that the agreement is indeed excellent.

4. Oxygen Electroreduction Intermediates and Their Interactions with Transition-Metal Surfaces

4.1. Atomic and Molecular Oxygen. Oxygen dissociation on Pt (111) is a thermally activated process that occurs above 150 K and evolves through chemical adsorbed precursors. The existence of these precursors has been verified both experimentally and theoretically.[33] Our studies of O_2 on the different surfaces are in agreement with the existence of these precursors. Local minima on the potential energy surface were found with initial configurations of O_2 parallel to the surface, while the studied vertical orientation is not bounded to the surface. The reference calculation for oxygen in the gas phase yielded a triplet as the most stable electronic structure with a O−O bonding distance of 1.23 Å and dissociation energy of 5.81 eV. Figure 2b shows O_2 adsorbed on the Pd skin surface of the $Pd_{0.75}Co_{0.25}$ alloy as well as changes on electron density upon adsorption. It is observed

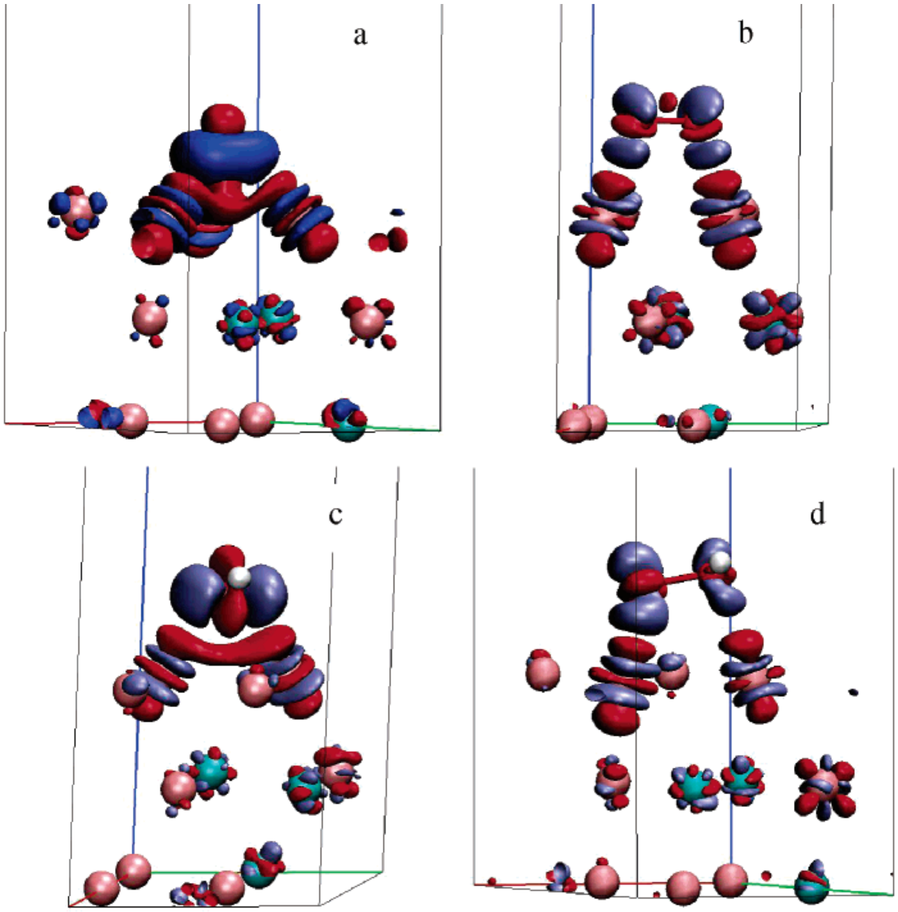


Figure 2. Electron density differences upon adsorption calculated according to eq 2, for (a) atomic oxygen, (b) molecular oxygen, (c) hydroxyl, and (d) hydroperoxyl on the $Pd_{0.75}Co_{0.25}$ (111) surface. Pd atoms are in pink, Co atoms are in light blue, oxygen are in red and hydrogen in white. Positive electron density differences are in ice blue, and negative electron density differences are in red. $\Delta\rho$ isosurfaces were calculated at $\pm$0.006 e/au^3.

Table 2. DFT Calculated Molecular Oxygen Adsorption Energies[a]

system	Δ*E* (eV)	distances (Å)		
		O1−M1	O2−M2	O−O
$Pt-O_2$	−0.431	2.06	2.07	1.37
$Pt_{0.75}Co_{0.25}-O_2$	−0.216	2.14	2.17	1.33
$Pd_{0.75}Co_{0.25}-O_2$	−0.408	2.07	2.08	1.33
$Pd-O_2$	−0.715	2.01	2.02	1.35

[a] The distance between the closest metal atom to each oxygen is tabulated together with the O−O bond distance.

Table 3. DFT Calculated Oxygen Adsorption Energies and Distances[a]

system (site)	Δ*E* (eV)	distances (Å)		
		O−M1	O−M2	O−M3
Pt−O (fcc hollow)	−3.727	2.08	2.08	2.08
Pt−O (hcp hollow)	−3.316	2.11	2.11	2.11
Pt−O (top)	−2.492	1.87		
$Pt_{0.75}Co_{0.25}$−O (fcc hollow)	−3.490	2.1	2.1	2.09
$Pt_{0.75}Co_{0.25}$−O (hcp Co-hollow)	−2.995	2.13	2.14	2.09
$Pt_{0.75}Co_{0.25}$−O (hcp hollow)	−2.855	2.09	2.09	2.15
$Pd_{0.75}Co_{0.25}$−O (fcc hollow)	−3.778	2.02	2.02	2.03
$Pd_{0.75}Co_{0.25}$−O (hcp Co-hollow)	−3.473	2.03	2.03	2.01
$Pd_{0.75}Co_{0.25}$−O (hcp hollow)	−3.464	2.01	2.02	2.03
Pd−O (fcc hollow)	−4.153	1.99	1.99	1.99
Pd−O (hcp hollow)	−3.962	2.01	2.01	2.01
Pd−O (top)	−2.646	1.8		

[a] In the hollow cases we report the distances to the three closest surface metal atoms.

Table 4. DFT Calculated Hydroxyl Adsorption Distances and Energies[a]

system (site)	Δ*E* (eV)	distances (Å)		
		O−M1	O−M2	O−M3
Pt−OH (top)	−2.081	2.02		
Pt−OH (bridge)	−1.989	2.22	2.21	
$Pt_{0.75}Co_{0.25}$−OH (top)	−1.870	2.05		
$Pt_{0.75}Co_{0.25}$−OH (bridge)	−1.791	2.24	2.27	
$Pd_{0.75}Co_{0.25}$−OH (top)				
$Pd_{0.75}Co_{0.25}$−OH (bridge)	−2.199	2.19	2.17	
Pd−OH (top)	−2.173	1.97		
Pd−OH (bridge)	−2.357	2.14	2.13	

[a] In the bridge adsorption cases the reported distances are the two closest surface metal atoms.

Table 5. DFT Calculated HO_2 Adsorption Distances and Energies[a]

system	Δ*E* (eV)	distances (Å)		
		O1−M1	O2−M2	O−O
$Pt-HO_2$	−1.043	2.86	2.06	1.43
$Pt_{0.75}Co_{0.25}-HO_2$	−0.881	2.9	2.08	1.43
$Pd_{0.75}Co_{0.25}-HO_2$	−0.973	2.41	2.04	1.46
$Pd-HO_2$	−1.218	2.32	2.01	1.47

[a] The last column shows the O−O distance in the adsorbate.

that there is a decrease in the electron density in the O−O bond region and an electron flow from the surface to the antibonding π states of each O atom. Qualitatively similar changes in the electronic density upon O_2 adsorption are obtained for a Pt(111) surface; these observations are in excellent agreement with previously reported calculations on a Pt surface.[34] The bond elongations signaling the weakening of the O−O bond in all studied surfaces are indicated in Table 2, and the value of 1.37 Å is in good agreement with experimental reports of the superoxo state on Pt(111) surfaces.[35]

In all surfaces in this study, the preferred adsorption site for atomic oxygen is found to be the *fcc* hollow site (in good agreement with previously reported results[36]), followed by *hcp* hollow, and then the top site (Table 3). O adsorbed on a top site becomes unstable in the alloy surfaces; in all cases it is found that it slips to a neighbor *hcp* hollow position with an underlying Co atom in the second layer of the surface. The energetic differences between adsorption in both hollow sites (and thus the barriers for O surface diffusion) are also altered by the presence of the second layer Co atoms; in fact, in both alloys this difference is higher than that of the pure metal, going from 0.411 eV in pure Pt to 0.635 eV in the $Pt_{0.75}Co_{0.25}$ alloy and from 0.192 eV on Pd to 0.314 eV in the $Pd_{0.75}Co_{0.25}$ alloy. Due to the strong O-metal bonding, the changes in electron density depicted in Figure 2a are the most dramatic among the different adsorbates showing increased electronic density along a plane parallel to the surface and depleted electronic density along a line perpendicular to the surface and changes on the surface metal atoms along the metal−adsorbate bonds. These features remain essentially present for all the adsorbates (symmetry axes for the electronic density changes parallel and perpendicular to the surface for the adsorbate and along the metal−adsorbate bonds).

4.2. Hydroxyl (OH). OH is perhaps one the most important OERR intermediates; experimental evidence suggests that it adsorbs strongly on the catalyst surface blocking active sites,[3] and its degree of interaction with different transition-metal surfaces is an important indication of their efficiency toward oxygen reduction.

From these calculations it was observed that in Pt-based surfaces the preferred adsorption site is on top, while the preferred adsorption site is a bridge for Pd surfaces. The top site is no longer stable on the Pd atoms of the $Pd_{0.75}Co_{0.25}$ alloy (as in the atomic oxygen case, the OH placed on top slips to neighbor bridge sites), and the adsorbate is slightly displaced from the top position in $Pt_{0.75}Co_{0.25}$. Energies and bond distances are provided in Table 4.

Figure 2c shows the geometry and electronic changes for OH adsorbed on the Pd surface of the $Pd_{0.75}Co_{0.25}$ alloy. Qualitatively the changes in electronic density are similar to those of the adsorbed atomic oxygen (Figure 2a), although less dramatic changes due to its lower binding strength with the surface.

At a higher OH concentration (half of an OH monolayer obtained as a product of the H_2O_2 dissociation on the surfaces), a hydrogen bond is formed between adsorbed neighbor OH species. This effect further stabilizes the hydroxyl groups giving adsorption energies per OH that are lower than at the low coverage by an amount close to the hydrogen bonding energy in water (23.3 kJ/mol or about 0.24

Table 6. DFT Calculated H_2O_2 Dissociation into Two Adsorbed OH Radicals[a]

system (site 1-site 2)	$\Delta E^{(1)}$ (eV)	$\Delta E^{(2)}$ (eV)	distances (Å) O1−M1	O2−M2	O2−M3	O2−M4	H-bond dist (Å)
Pt−2OH(top-top)	−2.056	−2.294	2.02	2.02			1.87
$Pt_{0.75}Co_{0.25}$−2OH (top-top)	−1.668	−2.100	2.04	2.04			1.78
$Pd_{0.75}Co_{0.25}$−2OH (top-top)	−1.795	−2.163	2	1.99			1.85
$Pd_{0.75}Co_{0.25}$−2OH (top-hollow)	−1.777	−2.154	2.01	2.09	2.31	2.74	2.14
Pd−2OH (top-top)	−2.144	−2.338	1.97	1.97			1.95

[a] $\Delta E^{(1)}$ is the binding energy calculated according to eq 1 with $n = 1$ and taking E_{H2O2} as E_A, whereas $\Delta E^{(2)}$ is calculated with $n = 2$ and E_{OH} as E_A. The values of $\Delta E^{(2)}$ may be compared with those in Table 4 to quantify the stabilization due to hydrogen bonding. The last column in this table gives the hydrogen bonding distance between neighbor hydroxyls.

eV see for example ref 37); Table 6 displays binding energies and geometries for the case of 0.5 monolayer OH adsorbed.

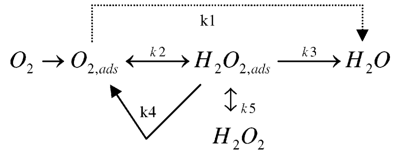
4.3. Hydroperoxyl (HO_2). In all the calculated Pt- and Pd-skin cases HO_2 is stable, and no dissociated form is found when starting from an undissociated initial condition. To calculate the binding energies, the HO_2 radical is taken as a reference in the gas phase, with a O−O bonding distance of 1.35 Å, O−H distance of 0.99 Å, and O−O−H angle of 104.96°. After adsorption, the molecule tilts on top of two metal atoms; this tilting is due to the weaker interaction of the OH side of the molecule with the surface. The adsorption energies are listed in Table 5, and Figure 2d shows the geometry over the Pd surface of the $Pd_{0.75}Co_{0.25}$ alloy and an isosurface of the electron density difference calculated according to eq 2. Not surprisingly, in this case, the features of the $\Delta\rho$ plot (Figure 2d) are comparable to the O_2 case; that is, there is an electron density gain in the antibonding π orbital associated with each oxygen atom (this gain is lower in the extreme of the molecule that bonds to hydrogen because of the stability of the O−H covalent bond).

4.4. Hydrogen Peroxide (H_2O_2). The most stable electronic configuration for H_2O_2 is a singlet with bond lengths 1.47 Å for O−O and 0.98 Å for O−H, angles about 100° for H−O−O, and a dihedral angle about 112°. When placed over a surface at about 2 Å, in its vacuum configuration, dissociation is observed in all cases whereas when placed above 2.6 Å, the molecule adsorbs without dissociation tilted with one of its oxygen atoms over a top position. Dissociation energies are reported in Table 6. Once dissociated, two OH species are observed in two possible configurations: one where both OH radicals are adsorbed on top of metal atoms and another where one OH adsorbs on top while the other is in a near-hollow position. On the Pd surface of the $Pd_{0.75}Co_{0.25}$ alloy, a top-top configuration seems to be slightly preferred (Table 6); others reported similar results on Pt(111).[38] In both cases, the adsorbed OH rotates forming a hydrogen bond with its neighbor adsorbed OH. The hydrogen bond distances are also reported in Table 6.

5. Free Energy Profiles

Oxygen electroreduction in acidic media can evolve through one or a combination of two different paths; a *direct* pathway that involves O−O bond breaking before a second proton is attached to the molecule and a *series* pathway that generates hydrogen peroxide as an intermediate. Schematically these two pathways are shown in Scheme 1.[1,39]

Scheme 1. Two Possible OERR Mechanisms According to Wroblowa et al.[39] [a]



[a] The k_i values are the kinetic rate constants associated with each process. The process associated with k_1 is the direct mechanism, whereas the series mechanism involves the formation of adsorbed H_2O_2 (rate constants k_2 and k_3).

In this section we study—via generation of free energy profiles—two representative mechanisms of the OERR, one for each pathway. Gibbs free energy changes for the elementary steps including zero point energies and entropies, and cell potential effects are estimated following the methodology introduced by Norskov et al.[20,21]

The ΔG for the overall reaction ($^1/_2O_2 + H_2 \rightarrow H_2O$) is calculated to be 2.13 eV that makes the theoretical results off by about 14% (the experimental value is 2.5 eV) compared with experiments; this discrepancy has been attributed in part to the difficulties in modeling the O_2 electronic structure.[21] The results shown below that introduce the electrode potential in the free energy profiles are based on the calculated ΔG for the overall reaction.

5.1. A *Direct* Pathway Mechanism. A simple mechanism associated with the *direct* pathway, where the O−O splits before the first hydrogenation, is summarized by eqs 3−5, and Figure 3 displays the ΔG profile for the cell at the equilibrium potential.

$$\frac{1}{2}O_2 + * \rightarrow O^* \tag{3}$$

$$O^* + (H^+ + e^-) \rightarrow OH^* \tag{4}$$

$$OH^* + (H^+ + e^-) \rightarrow H_2O + * \tag{5}$$

The surface with the highest d-band center and thus the most reactive one (Pd (111)) is the only one on which the current model predicts water dissociation (the reverse reaction to that shown in Figure 3) at the cell's equilibrium potential and at the studied oxygen coverage ($\theta = 0.25$). On the other hand, for Pt(111) at the equilibrium potential, at the oxygen coverage $\theta = 0.25$ and in the absence of water we find a slightly uphill path for water dissociation. Norskov et al.[20] calculations yield a thermodynamic barrier for water dis-

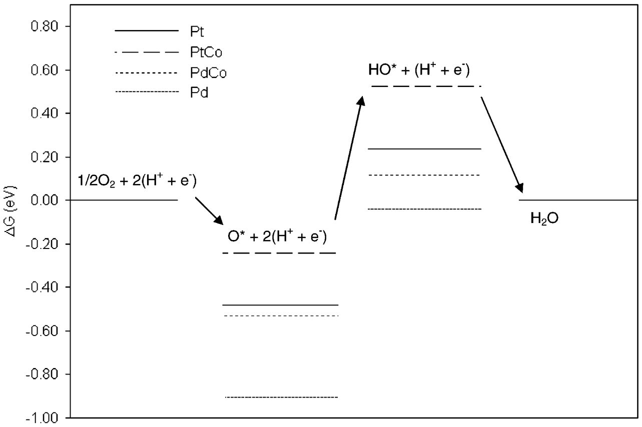


Figure 3. DFT calculated Gibbs free energy profile for a *direct* mechanism of the OERR on Pt, $Pt_{0.75}Co_{0.25}$ and $Pd_{0.75}Co_{0.25}$, and Pd surfaces at the cell equilibrium potential.

sociation on Pt (111) at the cell equilibrium potential and $\theta = 0.5$, while the barrier becomes downhill at an unspecified low coverage; we assume (based on their largest cell size) that this corresponds to a value of $\theta = 0.167$. This is in good agreement with our results that predict a thermodynamic barrier for an oxygen coverage intermediate between those two cases.

The Pt surface of the $Pt_{0.75}Co_{0.25}$ alloy shows Δ*G* values along the path that are above the Pt and $Pd_{0.75}Co_{0.25}$ surfaces, while the Δ*G* values corresponding to the Pd surface which has the strongest bonds with the intermediates is below. Based solely on the data obtained on the profile the highest thermodynamic barrier for this reaction mechanism under the specific conditions (low oxygen coverage, no water) is likely to be located in the first hydrogenation (eq 4).

5.2. A *Series* Mechanism. The Gibbs free energy profile for a series mechanism indicated by the chemical eqs 6−9 is shown in Figure 4 at the cell equilibrium potential. In this mechanism we assume that H_2O_2 is unstable in all surfaces, and thus the result of the second hydrogenation reaction is its immediate dissociation yielding two adsorbed hydroxides.

$$O_2 + * \rightarrow O_2* \tag{6}$$

$$O_2* + (H^+ + e^-) \rightarrow HO_2* \tag{7}$$

$$HO_2* + (H^+ + e^-) \rightarrow H_2O_2* \rightarrow 2OH* \tag{8}$$

$$2OH* + 2(H^+ + e^-) \rightarrow 2H_2O + 2* \tag{9}$$

At the cell equilibrium potential (1.23 V) there are a series of uphill thermodynamic barriers for the formation of water from hydrogen and oxygen. As in the first mechanism, the free energy profile under these conditions (low O coverage, no solvent) indicates that the highest thermodynamic barrier is the first hydrogenation (eq 7). Reactivity trends similar to the *direct* mechanism are found among the various surfaces, again following with the d-band center model. The difference in values for the thermodynamic barriers in both mechanisms is comparable signaling that both mechanisms might be active and operating in parallel.

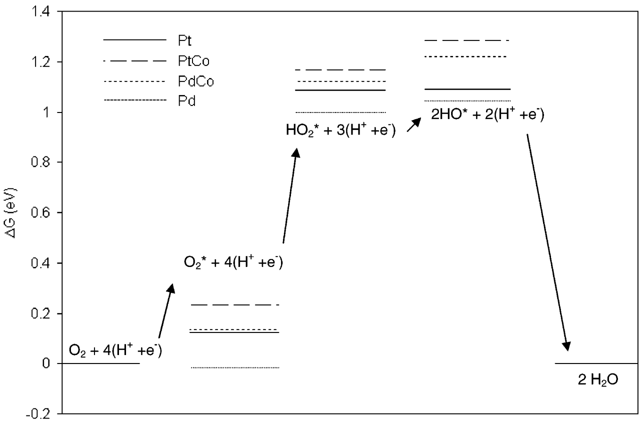


Figure 4. DFT calculated Gibbs free energy profile for a possible *series* mechanism of OERR in Pt, $Pt_{0.75}Co_{0.25}$ and $Pd_{0.75}Co_{0.25}$, and Pd surfaces at the cell equilibrium potential.

6. Conclusions

Binding energies of oxygen electroreduction intermediates and free energy profiles for *direct* and *series* reaction mechanisms on (111) surfaces of Pt, Pd, $Pd_{0.75}Co_{0.25}$, and $Pt_{0.75}Co_{0.25}$ indicate that the affinities toward the different ORR intermediates and reactivity trends on the studied surfaces follow the prediction of the d-band model. The study of the free energy profiles and the magnitude of the thermodynamic barriers in both mechanisms favors the hypothesis that the O_2 reduction mechanism to water might be operating in *parallel*. The highest thermodynamic barriers appear to be located in the first hydrogenation reaction for both mechanisms when the system is studied at conditions of $^1/_4$ of a monolayer oxygen coverage and in the absence of water.

Acknowledgment. This work was supported by the Department of Energy, grant DE-FG02-05ER15729. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Supercomputer time granted by the DoD Major Shared Resource Center (ARL MSRC) is gratefully acknowledged.

References

(1) Markovic, N. M.; Schmidt, T. J.; Stamenkovic, V.; Ross, P. N. Oxygen Reduction Reaction on Pt and Pt bimetallic Surfaces: A Selective Review. *Fuel Cells* **2001**, *1*, 105−116.

(2) Toda, T.; Igarashi, H.; Uchida, H.; Watanabe, M. Enhancement of the Electroreduction of Oxygen on Pt Alloys with Fe, Ni and Co. *J. Electrochem. Soc.* **1999**, *146*, 3750−3756.

(3) Stamenkovic, V.; Schmidt, T. J.; Ross, P. N.; Markovic, N. M. Surface segregation effects in electrocatalysis: kinetics of oxygen reduction on polycrystalline Pt3Ni alloy surfaces. *J. Electroanal. Chem.* **2003**, *554−555*, 191−199.

(4) Collman, J. P.; Wagenknecht, P. S.; Hutchison, J. E. Molecular Catalysts for Multielectron Redox Reactions of Small Molecules: The "Cofacial Metallodiporphyrin" Approach. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 1537−1554.

(5) Savadogo, O.; Lee, K.; Oishi, K.; Mitsushima, S.; Kamiya, N.; Ota, K.-I. New palladium alloys catalyst for the oxygen reduction reaction in an acid medium. *Electrochem. Commun.* **2004**, *6*, 105−109.

(6) Shao, M. H.; Sasaki, K.; Adzic, R. R. Pd−Fe nanoparticles as electrocatalysts for oxygen reduction. *J. Am. Chem. Soc.* **2006**, *128*, 3526−3527.

(7) Fernandez, J. L.; Walsh, D. A.; Bard, A. J. Thermodynamic Guidelines for the Design of Bimetallic Catalysts for Oxygen Electroreduction and Rapid Screening by Scanning Electrochemical Microscopy. M-Co (M: Pd, Ag, Au). *J. Am. Chem. Soc.* **2005**, *127*, 357−365.

(8) Lee, K.; Savadogo, O.; Ishihara, A.; Mitsushima, S.; Kamiya, N.; Ota, K. Methanol-Tolerant Oxygen Reduction Electrocatalyst Based on Pd-3D Transition Metal Alloys for Direct Methanol Fuel Cell. *J. Electrochem. Soc.* **2006**, *153*, A20-A24.

(9) Xu, Y.; Ruban, A.; Mavrikakis, M. Adsorption and Dissociation of O_2 on Pt−Co and Pt−Fe Alloys. *J. Am. Chem. Soc.* **2003**, *126*, 4717−4725.

(10) Roques, J.; Anderson, A. B. Electrode Potential Dependence Stages in OHads Formation on the Pt3Cr Alloy (111) Surface. *J. Electrochem. Soc.* **2004**, *151*, E340−E347.

(11) Balbuena, P. B.; Altomare, D.; Vadlamani, N.; Bingi, S.; Agapito, L. A.; Seminario, J. M. Adsorption of O, OH, and H_2O on Pt-Based Bimetallic Clusters Alloyed with Co, Cr, and Ni. *J. Phys. Chem. A* **2004**, *108*, 6378−6384.

(12) Wang, Y.; Balbuena, P. B. Design of oxygen reduction bimetallic catalysts: Ab-initio derived thermodynamic guidelines. *J. Phys. Chem. B* **2005**, *109*, 18902−18906.

(13) Panchenko, A.; Koper, M. T. M.; Shubina, T. E.; Mitchell, S. D.; Roduner, E. Ab Initio Calculations of Intermediates of Oxygen Reduction on Low-Index Platinum Surfaces. *J. Electrochem. Soc.* **2004**, *151*, A2016−A2027.

(14) Xu, Y.; Ruban, A. V.; Mavrikakis, M. Adsorption and dissociation of O_2 on Pt−Co and Pt−Fe alloys. *J. Am. Chem. Soc.* **2004**, *126*, 4717−4725.

(15) Adzic, R. R.; Wang, J. X. Configuration and site of O_2 adsorption on the Pt(111) electrode surface. *J. Phys. Chem. B* **1998**, *102*, 8988−8993.

(16) Wang, J. X.; Markovic, N. M.; Adzic, R. R. Kinetic analysis of oxygen reduction on Pt(111) in acid solutions: Intrinsic kinetic parameters and anion adsorption effects. *J. Phys. Chem. B* **2004**, 4127−4133.

(17) Paulus, U. A.; Wokaum, A.; Scherer, G. G.; Schmidt, T. J.; Stamenkovic, V.; Markovic, N. M.; Ross, P. N. Oxygen reduction on high surface area Pt-based alloy catalyst in comparison to well-defined bulk alloy electrodes. *Electrochim. Acta* **2002**, *47*, 3787−3798.

(18) Hammer, B.; Norskov, J. K. Electronic factors determining the reactivity of metal surfaces. *Surf. Sci.* **1995**, *343*, 211−220.

(19) Chorkendorff, I.; Niemantsverdriet, J. W. *Concepts of Modern Catalysis and Kinetics*; Wiley-VCH: 2003.

(20) Norskov, J. K.; Rossmeisl, J.; Logadottir, A.; Lindqvist, L. Origin of the Overpotential for Oxygen Reduction at a Fuel-Cell Cathode. *J. Phys. Chem. B* **2004**, *108*, 17886−17892.

(21) Rossmeisl, J.; Logadottir, A.; Norskov, J. K. Electrolysis of water on (oxidized) metal surfaces. *Chem. Phys.* **2005**, *319*, 178−184.

(22) Vanderbilt, D. Soft Self-Consistent Pseudopotentials in a Generalized Eigenvalue Formalism. *Phys. Rev. B* **1990**, *41*, 7892.

(23) Vanderbilt, D. Vanderbilt Ultra-Soft Pseudopotential Site. http://www.physics.rutgers.edu/∼dhv/uspp/ (accessed on July 19, 2006).

(24) Marzari, N.; Vanderbilt, D.; De Vita, A.; Payne, M. C. Thermal Contraction and Disordering of the Al (110) Surface. *Phys. Rev. Lett.* **1999**, *82*, 3296−3299.

(25) Broyden, C. G. The Convergence of a Class of Double-Rank Minimization Algorithms 2. The New Algorithm. *J. Inst. Math. Appl.* **1970**, *6*, 222−231.

(26) Fletcher, R. A. A new approach to variable metric algorithms. *Comput. J.* **1970**, *13*, 317−322.

(27) Goldfarb, D. A family of variable-metric methods derived by variational means. *Math. Comput.* **1970**, *24*, 23−26.

(28) Shanno, D. F. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* **1970**, *24*, 647−656.

(29) Scandolo, S.; Giannozzi, P.; Cavazzoni, C.; de Gironcoli, S.; Pasquarello, A.; Baroni, S. First-principles codes for Computational Crystallography in the Quantum-ESPRESSO package. *Z. Kristallogr.* **2005**, *220*, 574−579.

(30) Humphrey, W.; Dalke, A.; Schulten, K. VMD − Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(31) Wentzcovitch, R. M.; Martins, J. L.; Price, G. D. Ab Initio Molecular Dynamics with Variable Cell Shape: Application to $MgSiO_3$. *Phys. Rev. Lett.* **1993**, *70*, 3947−3950.

(32) Alexandre, S. S.; Mattesini, M.; Soler, J. M.; Yndurain, F. Comment on ‘‘Magnetism in Atomic-Size Palladium Contacts and Nanowires”. *Phys. Rev. Lett.* **2006**, *96*, 079701.

(33) Eichler, A.; Hafner, J. Molecular Precursors in the Dissociative Adsorption of O_2 on Pt (111). *Phys. Rev. Lett.* **1997**, *79*, 4481−4484.

(34) Grob, A.; Eichler, A.; Hafner, J.; Mehl, M. J.; Papaconstantopoulos, D. A. Unified picture of the molecular adsorption process: O_2/Pt(111). *Surf. Sci.* **2003**, *539*, L542−L548.

(35) Puglia, C.; Nilsson, A.; Hernnas, B.; Karis, O.; Bennich, P.; Martensson, N. Physisorbed, Chemisorbed and Dissociated O_2 on Pt(111) Studied by Different Core Level Spectroscopy Methods. *Surf. Sci.* **1995**, *342*, 119.

(36) Jacob, T.; Muller, R. P.; Goddard, W. A., III Chemisorption of Atomic Oxygen on Pt(111) from DFT Studies of Pt-Clusters. *J. Phys. Chem. B* **2003**, *107*, 9465−9476.

(37) Suresh, S. J.; Naik, V. M. Hydrogen bond thermodynamic properties of water from dielectric constant data. *J. Chem. Phys.* **2000**, *113*, 9727−9732.

(38) Michaelides, A.; Hua, P. A density functional theory study of hydroxyl and the intermediate in the water formation reaction on Pt. *J. Chem. Phys.* **2001**, *114*, 513−519.

(39) Wroblowa, H. S.; Yen-Chi-Pan; Razumney, G. Electroreduction of Oxygen a New Mechanistic Criterion. *J. Electroanal. Chem.* **1976**, *69*, 195−201.

CT6002059

J. Chem. Theory Comput. **2006,** *2,* 1395−1402

JCTC Journal of Chemical Theory and Computation

Computational Study of Aqueous Reactions in Tocopherol Regeneration

R. Lui[†] and A. L. Cooksy*[,†,‡]

Department of Chemistry, San Diego State University, San Diego, California 92182-1030, and Centro de Graduados e Investigacion, Instituto Tecnológico de Tijuana, Apdo Postal 1166, Tijuana, B.C. México

Received July 3, 2006

Abstract: The α-tocopheroxyl radical, resulting from the scavenging of the peroxyl radical by α-tocopherol (vitamin E) in the cell membrane, will further react with the peroxyl radical to form tocopherones. A computational study is presented of the aqueous proton-transfer reactions of the α-8a-(hydro-dioxy)tocopherone and α-8a-(methyl-dioxy)tocopherone to produce 1-benzopyrylium, the subsequent hydrolysis to 2H-1-benzopyran-6(8aH)-one, and the terminating rearrangement of 8a-hydroxytocopherone to 2,5-cyclohexadiene-1,4-dione. The alkyl tail of true tocopherol is replaced by methyl in these studies. Calculations of the structures and energies along the reaction pathways were first performed at the BP86/TZVP/DGA1 level with a COSMO solvent model, and additional solvent corrections from COSMO-RS theory were subsequently added. The proposed mechanism is found to be thermodynamically and kinetically feasible in water. The ion-mediated steps are found to have free-energies of activation under 3 kcal mol^{-1}, with kinetics likely to be diffusion-limited. The unimolecular rearrangement is the slow step, with a predicted reaction rate constant of 0.056 min^{-1} at 298 K, in excellent agreement with the experimental value of 0.046 min^{-1}. Kinetic and thermodynamic properties of the ion-mediated steps are shown to rely strongly on the use of an accurate continuum solvation model. Distinct entropy and enthalpy contributions are determined from the temperature dependence of the predicted free energies.

1. Introduction

Vitamin E is a fat-soluble vitamin, existing in nature as any of eight different tocols and tocotrienols. All of these structures consist of a chromanol heterocycle (Figure 1) with a hydrophilic hydroxyl group at the 6 position and a hydrophobic alkyl tail at the 2 position, the latter anchoring the vitamin in cell membranes. This dual solubility enhances its antioxidant activity by protecting the lipid bilayers of cell membranes against reaction with free radicals.[1] In the vitamin E family, α-tocopherol (α-TH), a trimethyl-substituted tocol, is found to be the most biologically significant form and has been widely used to study the activity of vitamin E.

To consume destructive lipid peroxyl (LOO•) radicals, α-TH transfers the hydroxyl hydrogen to generate LOOH either by direct hydrogen transfer or sequential electron−proton transfers. The resulting tocopherol radical (T•) is less reactive than LOO• or the lipid radical (L•), because the conjugated π system in the chromanol ring system confers the T• with extra stability. In general, it is believed that T• will regenerate back to TH by reaction in the cell with vitamin C. Research[2] has indicated the enhancement of TH regeneration in the presence of other cooperative antioxidants, in particular, with the water-soluble vitamin C.[3,4] However, some of the T• generated in antioxidation will not regenerate but will further react with another ROO• to form 8a-(alkyldioxy)tocopherones.

Several computational studies of the antioxidant activity of tocopherol were carried out in the early 1990s,[5−9] and that field is now being extensively revisited.[10−16] To our

---

* Corresponding author fax: (619) 594-4634; e-mail: acooksy@sciences.sdsu.edu.

[†] San Diego State University.

[‡] Instituto Tecnológico de Tijuana.

10.1021/ct600161w CCC: $33.50

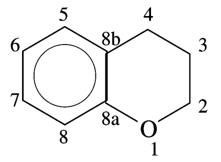


Figure 1. Position labels on the chromanol system.

knowledge, there has been no prior investigation by computational means of the subsequent reactions of tocopherol, including those that may lead to regeneration of the vitamin. One mechanism for the chemical recycling of tocopherol was proposed and studied by Liebler and co-workers,[17−19] with the steps illustrated in Scheme 1 for a model chromanol system. They proposed that the acid-catalyzed loss of the 8a-(alkyldioxy) in reaction 1 will lead to 1-benzopyrylium (α-T$^+$)[17] and that this can then be reduced by ascorbic acid directly back to α-TH. However, α-T$^+$ may also reversibly hydrolyze in reaction 2 to 8a-hydroxytocopherone (α-TOH), which then rearranges in reaction 3 to 2,5-cyclohexadiene-1,4-dione (α-TQ). The formation mechanism and relative stability in solution of the α-T$^+$ cation has recently been investigated in detailed experiments by Webster and co-workers.[20]

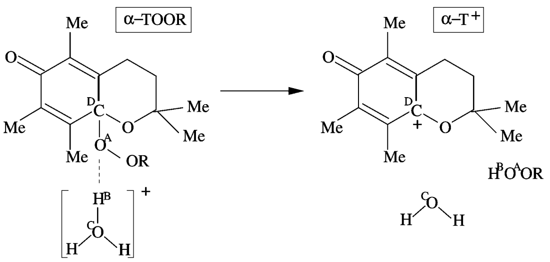
This work is a computational investigation of that mechanism. The specific reactions modeled are shown as reactions 1a, 1b, 2, and 3 in Scheme 1. The goal of this work has been to develop a computational model of these reactions to assess the feasibility of this proposed mechanism in an aqueous medium. Although this is not expected to be a major pathway in the cell membrane, these reactions may be significant in laboratory studies of vitamin E, particularly those that take place in the absence of the vitamin C or hydroquinone needed to regenerate the TH.

2. Methods

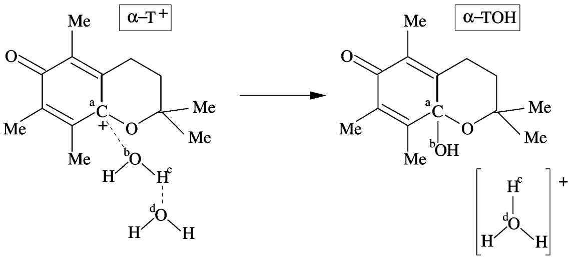
The chromanol ring of vitamin E consists of 12−15 heavy atoms (carbon or oxygen), depending on the number of methyl substituents. The tocopherol or tocotrienol then needs 16 additional heavy atoms to form the long alkyl tail. To reduce the conformational complexity of the system, these calculations focus on the chemically relevant chromanol ring system, replacing the long alkyl tail with a single methyl group. The resulting initial structure α-TH consists of only 16 heavy atoms, and there are then no more than six single-bond torsional modes connected to the chromanol system at any point in the reaction until the final ring-opening step in reaction 3.

Reaction 1 begins with our model 8a-(alkyl-dioxy)-tocopherone (α-TOOR), which is one possible product from the tocopherol antioxidizing reaction. It is simulated to interact with a hydronium ion (H_3O^+) in water, resulting in a α-T$^+$ and a corresponding alkyl-peroxide (HOOR). In Liebler et al.'s studies,[18] 8a-[(2,4-dimethyl-1-nitrilopent-2-yl)dioxy]tocopherone was used to study the hydrolysis and reduction of α-TOOR in acetonitrile/buffer mixtures. In the present work, the R substituents in the alkyl peroxide are chosen to be hydrogen (reaction 1a) and methyl (reaction 1b).

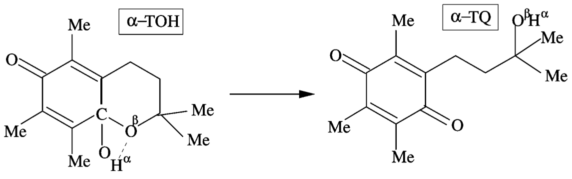
Scheme 1



Reaction 1(a, b): R=H for Reaction **1a**, R=Me for Reaction **1b**.



Reaction **2**.



Reaction **3**.

For reaction 2, α-T$^+$ reacts with two water molecules, generating an α-TOH and an H_3O^+. In general, α-TOOR, α-T$^+$, and α-TOH can be reduced by vitamin C back to α-TH, but the present study has been restricted to smaller and more computationally tractable reaction systems. For reaction 3, the α-TOH generated from reaction 2 will rearrange to α-TQ, by opening the chromanol ring as shown. Reaction 3 is the terminal reaction studied in this project, because the resulting α-TQ is not regenerated back to α-TH. Studies of the antioxidant activity of α-TQ indicate that α-TQ is reduced to α-tocopherylhydroquinone (α-TQH$_2$) by enzymes or other hydroquinones. The α-TQH$_2$, which can scavenge free radicals by hydrogen transfer, may then function as an antioxidant in lipids.[21−24]

Initial probes of the chemistry were carried out at the BP86 density functional level of theory,[25,26] using the COSMO solvent model[27] to incorporate the fundamental electrostatic interactions with water. The Dunning−Huzinaga TZVP basis set was used with DGA1 as the density fitting approximation.[28,29] The geometry optimization is the most time-consuming process, and the calculation has to be done for each point on the reaction surface.

The strong interactions inherent in aqueous ionic chemistry require a more accurate solvent model, and the results of the BP86/COSMO calculations were then corrected by the COSMO-RS method of Klamt et al.[30−32] This correction accounts for structural and thermal variations in the intermolecular forces and predicts relative aqueous ion free energies within roughly 1 kcal mol^{-1} of the experimental values. The COSMO-RS corrections were applied at 298 and 310 K.

Table 1. Summary of Parameters in Partial Optimizations

reaction	reaction coordinate	min	max	number of points
1a	r_{AB}	1.1 Å	1.9 Å	41
1b	r_{AB}	1.1 Å	1.8 Å	45
2	r_{ab}	1.4 Å	2.7 Å	47
3	$r_{\alpha\beta}$	1.0 Å	2.4 Å	18

In this process, the transition state of the reaction cannot be located by analytical methods, because the COSMO-RS correction is applied in a computational step distinct from the geometry optimization. Hence, for each reaction, one bond length was selected as the effective reaction coordinate (r), and the free energy was calculated at discrete points along this coordinate. At each of these points, corresponding to some fixed value of the reaction coordinate, the remaining geometric parameters were optimized to minimize the vibrational potential energy. At least 40 points along this coordinate were calculated for each of reactions 1a, 1b, and 2. Reaction 3, as explained below, is less sensitive to the solvent model and was comparably modeled with only 18 points. The parameters for these series are summarized in Table 1. The r values generally range from a minimum near the corresponding chemical bond length and a maximum near 2 Å. The maximum r value is limited by the COSMO solvent model's description of the molecular system using a single solvent cavity, an increasingly inappropriate representation as r exceeds the effective diameter of the solvent molecule.

For reactions 1a and 1b, the bond length between O^A and H^B (r_{AB}) was initially chosen as the effective reaction coordinate, anticipating that both of the other bonds heavily featured in the reaction, O^CH^B and O^AC^D, would be strongly coupled to r_{AB}. The reaction coordinate for 2 was r_{ab}, the separation between the bridging carbon 8a on the chromanol ring and the oxygen from a neighboring water molecule. In the unimolecular reaction 3, the reaction coordinate was the distance $r_{\alpha\beta}$ between the chromanol oxygen atom and a hydrogen atom transferred as the ring opens. In each case, the reaction coordinate decreases, approaching a chemical bond length, as the reaction progresses.

The free energies reported in this work are based on a combination of the electronic energy calculations and the COSMO-RS solvent free energy correction. This sum neglects the entropic contribution to the free energy from the internal vibrations of the molecule, and in some cases, this contribution is critical. The reaction free energies ΔG_v° include approximate vibrational contributions, obtained by carrying out a distinct harmonic frequency analysis on the isolated reactants and products of each reaction. We have not attempted to carry out the vibrational analysis on the transition states because, for reactions 1 and 2, the transition states are stationary points on the reaction diagram only after the COSMO-RS correction is applied. The typical harmonic analysis is invalid at nonstationary point geometries, and a numerical frequency calculation including the COSMO-RS correction is prohibitive.

The Gaussian 03 program suite[33] was used for all geometry optimizations, including the COSMO solvent model, and for the vibrational calculations. These calculations were carried out on a variety of Intel Pentium-based computers running Linux. With processor speeds of roughly 1−2.5 GHz, the geometry optimizations typically required 4−5 days of real time on a single dedicated computer node. Despite the simplified structure of our model tocopherol, convergence of the calculations at geometries near the transition state was often slowed by the difficulty in optimizing relative positions of weakly interacting chemical components. In particular, the potential energy is fairly insensitive to the orientation of the H_3O^+ in reaction 1. Minimum memory requirements were roughly 25 MW. The COSMO-RS calculations were carried out with the program COSMOtherm running on an Intel-processor PC.[34]

Table 2. Computed Free Energies and Enthalpies (kcal mol^{-1}), Entropies (cal K^{-1} mol^{-1}), and Equilibrium Constants

reaction	1a	1b	2	3
ΔG° (298 K)	−3.29	−1.73	−0.40	−5.67
ΔG_v° (298 K)	−4.72	−1.60	3.80	−11.13
K (298 K)	$2.90 \cdot 10^3$	14.9	1.63×10^{-3}	$6.37 \cdot 10^7$
$\Delta G^\ddagger$ (298 K)		1.82	2.80	21.50
ΔG° (310 K)	−4.12	−2.58	0.42	−5.67
ΔG_v° (310 K)	−5.44	−2.21	4.60	−11.13
K (310 K)	$6.84 \cdot 10^3$	36.1	5.71×10^{-4}	$3.18 \cdot 10^7$
$\Delta G^\ddagger$ (310 K)		1.44	2.90	21.50
ΔH°	17.9	18.2	−16.6	−5.67
ΔS°	69.2	66.5	−68.3	0.00

3. Results

The standard Gibbs free energies (G°) calculated from COSMOtherm at both 298 and 310 K were plotted against the selected reaction coordinate. From these curves, the free energies of activation were estimated. Overall ΔG°, ΔH°, and ΔS° values were calculated instead from the COSMO-RS-corrected free energies of the isolated reactants and products, with equilibrium constants approximated by the equilibrium relation

$$K = \exp[-\Delta G^\circ/(RT)] \tag{1}$$

and enthalpies and entropies of reaction estimated from the differences between results at 298 and 310 K. These results are summarized for all of the reactions in Table 2. The following sections discuss characteristics specific to the individual reactions.

Reaction 1. In reaction 1a, α-TOOH reacts with a H_3O^+ molecule to form α-T^+, hydroperoxide (H_2O_2), and a water molecule, while for reaction 1b, α-TOOMe reacts with a H_3O^+ to form α-T^+ and methylperoxide (HOOMe) with H_2O. In the lowest energy structures, both α-TOOH and α-TOOMe are bicyclic with the OOH/OOMe group perpendicular to the chromanol ring. For α-TOOH, the two oxygen atoms have partial negative charges, and the OOC group in the methylperoxyl are partially negative for α-TOOMe. In both reactants, the π-conjugated ring is slightly positive and the charge is broadly distributed around the ring. Within the peroxyl group, O^A has the most negative formal charge and is the likeliest site of attack by the hydronium to initiate the reaction.

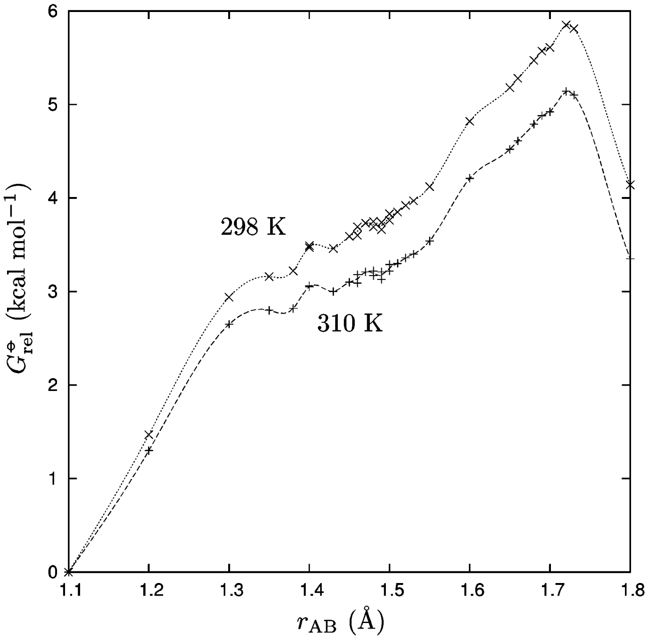


Figure 2. Reaction 1a: Relative standard Gibbs free energy versus r_{AB}.

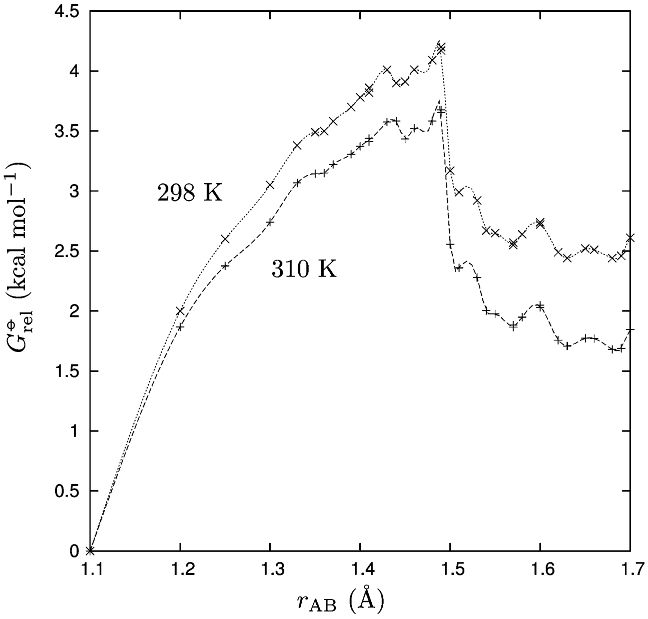


Figure 3. Reaction 1b: Relative standard Gibbs free energy as a function of O^AH^B separation.

The relative free energies are plotted in Figures 2 and 3. In these and subsequent free energy graphs, the combined energy of the reactants at maximum separation is chosen as the reference point. Note that the direction of the reaction in these graphs proceeds from right (large *r*) to left. The small variations in the curves are consistent with expected variations in the extent to which the convergence criteria are satisfied. In reaction 1a, the uncorrected COSMO calculations predict a fairly flat energy profile as r_{AB} initially decreases in these calculations. With the addition of the COSMO-RS correction, a shallow local minimum appears at $r_{\text{AB}} = 1.76$ Å. Here, the system has a free energy of −0.41 or −0.48 kcal/mol (at 298 and 310 K, respectively). The curve then suddenly jumps in free energy by roughly 2.5 kcal/mol. The discontinuity in Figure 2 occurs between r_{AB} values of 1.74 and 1.76 Å and is a common consequence of parametrizing the reaction coordinate using a single bond length. At this same point, the C−O bond length r_{AD} increases suddenly by 0.4 Å, indicating that, near the transition state, the true reaction coordinate depends heavily on this bond length. A survey of the free energy at 12 points as a function of r_{AB} and r_{AD} near this geometry confirms that the transition state for the reaction is at the geometry corresponding to $r_{\text{AB}} = 1.74$ Å in Figure 2. Similar difficulties involving the selection of the reaction coordinate do not appear in the other reactions studied.

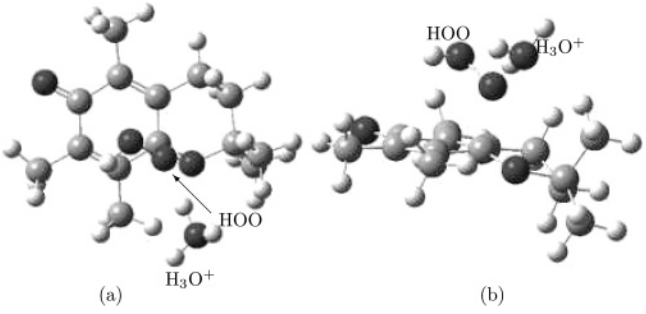


Figure 4. Approximate reaction 1a transition state geometry, viewed (a) from above the chromanol plane and (b) with the chromanol edge-on.

The transition state of reaction 1b is identifiable at $r_{\text{AB}} =$ 1.435 Å, with a relative free energy of 1.83 kcal/mol at 298 K and 1.44 kcal/mol at 310 K. Both reactions are mildly spontaneous, with negative ΔG° values of magnitudes less than 5 kcal/mol.

For both reactions 1a and 1b, Table 2 shows that $\Delta G^{\circ} < 0$ at both temperatures, with the equilibria lying strongly on the product side. However, a comparison of Figures 2 and 3 shows that reaction 1a has a greater drop in free energy than reaction 1b. This may be due to the hydrogen bond being more likely to form between H_2O_2 and water than between CH_3OOH and water. Both reactions have similar products, differing only in the nature of the peroxide. Because the methyl group is more electronegative than the hydrogen, the methyl group donates electrons to the oxygen less effectively, and this causes the oxygen to be less negatively charged. The charge on O^A is −0.31 *e* at $r_{\text{AB}} = 1.1$ Å in reaction 1a, slightly higher than the corresponding value of −0.28 for reaction 1b, while the charges at H^B are the same in both reactions. This results in a reduced hydrogen bond strength for the methylperoxide and, therefore, less stabilization in an aqueous solvent. Hydrogen bonds usually contribute to stronger intermolecular forces, or a lower G° in this case.

The electron-donating effect from the methyl group is also reflected in the bond strength of the O^AC^D bond, which requires a smaller r_{AB} value before breaking. In the transition state, bonds are breaking between O^A and C^D and between H^B and O^C, while a new bond is being formed between O^A and H^B. The transition state geometry for reaction 1a is shown in Figure 4 and lies at approximately $r_{\text{AB}} = 1.74$ Å. For reaction 1b, the transition state is at $r_{\text{AB}} = 1.47$ Å.

Reaction 1b thermodynamically favors the product side. This is driven by the formation of α-T^+, which stabilizes the charge in the π-conjugated system, and the increase in entropy by the increase of the number of moles in the product. The effects of the temperature difference are small, with $G°$ values shifting by less than 1 kcal/mol between reactions at both temperatures. Without the COSMO-RS correction, a curve with no barrier is obtained. The COSMO-RS correction has its greatest effect in these calculations by stabilizing the H_3O^+ moiety. In contrast to the other principal cation, α-T$^+$, for which the ring's π-conjugated system contributes charge delocalization, the charge on H_3O^+ is strongly localized and gains most from a more adaptable solvent model.

Applying the conditions pH = 3.6 and 298 K from Liebler et al.'s study,[18] the rate constant can in principle be calculated from the free energy of activation, $\Delta G^\ddagger$, using the Eyring equation from elementary transition state theory. The Eyring equation for a bimolecular reaction has a general form

$$k = \frac{k_B T}{hC} \exp[-\Delta G^\ddagger/(RT)]$$

where C is 55.6 mol L^{-1} for aqueous reactions. To put this in the form of the first-order rate constant quoted by Liebler et al., this value is multiplied by the H_3O^+ concentration. The first-order rate constant of reaction 1b at 298 K, k'_{1b}, is then calculated to be 8.0×10^7 min^{-1}. This value greatly exceeds the diffusion-limited collision rate, on the of order 1 min^{-1}, indicating that Liebler et al.'s value of 0.296 min^{-1} arises fully from diffusion-limited kinetics.

Reaction 2. In reaction 2, α-T$^+$ reacts with two water molecules to form α-TOH and an H_3O^+. In the optimized α-TOH structure, the attached OH group is perpendicular to the π-conjugated ring, while the other side of the chromanol ring is bent. As was the case for reaction 1b, the COSMO-RS correction is again critical to locating a transition state. In the absence of this correction, the free energy climbs without barrier from reactants to products, as shown for the COSMO curve in Figure 5, and indeed the product instability would then be predicted to halt the reaction at this step. Once the correction is added, the relative free energy of the products drops by 9 kcal mol^{-1}, again driven by the differential stabilization of H_3O^+ relative to α-T$^+$. The transition state appears at $r_{ab} \approx 1.9$ Å, as the oxygen atom of one water molecule moves toward the C^a of the ring while a hydrogen rotates to hydrogen-bond to an adjacent water molecule (H^c−O^d in Scheme 1), with the geometry shown in Figure 6. The O^{b}H^c bond lengthens significantly at $r_{ab} = 1.6$ Å, indicating that the proton shift of H^c from O^b to O^d occurs before the transition state is reached. The optimized value r_{ab} of α-TOH is 1.45 Å, leading to a local minimum in Figure 5 as this distance approaches 1.40 Å.

One could expect the charge in α-T$^+$ to be delocalized along the π-conjugated system of the chromanol ring. However, the charges around the chromanol of α-T$^+$ predicted by the electronic structure calculations, listed in Table 3, predict the positive charge to be concentrated at position 8a in Figure 1. Therefore, it is reasonable to assume that the electronegative O^b approaches this site to initialize the reaction. As shown in Figure 7, the charge on the ring system drops smoothly throughout the reaction, as the charge transfers briefly to the OH group (peaking at $r_{ab} = 1.6$ Å) before its redistribution ultimately to the hydronium group.

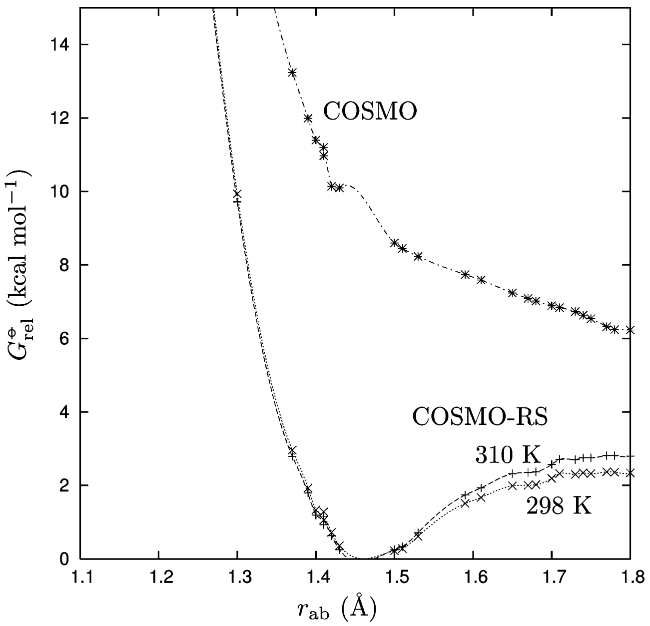


Figure 5. Reaction 2: Relative standard Gibbs free energy as a function of C^{a}O^b separation.

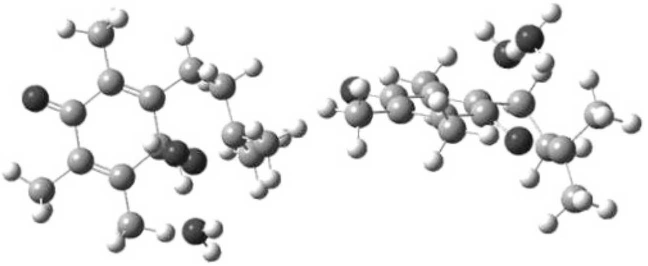


Figure 6. Approximate reaction 2 transition state geometry, viewed from two angles.

Table 3. Relative Charges around the Chromanol Ring in α-T$^+$

atom	relative charge	atom	relative charge
O1	−0.07	C2	−0.02
C3	0.12	C4	0.13
C5	−0.03	C6	0.21
C7	0.06	C8	−0.14
C8a	0.36	C8b	0.08

Unlike reaction 1, the results in Table 2 show that reaction 2 has $\Delta G° > 0$ at both temperatures, and the predicted equilibrium for this hydration reaction lies strongly on the reactant side. The entropy decrease, driven by the loss of free water, overwhelms the exothermicity which results from the C−O bond formation. However, for both reactions 1 and 2, the free energy of activation $\Delta G^\ddagger$ is less than 3 kcal mol^{-1}, indicating that the kinetics of these reactions will be determined instead by the diffusion-limited collision rates in solution, in agreement with Liebler et al.'s qualitative observations.

Reaction 3. The relative free energies for reaction 3 are plotted in Figure 8 as a function of the H$^\alpha$O$^\beta$ separation. In

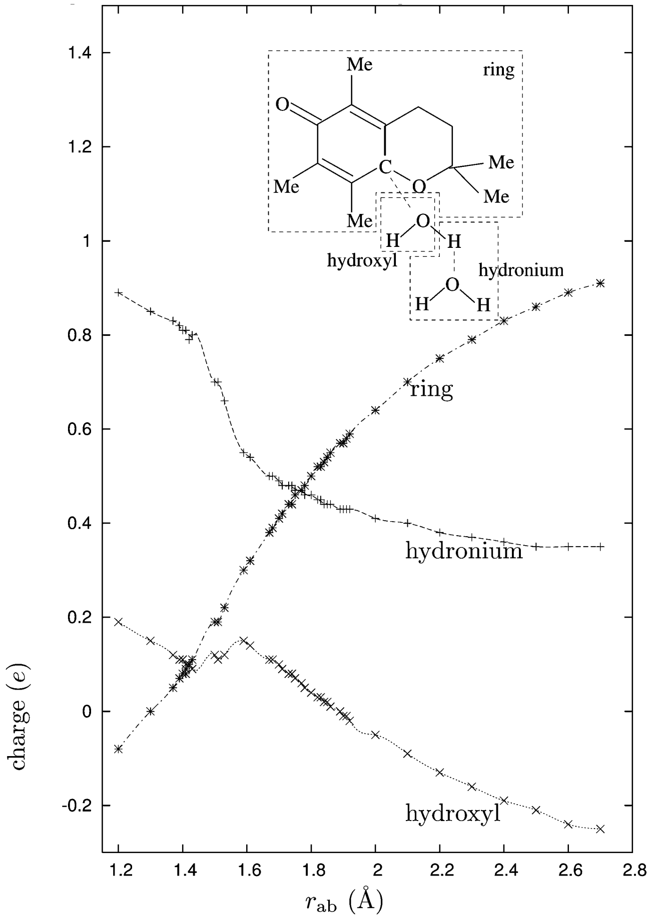


Figure 7. Reaction 2: Charges on various species.

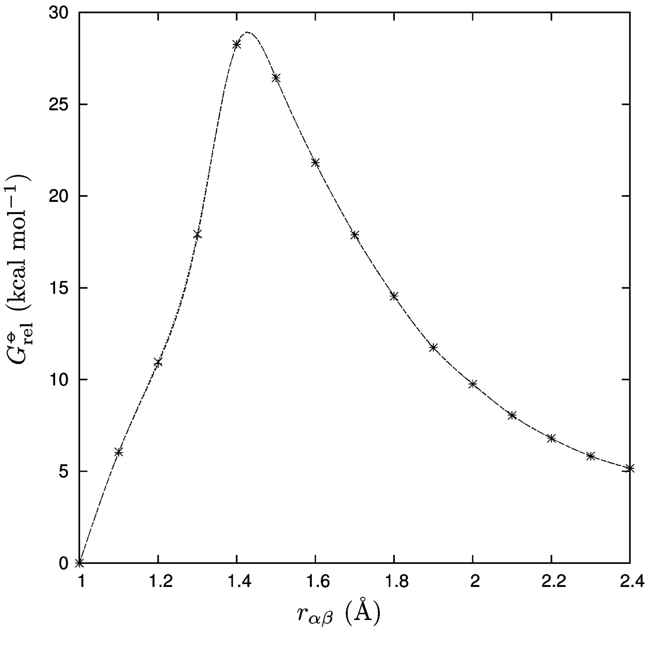


Figure 8. Reaction 3: Relative standard Gibbs free energy as a function of $H^{\alpha}O^{\beta}$ separation. Results at both 298 and 310 K are graphed but not visibly distinguishable at this scale.

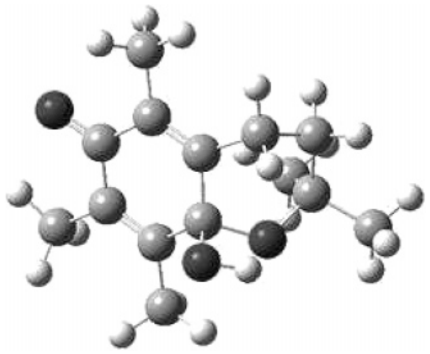


Figure 9. Approximate reaction 3 transition state geometry.

reaction 3, α-TOH rearranges to α-TQ by opening the ring. During the rearrangement, bonds are broken between O^{δ} and H^{α} and between C^{γ} and O^{β}. At the same time, a bond forms between O^{β} and H^{α} and a double bond between C^{γ} and O^{δ}. The transition state structure, with H^{α} pointing toward the chromanol O^{δ}, is shown in Figure 9. At this geometry, where $r_{\alpha\beta} \approx 1.4$ Å, the optimized values of $r_{\gamma\beta}$ and $r_{\alpha\delta}$ increase sharply, indicating that those bonds are breaking. The long tail attached to the quinone is then free to move and rotate away from the remaining monocycle. At the same time, the $r_{\gamma\delta}$ drops from approximately 1.44 Å to 1.25 Å, which corresponds to the change from a single bond to a double bond.

Table 2 indicates that the reaction equilibrium lies strongly on the product side. Unlike the previous reactions, the COSMO-RS correction exerts very little effect on the free energies for reaction 3 at either temperature; the differences with and without the correction average to only 0.20 kcal/mol across the sampled points. The vibrational correction to the free energy is also much greater for reaction 3 than for the other two reactions, as expected from the substantial increase in vibrational entropy associated with the ring opening.

Employing the Eyring equation for a unimolecular reaction, the rate constant of reaction 3, k_3, is determined to be 0.056 min^{-1}. A comparison with the value of 0.046 min^{-1} from Liebler et al. for the corresponding reaction shows that the results are compatible. This encouraging agreement is likely to be in some part fortuitous, because no vibrational correction has been applied to the transition state. However, the Eyring equation should be more successful here than in the higher-order reaction 1 and reaction 2, because fewer assumptions are necessary in the unimolecular case. In addition, the insensitivity of the free energy curve to solvent corrections indicates that the neglect of solvent interactions (such as diffusion and cage effects) in the Eyring equation should be less of a problem for reaction 3.

4. Conclusion

Neglecting vibrational corrections, the formation of α-TQ from α-TOOH is predicted to have an overall equilibrium constant of 3.01×10^{8} at 298 K and 1.24×10^{8} at 310 K. When the reactant in step 1 is replaced with α-$TOOCH_3$, the overall equilibrium constants are estimated to be 1.54×10^{6} and 6.55×10^{5} for temperatures 298 and 310 K, respectively. These results indicate that the overall reaction of α-TOOR to α-TQ is thermodynamically favorable, lying strongly on the product side in the absence of additional reactants such as vitamin C.

Liebler et al. measure the formation of α-TQ from α-TOOR as having an effective first-order reaction rate constant $k_{\text{overall}} = 0.036\ \text{min}^{-1}$.[18] The formation of α-TQ takes place via three sequential reactions:

$$\alpha\text{-TOOR} + H_3O^+ \xrightarrow{k_1} \alpha\text{-T}^+ + \text{ROOH} \quad (2)$$

in which the effective first-order rate constant k'_1 was

$$\alpha\text{-T}^+ + 2H_2O \xrightarrow{k_2} \alpha\text{-TOH} + H_3O^+ \quad (3)$$

$$\alpha\text{-TOH} \xrightarrow{k_3} \alpha\text{-TQ} \quad (4)$$

estimated to be $0.296\ \text{min}^{-1}$. The rate of hydrolysis of α-T^+ to α-TOH was too fast to determine, whereas the rate of arrangement of α-TOH to α-TQ, k_3, was found to be 0.046 min^{-1}.

Because both reactions 1 and 2 have small free energy barriers of around 4 kcal/mol, they are both kinetically facile at these temperatures. Of these, the effective rate constant k'_2 is predicted to be greater than k'_1 because, in the pseudo-first-order limit, k'_2 is proportional to the H_2O concentration and k'_1 to the much lower H_3O^+ concentration:

$$k'_2 \approx [H_2O]k_2;\ k'_1 \approx [H_3O^+]k_1$$

This result, coupled with the relatively high activation energy barrier for reaction 3, predicts the ordering of the rate constants to be $k_3 < k'_1 < k'_2$. This ordering is consistent with the results from Liebler et al.'s work, in which the rearrangement of α-TOH to α-TQ (reaction 3) is the rate-determining step of the overall reaction, and reaction 2 is the fast step.

The COSMO-RS solvent correction has proven to be a significant factor in predicting properties of this reaction pathway. The relative instability of H_3O^+ prior to the COSMO-RS correction would otherwise appear to halt the reaction at step 2. At present, this is only a qualitative result, because the activation barriers in the ion-mediated steps are not determined experimentally. Future work will attempt to assess the quantitative success of these methods in modeling ion-mediated biochemical systems.

Acknowledgment. The authors thank Dr. Andreas Klamt for kindly suppling the COSMOtherm license during the course of this project and for comments on the manuscript, William Root for extensive use of his computing facilities, and Prof. Pete Kovacic for valuable guidance. The authors are grateful to Dong Xu and Olga Noris for technical support. Funding was provided by the Blasker Fund of the San Diego Foundation and Grant CHE-0216563 of the National Science Foundation.

Supporting Information Available: Numerical free energy values and the geometries of the reaction endpoint and transition state species. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Traber, M. G.; Kayden, H. J. *Am. J. Clin. Nutr.* **1987**, *46*, 488−495.

(2) Buettner, G. R. *Arch. Biochem. Biophys.* **1993**, *300*, 535−543.

(3) Bisby, R. H.; Ahmed, S. *Free Radical Biol. Med.* **1989**, *6*, 231−239.

(4) Wang, X.; Quinn, P. J. *Mol. Membr. Biol.* **2000**, *17*, 143−156.

(5) Babbs, C. F.; Steiner, M. G. *Free Radical Biol. Med.* **1990**, *8*, 471−485.

(6) Nagaoka, S.; Okauchi, Y.; Urano, S.; Nagashima, U.; Mukai, K. *J. Am. Chem. Soc.* **1990**, *112*, 8921−8924.

(7) Nagaoka, S.; Kuranaka, A.; Tsuboi, H.; Nagashima, U.; Mukai, K. *J. Phys. Chem.* **1992**, *96*, 2754−2761.

(8) Nagaoka, S.; Sawada, K.; Fukumoto, Y.; Nagashima, U.; Katsumata, S.; Mukai, K. *J. Phys. Chem.* **1992**, *96*, 6663−6668.

(9) Nagaoka, S.; Mukai, K.; Itoh, T.; Katsumata, S. *J. Phys. Chem.* **1992**, *96*, 8184−8187.

(10) Wright, J. S.; Johnson, E. R.; DiLabio, G. A. *J. Am. Chem. Soc.* **2001**, *123*, 1173−1183.

(11) Leopoldini, M.; Marino, T.; Russo, N.; Toscano, M. *J. Phys. Chem. A* **2004**, *108*, 4916−4922.

(12) Rhodes, C. J.; Tran, T. T.; Denton, P.; Morris, H. *Free Radical Res.* **2002**, *36*, 89−90.

(13) Lucarini, M.; Mugnaini, V.; Pedulli, G. F.; Guerra, M. *J. Am. Chem. Soc.* **2003**, *125*, 8318−8329.

(14) Setiadi, D. H.; Chass, G. A.; Koo, J. C. P.; Penke, B.; Csizmadia, I. G. *THEOCHEM* **2003**, *666*−*667*, 439−443.

(15) Espinosa-Garcia, J. *Chem. Phys. Lett.* **2004**, *388*, 274−278.

(16) Rhodes, C. J.; Tran, T. T.; Morris, H. *Spectrochim. Acta, Part A* **2004**, *60*, 1401−1410.

(17) Liebler, D. C.; Burr, J. A. *Biochemistry* **1992**, *31*, 8278−8284.

(18) Liebler, D. C.; Kaysen, K. L.; Kennedy, T. A. *Biochemistry* **1989**, *28*, 9772−9977.

(19) Liebler, D. C.; Baker, P. F.; Kaysen, K. L. *J. Am. Chem. Soc.* **1990**, *112*, 6995−7000.

(20) Lee, S. B.; Lin, C. Y.; Gill, P. M. W.; Webster, R. D. *J. Org. Chem.* **2005**, *70*, 10466−10473.

(21) Liebler, D. C.; Burr, J. A. *Lipids* **2000**, *35*, 1045−1047.

(22) Kohar, I.; Baca, M.; Suana, C.; Stocker, R.; Southwell-Keely, P. T. *Free Radical Biol. Med.* **1995**, *19*, 197−207.

(23) Siegel, D.; Bolton, E. M.; Burr, J. A.; Liebler, D. C.; Ross, D. *Mol. Pharmacol.* **1997**, *52*, 300−305.

(24) Shi, H.; Nogichi, N.; Niki, E. *Free Radical Biol. Med.* **1999**, *27*, 334−346.

(25) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(26) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822−8824.

(27) Klamt, A.; Schürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.

(28) Dunning, T. H. *J. Chem. Phys.* **1971**, *55*, 716−723.

(29) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560−571.

(30) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

(31) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(32) Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. *J. Phys. Chem. A.* **2003**, *107*, 9380−9386.

(33) 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.: Wallingford, CT, 2004.

(34) Klamt, A.; Eckert, F. *COSMOtherm*, version 2.1; COSMOlogic GmbH & Co., KG: Leverkusen, Germany, 2005.

CT600161W

J. Chem. Theory Comput. **2006,** *2,* 1403−1415 **1403**

JCTC Journal of Chemical Theory and Computation

Ligand Field Effects on the Aqueous Ru(III)/Ru(II) Redox Couple from an All-Atom Density Functional Theory Perspective

Regla Ayala and Michiel Sprik*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Received May 12, 2006

Abstract: The $[RuCl_6]^{4-}$(aq) → $[RuCl_6]^{3-}$(aq) + e^- and $[Ru(CN)_6]^{4-}$(aq) → $[Ru(CN)_6]^{3-}$(aq) + e^- half redox reactions are investigated using density functional based ab initio molecular dynamics methods. The aim is to understand at a microscopic level how the difference in π-bonding of these ligands is reflected in the redox chemistry. To this end, we have computed the redox and reorganization free energies using a method derived from the Marcus theory of electron transfer. The resulting estimate of the free energy change of the full redox reaction between the two coordination complexes is compared to experiment. Our findings indicate that ligand character has an important effect on the vertical ionization chemistry but less on the relaxation of the system after removal or addition of electrons. This enables us to correlate the redox free energies with the HOMO energy levels of the combined solute + solvent system and analyze the redox chemistry in terms of the corresponding energy level diagram.

1. Introduction

The response to removal or excitation of electrons of molecules in solution is determined ultimately by changes in total energy. However the microscopic understanding of such a process has been largely based on models of effective one-electron states of the solute species. The ligand field theory for transition-metal coordination complexes is a good illustration of the power of the orbital picture. Many of the chemical and physical properties of these systems can be rationalized in terms of energy levels and symmetries of the one-electron orbitals. The particular question that will concern us here is the effect of a change in the chemical nature of the ligands. Among various model systems used to study this question, ruthenium(II) and ruthenium(III) coordination complexes occupy a special place. The reason for this popularity is the huge variety of stable low spin complexes formed by these transition-metal ions and the corresponding wide range of accessible optical and electrochemical properties.[1−4]

The search for correlations between molecular orbitals (MOs) and electrochemical and optical characteristics of transition metal complexes was initially based on semiempirical electronic structure methods (see for example ref 4 and references therein). The smaller coordination complexes can now also be treated using self-consistent field methods. In particular density functional theory (DFT) has proven to be an efficient and (comparatively) reliable tool for electronic structure computation of transition metal complexes.[5,6] The effect of the solvent is usually modeled by a dielectric continuum[6−8] or cancels when special properties of similar complexes are compared (for example pK_a's of aqueous cations).[9] DFT based ab initio molecular dynamics (MD) methods[10] should "in principle" allow us to take a more fundamental approach and treat the redox active solute and solvent at the same level of electronic structure theory. Recent studies of aqueous trivalent aqua cations (Al^{3+}[11], Y^{3+}[12], Fe^{3+}[13],) indicate that this integral "ab initio" MD approach is capable of giving a realistic description of the hydration structure and dynamics of these species. Also the reactivity of the Fe^{3+} aqua-ion toward hydrogen peroxide has already been investigated.[14] However, in the present context such a scheme is only meaningful provided we are also able to compute redox potentials for these model systems. We have recently shown that for a subset of redox

---

* Corresponding author e-mail: ms284@cam.ac.uk.

10.1021/ct600169e CCC: $33.50

reactions it is in fact possible to obtain reasonably accurate estimates of the reaction free energy change.[15−18] Our method is an ab initio MD implementation of the method originally developed by Warshel and co-workers[19−21] for molecular dynamics simulation of electron transfer in solution. The aqueous ruthenium cation, being a textbook example of a redox system in the linear response regime where Marcus theory applies, has played a crucial role in the development and validation of our method[15,17] (for a related study of the electronic absorption of this system see ref 22).

In the present work we apply the same ab initio MD approach to two aqueous ruthenium coordination complexes, $[Ru(CN)_6]^{4-}$ and $[RuCl_6]^{4-}$. These complexes can be regarded as a minimal model for the study of the effect of exchange of a ligand with π-acceptor (CN^-) for a ligand with π-donor character (Cl^-) (see for example refs 23 and 24). More specifically, the objective is to investigate the nature of the, in principle, extended DFT one-electron (Kohn−Sham) states produced by the ab initio MD calculation and interpret or, to use a spectroscopic analogy, assign these states in terms of ligand field theory. The first question is how the traditional picture of π-bonding is reflected in the actual Kohn−Sham orbitals of the extended solute + solvent system obtained from the ab initio MD trajectory. The next step is to correlate the orbital structure with the redox and reorganization free energies as computed for the $[RuCl_6]^{4-}(aq) \rightarrow [RuCl_6]^{3-}(aq) + e^-$ and $[Ru(CN)_6]^{4-}(aq) \rightarrow [Ru(CN)_6]^{3-} + e^-$ half reactions. As could be expected in view of the success of ligand field theory, our calculations verify that the occupied orbitals of the combined solute + solvent system conform to the ligand field theory picture. More surprising is that, for this particular system, the agreement extends even to a semiquantitative level, in that the difference in energy between the highest occupied molecular orbital (HOMO) of $[Ru(CN)_6]^{4-}$ and $[RuCl_6]^{4-}$ is found to be consistent with the computed reaction free energy change of the full $[RuCl_6]^{4-} + [Ru(CN)_6]^{3-} \rightarrow [RuCl_6]^{3-} + [Ru(CN)_6]^{4-}$ redox reaction. The final test will be to compare this value to experimental estimates.

The organization of the paper is as follows. Section 2 briefly summarizes the relevant ligand field theory and the Marcus theory based method for the computation of free energies. Technical details and parameters defining the density functional MD technique are also given in this section. Results are presented and discussed in section 3. Section 4 contains some concluding remarks putting this work in the more wider context of the computational study of the redox reactions.

2. Theory and Method

2.1. Ligand Field Theory. Coordination complexes of ruthenium in oxidation states II and III are very popular with both experimentalists and theorists. One reason is the preference of ruthenium for low spin octahedral structures, such as the $[RuCl_6]^{3/4\,-}$ and $[Ru(CN)_6]^{3/4\,-}$ anions studied here. The central metal ion in $Ru(II)L_6$ has formally a $4d^6$ configuration, which under low spin conditions, leads to full occupation of the t_{2g} levels of the complex (t_{2g}^6) and therefore to singlet spin multiplicity. Similarly, $Ru(III)L_6$ compounds with one electron less (t_{2g}^5) are doublets. Another attractive feature is that most of these complexes are textbook examples of redox active species following outer-sphere redox kinetics as described by Marcus theory.[25−30] Perhaps the most well-known are Ru hexa-amines, $[Ru(NH_3)_6]^{2/3+}$, with one or more ammonia ligands replaced by aromatic molecules with nitrogen heteroatoms (pyridine, pyrazine, etc).[2] The well-behaved Marcus character of these complexes was also the reason we chose aqueous Ru hexahydrate ($[(Ru(H_2O)_6]^{2/3+}$) as the model system for the development and validation of the Marcus theory based ab initio MD methodology for the computation of redox potentials.[15,17]

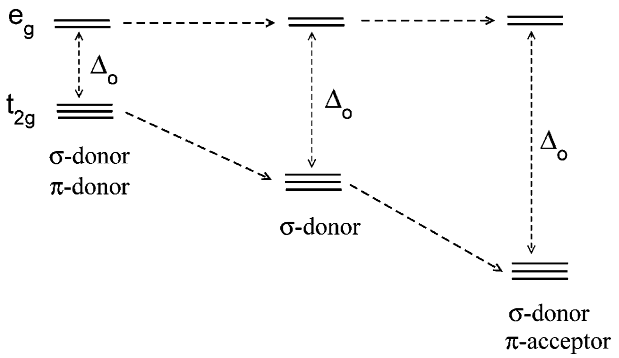


Figure 1. Schematic MO diagram of the response of the ligand field splitting Δ_o of an octahedral coordination complex ML_6 to change the π-bonding character of ligands L.

The focus of the present application is less on the central metal atom, which is chosen for convenience, but on the Cl^- and CN^- ligands. Cl^- and CN^- are at opposite ends of the spectrochemical series giving rise to a different splitting of the manifold of *d*-orbitals. The ligand splitting due to 6-fold CN^- coordination can easily be in the 3 eV range or more, while the splitting for Cl^- is usually about half this value.[23] The standard explanation is that Cl^- is a π-donor ligand, while CN^- acts as a π-acceptor. The difference is in the mixing with the filled t_{2g} manifold (see Figure 1). These orbitals are not participating in the σ-bonding and retain predominantly metallic *d* orbital character. In the case of π donation from a lower lying ligand π orbital (a 3*p* lone pair for Cl^-) the t_{2g} orbitals acquire some antibonding character and are pushed up. This closes the gap with the (unoccupied) σ-antibonding e_{2g} levels which are not affected by π-bonding. Conversely, in the case of π-acceptor ligands, the t_{2g} states mix with a higher lying empty π^* orbital, become (weakly) bonding, and are as a result pushed down opening the gap with the e_{2g} manifold.

The orbital picture sketched above makes no mention of the environment of the complex. When the complex is embedded in a dielectric continuum or is part of a static ordered crystal, this is presumably not necessary, at least at the qualitative or semiempirical level. However, what will be the effect of the multitude of electronic levels and the thermal fluctuations in solution? The very fact that ligand field theory is also used in solution chemistry suggests that the orbital picture still applies. We will see that this expectation is justified, although the thermal fluctuations do add an extra dimension to the problem which can even be exploited in calculations (see section 2.2). Moreover, the solvent effect, in a quantitative sense, can be substantial. This holds in particular for multiply charged anions such as $[Ru(CN)_6]^{3/4-}$ and $[RuCl_6]^{3/4-}$ which are unstable in the gas phase. At this point it should be mentioned that, while the Ru(II/III) hexacyanides are stable aqueous species, the Ru(II)/(III) chlorides are extremely labile as a result of rapid ligand−water exchange. In fact the hexachloride is not known in aqueous solution. However, on the ab initio MD time scale (≈10 ps) $[RuCl_6]^{3/4-}$ complexes can easily survive, and the use of this complex rather than a stable existing π-donor is a compromise imposed by the limitations in model system size (see also section 2.3). Fortunately for ruthenium complexes a comprehensive set of ligand additivity rules is available,[3,4] enabling us to make a reasonably accurate estimation of the hypothetical experimental reduction potential for the aqueous $[RuCl_6]^{3/4-}$ couple. This is what we will rely on in this computational study.

2.2. Free Energy Calculations. Marcus theory provides an ideal framework for the application of MD simulation to electron transfer (ET). Marcus theory[25−30] connects the ET driving force (reaction free energy change) and the free energy of activation with the diabatic total energy of the reacting species coupled to a solvent bath. The solvent is assumed to respond linearly to a change of charge of the solute. These assumptions lead to two overlapping quadratic free energy surfaces with equal curvature with respect to the reaction coordinate. The ET process occurs at the crossing point of the free energy surfaces. Warshel has converted this formulation in an efficient and elegant MD method for the study of ET reactions.[19,21] The key feature of the method is that it employs two potential energy surfaces: one potential energy surface (PES) for the reactant and one for the product system. The method has been applied in numerous model potential studies, either purely classical, with the potential energy surfaces derived from a force field model, or quantum with a Hamiltonian based on a empirical valence bond scheme.

The extension of the two surface approach to DFT should be in principle straightforward but is, in practice, seriously hampered by the well-known shortcomings in the DFT treatment of charge transfer. The implementation for half reactions $R \rightarrow O + e^-$, however, is less precarious from the DFT perspective, as we have shown in a series of recent papers.[15−18,31−34] The central quantity in this approach is the vertical ionization energy ΔE defined as the difference between the total energies $E_M(\mathbf{R}^N)$ of the oxidized (M = O) and reduced (M = R) system at fixed atomic configuration $\mathbf{R}^N$.

$$\Delta E(\mathbf{R}^N) = E_O(\mathbf{R}^N) - E_R(\mathbf{R}^N) \tag{1}$$

The key assumption of Marcus theory, in its microscopic formulation, is that the distribution of equilibrium fluctuations of ΔE in the reduced and oxidized states is Gaussian.[19] If this condition is satisfied, then the free energy change of oxidation

$$\Delta A = A_O - A_R \tag{2}$$

can be directly computed from the mean and variance of the vertical energy gap ΔE in oxidized and reduced states according to

$$\Delta A = \Delta E_O + \lambda \tag{3}$$

$$= \Delta E_R - \lambda \tag{4}$$

The MD estimates of ΔE_O and ΔE_R are computed by averaging the vertical energy gap of eq 1 over an equilibrium trajectory in the O respectively R state. λ in eqs 3 and 4 is obtained from the corresponding variance (second moment) $\sigma_M{}^2$, M = O, R of the gap fluctuations

$$\lambda = \frac{\sigma_O{}^2}{2k_BT} = \frac{\sigma_R{}^2}{2k_BT} \tag{5}$$

where k_B is the Boltzmann constant, and T is the temperature at which the MD simulations are carried out. The quantity λ can be equated (in the linear regime) with the free energy of reorganization (for more details of the development followed to obtain the above expressions see ref 16 and references therein). Linking the Marcus reorganization free energy to the variance of the gap fluctuations, eq 5 implies that this width is independent of the oxidation state. This equality is not an additional assumption but is a special property of Gaussian energy gap fluctuation.[16,19,35] In practice this symmetry is used as a criterion for the validity of the Marcus theory for a given system. It also allows us to eliminate the width in the expression for the oxidation free energy. Adding and subtracting eqs 3 and 4 gives

$$\Delta A = \frac{1}{2}(\Delta E_R + \Delta E_O) \tag{6}$$

$$\lambda = \frac{1}{2}(\Delta E_R - \Delta E_O) \tag{7}$$

It will be clear that the pairs of eqs 3, 4 and 6, 7, while mathematically equivalent, have a very different status from a MD point of view. A MD approach based on eqs 6, 7 requires two equilibrium trajectories, one on the oxidized PES ($E_O(\mathbf{R}^N)$ in eq 1) and one on the reduced PES ($E_R(\mathbf{R}^N)$). Equations 3 and 4 claim that the same information can be obtained from a simulation on just a single PES (either O or R), at the expense, however, of having to converge the second moment of the fluctuations of the energy gap (eq 5), which usually takes a significantly longer run than needed for an estimate of the expectation values ΔE_M. This is why the two-state scheme is preferred in practice. However, for reasons explained below, the free energy computations of the present study are based on the one-state scheme.

2.3. Model System Composition and Size. Model systems in the half reaction scheme as implemented in previous studies[15−18,31−34] consist of a periodically replicated cubic box containing 30−50 solvent molecules and a single redox active solute. No explicit counterions are included. Neutrality is restored by the background of a uniform charge distribution implicit in the Ewald treatment of long-range interactions in periodic systems. The same approach is also used in the present study. The length of the cubic cell is $L = 11.44$ Å. The cell contained one $[RuCl_6]^{4-}$ ($[Ru(CN)_6]^{4-}$) metal

Table 1: Reaction Free Energies in Units of eV of Four Model Redox Reactions Compared to Experiment[a]

redox reaction	solvent	ΔA(calc)	ΔG(exp)	ref
$Cu^{1+} + Ag^{2+} \rightarrow Cu^{2+} + Ag^{1+}$	water	−1.7	−1.82	31
$RuO_4^{2-} + MnO_4^{1-} \rightarrow RuO_4^{1-} + MnO_4^{2-}$	water	−0.3	−0.04	16
$TH^{\bullet+} + TTF \rightarrow TH + TTF^{\bullet+}$	acetonitrile	−0.9	−0.93	33
$DQ^{\bullet-} + BQ \rightarrow DQ + BQ^{\bullet-}$	methanol	−0.43	−0.46	34
$RuCN_6^{3-} + RuCl_6^{4-} \rightarrow RuCN_6^{4-} + RuCl_6^{3-}$	water	−1.4	−1.78[a]	tw

[a] The first two reactions involve transition-metal aqua ions. TH (thianthrene) and TTF (tetrathiafulvalene) are two organosulfur compounds which can be oxidized to stable radical cations. BQ (benzoquinone) and DQ (duroquinone) are small quinones forming radical anions. The final column gives the reference to the original papers. The last line compares the result obtained in this work (tw) to a hypothetical experimental value using the ligand additivity rules of ref 3 (see section 3.2.4).

Table 2: Vertical Ionization Energy ΔE (Eq 1) in eV of Vacuum $[RuCl_6]^{4-}$ and $[Ru(CN)_6]^{4-}$ Metal Complexes as a Function of the Plane Wave Cutoff (in Ry)[a]

cutoff	$[RuCl_6]^{4-}$	$[Ru(CN_6]^{4-}$
70	−9.536	−8.376
75	−9.541	−8.375
80	−9.546	−8.376
85	−9.548	−8.377
90	−9.548	−8.376
95	−9.546	−8.374

[a] The complexes have been stabilized by placing them at the center of a repulsive confining potential of cubic symmetry (eq 9). The t_{2g} (HOMO) manifold in this potential is at positive energy (see Figure 2) leading to negative ionization energy.

complex and 45 (48) water molecules. The number of water molecules was chosen to guarantee vanishing mean total pressure as determined from classical simulations using parametrized interaction potentials.

Clearly the finite size and periodic boundary effects in a system of such small dimensions will be huge. Interactions with periodic images and background scale with the square of the net charge of the system. For a net charge of −4 in the reduced state these energies are in the order of 10 eV. Rather than trying to correct for such large periodic boundary effects we regard them as part of the artificial electrochemical half cell defined by our model system. Reference potential and effective ionic strength in this cell have no experimental counterpart, and half reaction energies computed from eqs 3, 4 or eq 6 have, therefore, no direct experimental significance. Reference potentials, in principle, should cancel in full redox reactions. However free energies of full reactions obtained by subtracting half reaction energies are still subject to major uncertainties related to system size effects and can also not be compared to experimental reaction free energies.

The exception is a special class of isocoulombic reactions. In this type of redox reaction the species in reactant and product have the same charges. This can be expressed schematically as

$$X^m + Y^{m+1} \rightarrow X^{m+1} + Y^m \tag{8}$$

If the species X and Y also have approximately the same spatial dimension, they will look from a distance rather similar to the solvent and image charges and the long range errors largely cancel. While there must be fundamental reasons for this compensation, the best evidence we can offer at the moment is the accuracy of the results we have obtained so far with this approach. These results are summarized in Table 1. Agreement with experimental standard potentials is good. Discrepancies are in the 100 meV range except for the ruthenate/manganate reaction, where the error is 300 meV (note that this number is still at least an order of magnitude smaller than the contributions of boundary effects to the half reaction energies).

2.4. Electronic Structure and MD Method. The electronic structure calculation methods employed here are similar to the approach used in previous ab initio MD studies of redox reactions in aqueous solutions involving transition metal ions.[15−18,31,32] The calculations were carried out using the constant-volume ab initio molecular dynamics method (Car−Parrinello scheme),[10] as implemented in the CPMD code.[36] Kohn−Sham (KS) DFT calculations were performed within the BLYP[37,38] approximation to the exchange-correlation energy. The oxidized state ($[RuCl_6]^{3-}$ and $[Ru(CN)_6]^{3-}$) was treated in the local spin density approximation constrained to doublet spin multiplicity. Norm conserving pseudopotentials (PP) according to Troullier−Martins[39] previously tested for this type of systems[15−17,40] were employed. For the construction of the Ru ion PP the electron configuration [Kr] $4d^7$ of Ru^+ is taken as a reference state, and the 4*s*, 4*p*, and 4*d* electrons are treated as valence. The pseudization radii are 1.1 au for the *s* channel, 1.2 for *p*, and 1.24 for *d*. The PPs for C, N, O, and Cl are standard atom Troullier−Martins potentials with valence *s* and *p* electrons only. The are pseudization radii $r_s = r_p = 1.23$ for C, $r_s = r_p = 1.12$ for N, $r_s = r_p = 1.05$ for O, and $r_s = 1.57$, $r_p = 1.52$ for Cl (all in a.u.). The Kleinman−Bylander scheme[41] was used for the calculation of the nonlocal part of the O, Cl, C, and N potentials, whereas the transition metal atom was treated using Gauss−Hermite integration.

The KS orbitals were expanded in a plane wave (PW) basis set (Γ point only) with the kinetic energy cutoff of 70 Ry. Earlier calculations[15−17,40] on ionic solutions of ruthenium−metal complexes showed that this cutoff is sufficient for achieving a good convergence of energies and structural properties. Nevertheless, this value has been checked by computing the vertical ionization energy of the process $[ML_6]^{4-} \rightarrow [ML_6]^{3-}$ ($M \equiv Ru$ and $L \equiv Cl$ and CN) for the isolated (gas-phase) complex. These results are collected in Table 2 and confirm that a 70 Ry basis set is sufficiently large for the systems under study. The gas-phase calculations have been carried out in a cubic box of the same size as the MD cell used in the solution simulations (L=11.44 Å, see section 2.3). The interactions between periodic images and charge compensating homogeneous background were elimi-

Table 3: Structural Parameters of Isolated $[RuCl_6]^{4-}$ and $[Ru(CN)_6]^{4-}$ Complexes and in Aqueous Solution[a]

vacuum		solution	
$[RuCl_6]^{4-}$	$[Ru(CN)_6]^{4-}$	$[RuCl_6]^{4-}$	$[Ru(CN)_6]^{4-}$
Ru−Cl 2.63	Ru−C 2.06	Ru−Cl 2.52 (0.08)	Ru−C 2.05 (0.06)
	Ru−N 3.24	Ru−O 4.62	Ru−N 3.21 (0.06)
		Ru−H 3.71	N−H 1.8
		Cl−H 2.15	
Cl−Ru−Cl 180.	C−Ru−C 180.	Cl−Ru−Cl 174. (3.)	C−Ru−C 173. (3.)
	Ru−C−N 180.		Ru−C−N 171. (4.)

[a] Distances are in Å, angles in degrees, and uncertainties in parentheses.

nated using screening methods[42] developed for this purpose. Of course, removing all interaction with the environment is not possible in the case of multiply charged anions, since these species are unstable in a vacuum. However, because we are only interested in relative energy we have considerable freedom in the design of a convenient artificial stabilizing potential. We opted for an external potential of the form

$$V(r) = \frac{1}{2}\exp(-r) \tag{9}$$

applied at the vertices of a cube with the same size as the MD cell centered at the metal atom. This potential has the effect of confining the charge density preventing the electrons to escape. The geometry of the reduced complexes (the 4-anions) was optimized under the constraint of this potential. The resulting minimum energy structures of the $[RuL_6]^{4-}$ (L ≡ Cl or CN) metal complexes kept their octahedral symmetry with Ru−Cl, Ru−C, and Ru−N distances of 2.63, 2.06, and 3.24 Å, respectively. These and other geometry parameters are listed in Table 3. Again, these values cannot be directly compared to experimental crystal data or other theoretical estimations because they depend on our choice of external potential. We verified that the structure with CN^- bound to the ruthenium ion through the N atom is higher in energy than the C atom coordinated structure. The vertical ionization potentials of Table 2 have been evaluated for these equilibrium geometries.

The CPMD runs were carried out using a fictitious electronic mass of 500 au and a time step of 5 a.u. (0.1209 fs). Temperature was controlled by velocity scaling with a target temperature of 300 K. Every 5000 MD steps the wave function was reoptimized ("quenched") to the ground state, and the CP dynamics were restarted from the optimized wave function. This MD protocol is standard in Car−Parrinello dynamics. While adequate for the simulation of the reduced state (the solutions containing $[RuL_6]^{4-}$ L ≡ Cl, CN), we were unable to generate stable MD trajectories for the oxidized complexes using this scheme. The explanation of this instability is that the oxidized systems are open shell systems with an effectively vanishing HOMO−LUMO gap for the minority spin. The dynamical electronic optimization scheme (extended Lagrangian) used in Car−Parrinello simulation has difficulty dealing with such systems, diverging from the Born−Oppenheimer as time proceeds. This effectively prevented us from using the two-state method of eqs 6 and 7 for the estimation of the redox and reorganization free energies. These quantities have been computed instead from a trajectory of the (closed shell) reduced systems only using eqs 4 and 5.

Trajectories were around 8 ps long. Structures were saved every 10 steps for further analysis. The vertical energy gap ΔE was computed as a difference of the total energies of the $[RuL_6]^{3-}$ and $[RuL_6]^{4-}$ (L ≡ Cl, CN) complexes at the same solvent configuration generated by the ab initio MD simulation of the aqueous $[RuL_6]^{4-}$ system. At each of these configurations the total energies were computed using standard electronic optimization methods. Closed shell system energies were converged within an accuracy of 10^{-5} a.u. The convergence for the open shell system was somewhat less strict (3.5×10^{-5} a.u.).

3. Results and Discussion

3.1. Isolated Complex. The confining potential of eq 9, which was introduced in section 2.4 to test the convergence of the basis set, has the same cubic symmetry as the ligand field established by the octahedral coordination. While artificial, this system can be useful for a semiquantitative exploration of the effect of interchanging ligands in the absence of solvent. We are particularly interested in the effect on the splitting of the manifold of *d*-orbitals. The KS orbital energies in the equilibrium geometry as given in Table 3 are shown in Figure 2. Cubic symmetry labels are assigned according to the shape of the partial electron densities. From these MO diagrams, we can conclude that there is a direct correlation between the softness of the ligand and the splitting of the manifold of *d*-orbitals. We find a $e_{2g} - t_{2g}$ orbital energy gap of 0.9 eV for the Cl^- ligand and 2.3 eV for the CN^- ligand, giving a ratio of 2.6 (for further discussion see section 3.2.2).

Not only the gap between t_{2g} and e_{2g} levels in Figure 2 responds as predicted by ligand field theory but also the relative position of the HOMO levels (the t_{2g} states) reflects the difference in ligation by Cl^- and CN^-. π-acceptor bonds should be more stable than π-donor bonds, and indeed the HOMO of $[Ru(CN)_6]^{4-}$ is situated below the HOMO of $[RuCl_6]^{4-}$. Note that the value of the HOMO energy is positive ($\epsilon_{HOMO} > 0$). This is because the potential eq 9 stabilizing the intrinsically unstable anions is purely repulsive (particle in the box). This is also the explanation for the negative sign of vertical ionization potentials of Table 2. In fact the absolute values of the ionization potentials, $-IP = 9.5$ eV for $[RuCl_6]^{4-}$ and $-IP = 8.4$ eV for $[Ru(CN)_6]^{4-}$, differ only by about 1 eV from the HOMO energies of $\epsilon_{HOMO} = 10.1$ eV respectively $\epsilon_{HOMO} = 9.3$ eV in Figure 2. This

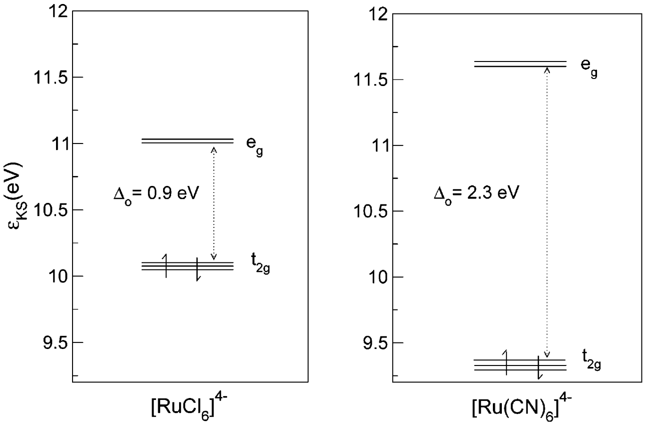


Figure 2. Kohn−Sham (KS) orbital energies of the $[RuCl_6]^{4-}$ (left) and $[Ru(CN)_6]^{4-}$ (right) complex in "vacuum". Orbital energies have been computed for optimized geometries under the influence of a confining external potential of the form $V = 1/2\ \exp(-r)$ applied at the vertices of a cube centered on the metal complexes (see section 2.4). Splitting of the t_{2g} and e_{2g} manifolds is schematic and shows the number of degenerate states. Similarly the up and down half arrows indicate full occupation of the manifold.

kind of agreement between ionization potential and HOMO energy is exceptional for the BLYP functions. While in exact DFT (minus) the vertical IP and the energy of the HOMO orbital should be rigorously identical for stable closed shell molecules,[43,44] approximate density functionals such as BLYP usually give $\epsilon_{HOMO} \approx -IP/2$ (see for example ref 45, we return to the issue of the comparison of energy levels and vertical ionization energies in section 3.2.5).

3.2. Aqueous Complex. *3.2.1. Structure and Solvation.* The radial distribution functions (RDFs) for the water−metal complexes are shown in Figure 3. As a first observation, we note that the metal water RDFs (Ru−O and Ru−H) are more structured for $[RuCl_6]^{4-}$ than for $[Ru(CN)_6]^{4-}$. In fact, while the Ru−O and Ru−H RDFs for the $[RuCl_6]^{4-}$ complex show pronounced first maxima at 4.62 and 3.71 Å, it is not easy to identify any clear maximum for the $[Ru(CN)_6]^{4-}$ complex. The ligand solvent correlations as probed by the (Cl, N)−H RDFs in both the $[RuCl_6]^{4-}$ and $[Ru(CN)_6]^{4-}$ complexes show the characteristic first peaks for hydrogen bonding of water to anions. The N−H peak is somewhat stronger and at closer distance, 1.8 Å, compared to Cl−H (2.15 Å) in accordance with the difference in atomic radius. All these geometrical parameters are summarized in Table 3. Analyzing the structure in solution in more detail we see that distortions from octahedral symmetry for $[RuCl_6]^{4-}$ are only minor. The average Ru−Cl distance in the $[RuCl_6]^{4-}$ complex is 2.52 ± 0.08 Å, and the average Cl−Ru−Cl angle for Cl^- ions in the trans position is 174 ± 3°. The average Ru−C and Ru−N distances in the $[Ru(CN)_6]^{4-}$ complex are 2.05 ± 0.05 and 3.21 ± 0.06 Å, respectively, and the average C−Ru−C angle is 173 ± 3°. The important difference between the two complexes are the extra degrees of freedom due to the dimeric structure of the CN^- anion. The Ru−C−N angle undergoes substantial fluctuations in solution (see Figure 4) with an average of 171 ± 9°. The Ru, C, and N atoms are therefore (on average) not aligned, which, considering the linear Ru−C−N geometry in a vacuum, presumably means that the bending optimizes hydrogen bonding.

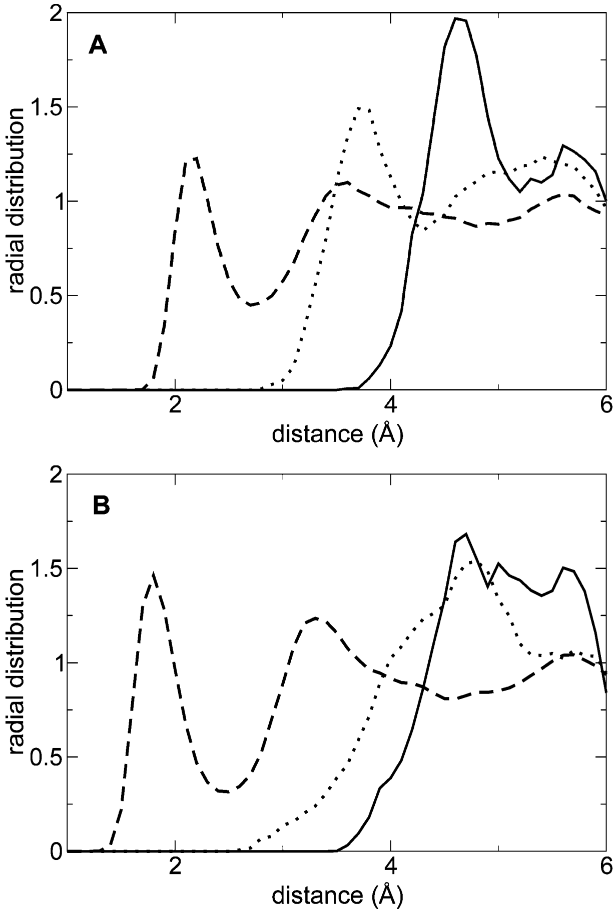


Figure 3. A: Ru−O (solid line), Ru−H (dotted line), and Cl−H (dashed line) radial distributions functions of the aqueous $[RuCl_6]^{4-}$ complex obtained from ab initio molecular dynamics simulations at 300 K. B: Ru−O (solid line), Ru−H (dotted line), and N−H (dashed line) radial distributions functions of the aqueous $[Ru(CN)_6]^{4-}$ complex obtained from ab initio molecular dynamics simulations at 300 K.

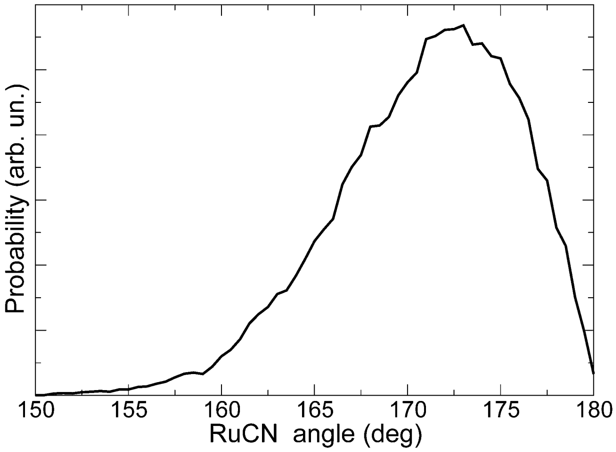


Figure 4. Distribution of the RuCN angle averaged over the aqueous $[Ru(CN)_6]^{4-}$ trajectory for which the RDF is shown in Figure 3B.

3.2.2. One-Electron Orbitals and Energies. A plot of the t_{2g} and e_g energy level scheme of the $[RuCl_6]^{4-}$ and

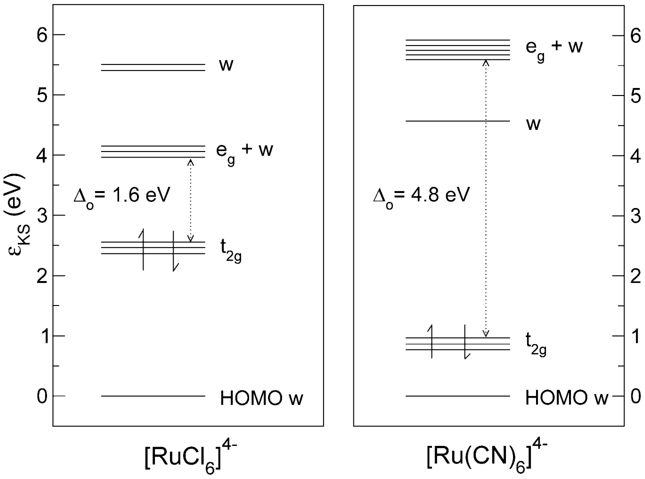


Figure 5. Kohn−Sham (KS) orbital energies of the $[RuCl_6]^{4-}$ (left) and $[Ru(CN)_6]^{4-}$ (right) complex in aqueous solution. The orbital energies have been computed for a representative configuration sampled from the ab initio molecular dynamics trajectory. The octahedral orbital symmetry labels have been assigned according to the shape of the partial electron densities (see also Figure 6). Splitting of the manifolds is schematic (see caption for Figure 2). Half arrows indicate full occupation of the HOMO. While the t_{2g} states are clearly recognizable as such (see Figure 6A), the empty e_{2g} levels are strongly mixed with solvent states (indicated by w) and have as a result very irregular contours (see Figure 6C). Δ_o is the effective octahedral crystal field splitting as obtained from the energy gap between the t_{2g} and e_{2g} manifold. The pure solvent state inserted between the t_{2g} and e_{2g} levels of $[Ru(CN])_6]^{4-}$ is depicted in Figure 6B.

$[Ru(CN)_6]^{4-}$ complexes in solution is shown in Figure 5. Also given are the energies of the highest occupied and lowest unoccupied states of the solvent. The MO diagrams are obtained by diagonalizing the KS matrix of a representative configuration sampled from the ab initio MD simulations. As KS energies in periodic systems are subject to an unphysical and system dependent shift, all one-electron energies are given relative to the HOMO of the water molecules, i.e., the top of the valence solvent band (for a rough estimate of this energy for the computational model of liquid water used here see ref 46). Using the water valence band edge as an energy reference we are assuming that its energy is not affected by the presence of a single solute. This is a rigorous property in the limit of an infinite number of solvent molecules but is an approximation in our finite size system. Comparison between different snapshots along the trajectories indicates that fluctuations in the HOMO energy are in the order of a few tenths of an eV. Symmetry labels are assigned according to the shape of the partial electron densities of the levels on the basis of the MO diagram of an octahedral ML_6 complex.

The key characteristic of the ligand field is the energy gap Δ_o induced between the e_{2g} and t_{2g} levels. The Δ_o values for the Cl^- and CN^- ligands in Figure 5 are ≈1.6 and 4.8 eV, respectively. A related piece of information contained in Figure 5 is the position of the t_{2g} and e_{2g} levels of the two species relative to the solvent states and hence, indirectly, relative to each other. We see that for both complexes the t_{2g} manifold lies well above the valence band of the solvent. Graphical rendering shows orbitals clearly recognizable as hybrid metal ligand orbitals with t_{2g}-like contours localized on the complex (Figure 6A). t_{2g} states can therefore be compared to impurity states in the gap of the solvent. This feature of the solute HOMO is shared by all aqueous transition-metal coordination complexes we have investigated so far.[16,17,22,31,47] As regards the quantative elevation of the t_{2g} levels above the water valence band, we see that for $[Ru(CN)_6]^{4-}$ the separation from the water HOMO is only ≈0.9 eV. For $[RuCl_6]^{4-}$ the solvent−solute gap is ≈2.5 eV, so substantially larger than for $[Ru(CN)_6]^{4-}$. In accordance with Figure 1 it indeed seems that in solution the t_{2g} manifolds are shifted in opposite directions, up for π-donors and down for π-acceptors (taking the valence band edge of the solvent as reference). The result is an ordering of the t_{2g} levels, $[RuCl_6]^{4-}$ (π-donor) above $[Ru(CN)_6]^{4-}$ (π-acceptor), which is also consistent with the qualitative scheme of Figure 1. Making this comparison we are of course tacitly assuming that the σ-bonding in the two complexes is similar. This assumption may be justified in view of Cl^- and CN^- ligands occupying equivalent positions in the nephelauxetic series (see for example ref 24).

Extrapolating the argument above to the empty e_{2g} manifold, one might expect on the basis of Figure 1 that these states should be rather stable and comparatively insensitive to a change of ligands. However, the e_{2g} levels, as obtained in our calculation for $[Ru(CN)_6]^{4-}$, are significantly higher in energy (>1.5 eV) compared to $[RuCl_6]^{4-}$. e_{2g} states are σ-antibonding, and it is therefore perhaps not surprising that they should be more sensitive to substitution of Cl^- by CN^- than t_{2g} states which are nonbonding for σ interactions. A complication is that the assignment is considerably more uncertain for the e_{2g} levels than it was for the t_{2g}. The reason is that mixing with virtual solvent states is much stronger, in particular with the solvent LUMO. In pure water model systems of the size used here this state appears as a discrete energy level ≈4.5 eV above the water HOMO ($1b_1$) band.[22,47,48] This state is effectively degenerate with the $RuCl_6{}^{4-}$ e_{2g} states. Interactions are strong, and it is in practice not possible to distinguish between the three levels in the manifold assigned as e_{2g} + w in Figure 5.

The water LUMO is also the dominant component in the (single) state observed in the orbital diagram of $[Ru(CN)_6]^{4-}$ at approximately the same energy (4.5 eV) as for $[RuCl_6]^{4-}$. However, while there is some hybrization with CN^- orbitals, this state has negligible metal *d* character (Figure 6B). Metal *d* character only appears at higher energies (>5.5 eV). This prompted us to locate the e_{2g} levels there, although none of these states has a clear resemblance to textbook e_{2g} orbitals and all of them are mixed with empty delocalized (“conduction”) states of the solvent (Figure 6C). There is general agreement that the gap in the density of states of liquid water is underestimated by at least 2 eV by the BLYP functional.[22,47,48] Most of the phenomenology of the e_{2g} levels described above must therefore be considered an artifact of the DFT approximation employed in this calculation.[49] This

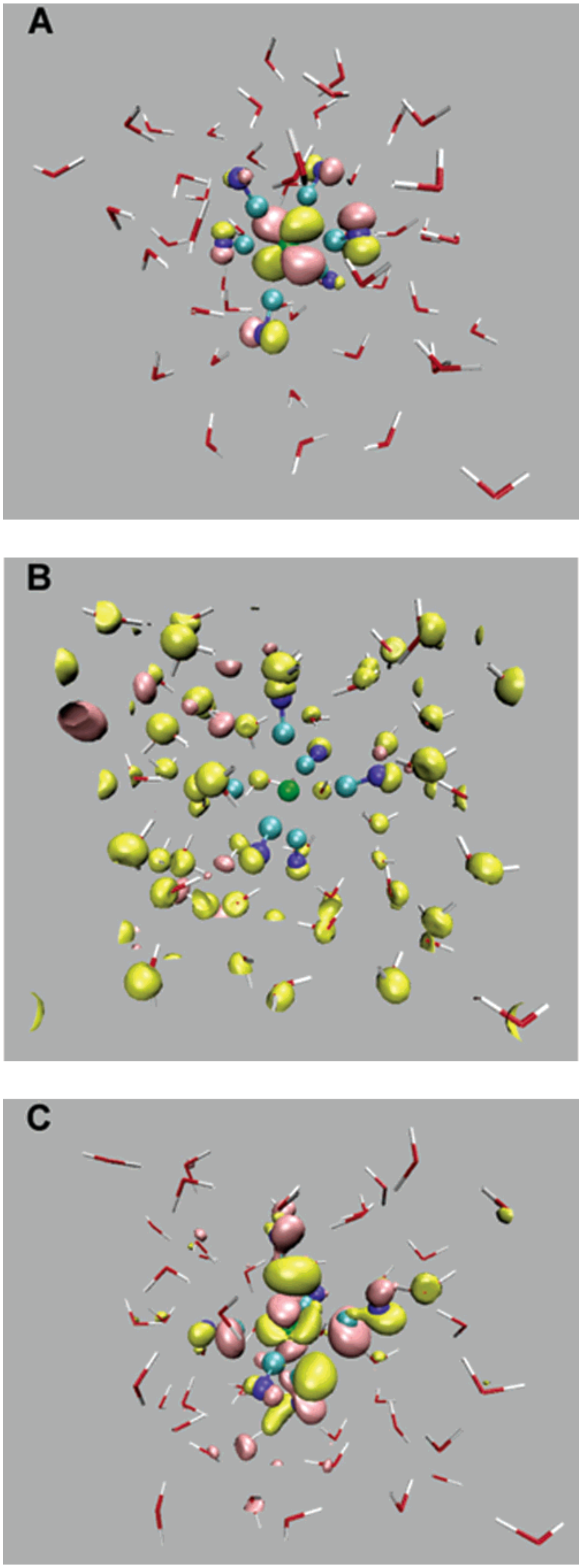


Figure 6. Isosurfaces of molecular orbitals of aqueous $[Ru(CN)]_6]^{4-}$. A: t_{2g} state (HOMO), localized on complex. B: LUMO, mainly solvent. C: e_{2g}-like (LUMO+1), complex + solvent. For the corresponding orbital energies see Figure 5.

Table 4: Literature Values for Crystal Field Splitting Parameter Δ_o in eV for Several Octahedral $M^{2/3+}$ L_6, L = Cl^-, CN^- Transition Metal Compounds[a]

	conf	Cl^-	CN^-
Fe^{2+}	$3d^6$		4.1
Co^{3+}	$3d^6$		4.2
Rh^{3+}	$4d^6$	2.5	
Ir^{3+}	$5d^6$	3.1	
Ru^{2+}(tw)	$4d^6$	1.6	4.8

[a] Data were taken from ref 24. The last line gives the estimates calculated for Ru^{2+} in this work (tw) from the gap in the KS one-electron levels in aqueous solution.

also represents a major source of uncertainty in our estimation of the crystal field parameter Δ_o.

Direct validation of the computed values of Δ_o by experiment is difficult because of the lack of data for Ru complexes. However, the CN^-/Cl^- ratio of 3 for the Δ_o in Figure 5 substantially exceeds the estimate of 2.2 compiled by Jorgensen using spectrochemical data of a series of coordination complexes.[23] Trends for hexachlorides and cyanides of related group VIII and IX transition metals can give a more detailed indication of the accuracy of our results. In Table 4 we have collected some relevant examples. The cyanide data suggest that our value of 4.8 eV for the $[Ru(CN)_6]^{4-}$ could in fact be realistic. Because Δ_o increases when going down from first to second row, the experimental numbers in the table must therefore be considered as a lower bound for the Ru complex. Our value of 1.6 eV for $[RuCl_6]^{4-}$ is more suspect and is most likely a significant underestimation, which would be consistent with the overestimation of the Δ_o CN^-/Cl^- ratio.

The discussion above is based on a direct comparison of KS orbital energy difference $\Delta\epsilon = \epsilon_{eg} - \epsilon_{t2g}$ to experimental crystal field parameters Δ_o obtained from electronic absorption spectroscopy. However, it is well-known that transitions between $t_{2g}{}^6$ and $t_{2g}{}^5e_g$ configurations give rise to two spectroscopic terms at different energies corresponding to excitation from the ${}^1A_1$ ground state to a ${}^1T_1$ and ${}^1T_2$ state. This introduces a further uncertainty in our estimation of Δ_o. This question was investigated in some detail in a time dependent density functional theory (TDDFT) study of the optical spectrum of the $Ru^{2+}(H_2O)_6$ complex in aqueous solution.[22] Analyzing the TDDFT absorption spectrum using the Sugano-Tanabe method (see e.g. ref 24) we obtained a octahedral crystal parameter of $\Delta_o = 2.23$ eV. This value is close to the KS energy gap of $\Delta\epsilon = 2.1$ eV we found for this system confirming that the KS energy gap is a reliable measure for crystal field parameter. Subjecting the experimental absorption spectrum to the same analysis gives $\Delta_o = 2.73$ eV. Since the DFT technology (BLYP functional, plane-wave basis set, and pseudopotentials) used in the Ru aqua-ion study is identical to the approach applied here, the discrepancy of ≈ 0.5 eV can be considered as a fair estimate of the error in our calculation of Δ_o. We also note that the BLYP value for Δ_o of the hexahydrate is too low, which is consistent with the current results for $[RuCl_6]^{4-}$ (see above).

3.2.3. Spin and Electron Densities. The density of unpaired electrons is observable as spin density by means of EPR techniques.[50,51] Spin density in DFT is defined as ρ_β-ρ_α, ρ_β,

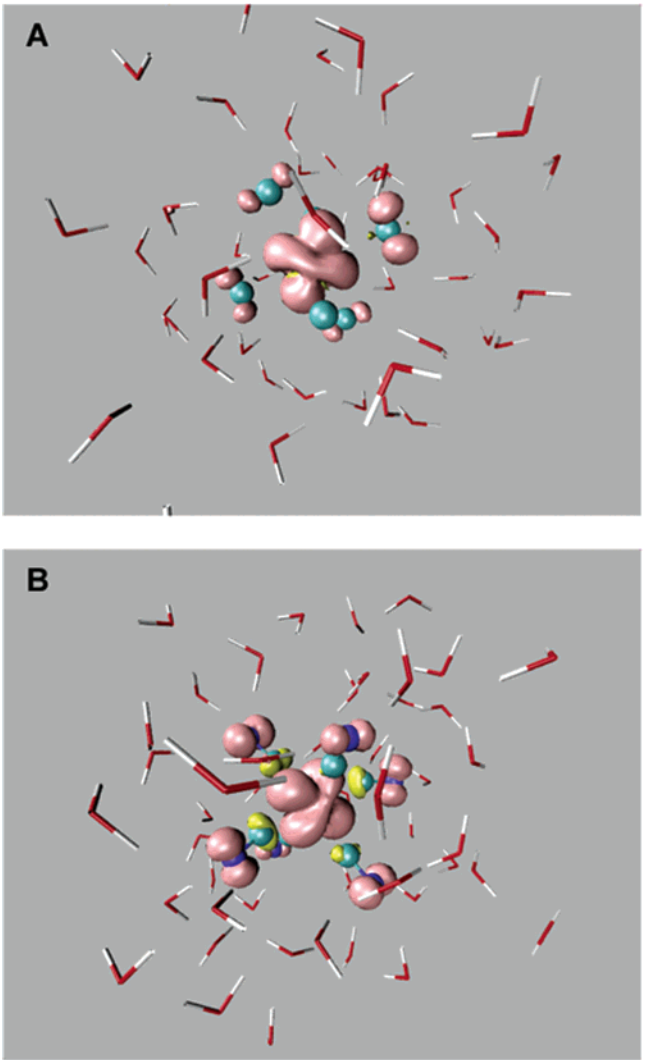


Figure 7. Isosurfaces corresponding to the spin density defined as ρ_β-ρ_α, with ρ_β and ρ_α being the electronic density of the β (minority spin) and α (majority spin) electrons of the aqueous $[RuCl_6]^{3-}$ complex (A) and $[Ru(CN)_6]^{3-}$ complex (B). The yellow isosurface indicates an excess of ρ_β of 0.002 a.u.$^{-3}$, and the pink isosurface indicates an excess of ρ_α of 0.002 a.u.$^{-3}$.

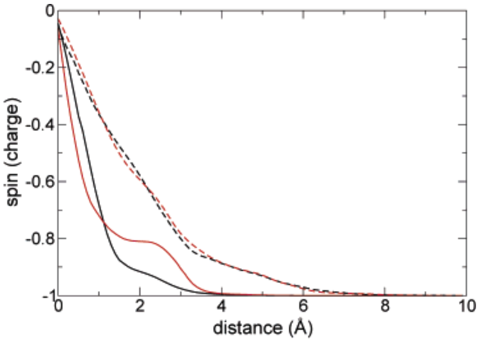


Figure 8. Radial extension of the electron hole in the aqueous $[RuCl_6]^{3-}$ (black) and $[Ru(CN)_6]^{3-}$ (red) complexes. Given are the radial integrals of the average spin density according to eq 10 (solid lines) and vertical charge difference densities (dashed lines) as a function of the distance to the ruthenium ion. For snapshots of the corresponding densities see Figures 7 and 9.

and ρ_α being the electronic density of the β and α electrons as obtained from an unrestricted KS calculation. The Ru(III) coordination complexes are doublets in a t_{2g}^5 configuration. The spin density therefore probes the spatial extent of a hole in the t_{2g} manifold. It contains information about the π-bonding of metal and ligand orbitals and may therefore be able to differentiate between the two complexes. In Figure 7 we show the spin density of the $[RuCl_6]^{3-}$ and $[Ru(CN)_6]^{-3}$ metal complexes in solution after vertical ionization. It can be seen that the spin density is delocalized over metal and ligand atoms but is not spilling out to the solvent molecules consistent with the orbital picture discussed in section 3.2.2. The most pronounced accumulation of spin density is concentrated on the ruthenium atom with contours of t_{2g} shape (compare Figure 6A). The smaller fraction of the spin density residing on the ligands also clearly reflects the π orbital character of the hole. Note that in the case of $[Ru(CN)_6]^{3-}$ the spin density on the C atoms has a sign opposite to the majority spin. Technically, this effect arises as a consequence of spin polarization in the unrestricted KS scheme and may or may not be a DFT artifact (for a discussion of these issues see for example refs 50 and 51).

For a more quantitative characterization we have investigated the average spin density $\zeta(r)$ as a function of the distance r to the ruthenium metal. Rather than the radial distribution $\zeta(r)$ we plot in Figure 8 the radial integral $n_\sigma(r)$ formally defined as

$$n_\sigma(r) = 4\pi \int_0^r dr'(r')^2\zeta(r') \tag{10}$$

For doublet states we should have in our convention $n_\sigma(r \rightarrow \infty) = -1$. From Figure 8 we can conclude that the spin density is more spread out in $[Ru(CN)_6]^{3-}$ than in $[RuCl_6]^{3-}$. This observation can be interpreted as evidence of a "bigger hole" in the case of the $[Ru(CN)_6]^{3-}$ complex due to the π-acceptor character of the CN^- ligands, which allows for net electron transfer to the ligands. Comparison to the RDFs in Figure 3 confirms that the spin density is largely confined to the complex. For both complexes the integrated spin density has reached -1 at the minimum Ru$-$O distance $r_{cO} \approx 3.5$ Å as estimated from the corresponding RDF in Figure 3. However, there is a small overlap between the radial spin density and Ru$-$H RDF. If spin density beyond a coordination radius $r_{cH} = 3.0$ Å can be regarded as effectively transferred to the solvent, the results for $[Ru(CN)_6]^{3-}$ would indicate that about 10% of the hole resides on the water H atoms. This interpretation, however, is far from certain, and a more detailed analysis of the hydrogen bonding to the nitrogen atoms is necessary.

A further quantity of interest is the vertical electronic density hole, defined as the electron density of the oxidized state (ρ_O) minus the electronic density of the reduced state (ρ_R) at the same atomic configuration (solute + solvent). While spin density and electronic density holes are not independent quantities, it has been frequently observed that they highlight rather different aspects of the bonding. This is nicely illustrated by Figure 9 which shows the electron density hole for the same configuration as used in Figure 7 for the visualization of the spin density. In the case of

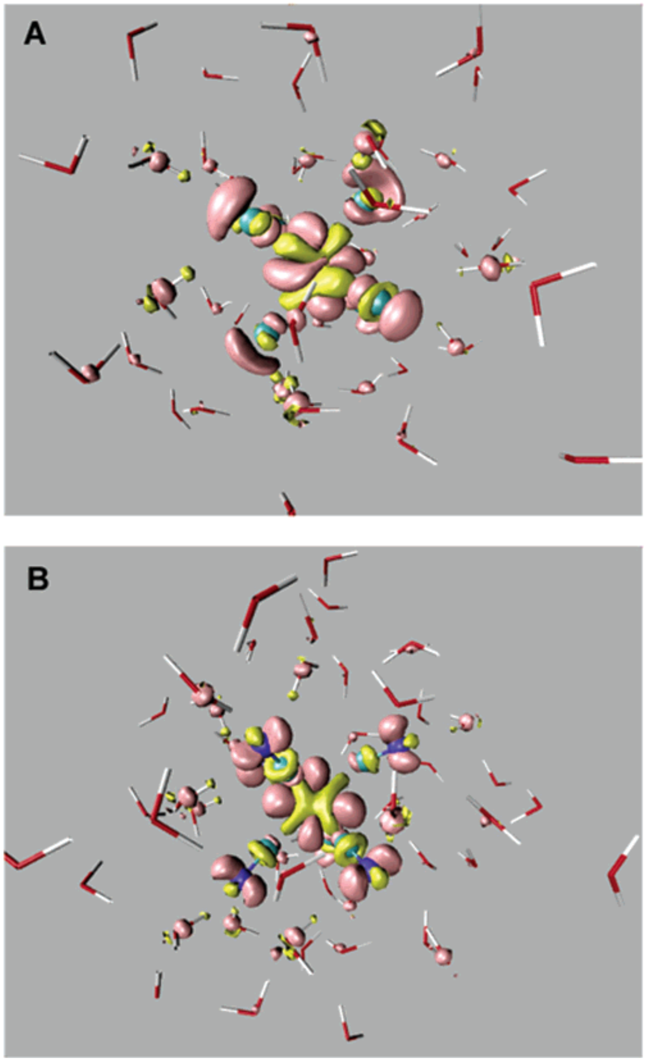


Figure 9. Isosurfaces corresponding to the vertical charge difference density defined as $\rho_O - \rho_R$, with ρ_O and ρ_R being the electronic density of the oxidized and reduced states at the same solvent configuration. Part A shows the result of an instantaneous configuration of the aqueous $[RuCl_6]^{3-}$ complex and part B for the $[Ru(CN)_6]^{3-}$ complex. The yellow isosurface indicates an excess of ρ_O of 0.002 a.u.$^{-3}$, and the pink isosurface indicates an excess of ρ_R of 0.002 a.u.$^{-3}$. Note the change in polarization of the solvent.

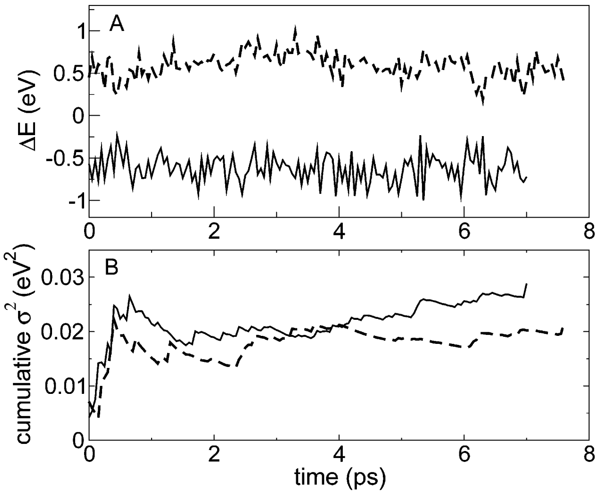


Figure 10. Vertical ionization energy ΔE (eq 1) obtained from ab initio molecular dynamics simulations of the aqueous $[RuCl_6]^{4-}$ complex (solid line) and $[Ru(CN)_6]^{4-}$ complex (dashed line). Part A shows the time series of ΔE and part B the cumulative variance.

$[Ru(CN)_6]^{4-}$ the main features reflect the spin density (Figure 7) except for the contribution of the solvent which is due to changes in the polarization (see below). It can be seen that most of the charge removed by (vertical) ionization comes from π bonds. The picture for $[RuCl_6]^{4-}$, in contrast, shows very little π character. If anything, the electron density hole for the chloride resembles the charge density of σ bonds. These results support an electronic structure model in which the bonding in Cl^- and CN^- compounds has a rather different origin.

For a quantitative analysis we introduce a radially integrated electron hole density $n_\rho(r)$ similar to the radial spin integral of eq 10. Again $n_\rho(r \rightarrow \infty) = -1$, hence $n_\rho(r)$ and $n_\sigma(r)$ can be directly compared (Figure 8). Unlike the spin densities, we find that the electronic difference densities are virtually identical for the two complexes. The radial dependence also shows little structure distinguishing between metal and ligand. The discontinuity in the derivative ("bend") at $r \approx 3.5$ Å coincides with the radius of minimum approach r_{cO} between the metal center and the solvent oxygen atoms in the RDFs of Figure 3. This confirms that $r = 3.5$ Å marks the transition between solute and solvent. Using this distance as a measure of the size of the complex, Figure 8 seems to suggest that around 85% of the electron density hole is on the solute (for both complexes). However, going back to Figure 9 we see that a good fraction of the charge density hole in the solvent can be explained by the electronic dielectric response of the solvent to the oxidation of the solute. Because of polarization by the negatively charged solute, solvent molecules that are not fully included inside the integration sphere will contribute a net charge to the spherical integral with a sign opposite to the charge of the ionic solute at the center of the sphere. The result is a sphere with a (reduced) screened charge, which will of course adjust to a change of charge of the solute when oxidized, contributing in this way to the charge density hole. This argument also underlines our observation about differences in the information contained in spin and charge density holes.

3.2.4. Thermodynamic Properties. The time evolution of the vertical oxidation energy ΔE is displayed in Figure 10. This quantity was computed according to eq 1 as the difference of the total energies of the $[RuL_6]^{3-}$ and $[RuL_6]^{4-}$ ($L \equiv$ Cl, CN) complexes. For given solute + solvent configuration the wave function of the oxidized and reduced states were quenched to the Born−Oppenheimer surface (eliminating possible inaccuracies related to Car−Parrinello dynamics). The configurations were sampled from a trajectory of the reduced complex ($[RuL_6]^{4-}$). Averages ΔE_R show a clear dependence on the ligands. We find $\Delta E_R = -0.62$ eV for Cl^- and $\Delta E_R = 0.59$ eV for CN^-. The fluctuations of the vertical energy gap are more similar, giving $\sigma_R^2 = 0.03$ eV^2 for Cl^- and $\sigma_R^2 = 0.02$ eV^2 for CN^-. Converting variances to reorganization free energies using eq 5 we obtain $\lambda = 0.6$ eV for the $[RuCl_6]^{3/4-}$ reaction and $\lambda = 0.4$ eV for the $[Ru(CN)_6]^{3/4-}$ reaction. These results are summarized in

Table 5: Summary of the Results Obtained for the Energetics and Thermochemistry by Averaging over Trajectories of the Reduced Systems[a]

	ΔE_R	λ	ΔA
$[Ru(CN)_6]^{3/4-}$	0.59	0.4	0.2
$[RuCl_6]^{3/4-}$	−0.62	0.6	−1.2

[a] ΔE_R is the vertical ionization energy (eq 1), λ is the reorganization energies computed from the gap fluctuations according to eq 5, and ΔA is the resulting estimate of the free energy of oxidation using eq 4.

Table 5 for later reference. Inserting the estimates for ΔE_R and λ in eq 4 leads to a $\Delta A = -1.2$ eV for oxidation of the $[RuCl_6]^{4-}$ complex and $\Delta A = 0.2$ eV for the $[Ru(CN)_6]^{4-}$ complex. Recalling that free energies of oxidation (eq 2) have the same sign as reduction potentials, we conclude that according to our results $[Ru(CN)_6]^{3-}$ is a much stronger oxidant than $[RuCl_6]^{3-}$. Indeed the difference in half reaction energies predicts a strongly exergonic reaction free energy change of $\Delta\Delta A = -1.4$ eV for the full $[RuCl_6]^{4-} + [Ru(CN)_6]^{3-} \rightarrow [RuCl_6]^{3-} + [Ru(CN)_6]^{4-}$ redox reaction.

Before attempting a comparison to experiment we first comment on the statistical uncertainties in our estimate for the oxidation free energy. As discussed in section 2.2 the drawback of the use of eq 4 is the slow convergence of the time average of the variance and hence the reorganization free energy (eq 5). Indeed, as can be seen from Figure 10B the statistical uncertainty in the estimate for the variance is in the order of 25%. For an independent assessment of this error we have performed a 1 ps run of the $[RuCl_6]^{3-}$ system at a reduced time step and suppressing the instability of the molecular dynamics of this open shell system by Nosé thermostats applied to both the ionic and fictitious electronic degrees of freedom (unfortunately this approach was considerably less effective for the stabilization of the open-shell Ru cyanide complex). The value of the vertical gap by averaging over this short trajectory of the chloride complex is $\Delta E_O = -1.3$ eV. Substituting this together with the ΔE_R of the 8 ps run (Table 5) in eq 6 we find an oxidation free energy $\Delta A = -1.0$ eV. The discrepancy with value of $\Delta A = -1.2$ eV computed using eq 4 suggests that the error in our estimate of redox free energy is approximately 0.2−0.3 eV.

The standard reduction potential of the aqueous $[Ru(CN)_6]^{4/3-}$ couple (as measured against the standard hydrogen electrode) is 0.9 V.[52] The oxidative nature of the cyanide complex as found in our calculation is therefore supported by experiment. However, as mentioned in section 2.3 the half reaction free energies computed by our method cannot be directly compared to experimental reduction potentials. Only free energies changes of full reactions are meaningful and then only for isocoulombic reactions (see eq 8). The $[Ru(CN)_6]^{3/4-}$ and $[RuCl_6]^{3/4-}$ reactions satisfy this condition, but, unfortunately, the $[RuCl_6]^{4/3-}$ metal complexes are not stable in aqueous solution against ligand-solvent exchange. Nonaqueous solvents, however, stabilize a large variety of cyanide and chloride complexes mixed with a host of other ligands.[2,3] Lever has used this vast database to parametrize a simple linear ligand additivity model capable of describing the experimental redox potentials with remarkable accuracy.[3,4] The model can be formulated by the following expression for the observed reduction Ru(III)/Ru(II) potential E_{obs} in units of V

$$E_{obs} = \sum_i E_L(L_i) + C \qquad (11)$$

where the summation is over the various ligands L_i of the complex. C is a constant depending on the reference electrode. The E_L parameters given in ref 3 for CN^- and Cl^- are $E_L(CN^-) = 0.12$ V and $E_L(Cl^-) = -0.24$ V. This would yield according to eq 11 a free energy for the full $[RuCl_6]^{4-} + [Ru(CN)_6]^{3-} \rightarrow [RuCl_6]^{3-} + [Ru(CN)_6]^{4-}$ redox reaction of $\Delta\Delta A_{obs} = -1.56$ eV. While our computed value of $\Delta\Delta A = -1.4$ eV is very close, the two redox potentials are not equivalent because the E_L coefficients are strictly valid only for reactions in acetonitrile solution. Reference 3 also gives an empirical rule for the conversion to aqueous solution, amounting to multiplication by a factor of 1.14. Applying this correction the predicted value increases to $\Delta\Delta A_{obs} = -1.78$ eV. Using this value as our experimental reference we conclude that the calculation underestimates the absolute reaction free energy change by ≈0.4 eV.

3.2.5. Redox Potentials and One-Electron Energy Levels. We are now ready to return to the issue raised in the Introduction, namely the question of a possible correlation between one-electron energies and the redox potential. As mentioned in section 3.2.2 such a correlation exists for stable finite systems in a vacuum. The HOMO energy (ϵ_{HOMO}) as computed for the (unknown) exact density functional is rigorously equal to the (minus) vertical ionization energy (IP).[43,44] However, for the approximate density functionals of the type used here (generalized gradient corrections) this relation is far from satisfied: $-\epsilon_{HOMO}$ is found to be considerably smaller than IP, usually only half this value.[45] The first question is, therefore, how the energy of the redox active orbitals of the reduced extended system compares to the vertical energy gap. Taking, as explained in section 3.2.2, the top edge of the solvent valence band as reference and subtracting the t_{2g} energy of $[RuCl_6]^{4-}$ from the corresponding energy for $[Ru(CN)_6]^{4-}$ we obtain a value of $\Delta\epsilon_{t2g} = -1.6$ eV. This number should be compared to the difference in vertical energy gap (Table 5) $\Delta\Delta E_R = 1.2$ eV. We find that in our solution model system $-\Delta\epsilon_{HOMO}$ differs from $\Delta\Delta E_R$ by about 30%. While this discrepancy is appreciable, it is significantly smaller than the mismatch observed for BLYP for ionization under vacuum conditions. We note that our calculations reported in ref 16 of the $MnO_4^{2-} + RuO_4^{1-} \rightarrow MnO_4^{1-} + RuO_4^{2-}$ aqueous redox reaction gave a very similar result, namely $-\Delta\epsilon_{HOMO} = 0.7$ eV versus $\Delta\Delta E_R = 0.4$ eV.

Comparing next the KS HOMO energy gap $\Delta\epsilon_{HOMO} = -1.6$ eV to the free energy change $\Delta\Delta A = -1.4$ eV computed for the full (reverse) redox reaction, we find an even better agreement. KS energy gaps and reaction free energies have a rather different thermodynamic status, and to understand the relation between these two quantities we go back to the discussion in section 3.2.4 and recall our observation that ligand character has a far more pronounced effect on the vertical ionization than on the relaxation of the system after the ET process. We found the reorganization free energies to be rather similar ($\Delta\lambda = -0.2$ eV, see Table 5). To see what this implies for the thermochemistry we substitute eq 4 in the expression for the reaction free energy of the full reaction and obtain

$$\Delta\Delta A = \Delta\Delta E_R - \Delta\lambda \tag{12}$$

Setting $\Delta\lambda \approx 0$ gives $\Delta\Delta A \approx \Delta\Delta E_R$. On the condition that the KS HOMO energy gap is a good approximation to the vertical ionization difference, the correspondence with the redox free energy is therefore a consequence of cancellation of the reorganization energies.

4. Concluding Remarks

The density functional MD simulations reported in this paper confirm that the higher oxidation state of the metal ion, Ru(III), is more stable when chelated by Cl^- than by CN^- ligands. This is in agreement with the conventional ligand field picture sketched in section 2.1. Due to the differences in π-bonding, oxidation of $[Ru(CN)_6]^{4-}$ amounts to removing an electron from a bonding t_{2g} HOMO. In the $[RuCl_6]^{4-}$ complex this orbital has changed character to π-antibonding, and we can expect that removing an electron from this orbital should be energetically less costly. We also found, that within a margin of a few tenths of eV, the gap between the HOMO energy levels of reactants could be used as an estimate of the free energy change of the full redox reaction. This near quantitative correlation between the orbital level scheme and thermochemistry was traced to a combination of cancellation of reorganization effects and the relatively close agreement between vertical ionization energy differences and KS energy gaps. Of these two observations, the latter is perhaps the more surprising. We note, however, that this agreement, while exceptional for vacuum systems, is not uncommon in extended metallic systems, where the work function can often be estimated to a very good approximation by the Fermi energy as computed for a slab of material using common approximate density functionals.[53] This suggests that the degree of localization of the redox active orbital may play an important role. This explanation, however, must remain at the moment rather speculative, and it would be useful to investigate whether similar conditions apply to other related coordination compounds and, if so, why. To conclude we comment on the accuracy of our result for redox free energy. Clearly an absolute error of 0.4 eV is larger than one would like (note, however, that this is the most unfavorable case in Table 1). There are three main sources for this error: insufficient convergence of statistics, size effects, and the level of approximation of the density functional calculations. Run length was of particular concern in the present calculation as it required the estimation of the second moment of fluctuations for the computation of reaction and reorganization free energy (eqs 5 and 4). This statistical uncertainty can be reduced substantially by the use of the two surface expressions (eqs 6 and 7). This was not feasible in the present Car−Parrinello simulation because of the near degenerate open shell character of one of the oxidation states. These difficulties can however be eliminated by switching to Born−Oppenheimer dynamics.[33] While size effects on half reaction energies are huge, they cancel to a large extent for the full reaction studied here. The chemistry controlling redox reactions is essentially short range. Figure 8 gave a nice illustration supporting this claim. The difference charge density determining the long range electrostatic component of the reaction energy is virtually the same for the chloride and cyanide complex. We are therefore fairly confident that size effects are not the main cause of the error. Further investigations are however needed.

The performance of the DFT is more difficult to assess. If the correlation between redox potentials and HOMO energy levels can be taken seriously, which seems a possibility at least for our system, then we can use the one-electron density of states to analyze the redox chemistry. Such an analysis suggests that the $[RuCl_6]^{4-}$ HOMO levels (the t_{2g} manifold in Figure 5) are too low in energy. Moving these levels up while leaving the t_{2g} levels of $[Ru(CN)_6]^{4-}$ where we found them (Figure 5) would increase the redox free energy bringing the DFT estimation in better agreement with experiment. Assuming that also the empty e_{2g} stay fixed, a higher t_{2g} energy would also reduce the crystal field splitting for the hexachloride complex, which came out too high in our calculation (section 3.2.2). We see, however, for the moment no convincing reason why the HOMO energy of the aqueous ruthenium chloride complex should be less reliable than for the cyanide complex. Increased interaction with the empty solvent states which we know are too close in energy to the occupied states could be a possibility. We are however optimistic that the more technical aspects of the simulation, sampling and system size, can be improved and related errors reduced making a detailed analysis of the relation between thermochemistry and electronic structure both meaningful and helpful for the understanding of the performance of DFT in condensed molecular systems.

Acknowledgment. This research is sponsored by EPSRC and HPC-Europe. R.A. is grateful to the Spanish government (MEC) and Marie Curie Fellowship for financial support. R.A. thanks the molecular modeling group at ICMAB, specially Lourdes F. Vega, for the warm hospitality during her stay at Barcelona. Part of the computations was carried out on BSC facilities.

References

(1) Bargeletti, F.; Juris, A.; Balzani, V.; Belser, P.; von Zelewsky, A. *Inorg. Chem.* **1987**, *26*, 4115−4119.

(2) Juris, A.; Balzani, V.; Barigelleti, F.; Campagna, S.; Belser, P.; von Zelewsky, A. *Coord. Chem. Rev.* **1988**, *84*, 85−277.

(3) Lever, A. B. P. *Inorg. Chem.* **1990**, *29*, 1271−1285.

(4) Lever, A. B. P.; Dodsworth, E. S. In *Inorganic Electronic Structure and Spectroscopy*; Solomon, E. I., Lever, A. B. P., Eds.; John Wiley and Sons: New Jersey, 1999; Vol. II, Chapter 4, p 227.

(5) Siegbahn, P. E. M. *Adv. Chem. Phys.* **1996**, *93*, 333−387.

(6) Li, J.; Fisher, C. L.; Chen, J. L.; Bahford, D.; Noodleman, L. *Inorg. Chem.* **1996**, *35*, 4694−4702.

(7) Martinez, J. M.; Pappalardo, R. R.; Sánchez Marcos, E.; Mennucci, B.; Tomasi, J. *J. Phys. Chem. B* **2002**, *106*, 1118−1123.

(8) Angelis, F. D.; Tilocca, A.; Selloni, A. *J. Am. Chem. Soc.* **2004**, *126*, 15024−15025.

(9) Rustad, J. R.; Dixon, D. A.; Rosso, K. M.; Felmy, A. R. *J. Am. Chem. Soc.* **1999**, *121*, 3234−3235.

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

(11) Ikeda, I.; Hirata, M.; Kimura, T. *J. Chem. Phys.* **2003**, *119*, 12386−12392.

(12) Ikeda, I.; Hirata, M.; Kimura, T. *J. Chem. Phys.* **2005**, *122*, 024510.

(13) Amira, S.; Spångberg, D.; Zelin, V.; Probst, M.; Hermansson, K. *J. Phys. Chem. B* **2005**, *109*, 14235−14242.

(14) Ensing, B.; Buda, F.; Blochl, P. E.; Baerends, E. J. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3619−3627.

(15) Blumberger, J.; Sprik, M. *J. Phys. Chem. B* **2005**, *109*, 6793−6804.

(16) Tateyama, Y.; Blumberger, J.; Sprik, M.; Tavernelli, I. *J. Chem. Phys.* **2005**, *122*, 234505.

(17) Blumberger, J.; Sprik, M. *Theor. Chem. Acc.* **2006**, *115*, 113−126.

(18) Blumberger, J.; Tavernelli, I.; Klein, M. L.; Sprik, M. *J. Chem. Phys.* **2006**, *124*, 064507.

(19) Warshel, A. *J. Phys. Chem.* **1982**, *86*, 2218−2224.

(20) Hwang, J.-K.; Warshel, A. *J. Am. Chem. Soc.* **1987**, *109*, 715−720.

(21) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682−8692.

(22) Bernasconi, L.; Sprik, M. *J. Phys. Chem. B* **2005**, *109*, 12222−12226.

(23) Jorgensen, C. K. *Modern aspects of ligand field theory;* North-Holland: Amsterdam, 1971.

(24) Bersuker, I. B. *Electronic structure and properties of transition metal compounds*; Wiley & Sons: 1996.

(25) Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 966−978.

(26) Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 979989.

(27) Marcus, R. A. *J. Chem. Phys.* **1957**, *26*, 867−872.

(28) Marcus, R. A. *J. Chem. Phys.* **1965**, *43*, 679−701.

(29) Marcus, R. A.; Sutin, N. *Biochim. Biophys. Acta* **1985**, *811*, 265322.

(30) Marcus, R. A. *Rev. Mod. Phys* **1993**, *65*, 599−610.

(31) Blumberger, J.; Bernasconi, L.; Tavernelli, I.; Vuilleumier, R.; Sprik, M. *J. Am. Chem. Soc.* **2004**, *126*, 3928−3938.

(32) Blumberger, J.; Sprik, M. *J. Phys. Chem. B* **2004**, *108*, 6529−6535.

(33) VandeVondele, J.; Lynden-Bell, R.; Meijer, E. J.; Sprik, M. *J. Phys. Chem. B* **2006**, *110*, 3614−3623.

(34) VandeVondele, J.; Sulpizi, M.; Sprik, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 1936−1938.

(35) Tachiya, M. *J. Phys. Chem.* **1989**, *93*, 7050−7052.

(36) CPMD Version 3.9.2, Copyright IBM Corp 1990−2001, Copyright MPI fur Festkorperforschung Stuttgart 1997−2005.

(37) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(38) Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B* **1988**, *37*, 785−789.

(39) Troullier, N.; Martins, J. *Phys. Rev. B* **1991**, *43*, 1993−2006.

(40) Blumberger, J.; Tateyama, Y.; Sprik, M. *Com. Phys. Comm.* **2005**, *169*, 256−261.

(41) Kleinman, L.; Bylander, D. *Phys. Rev. Lett.* **1982**, *48*, 1425−1428.

(42) Martyna, G.; Tuckerman, M. *J. Chem. Phys.* **1999**, *110*, 2810−2821.

(43) Almbladh, C. O.; von Barth, U. *Phys. Rev. B* **1985**, *31*, 3231−3244.

(44) Chong, D. P.; Gritsenko, O. V.; Baerends, E. J. *J. Chem. Phys.* **2002**, *116*, 1760−1772.

(45) Grüning, M.; Gritsenko, O. V.; van Gisbergen, S. J.; Baerends, E. J. *J. Chem. Phys.* **2002**, *116*, 9591−9601.

(46) Hunt, P.; Sprik, M. *Chem. Phys. Chem.* **2005**, *6*, 1805−1808.

(47) Bernasconi, L.; Blumberger, J.; Sprik, M.; Vuilleumier, R. *J. Chem. Phys.* **2004**, *121*, 11885−11899.

(48) Pendergast, D.; Grossman, J. C.; Galli, G. *J. Chem. Phys.* **2005**, *123*, 014501.

(49) An added complication is that the Γ point-only *k* space sampling as applied here is likely to be not adequate for the delocalized virtual solvent states.[48] Introduction of more *k* points may therefore affect the energies of these states and, hence, the mixing with the localized virtual states of the solute.

(50) Improta, R.; Barone, V. *Chem. Rev.* **2004**, *104*, 1231−1245.

(51) Remenyi, C.; Kaupp, M. *J. Am. Chem. Soc.* **2005**, *127*, 11399−11413.

(52) Bard, A.; Parsons, R.; Jordan, J. *Standard Potentials in Aqueous Solutions;* IUPAC: New York, 1985.

(53) Lozovoi, A. Y.; Alavi, A. *Phys. Rev. B* **2003**, *68*, 245416.

CT600169E

1416 *J. Chem. Theory Comput.* **2006,** *2,* 1416−1428

JCTC Journal of Chemical Theory and Computation

Intrinsic Viscosity of Proteins and Platonic Solids by Boundary Element Methods

David K. Hahn and Sergio R. Aragon*

Department of Chemistry and Biochemistry, San Francisco State University,
1600 Holloway Ave., San Francisco, California 94132

Received February 15, 2006

Abstract: The boundary element (BE) method is used to implement a very precise computation of the intrinsic viscosity for rigid molecules of arbitrary shape. The formulation, included in our program BEST, is tested against the analytical Simha formula for ellipsoids of revolution, and the results are essentially numerically exact. Previously unavailable, very precise results for a series of Platonic solids are also presented. The formulation includes the optional determination of the center of viscosity; however, for globular proteins, the difference compared to the computation based on the centroid is insignificant. The main application is to a series of 30 proteins ranging in molecular weight from 12 to 465 kD. The computation starts from the crystal structure as obtained from the Protein Data Bank, and a hydration thickness of 1.1 Å obtained in previous work with BEST was used. The results (extrapolated to an infinite number of triangular boundary elements) for the proteins are separated into two groups: monomeric and multimeric proteins. The agreement with experimental measurements of the intrinsic viscosity in the case of monomeric proteins is excellent and within experimental error of 5%, demonstrating that the solution and crystal structure are hydrodynamically equivalent. However, for some multimeric proteins, we observe strong systematic deviations around −20%, which we interpret as a systematic deviation of the solution structure from the crystal structure. A possible description of the structural change is deduced by using simple ellipsoid model parameters. A method to obtain intrinsic viscosity values for proteins to 1−2% accuracy (better than experimental error) on the basis of a single BE computation (avoiding the need for an extrapolation on the number of surface triangles) is also presented.

I. Introduction

The intrinsic viscosity, $[\eta]$, is simply the initial fractional slope obtained when the solution viscosity, η, is plotted against the concentration, c:

$$[\eta] = \lim_{c\to 0}\frac{\eta - \eta_o}{\eta_o c} \tag{1}$$

where η_o is the viscosity of the pure solvent. Thus, $[\eta]$ reflects the increase in viscosity brought about by the addition of an infinitesimal amount of solute to a pure solvent. Measurement of $[\eta]$ provides a simple and inexpensive way of obtaining information about molecular shape in solution, allowing phenomena such as protein denaturation[1] and oligomerization[2] to be examined.

Accurate computation of $[\eta]$ for macromolecules requires microscopic detail in the representation of the molecular surface. To achieve microscopic detail, the surface may be modeled in one of two ways: by a collection of small spherical beads (the hydrodynamic bead model) or by an array of flat triangular platelets (the boundary element, BE, method). While the bead model is well-known and has been previously applied to the computation of $[\eta]$ for proteins,[3] the problem can be formulated exactly as an integral equation that can be solved more accurately and precisely by the BE method.

The BE treatment by Zhou[4] used an approximate expression for $[\eta]$ accurate to within 2% for ellipsoids with axial ratios between $^1/_4$ and 4. Its application to globular proteins

---

* Corresponding author fax: 415-338-2384; e-mail: aragons@sfsu.edu.

10.1021/ct600062y CCC: $33.50

showed that a molecular surface based on the X-ray crystal structure is needed to obtain a description of hydration consistent with experimental protein hydration levels.[5] Allison[6] has since derived an exact BE expression for the intrinsic viscosity of a particle of arbitrary charge, situated in an arbitrary shear field and charge distribution. An application of this expression to the viscosity of ellipsoids[7] achieved agreement to within 1.0% of the exact value for axial ratios between $^{1}/_{10}$ and 10, even though a small number of platelets ($<$1000) were used to represent the surface.

In a previous paper,[8] hereafter referred as I, a thorough study of the translational and rotational diffusion coefficients of 41 globular proteins was presented. In I, the accuracy of a precise implementation of the boundary element method was clearly demonstrated. In this paper, the accuracy of our implementation is demonstrated with regard to the intrinsic viscosity of ellipsoids, polyhedrons, and globular proteins. Because of the rigorous nature of its implementation, the method's accuracy is limited only by the faithfulness of the representation of the hydrodynamic surface. Thus, when the hydrodynamic surface is constructed from crystallographic data, significant disagreement with an intrinsic viscosity experiment is indicative of either a difference in conformation between the crystal and solution phases or experimental error.

II. Theory

The hydrodynamic resistance problem for a particle translating with velocity $\mathbf{v}(\mathbf{y})$ at point y in a quiescent fluid can be formulated exactly as an integral over the surface of the molecule:[9]

$$\mathbf{v}(\mathbf{y}) = \int_{S_\mathrm{p}} \overleftrightarrow{\mathbf{T}}(\mathbf{x},\mathbf{y})\, \mathbf{f}(\mathbf{x})\, \mathrm{d}S_x \qquad (2)$$

where $\mathbf{f}(\mathbf{x})$ is the stress force at point x on particle surface S_p that must be solved for and $\overleftrightarrow{\mathbf{T}}(\mathbf{x},\mathbf{y})$, the Oseen hydrodynamic interaction tensor, is expressed as

$$\overleftrightarrow{\mathbf{T}}(\mathbf{x},\mathbf{y}) = \frac{1}{8\pi\eta_\mathrm{o}|\mathbf{x}-\mathbf{y}|}\left[\overleftrightarrow{\mathbf{I}} + \frac{(\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^2}\right] \qquad (3)$$

An approximate solution to eq 2 may be obtained by converting it into a matrix equation. This is accomplished by discretizing the surface of the molecule into triangular boundary elements.[10] The key assumption is that the surface stress force has a constant value within each of the N boundary elements

$$\mathbf{v}(\mathbf{y}_k) = \sum_{j=1}^{N} \int_{\Delta_j} \overleftrightarrow{\mathbf{T}}(\mathbf{x},\mathbf{y})\, \mathrm{d}S_x\, \mathbf{f}_j \qquad (4a)$$

$$= \sum_{j=1}^{N} \overleftrightarrow{\mathbf{G}}_{kj}\cdot\mathbf{f}_j \qquad (4b)$$

where Δ_j signifies integration over boundary element j (a triangle in this work). The resulting system of equations may then be solved for $\mathbf{f}_j$ using standard algebraic techniques. To compute different transport properties, one selects appropriate flows that specify the particular form of $\mathbf{v}(\mathbf{y})$. For the intrinsic viscosity, one selects five elementary shear flows[6]

$$\mathbf{v}_\mathrm{o}^{(l)}(\mathbf{y}) = \frac{1}{2}\mathbf{E}^{(l)}(\mathbf{y}-\mathbf{d}) \qquad (5a)$$

where $\mathbf{d}$ is the center of viscosity (the point where the dissipation of energy is minimized) and

$$\mathbf{E}^{(1)} = \gamma(\mathbf{e}_1\mathbf{e}_2 + \mathbf{e}_2\mathbf{e}_1) \qquad (5b)$$

$$\mathbf{E}^{(2)} = \gamma(\mathbf{e}_1\mathbf{e}_3 + \mathbf{e}_3\mathbf{e}_1) \qquad (5c)$$

$$\mathbf{E}^{(3)} = \gamma(\mathbf{e}_2\mathbf{e}_3 + \mathbf{e}_3\mathbf{e}_2) \qquad (5d)$$

$$\mathbf{E}^{(4)} = \gamma(\mathbf{e}_1\mathbf{e}_1 - \mathbf{e}_2\mathbf{e}_2) \qquad (5e)$$

$$\mathbf{E}^{(5)} = \gamma(\mathbf{e}_1\mathbf{e}_1 - \mathbf{e}_3\mathbf{e}_3) \qquad (5f)$$

with γ being the shear gradient and $\mathbf{e}_\alpha$ being a unit vector along axis α in the laboratory frame of reference.

For convenience, the intrinsic viscosity may be expressed as the ratio between a dimensionless function of particle shape known as the viscosity factor, ξ, and the density of the particle, that is, $[\eta] = \xi/\rho$. When the viscosity factor is formulated in terms of the boundary element method, one obtains

$$\xi = -\frac{1}{\eta_\mathrm{o}\gamma V_\mathrm{p}}\int_{S_\mathrm{p}} (\mathbf{y}-\mathbf{d})_b \mathbf{f}_a\, \mathrm{d}S_x \qquad (6)$$

where subscripts a and b denote vector components along the flow direction and the direction of greatest shear, respectively, and V_p is the volume of the particle. Discretizing the surface integral and then averaging over all orientations of the particle yields[6]

$$\xi = \frac{1}{5}(\xi_{12}^{(1)} + \xi_{13}^{(2)} + \xi_{23}^{(3)}) + \frac{1}{15}(\xi_{11}^{(4)} + \xi_{33}^{(4)} - 2\xi_{22}^{(4)} + \xi_{11}^{(5)} + \xi_{22}^{(5)} - 2\xi_{33}^{(5)}) \qquad (7a)$$

where

$$\xi_{\alpha\beta}^{(k)} = -\frac{1}{2\eta_\mathrm{o}\gamma V_\mathrm{p}}\left[\sum_{j=1}^{N} A_j(\mathbf{y}_j - \mathbf{d})(\mathbf{e}_\alpha\mathbf{e}_\beta + \mathbf{e}_\beta\mathbf{e}_\alpha)\mathbf{f}_j^{(k)}\right] \qquad (7b)$$

where the summation runs over the surface elements and superscript k represents the five elementary shear fields for which the viscosity factor is evaluated. In eq 7b, A_j is the area of boundary element j, the stress force of shear field k on surface element j is denoted $\mathbf{f}_j^{(k)}$, and $\mathbf{y}_j$ represents the incenter of triangle j in the lab reference frame. The incenter of a triangle is the center of a circle that is inscribed within the triangle, and it differs, in general, from the centroid.

III. Computational Method

The theory presented above has been implemented in a Fortran 90 routine for calculating the intrinsic viscosity of an uncharged, arbitrarily shaped particle and was added to BEST,[11] a hydrodynamics program based on the method of Youngren and Acrivos[10] for numerically solving the integral form of the Stokes equations. Preliminary calculations on

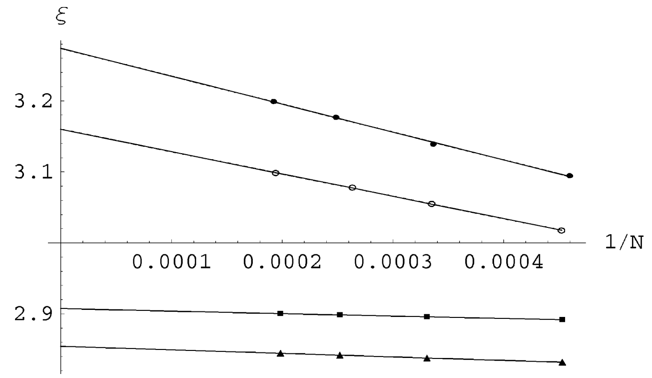


Figure 1. Graph of the viscosity factor as a function of the inverse triangle number, extrapolated to an infinite triangle number, for catalase (●), β-trypsin (○), an oblate ellipsoid of revolution with an axial ratio of 2 (▲), and a prolate ellipsoid of revolution with an axial ratio of $^1/_2$ (■). The protein data are fitted to a straight line, while the ellipsoid data are fitted to a quadratic line. A hydration layer thickness of $\delta = 1.1$ Å was used for calculating the protein viscosity factors.

ellipsoids showed that double precision accuracy would be needed in order to obtain results that correspond to experimental observations within the limits of experimental error. As discussed in some detail previously,[11] the accuracy of a BE calculation strongly depends on minimizing the error arising from the approximations used when integrating the Oseen tensor over the boundary elements of the tessellated surface. For this reason, double precision was assigned to all constants, variables, and operations in BEST, and the calculation of the Oseen tensor integral was increased to double precision accuracy using 21st order Gaussian quadrature. Double precision accuracy also eliminates round-off error from the BE calculation. Round-off error is manifested by a significant drop in the graph of the transport properties versus $1/N$ near $N = 2500$ when single precision accuracy is used.[11] No such drop is apparent in Figure 1.

Because $[\eta]$ is experimentally observed at the center of viscosity, the Fortran 90 routine computes the viscosity factor at this center or at the centroid of the molecule. The derivation of the center of viscosity within the framework of the BE method is provided in Appendix A, and all calculated values of $[\eta]$ reported herein were obtained at that center, unless otherwise noted. The program BEST calls LAPACK[12] routines including a set of basic linear algebra subroutines that have been hardware-optimized for the Opteron—the AMD core math library. The computation time for a given number of triangles varies from 2 to 20 min on dual processor AMD Opteron 248 servers with 4−16 GB of memory or a similar Itanium 2 workstation.

IV. Results and Discussion

We have performed several computations in order to demonstrate the accuracy and precision of the boundary element implementation of the intrinsic viscosity computation.

We first discuss the results as applied to ellipsoids of revolution for which analytical formulas are available for comparison. Then, we show new computations for a series of Platonic solids to demonstrate that the presence of sharp corners presents no difficulty for the implementation. Finally, we present our computations for a set of globular proteins and find strong evidence for differences in solution structure compared to the crystal for some multimeric proteins.

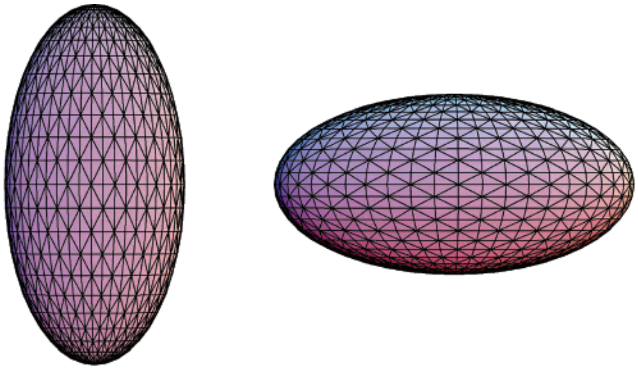


Figure 2. Triangulation of a prolate revolution ellipsoid, axial ratio $^1/_2$, and an oblate revolution ellipsoid, axial ratio 2, into 2208 platelets.

IV.A. Ellipsoids of Revolution. The accuracy of our implementation of the BE method was determined through comparison with the exact analytical solution for the viscosity factor of an ellipsoid. To implement the BE method, the rigid ellipsoids were represented by a collection of flat, triangular surface elements (Figure 2). The tessellations for the ellipsoids and polygons discussed below were done by a program we constructed using Mathematica (Wolfram Research) because BEST requires a triangulation as input. The error in shape and surface area that arises when a curved surface is triangulated was removed by multiplying the 3 × 3 blocks represented by $\overleftrightarrow{\mathbf{G}}_{kj}$ in eq 4b by the appropriate correction factors.[11] The viscosity factor, $\xi(N)$, was computed with BEST over a range of values for N (specifically, $N = 2208$, 3014, 3968, and 5040), and ξ was obtained from a polynomial least-squares fit of a plot of $\xi(N)$ versus $1/N$ extrapolated to an infinite triangle number. This is illustrated in Figure 1 for a prolate and oblate ellipsoid of revolution having an axial ratio of $p = {}^1/_2$ and 2, respectively. The extrapolation removes the discretization error, which arises because the surface stress force is approximated as a constant within each of the N boundary elements.

Table 1 compares the computed values of ξ to the exact values obtained from the Simha formula.[13] The error in the calculated value is below 0.01% for both the prolate and the oblate cases for all axial ratios less than or equal to 10. These results may be further improved by using an adaptive grid to triangulate the surface so that more triangles are situated in regions of high curvature and fewer are situated in regions of low curvature, demonstrating that our method is essentially numerically exact. The improvement compared to the previous work of Allison[7] is notable. However, the error in the present approach is 2 orders of magnitude smaller than the experimental uncertainty[14] in $[\eta]$, which is at least 1%, so improving our present values is not worthwhile. When applied to globular proteins, the accuracy of BEST is guaranteed if a well-triangulated, faithful representation of the hydrodynamic surface is employed.

Table 1. Viscosity Factor of Ellipsoids of Revolution and the Sphere

p	ξ^a	ξ(BEST)	% error
$^1/_{10}$	13.6343	13.6341	−0.0018
$^1/_9$	11.8043	11.8037	−0.0047
$^1/_8$	10.1026	10.1023	−0.0031
$^1/_7$	8.53274	8.53240	−0.0039
$^1/_6$	7.09876	7.09838	−0.0053
$^1/_5$	5.80621	5.80587	−0.0058
$^1/_4$	4.66332	4.66306	−0.0055
$^1/_3$	3.68488	3.68473	−0.0041
$^1/_2$	2.90761	2.90751	−0.0035
1	2.5	2.50004	0.0016
2	2.85437	2.85443	0.0019
3	3.43083	3.43080	−0.0001
4	4.05933	4.05940	0.0018
5	4.70821	4.70821	0.0000
6	5.36720	5.36720	0.0000
7	6.03194	6.03196	0.0003
8	6.70027	6.70083	0.0083
9	7.37099	7.37112	0.0018
10	8.04337	8.04359	0.0027

Table 2. Viscosity Factor of Platonic Solids

		ξ (BEST)	
structure	S/V^a	uniform	edged
tetrahedron	9.0000	4.172	4.210
cube	5.1962	3.096	3.118^b
octahedron	5.1962	2.996	3.016
dodecahedron	3.7752	2.683	2.691
icosahedron	3.3887	2.632	2.636

^{a} Surface-to-volume ratio. ^{b} Estimated from the tetrahedron behavior.

IV.B. Platonic Solids. The Platonic solids are of interest because their sharp corners provide a severe test of the robustness of the boundary element method implemented in this work. Both uniform and edge-enhanced triangulations of the polyhedron surface, Figure 3, were used to calculate ξ. The uniform triangulation tessellates the surface by repeatedly subdividing each triangular facet into four equilateral triangles. In this study, up to $N = 10\,000$ was used, requiring a machine with at least 8 GB of memory. The edge-enhanced tessellation requires at least one initial uniform triangulation in order to obtain interior and exterior triangles on each facet, whereupon only the exterior triangles are subdivided in each subsequent refinement. For polyhedrons composed of *n*-polygonal facets rather than triangular facets, such as the cube and dodecahedron, an initial tessellation of each facet into *n* triangles is also required. A polynomial least-squares fit of a plot of $\xi(N)$ as a function of $1/N$ yielded ξ from the extrapolation to infinite *N*. The edge-enhanced tessellation is expected to produce more accurate results because this kind of tessellation allows one to define the corners and edges more precisely with a given number of data points at large values of *N*, and these are the regions where the surface stress varies more rapidly.

Use of the edge-enhanced tessellation is justified by the larger variation in the surface stress force near the edges and vertexes of each facet, which requires a greater density of triangles; the surface stress force is nearly constant toward the center of each facet. These circumstances may be inferred from a comparison of the results in Table 2. For example, the difference between the edged and uniform results is largest for the tetrahedron. Because this is the Platonic solid with the most prominent vertexes, it has the largest variation in $\mathbf{f}(\mathbf{x})$ on its surface, and the result from the uniform tessellation has the largest error. Conversely, the icosahedron has the least prominent vertexes of all of the Platonic solids, and the difference between the edged and uniform results is the smallest. The extrapolations versus $1/N$ for the polyhedra show linear behavior as in Figure 1, with no significant differences between the uniform and edge-enhanced tessellations. (The edge-enhanced tessellation for the surface of the cube leads to linear dependence in $[\overleftrightarrow{\mathbf{G}}_{kj}]$, so that the value in Table 2 is estimated from the behavior of the octahedron.)

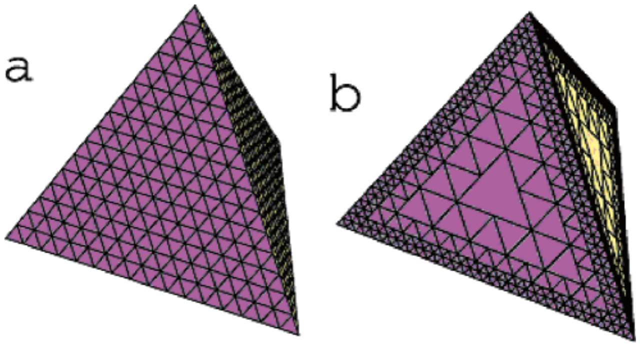


Figure 3. (a) Uniform triangulation of the tetrahedron into 1024 platelets. (b) Edge-enhanced triangulation of the tetrahedron into 1756 platelets.

Polyhedra have biological and chemical significance; HIV-1 and many other small viruses have icosahedral symmetry,[15] for example, and of course C_{60} is a truncated icosahedron. It does not appear that the viscosity factor of any polyhedron has been obtained previously, in either numerical or analytical form, so these very accurate values are new. (The viscosity of structures composed of beads situated at the vertexes of a polyhedron has been determined,[16] but this is not an appropriate comparison to the present results.) However, the size of ξ should at least reflect the surface-to-volume ratio, *S*/*V*, of a structure. A decrease in *S*/*V*, for example, causes a decrease in the amount of friction generated with the surrounding fluid and a concomitant decrease in ξ. Table 2 shows that BEST does in fact predict values for ξ that are consistent with the *S*/*V* of the polyhedrons. In addition, the value of ξ appropriately converges on 2.5, the exact value for a sphere,[17] which has the smallest possible surface-to-volume ratio. The estimated uncertainty in the edged values of the viscosity factor is 0.03%, except for the estimated value of the cube for which the uncertainty is no more than 0.1%. All of the viscosity factors presented below were obtained from extrapolations versus $1/N$. The viscosity factor was also calculated for two nonplatonic polyhedrons. For the truncated icosahedron, we obtained $\xi = 2.546$, and for the icosidodecahedron, $\xi = 2.588$.

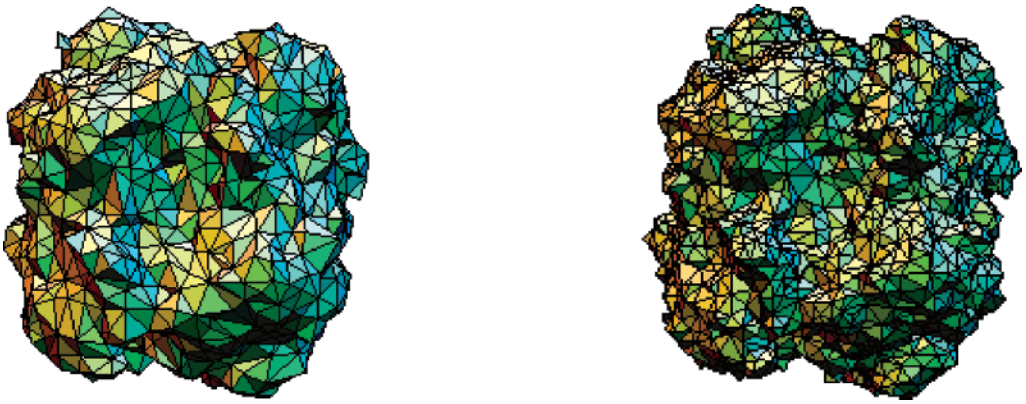


Figure 4. Surface of catalase (4BLC) triangulated with Msroll and processed with Coalesce to yield 2248 triangles (left) and 4154 triangles (right).

IV.C. Proteins. To extend the study to proteins, molecular surfaces were constructed from crystallographic data using MSROLL.[18] Coalesce,[11] which reduces N by merging small and slender triangles, was then applied to make the triangulations more suitable for the BE method. Figure 4 shows the triangulated surface for the large protein catalase. Hydration was modeled by uniformly increasing the van der Waals radius of all of the atoms in the PDB file (the inflation model of solvation) before creating the molecular surface. The optimum value of the hydration layer thickness, $\delta = 1.1 \pm 0.1$ Å, was determined in I by minimizing the disagreement between the computed and experimental values of the translational diffusion coefficient (D_t) for a set of four well-studied proteins, namely, ribonuclease, lysozyme, myoglobin, and chymotrypsinogen a.

To optimize δ, $D_t(N)$ and $[\eta(N)]$ were computed in the region where N is large enough for the plots against $1/N$ to be linear. The transport properties were then obtained from a straight-line fit extrapolated to an infinite triangle number.

The location of the linear region depends on how the molecular surface is constructed. More precisely, it is sensitive to the number of triangles eliminated by Coalesce, because the surface roughness decreases as the number of eliminated triangles increases, and this affects the curvature in a plot of the transport properties against $1/N$.

In turn, the number of eliminated triangles depends on the number of triangles in the molecular surface obtained from MSROLL, which is controlled by a fineness parameter.

For most of the proteins in this study, the MSROLL molecular surface contained at least 20 000 triangles, and the region of linearity extended down to approximately $N = 2000$ after processing by Coalesce. This is demonstrated in Figure 1, which shows the linear fit to the plot of $\xi(N)$ against $1/N$ using data obtained for β-trypsin and catalase.

Processing the MSROLL triangulation with Coalesce also reduces the amount of scatter in a plot of the transport properties against $1/N$. Combined with the high precision of BEST, there results an excellent straight-line fit for small proteins such as β-trypsin, as Figure 2 shows. The scatter tends to increase with increasing particle size, but even for β-galactosidase, the largest protein in the data set, there is only a 0.1% statistical error on the intercept, and a small number of points are needed for confidence. (For very large particles, the data set may contain a data point with a discordant value. In such instances, a “Q test” should be performed on the residuals of a linear regression analysis of the data to determine whether the suspect value should be retained.[19] Data points that do not pass the Q test can be discarded.)

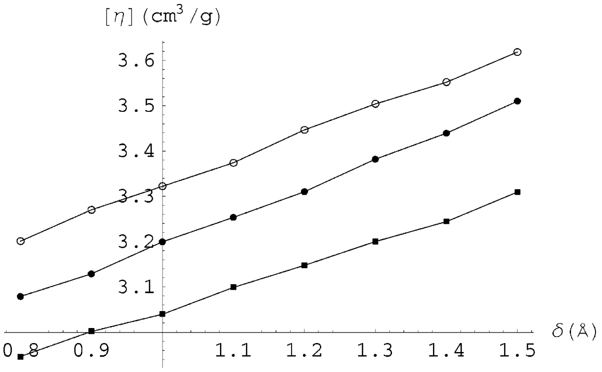


Figure 5. Graph of the intrinsic viscosity as a function of the hydration layer thickness for myoglobin (○), lysozyme (●), and chymotrypsinogen (■).

The transport properties are smooth functions of δ when $[\eta(N)]$ values are calculated in the linear region, as Figure 5 shows. $[\eta]$ varies by approximately 0.06 cm^3g^{-1} for every 0.1 Å increase in δ. An optimum value of 1.1 Å is also obtained for δ when minimizing the discrepancy of $[\eta]$ for a small test set of proteins, but with a larger uncertainty, compared to the value obtained in I. Thus, we find that a single value of the hydration thickness is suitable for all transport properties, including the intrinsic viscosity. Zhou[5] obtained a smaller hydration layer thickness, 0.9 Å, which we attribute to differences in tessellation methods. While MSROLL uses spherical, toroid, and saddle regions to represent the molecular surface, Zhou used only spherical polygons, which resulted in undersized excluded molecular volumes. The smaller hydration layer obtained by Zhou compensates for this because it results in a rougher molecular surface and, therefore, a larger viscosity factor, allowing the calculated and experimental values of $[\eta]$ to match each other.

Table 3 compares calculated and literature values of $[\eta]$ and D_t for 30 globular proteins. For each protein, only crystallographic, viscometric, and diffusion data obtained

Table 3. Intrinsic Viscosity and Translational Diffusion Coefficient of Native Proteins

			$[\eta]$ (cm^3/g)				D_t (10^{-7}cm^2/s)			
protein	s[a]	mass (kDa)	calcd	exptl	ref	Δ[b]	calcd	exptl	ref	Δ[b]
cytochrome C (1HRC)	1	12.4	3.07	2.74	21	13	11.63	11.1−12.1	22−25	0
ribonuclease A (7RSA)[c]	1	13.7	3.52	3.30, 3.50	26, 27	3	10.84	10.68	28	2
α-lactalbumin (1HFX)	1	14.2	3.41	3.01, 3.4	29, 30	6	10.84	10.57, 10.6	31, 32	2
lysozyme (2CDS)	1	14.3	3.22	2.66−3.00	33−37	12	11.04	10.6, 11.2	38, 36	1
myoglobin (1MBO)	1	17.2	3.37	3.25	39	2	10.24	10.4, 10.5	40, 41	−2
soyb. tryp. inhib. (1AVU)	1	20.1	3.18	2.8	42	14	9.88	9.8	43	1
β-trypsin (1TPO)	1	23.3	3.08	3.1	44	0	9.50	9.3	45	2
trypsinogen (1TGN)	1	24.0	3.00	2.96	46	1	9.49	9.68	47	−2
α-chymotrypsin (4CHA)	1	25.2	3.25	3.00	44, 48	8	9.08	10.20	48	−11
chymotrypsinog. A (2CGA)	1	25.7	3.20	2.5, 3.13	49, 50	4	9.16	9.23	51	−1
carbonic anhyd. B (2CAB)	1	28.8	3.07	2.76, 3.2, 3.7	52, 37, 53	−5	8.84	8.89	52	−1
Zn-α2-glycoprotein (1ZAG)	1	32.6	4.88	5.0	54	−2	7.32	6.5	54	12
pepsin (4PEP)	1	34.5	3.33	3.09, 3.35	55, 56	3	8.10	8.01, 8.71	57, 56	−3
G-ADP actin (1J6Z)	1	43.0	3.56	3.7	58	−3	7.39	7.15, 7.88	59, 60	−2
Taka-amylase A (6TAA)	1	52.5	3.15	3.3	61	−3	7.22	7.37	62	−2
human serum Alb. (1AO6)	1	66.5	4.12	3.9, 4.2, 4.73	63−65	−3	6.07	5.9, 6.1, 6.3	66, 64, 67	−1
ovotransferrin (1OVT)	1	76.0	3.83	3.8	68	0	6.03	5.9	69	2
lactotransferrin (1LFG)	1	77.1	3.97	4.0	70	0	5.87	5.6	71	5
superoxide dismu. (2SOD)	2	32.5	3.57	3.3	72	9	8.10	8.27	72	−2
β-lactoglobulin (1BEB)[c]	2	36.7	3.65	3.4−4.2	73−76, 37	−5	7.74	7.3	77, 31	5
α-chymotrypsin (4CHA)	2	50.4	3.31	4.1, 4.25	78, 79	−21	7.17	7.1, 7.40	80, 79	−1
concanavalin (1GKB)	2	51.4	3.95	4.1	81	−2	6.72	6.2	82	8
triosephos. isom. (8TIM)	2	53.2	3.59	3.75	83	−4	6.88	6.76	83	2
ricin (2AAI)	2	61.5	3.33	2.96	84	13	6.61	6.0	85	10
oxyhemoglobin A (1HHO)	4	63.2	2.89	2.77	20	4	7.03	6.78	86	4
alkaline phosphat. (1ALK)	2	94.6	3.09	3.4	87	−7	5.96	5.7	88	4
citrate synthase (1CTS)	2	98.0	3.20	3.95	89	−20	5.82	5.8	89	0
inorganic pyrophos. (1FAJ)	6	117.3	2.93	4.0	90	−28	5.62	5.7	90	−2
aldolase (1ADO)	4	157.1	3.84	3.4, 4.0, 4.04	91−93	0	4.66	4.29−4.8	94−97	4
catalase (4BLC)	4	235.7	3.08	3.9	98, 99	−21	4.49	4.1	100, 101	10
β-galactosidase (1BGL)	4	465.8	3.84	3.78	102	2	3.26	3.13	102	4

[a] Number of subunits. [b] The percent difference between the calculated and experimental value determined from the average of the experimental values. [c] Heavy atoms only.

from the same species are considered. (A possible exception exists for trypsin and pepsin, for which the source of the protein in the viscosity experiment was not specified.) This precaution is taken because $[\eta]$ may depend on the source of the protein, even if that protein's molecular weight does not significantly differ from one strain to the next.[20]

The present computational model for investigating protein viscosity contains three assumptions: the crystal- and solution-phase geometries of the protein are identical, the thickness of the hydration layer is uniform, and the hydration layer is smooth on the measurement time scale. In light of the accuracy demonstrated in Table 1, any significant discrepancy between theory and measurement may be attributed to either a failure of one of these assumptions or experimental error.

For almost all of the monomeric proteins in Table 3, the assumptions of the computational model appear to be valid, as it is possible to find at least one value for $[\eta]$ from the literature that is within 5% of the calculated value. For several of the oligomeric proteins, however, the value of $[\eta]$ is significantly underestimated by the calculation. A possible explanation is that the assumption regarding the geometry is not valid for these proteins. The subunits of oligomeric proteins are typically bound together by noncovalent forces that may weaken in solution, resulting in a greater separation compared to the crystal phase, a more elongated shape, and a larger value of $[\eta]$. Support for this hypothesis is particularly apparent in the comparison between the results for the α-chymotrypsin monomer and dimer entries in Table 3. (A single chain from the PDB file 4CHA was used to represent the monomer geometry.) The computed value of the monomer intrinsic viscosity exceeds the measured value by only 8%, while the computed value of the dimer intrinsic viscosity undershoots the measured values by about 20%.

A closer examination of the results in Table 3 reveals that four multi-subunit proteins have large underestimations of $[\eta]$ that generally occur for oligomers with a weak noncovalent contact between subunits in solution. The dimers of α-chymotrypsin[103] and citrate synthase,[104] for example, undergo a reversible equilibrium between the associated and dissociated species, which is typical of oligomeric proteins with weakly bound subunits, and the intrinsic viscosity of both of these proteins is underestimated by about 20%. Hexameric inorganic pyrophosphatase of *Escherichia coli* also appears to have a weak noncovalent interchain bond; under mildly acidic conditions, it dissociates into two trimers.[105] The intrinsic viscosity for this protein is underestimated by nearly 30%.

For catalase and oxyhemoglobin, however, the strength of the subunit contact does not correlate with the error in $[\eta]$. Tetrameric catalase is stable over a wide range of pH values,[98] while $[\eta]$ is underestimated by 19% for this protein. Conversely, the physicochemical behavior of oxyhemoglobin does not suggest strong subunit contacts,[111] while the calculated value of $[\eta]$ for this protein is accurate to within the limits of experimental uncertainty. Nevertheless, the precision of the computations is strong evidence for discrepancies in the solution and crystal structure for the four proteins whose intrinsic viscosity is underestimated by 20% or more. It is also possible that there is some difference in structure in going to solution for the extra cases where the discrepancy is greater than 10%, as in the cases of ricin among the oligomeric proteins and cytochrome C and soybean tripsin inhibitor among the monomeric proteins. The case of lysozyme is unclear because the experimental values have such a wide range.

A much smaller discrepancy in $[\eta]$ is obtained for most of the oligomeric proteins with a strong noncovalent interaction between subunits. Superoxide dismutase,[106] aldolase,[107] concanavalin,[108] alkaline phosphatase,[109] and triosephosphate isomerase[110] all have a strong solution-phase noncovalent contact between subunits, on the basis of physicochemical studies, and the computed intrinsic viscosity for these proteins agrees with experimental results to within 10%. The small discrepancy in $[\eta]$ for oxyhemoglobin conforms to the quaternary structure of carbonmonoxy hemoglobin in solution, which has been characterized by NMR measurements as an intermediate between two known crystal structures.[112] These crystal structures differ by a 13.3° rotation of two subunits (an $\alpha\beta$ dimer) relative to the other two, with little change in the center of mass separation between the subunits.[113] Ricin is the only oligomeric protein in the data set with an interchain disulfide bond.[114] The positive error in the calculated value for the intrinsic viscosity of this protein is partially due to this covalent interaction, yet it too may have a change in conformation, as described below.

IV.D. Qualitative Geometry Changes in Oligomeric Proteins. Our precise treatment of the intrinsic viscosity demonstrates that an accurate description of the real shape of a macromolecule is necessary to obtain reasonable quantitative accuracy. Nevertheless, if we desire only qualitative information, then an ellipsoidal model could be used because it is simple enough to be analytically solved. This does not imply that we believe ellipsoidal models are useful for anything other than a qualitative exploration of shape effects. A qualitative description of the difference between the crystal structure and the solution-phase geometry may be determined from data in Table 3 by modeling the proteins as ellipsoids of revolution. From the relation $[\eta] \propto V_p\,\xi(\rho)$ for an ellipsoid of axial ratio ρ and volume V_p and having a Simha factor of $\xi(\rho)$, one obtains

$$\xi(\rho') = (1 + q)(V_p/V_p')\,\xi(\rho) \tag{8}$$

where the prime denotes a measured value and unprimed denotes a calculated value and where q is the fractional error in $\xi(\rho)$. Similarly, $D_t \propto G(\rho)/a$ for an ellipsoid of length a having a shape factor of $G(\rho)$, so that

$$G(\rho') = (1 - r)(a'/a)\,G(\rho) \tag{9}$$

where r is the fractional error in $G(\rho)$. Multiplying $\xi(\rho')$ by $G(\rho')$ and assuming $V_p/V_p' = a/a'$, one obtains

$$\xi(\rho')\,G(\rho') = (1 + q)(1 - r)\,\xi(\rho)\,G(\rho) \tag{10}$$

which may then be solved for ρ'. The geometry change is then deduced from the change in ρ, a, and V_p consistent with the signs of p and q. For example, the α-chymotrypsin dimer has a computed viscosity factor corresponding to an ellipsoid with $\rho = 0.38$, while $q = -0.20$ and $r = 0.03$ from Table 3, so that the solution of eq 10 yields $\rho' = 0.32$. The comparison between ρ and ρ' indicates that the protein is less spherical in solution. The behavior of $V_p\,\xi(\rho)$ and $G(\rho)/a$ as a function of ρ then dictates that the protein must also expand, in addition to becoming less spherical, to satisfy the condition imposed by the values of q and r.

Table 4. Qualitative Solution-Phase Geometry Change

	Δ				
protein	$[\eta]$	D_t	r	ρ¢	inference
cytochrome C[a]	>0	0	0.45	0.51	shrink and more spherical
lysozyme	>0	0	0.40	0.43	shrink and more spherical
soybean trypsin inhibitor	>0	0	0.36	0.42	shrink and more spherical
ricin	>0	>0	0.37	0.46	more spherical
α-chymotrypsin	<0	0	0.38	0.32	expand and less spherical
citrate synthase	<0	0	0.39	0.34	expand and less spherical
inorganic pyrophosphatase	<0	0	0.45	0.36	expand and less spherical
catalase	<0	>0	2.9	2.4	expand and/or less spherical

[a] See text.

An oblate model was used for catalase, and a prolate model was used for the remaining proteins (by examination of their rotational diffusion tensors in paper I). Table 4 provides a summary of the geometry changes for each of the five cases encountered in the present set of proteins. Cytochrome C is an interesting case. Along with lysozyme and soybean trypsin inhibitor, they are the only 3 of 18 monomeric proteins that show a significant positive discrepancy. As mentioned previously, the experimental data for lysozyme have a broad range, so the discrepancy could be as little as 7%, so no change in going to solution may be present. Cytochrome C, unlike the other two, has a very asymmetric rotational diffusion tensor and cannot be represented as an ellipsoid of revolution. The crystal structure has "three major and two minor helical elements interconnected by strands of polypeptide chain and folded into a roughly globular shape within which a (hydrophobic) heme pocket is formed."[115] If one insists on the assignment of a prolate shape, then the discrepancies predict that the protein will become more spherical in solution, as shown in Table 5. However, if we consider it to be oblate, we conclude the opposite: it becomes less spherical ($\rho = 2.83$ and $\rho' = 3.59$). The ellipsoid model does not work even qualitatively for this protein.

The ellipsoid model predicts that ricin becomes more spherical in solution without undergoing a change in size, suggesting torsional motion around the interchain disulfide bond without a change in the distance between the subunits. In addition to the disulfide bond, a strong noncovalent interaction is known to exist between ricin's A and B chains,[116] so it is surprising that this protein shows a moderate positive discrepancy in both the intrinsic viscosity and the translational diffusion coefficient.

It is noteworthy that the crystal- and solution-phase geometries may differ and yet the translational diffusion coefficients may not show a large change. For the majority of the oligomers, the error in D_t is positive, larger than that of the monomers, and smaller than that of $[\eta]$. These results are anticipated from a simple ellipsoid model; for an ellipsoid of length a and axial ratio $\rho = b/a$, $D_t \propto G(\rho)/a$. Because the shape function $G(\rho)$ is more slowly varying than a in the region relevant to globular proteins ($\rho \approx {}^1/_2$), a positive error in D_t is predicted by the ellipsoid model if a is too small. However, the errors in $G(\rho)$ and a at least partially cancel each other out when D_t is calculated. Because $[\eta] \propto V_p\ \xi(\rho)$ for an ellipsoid of volume V_p, any error in V_p will be amplified by the accompanying error in $\xi(\rho)$, so that a larger error arises when $[\eta]$ is calculated.

V. Computational Shortcuts

Two aspects of the very precise method lead to large computer memory requirements, and large computation times. Because the solution of a linear system is an N^3 process, using large numbers of triangles slows the computation considerably, while computing the center of viscosity requires the even more time-consuming (and doubling the memory requirement) explicit inverse matrix computation. In this section, we demonstrate that both of these problems can be circumvented. First, we show that the centroid is quite sufficient so that the center of viscosity is not required for protein computations. Second, we obtain a simple linear extrapolation equation that allows the computation of the viscosity factor from a single BE calculation at a moderate number of triangles, eliminating the need to compute at many values of the triangle number and extrapolating to an infinite number of triangles.

Table 5. Protein Intrinsic Viscosity from a Single BE Computation

protein	N	$[\eta(N)]$	$[\eta]$	% error
cytochrome C (1HRC)	2316	3.029	3.168	3.21
ribonuclease A (7RSA)	2560	3.394	3.535	0.43
lysozyme (2CDS)	2580	3.140	3.270	1.54
α-lactalbumin (1HFX)	2558	3.245	3.380	−0.88
myoglobin (1MBO)	2774	3.230	3.354	−0.47
soybean trypsin inhibitor (1AVU)	2900	3.032	3.143	−1.18
β-trypsin (1TPO)	2428	2.957	3.087	0.22
trypsinogen (1TGN)	2678	2.890	3.005	0.16
α-chymotrypsin (4CHA)	2566	3.081	3.209	−1.25
chymotrypsinogen A (2CGA)	2214	3.080	3.229	0.90
carbonic anhydrase B (2CAB)	2500	2.959	3.086	0.51
Zn-α2-glycoprotein (1ZAG)	2170	4.624	4.853	−0.56
pepsin (4PEP)	2600	3.201	3.332	0.07
G-actin (1J6Z)	2218	3.369	3.532	−0.79
Taka-amylase (6TAA)	2396	3.018	3.152	0.07
human serum albumin (1AO6)	2630	3.874	4.031	−2.16
ovotransferrin (1OVT)	2426	3.644	3.805	−0.66
lactoferrin (1H76)	2316	3.757	3.930	−1.00
superoxide dismutase (2SOD)	2472	3.426	3.574	0.12
β-lactoglobulin (1BEB)	2206	3.458	3.626	−0.65
α-chymotrypsin (4CHA, dimer)	2670	3.150	3.276	−1.03
concanavalin (1GKB)	2180	3.774	3.960	0.25
triosephosphate isomerase (8TIM)	2672	3.422	3.558	−0.89
ricin (2AAI)	2204	3.294	3.454	3.73
oxyhemoglobin A (1HHO)	2712	2.708	2.814	−2.61
alkaline phosphatase (1ALK)	2724	3.074	3.194	3.36
citrate synthase (1CTS)	2812	3.123	3.241	1.28
inorganic pyrophosphatase (1FAJ)	2296	2.882	3.016	2.94
aldolase (1ADO)	2466	3.693	3.852	0.32
catalase (4BLC)	2994	2.858	2.959	−3.93

V.A. Intrinsic Viscosity Values from a Single BE Computation. For a machine with only 1 GB of RAM, the maximum number of triangles that memory storage permits is around 3000. The calculations detailed above have used machines with up to 16 GB of RAM. Thus, the question arises, can we obtain a useful transport property without requiring extrapolations including very large numbers of triangles? The data presented in Table 4 demonstrate that we can give an affirmative answer to the previous question. The slope of the extrapolations as a function of $1/N$ is not large, and the slope divided by the intercept does not vary widely across the protein data set. Thus, it is possible to estimate the extrapolated value to an infinite number of triangles by using Q, the average slope/intercept, over a protein data set. This implies that, given the value Q for each property, one can perform a single calculation with 2000−3000 triangles and obtain a value for a viscosity factor with a statistical error of about 1−2%. This is 2 to 4 times worse than the statistical error of the accurate extrapolations but still much better than the experimental error. Using regression data from plots of $[\eta(N)]$ versus $1/N$ for a set of proteins, an average slope/intercept, Q, is obtained, and the intrinsic viscosity can be quickly calculated as

$$[\eta] = [\eta(N)]/(1 + Q/N) \tag{11}$$

For the set of 30 proteins in Table 3, $Q = -102.31$. Table 5 shows the value of $[\eta(N)]$ at the given value of N, the extrapolation using eq 11, and its error for each of the 30 proteins. Agreement between the formula and the accurate extrapolated value has a standard deviation just over 1% and is at worst 4%. Because the typical uncertainty in a measurement of $[\eta]$ is 5%, eq 11 may be used with no significant loss in accuracy or precision with the consequent savings in effort and lesser hardware requirements. More precise values can be obtained from a single BE computation if N is larger than 3000.

V.B. Consequences of Calculating $[\eta(N)]$ at the Centroid. The calculation of $[\eta(N)]$ may be expedited by performing it at the centroid instead of the center of viscosity. Evaluation of the center of viscosity requires explicit inversion of the Oseen tensor matrix, which is the most time-consuming step in the calculation of $[\eta(N)]$. Neglecting this calculation allows the direct solution of eq 4—a much faster alternative that does not require storage of the inverse matrix. Because globular proteins are roughly spherical or ellipsoidal, the centroid should nearly coincide with the center of viscosity, and this is indeed the case for all of the proteins in this study. Typically, the computed distance between the center of viscosity and the centroid is less than 0.7 Å and tends to decrease with increasing N. For citrate synthase (1CTS), the distance between the viscosity center and centroid is unusually large, being 1.1 Å for $N = 2108$. However, only a 0.08% difference in $[\eta(N)]$ was found to arise when the calculation was performed at the centroid rather than the viscosity center. The difference in CPU time for the two calculations was 474 s on an AMD Opteron 64-bit processor. Our program BEST provides a flag to compute the viscosity factor at either center.

VI. Summary and Conclusions

We have implemented a very accurate boundary element method to determine the intrinsic viscosity of arbitrarily shaped objects. Our formulation is in numerically exact agreement with the Simha formula for ellipsoids of revolution. We are able to calculate high precision values of the intrinsic viscosity for sharp-cornered polyhedra, some of which are representative of viral shapes. In the application to proteins, we have found that the same uniform hydration layer thickness of 1.1 Å that was successfully used to model diffusion tensors in I also works for the intrinsic viscosity. Thus, a uniform hydration layer provides an excellent hydrodynamic model for the accurate computation of all transport properties of proteins. Because the intrinsic viscosity is a function of shape, independent of the size of an object, our study also allows us to compare the molecular shape in the crystal and in solution. For almost all monomeric proteins, the method agrees within experimental error with the measured intrinsic viscosities of proteins, demonstrating that the conformation in the crystal and that in solution are hydrodynamically indistinguishable.

For several multi-subunit proteins, the computational model has sufficient accuracy to strongly suggest that the crystalline- and solution-phase geometries are different. Most of the proteins whose subunits are bound together by weak noncovalent forces show a discrepancy consistent with a shape change in going to solution from the crystal. We have proposed a simple qualitative model to understand what type of shape change could account for the observed differences. A characterization of the possible shape changes is presently being pursued in our laboratory using implicit solvent molecular modeling with Amber 8 with encouraging initial results. That work will be reported elsewhere. Finally, we have also proposed a simple equation that permits the computation of the intrinsic viscosity to within 2% using a single BE computation, making these calculations possible in readily available computer hardware.

Note Added in Proof: Prof. J. Michael Schurr made us aware that the crystal structure may fail to include all the residues of a protein. This is actually the case for one third of the proteins studied here. However, the percentage of missing residues is small, typically around 2%, so that the conclusions of the paper are not affected.

Acknowledgment. This research was supported through a grant from the National Institutes of Health, MBRS SCORE Program, Grant S06 GM52588 to S.A.

Appendix A: Derivation of the Center of Viscosity

The center of viscosity is obtained by applying the condition to eq 7a and solving for **d**. The dependence of $\mathbf{f}_j^{(l)}$ on **d**

$$\frac{\partial \xi}{\partial \mathbf{d}} = 0 \tag{A1}$$

must first be made explicit, that is,

$$\mathbf{f}_{\mathrm{j}}^{(l)} = {\mathbf{U}_{\alpha\beta_j}}^{-1} \times \mathbf{y}_j + \mathbf{d} \times {\mathbf{U}_{\alpha\beta_{\mathrm{j}}}}^{-1} \tag{A2a}$$

$$= \mathbf{g}_{\alpha_j}^{(l)} + \mathbf{d} \times \mathbf{h}_{\alpha\beta_j}^{(l)} \tag{A2b}$$

where $\mathbf{U}_{\alpha\beta}$ is a matrix element of the Oseen tensor integral. Substituting eq A2b into eq 7a and performing the multiplication between the unit vectors then gives

$$\begin{aligned}\xi = \frac{1}{15}\sum_{\mathrm{j}} A_j \Big\{ \frac{3}{2}[(d_2 - r_{2j})(-g_{1_j}^{(1)} + d_1 \times h_{11_j}^{(1)} + d_2 \times h_{12_j}^{(1)} + \\ d_3 \times h_{13_j}^{(1)}) + (d_1 - r_{1_j})(-g_{2_j}^{(1)} + d_1 \times h_{21_j}^{(1)} + d_2 \times h_{22_j}^{(1)} + d_3 \times \\ h_{23_j}^{(1)}) + (d_3 - r_{3_j})(-g_{1_j}^{(2)} + d_1 \times h_{11_j}^{(2)} + d_2 \times h_{12_j}^{(2)} + d_3 \times \\ h_{13_j}^{(2)}) + (d_1 - r_{1_j})(-g_{3_j}^{(2)} + d_1 \times h_{31_j}^{(2)} + d_2 \times h_{32_j}^{(2)} + d_3 \times \\ h_{33_j}^{(2)}) + (d_3 - r_{3_j})(-g_{2_j}^{(3)} + d_1 \times h_{21_j}^{(3)} + d_2 \times h_{22_j}^{(3)} + d_3 \times \\ h_{23_j}^{(3)}) + (d_2 - r_{2j})(-g_{3_j}^{(3)} + d_1 \times h_{31_j}^{(3)} + d_2 \times h_{32_j}^{(3)} + d_3 \times \\ h_{33_j}^{(3)})] + (d_1 - r_{1_j})(-g_{1_j}^{(4)} + d_1 \times h_{11_j}^{(4)} + d_2 \times h_{12_j}^{(4)} + d_3 \times \\ h_{13_j}^{(4)}) - 2(d_2 - r_{2j})(-g_{2_j}^{(4)} + d_1 \times h_{21_j}^{(4)} + d_2 \times h_{22_j}^{(4)} + d_3 \times \\ h_{23_j}^{(4)}) + (d_3 - r_{3_j})(-g_{3_j}^{(4)} + d_1 \times h_{31_j}^{(4)} + d_2 \times h_{32_j}^{(4)} + d_3 \times \\ h_{33_j}^{(4)}) + (d_1 - r_{1_j})(-g_{1_j}^{(5)} + d_1 \times h_{11_j}^{(5)} + d_2 \times h_{12_j}^{(5)} + d_3 \times \\ h_{13_j}^{(5)}) + (d_2 - r_{2_j})(-g_{2_j}^{(5)} + d_1 \times h_{21_j}^{(5)} + d_2 \times h_{22_j}^{(5)} \times d_3 \times \\ h_{23_j}^{(5)}) - 2(d_3 - r_{3_j})(-g_{3_j}^{(5)} + d_1 \times h_{31_j}^{(5)} + d_2 \times h_{32_j}^{(5)} \times d_3 \times \\ h_{33_j}^{(5)}) \Big\}\end{aligned} \tag{A3}$$

where $\mathbf{r}_j = \mathbf{y}_j - \mathbf{d}$ is the incenter of boundary element j in the laboratory reference frame.

Solving eq A1 for **d** is straightforward and yields

$$\begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}^{-1} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

where

$$b_{11} = \frac{1}{15}\sum_j A_j[3h_{21_j}^{(1)} + h_{31_j}^{(2)} + 2(h_{11_j}^{(4)} + h_{11_j}^{(5)})]$$

Substituting of **d** into eq 7 then gives ξ.

$$\begin{aligned} b_{12} = \frac{1}{30}\sum_j A_j[-4h_{21_j}^{(4)} + 3(h_{11_j}^{(1)} + h_{22_j}^{(1)} + h_{32_j}^{(2)} + h_{31_j}^{(3)}) + \\ 2(h_{12_j}^{(4)} + h_{12_j}^{(5)} + h_{21_j}^{(5)})]\end{aligned}$$

$$b_{21} = b_{12}$$

$$b_{13} = \frac{1}{30}\sum_j A_j[-4h^{(5)}_{31_j} + 3(h^{(1)}_{23_j} + h^{(2)}_{11_j} + h^{(2)}_{33_j} + h^{(3)}_{21_j}) + 2(h^{(4)}_{13_j} + h^{(4)}_{31_j} + h^{(5)}_{13_j})]$$

$$b_{22} = \frac{1}{15}\sum_j A_j(-4h^{(4)}_{22_j} + 3(h^{(1)}_{12_j} + h^{(3)}_{32_j}) + 2h^{(5)}_{22_j}))$$

$$b_{23} = \frac{1}{30}\sum_j A_j[-4(h^{(4)}_{23_j} + h^{(5)}_{32_j}) + 3(h^{(3)}_{22_j} + h^{(3)}_{33_j} + h^{(1)}_{13_j} + h^{(2)}_{12_j}) + 2(h^{(4)}_{32_j} + h^{(5)}_{23_j})]$$

$$b_{31} = b_{13}$$

$$b_{32} = b_{23}$$

$$b_{33} = \frac{1}{15}\sum_j A_j(-4h^{(5)}_{33_j} + 3(h^{(2)}_{13_j} + h^{(3)}_{23_j}) + 2h^{(4)}_{33_j}))$$

$$c_1 = \frac{1}{30}\sum_j A_j(-3g^{(1)}_{2_j} - 3g^{(2)}_{3_j} - 2g^{(4)}_{1_j} - 2g^{(5)}_{1_j} - r_{1_j}[3(h^{(1)}_{21_j} + h^{(2)}_{31_j}) + 2(h^{(4)}_{11_j} + h^{(5)}_{11_j})] - r_{2j}[3(h^{(1)}_{11_j} + h^{(3)}_{31_j}) - 4h^{(4)}_{21_j} + 2h^{(5)}_{21_j}] - r_{3_j}[3(h^{(2)}_{11_j} + h^{(3)}_{21_j}) - 4h^{(5)}_{31_j} + 2h^{(4)}_{31_j})]$$

$$c_2 = \frac{1}{30}\sum_j A_j(-3g^{(1)}_{1_j} - 3g^{(3)}_{3j} - 4g^{(4)}_{2_j} - 2g^{(5)}_{2_j} - r_{1_j}(3(h^{(1)}_{22_j} + h^{(2)}_{32_j}) + 2(h^{(4)}_{12_j} + h^{(5)}_{12_j})] - r_{2j}[3(h^{(1)}_{12_j} + h^{(3)}_{32_j}) - 4h^{(4)}_{22_j} + 2h^{(5)}_{22_j})] - r_{3_j}[3(h^{(2)}_{12_j} + h^{(3)}_{22_j}) - 4h^{(5)}_{32_j} + 2h^{(4)}_{32_j}))$$

$$c_3 = \frac{1}{30}\sum_j A_j(-3g^{(2)}_{1_j} - 3g^{(3)}_{2_j} - 2g^{(4)}_{3j} + 4g^{(5)}_{3j} - r_{1_j}[3(h^{(1)}_{23_j} + h^{(2)}_{33_j}) + 2(h^{(4)}_{13_j} + h^{(5)}_{13_j})] - r_{2j}[3(h^{(1)}_{13_j} + h^{(3)}_{33_j}) - 4h^{(4)}_{23_j} + 2h^{(5)}_{23_j}] - r_{3_j}[3(h^{(2)}_{13_j} + h^{(3)}_{23_j}) - 4h^{(5)}_{33_j} + 2h^{(4)}_{33_j})]$$

References

(1) Tanford, C.; Kawahara, K.; Lapanje, S. Proteins as random coils. I. Intrinsic viscosities and sedimentation coefficients in concentrated guanidine hydrochloride. *J. Am. Chem. Soc.* **1967**, *89*, 729−736.

(2) Morel, J.-E.; Taouil, K.; D'hahan, N.; Aguilar, A.; Merah, Z.; Dalbiez, J.-P.; Bayol, P.; Guillo, N.; Patard, L.; Cabane, V.; Ferrari, M.; Picazo, G. F.; Hieu, H. D.; Francin, M. Dimerization of native myosin LC2(RLC)-free subfragment 1 from adult rabbit skeletal muscle. *Biochemistry* **1998**, *37*, 15129−15136.

(3) Garcia de la Torre, J.; Huertas, M. L.; Carrasco, B. Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure. *Biophys. J.* **2000**, *78*, 719−730.

(4) Zhou, H.-X. Calculation of translational friction and intrinsic viscosity. I. General formulation for arbitrarily shaped particles. *Biophys. J.* **1995**, *69*, 2286−2297.

(5) Zhou, H.-X. Calculation of translational friction and intrinsic viscosity. II. Application to globular proteins. *Biophys. J.* **1995**, *69*, 2298−2303.

(6) Allison, S. A. The primary electroviscous effect of rigid polyions of arbitrary shape and charge distribution. *Macromolecules* **1998**, *31*, 4464−4474.

(7) Allison, S. A. Low Reynolds number transport properties of axisymmetric particles employing stick and slip boundary conditions. *Macromolecules* **1999**, *32*, 5304−5312.

(8) Aragón, S. R.; Hahn, D. K. Precise boundary element computation of protein transport properties: Diffusion tensors, specific volume, and hydration. *Biophys. J.* **2006**, in press.

(9) Wegener, W. A. On an exact starting expression for macromolecular hydrodynamic models. *Biopolymers* **1986**, *25*, 627−637.

(10) Youngren, G. K.; Acrivos, A. Stokes flow past a particle of arbitrary shape: A numerical method of solution. *J. Fluid Mech.* **1975**, *69*, 377−402.

(11) Aragon, S. R. A precise boundary element method for macromolecular transport properties. *J. Comput. Chem.* **2004**, *25*, 1191−1205.

(12) Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Du Croz, J.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D. *LAPACK Users's Guide*, 3rd ed.; SIAM: Philadelphia, PA, 1999.

(13) Simha, R. The influence of Brownian motion on the viscosity of solutions. *J. Phys. Chem.* **1940**, *44*, 25−34.

(14) Harding, S. E. The intrinsic viscosity of biological macromolecules. Progress in measurement, interpretation and application to structure in dilute solution. *Prog. Biophys. Mol. Biol.* **1997**, *68*, 207−262.

(15) Madeley, C. R.; Field, A. M. *Virus Morphology*, 2nd ed.; Churchill Livingstone: New York, 1988.

(16) Garcia de la Torre, J.; Bloomfield, V. A. Hydrodynamic properties of macromolecular complexes. IV. Intrinsic viscosity, with application to once-broken rods and multisubunit proteins. *Biopolymers* **1978**, *17*, 1605−1627.

(17) Einstein, A. Berichtigung zu meiner Arbeit: "Eine neue Bestimmung der Molekulardimensionen". *Ann. Phys.* **1911**, *34*, 591−597.

(18) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548−558.

(19) Shoemaker, D. P.; Garland, C. W.; Nibler, J. W. *Experiments in Physical Chemistry*, 5th ed.; McGraw-Hill: New York, 1989.

(20) Monkos, K. Viscometric study of human, bovine, equine and ovine haemoglobin in aqueous solution. *Int. J. Biol. Macromol.* **1994**, *16*, 31−35.

(21) Quershi, S. H.; Moza, B.; Yadav, S.; Ahmad, F. Configurational and thermodynamic characterization of the molten globule state occurring during unfolding of cytochrome-c by weak salt. *Biochemistry* **2003**, *42*, 1684−1695.

(22) Fling, M.; Horowitz, N. H.; Heinemann, S. F. The isolation and properties of crystalline tyrosinase from neurospora. *J. Biol. Chem.* **1963**, *238*, 2045−2053.

(23) Larew, L.; Walters, R. W. A kinetic, chromatographic method for studying protein hydrodynamic behavior. *Anal. Biochem.* **1987**, *164*, 537−546.

(24) Walters, R. W.; Graham, J. F.; Moore, R. M.; Anderson, D. J. Protein diffusion coefficient measurements by laminar flow analysis: Method and applications. *Anal. Biochem.* **1984**, *140*, 190−195.

(25) Clark, S. M.; Leaist, D. G.; Konermann, L. Taylor dispersion monitored by electrospray mass spectrometry: A novel approach for studying diffusion in solution. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1454−1462. Converted from $D_t = 13.5$ cm^2/s at 24 °C.

(26) Buzzell, J. G.; Tanford, C. The effect of charge and ionic strength on the viscosity of ribonuclease. *J. Phys. Chem.* **1956**, *60*, 1204−1207.

(27) Kupke, D. W.; Hodgkins, M. G.; Beams, J. W. Simultaneous determination of viscosity and density of protein solutions by magnetic suspension. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2258−2262.

(28) Creeth, J. M. Studies of free diffusion in liquids with the Rayleigh method. III. The analysis of known mixtures and some preliminary investigations with proteins. *J. Phys. Chem.* **1958**, *62*, 66−74. Converted from $D_t = 12.11$ cm^2/s at 25 °C.

(29) Wetlaufer, D. B. Osmometry and general characterization of alpha-lactalbumin. *C. R. Trav. Lab. Carlsberg* **1961**, *32*, 125−138.

(30) Dolgikh, D. A.; Gilmanshin, R. I.; Braznikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; Venyaminov, S. Y.; Ptitsyn, O. B. Alpha-lactalbumin: Compact state with fluctuating tertiary structure. *FEBS Lett.* **1981**, *136*, 311−315.

(31) Polson, A. Über die berechnung der gestalt von proteinmolekülen. *Kolloid Z.* **1939**, *88*, 51−61.

(32) Gordon, W. G.; Semmett, W. F. Isolation of crystalline α-lactalbumin from milk. *J. Am. Chem. Soc.* **1953**, *75*, 328−330.

(33) Monkos, K. Concentration and temperature dependence of viscosity in lysozyme aqueous solutions. *Biochim. Biophys. Acta* **1997**, *1339*, 304−310.

(34) Kamiyama, T.; Morita, M.; Kimura, T. Rheological study of lysozyme in dimethyl sulfoxide + water solution at 298.15 K. *J. Chem. Eng. Data* **2004**, *49*, 1350−1353.

(35) Luzzati, A.; Champagne, M. The molecular weight and dimensions of the lysozyme molecule in solution. *C. R. Hebd. Seances Acad. Sci.* **1957**, *244*, 2930−2932.

(36) Sophianopoulos, A. J.; Rhodes, C. K.; Holcomb, D. N.; van Holde, K. E. Physical studies of lysozyme. I. Characterization. *J. Biol. Chem.* **1962**, *237*, 1107−1112.

(37) Bouthier, M.; Quaranta, C.; Savary, J.; Reynaud, J. Détermination du nombre de viscosité limite des proteins. Application aux anhydrases carboniques erythrocytaires humaines. *J. Chim. Phys. Phys.-Chim. Biol.* **1976**, *73*, 776−782.

(38) Dubin, S. B.; Clark, N. A.; Benedek, G. B. Measurements of the rotational diffusion coefficient of lysozyme by depolarized light scattering. I. Configuration of lysozyme in solution. *J. Chem. Phys.* **1971**, *54*, 5158−5164.

(39) Harding, S. E. Viscometric parameters for myoglobin. *IRCS Med. Sci.* **1980**, *8*, 610.

(40) Ehrenberg, A. Determination of molecular weights and diffusion coefficients in the ultracentrifuge. *Acta Chem. Scand.* **1957**, *11*, 1257−1270.

(41) Riveros-Moreno, V.; Wittenberg, J. B. The self-diffusion coefficients of myoglobin and hemoglobin in concentrated solutions. *J. Biol. Chem.* **1972**, *247*, 895−901.

(42) Edelhoch, H.; Steiner, R. F. Structural transitions of soybean trypsin inhibitor. II. The denatured state in urea. *J. Biol. Chem.* **1963**, *238*, 931−938.

(43) Rackis, J. J.; Sasame, H. A.; Mann, R. K.; Anderson, R. L.; Smith, A. K. Soybean trypsin inhibitors: Isolation, purification and physical properties. *Arch. Biochem. Biophys.* **1962**, *98*, 471−478. Calculated from the Svedberg equation using $s_{20,\mathrm{w}} = 2.29$ S and $\bar{\nu} = 0.735$ cm^3/g and assuming solvent density $\rho = 1.0$ g/cm^3.

(44) Harris, J. L. Effect of urea on trypsin and alpha-chymotrypsin. *Nature* **1956**, *177*, 471−473.

(45) Cunningham, L. W., Jr.; Tietze, F.; Green, N. M.; Neurath, H. Molecular kinetic properties of trypsin and related proteins. *Discuss. Faraday Soc.* **1953**, *13*, 58−67.

(46) Kay, C. M.; Smillie, L. B.; Hilderman, F. A. The molecular weight of trypsinogen. *J. Biol. Chem.* **1961**, *236*, 118−121.

(47) Tietze, F. Molecular-kinetic properties of crystalline trypsinogen. *J. Biol. Chem.* **1953**, *204*, 1−11.

(48) Schwert, G. W.; Kaufman, S. The molecular size and shape of the pancreatic proteases. III. α-chymotrypsin. *J. Biol. Chem.* **1951**, *190*, 807−816.

(49) Tanford, C. Protein denaturation. *Adv. Protein Chem.* **1968**, *23*, 121−282.

(50) Schwert, G. W. The molecular size and shape of the pancreatic proteases. II. Chymotrypsinogen. *J. Biol. Chem.* **1951**, *190*, 799−806. Calculated from $[\eta] = \xi\bar{\nu}$, where ξ is the viscosity factor and $\bar{\nu}$ is the partial molar volume.

(51) Wilcox, P. E.; Kraut, J.; Wade, R. D.; Neurath, H. The molecular weight of alpha-chymotrypsinogen. *Biochim. Biophys. Acta* **1957**, *24*, 72−78.

(52) Armstrong, J. M.; Myers, D. V.; Verpoorte, J. A.; Edsall, J. T. Purification and properties of human erythrocite carbonic anhydrases. *J. Biol. Chem.* **1966**, *241*, 5137−5149.

(53) Wong, K.-P.; Tanford, C. Denaturation of bovine carbonic anhydrase b by guanidine hydrochloride. A process involving sequential conformational transitions. *J. Biol. Chem.* **1973**, *248*, 8518−8523.

(54) Bürgi, W.; Schmid, K. Preparation and properties of Zn-alpha2-glycoprotein of normal human plasma. *J. Biol. Chem.* **1961**, *236*, 1066−1074. Calculated from the Svedberg equation using $s_{20,\mathrm{w}} = 3.1$ S and $\bar{\nu} = 0.706$ cm^3/g and assuming solvent density $\rho = 1.0$ g/cm^3.

(55) Blumenfeld, O. O.; Léonis, J.; Perlmann, G. E. The effect of guanidine hydrochloride on crystalline pepsin. *J. Biol. Chem.* **1960**, *235*, 379−382.

(56) Edelhoch, H. The denaturation of pepsin. I. Macromolecular change. *J. Am. Chem. Soc.* **1957**, *79*, 6100−6109.

(57) Neurath, H.; Cooper, G. R.; Erickson, J. O. The shape of protein molecules. II. Viscosity and diffusion studies of native proteins. *J. Biol. Chem.* **1941**, *138*, 411−436.

(58) Cohen, L. B. Viscosity of g-ADP and g-ATP actin. *Arch. Biochem. Biophys.* **1966**, *117*, 289−295.

(59) Lanni, F.; Ware, B. R. Detection and characterization of actin monomers, oligomers, and filaments in solution by measurement of fluorescence photobleaching recovery. *Biophys. J.* **1984**, *46*, 97−110.

(60) Newman, J.; Estes, J. E.; Selden, L. A.; Gershman, L. C. The presence of oligomers at subcritical actin concentrations. *Biochemistry* **1985**, *24*, 1538−1544.

(61) Takagi, T.; Isemura, T. Extent of renaturation of reduced taka-amylase a before reformation of disulfide bonds. *Biochim. Biophys. Acta* **1966**, *130*, 233−240.

(62) Isemura, T.; Fujita, S. Physicochemical studies on taka-amylase a. I. Size and shape determination by the measurement of sedimentation constant, diffusion constant, and viscosity. *J. Biochem. (Tokyo)* **1957**, *44*, 443−450.

(63) Muzammil, S.; Kumar, Y.; Tayyab, S. Molten globule-like state of human serum albumin at low pH. *Eur. J. Biochem.* **1999**, *266*, 26−32.

(64) Oncley, J. L.; Scatchard, G.; Brown, A. Physical-chemical characteristics of certain of the proteins of normal human plasma. *J. Phys. Colloid Chem.* **1947**, *51*, 184−198.

(65) Monkos, K. On the hydrodynamics and temperature dependence of the solution conformation of human serum albumin from viscometry approach. *Biochim. Biophys. Acta* **2004**, *1700*, 27−34.

(66) Pedersen, K, O. *Ultracentrifugal Studies on Serum and Serum Fractions*; Almqvist and Wiksells: Upsala, 1945.

(67) Charlwood, P. A. Sedimentation and diffusion of human albumins. I. Normal human albumins at low concentration. *Biochem. J.* **1952**, *51*, 113−118.

(68) Phelps, R. A.; Cann, J. R. On the modification of conalbumin by acid. II. Effect of pH and salt concentration on the sedimentation behavior, viscosity and osmotic pressure of conalbumin solutions. *Arch. Biochem. Biophys.* **1956**, *61*, 51−71.

(69) Yajima, H.; Yamamoto, H.; Nagaoka, M.; Nakazato, K.; Ishii, T.; Niimura, N. Small-angle neutron scattering and dynamic light scattering studies of N- and C-terminal fragments of ovotransferrin. *Biochim. Biophys. Acta* **1998**, *1381*, 68−76. Converted from $D_t = 6.7 \times 10^{-7}$ cm^2/s at 25 °C.

(70) Léger, D.; Verbert, A.; Loucheux, M.-H.; Spik, G. Etude de la moléculaire de la lactotransferrine et de la serotransferrine humaines. *Ann. Biol. Anim. Biochem. Biophys.* **1977**, *17*, 737−747.

(71) Querinjean, P.; Masson, P. L.; Heremans, J. F. Molecular weight, single-chain structure and amino acid composition of human lactoferrin. *Eur. J. Biochem.* **1971**, *20*, 420−425.

(72) Wood, E.; Dalgleish, D.; Bannister, W. Bovine erythrocite cupro-zinc protein. 2. Physicochemical properties and circular dichroism. *Eur. J. Biochem.* **1971**, *18*, 187−193.

(73) Bunville, L. G. Ph.D. Thesis, Iowa State University, 1959.

(74) McKenzie, H. A.; Sawyer, W. H. Effect of pH on β-lactoglobulin. *Nature* **1967**, *214*, 1101−1104.

(75) Bull, H. B.; Currie, B. T. Osmotic pressure of β-lactoglobulin solutions. *J. Am. Chem. Soc.* **1946**, *68*, 742−745.

(76) Fox, K. K.; Holsinger, V. H.; Posati, L. P.; Pallansch, M. J. Separation of beta-lactoglobulin from other milk serum proteins by trichloroacetic acid. *J. Dairy Sci.* **1967**, *50*, 1363−1367.

(77) Ogston, A. G. The Guoy diffusiometer; further calibration. *Proc. R. Soc. London* **1949**, *196*, 272−285.

(78) Samsonov, G. V.; Ponomareva, R. B.; Bolotina, I. A. Investigation of the physical-chemical properties of α-chymotrypsin and its B and C chain. *Biofizika* **1965**, *10*, 520−522.

(79) Schwert, G. W.; Kaufman, S. The molecular size and shape of the pancreatic proteases. III. α-Chymotrypsin. *J. Biol. Chem.* **1951**, *190*, 807−816.

(80) Kunitz, M.; Northrop, J. H. Crystalline chymo-trypsin and chymo-trypsinogen: I. Isolation, crystallization, and general properties of a new proteolytic enzyme and its precursor. *J. Gen. Physiol*. **1935**, *18*, 433−458.

(81) McCubbin, W. D.; Kay, C. M. Molecular weight studies on concanavalin A. *Biochem. Biophys. Res. Commun.* **1971**, *44*, 101−109.

(82) Huet, M.; Claverie, J.-M. Sedimentation studies of the reversible dimer−tetramer transition kinetics of concanavalin A. *Biochemistry* **1978**, *17*, 236−241.

(83) McVittie, J. D.; Esnouf, M. P.; Peacocke, A. R. The denaturation of chicken-muscle triosephosphate isomerase in guanidinium chloride. *Eur. J. Biochem.* **1977**, *81*, 307−315. D_t calculated from the Svedberg equation using $s_{20,w} =$ 3.75 S and $\bar{\nu} = 0.740$ cm^3/g and assuming solvent density $\rho = 1.0$ g/cm^3.

(84) Frénoy, J.-P. Effect of physical environment on the conformation of ricin. Influence of low pH. *Biochem. J.* **1986**, *240*, 221−226.

(85) Kabat, E. A.; Heidelberger, M.; Bezer, A. E. A study of the purification and properties of ricin. *J. Biol. Chem.* **1947**, *168*, 629−639.

(86) Sanders, A. H.; Purich, D. L.; Cannell, D. S. Oxygenation of hemoglobin: Correspondence of crystal and solution properties using translational diffusion constant measurements. *J. Mol. Biol.* **1981**, *147*, 583−595.

(87) Reynolds, J. A.; Schlesinger, M. J. Conformational states of the subunit of *Escherichia coli* alkaline phosphatase. *Biochemistry* **1967**, *6*, 3552−3559.

(88) Altman, P. L.; Dittmer, D. S. *Biology Data Book*, 2nd ed.; FASEB: Bethesda, Maryland, 1972; Vol. I. Calculated from the Svedberg equation using $s_{20,w} = 6.1$ S and $\bar{\nu} = 0.725$ cm^3/g and assuming solvent density $\rho = 1.0$ g/cm^3.

(89) Wu, J.-Y.; Yang, J. T. Physicochemical characterization of citrate synthase. *J. Biol. Chem.* **1970**, *24*, 212−218.

(90) Wong, S. C. K.; Hall, D. C.; Josse, J. Constitutive inorganic pyrophosphatase of *Escherichia coli*. III. Molecular weight and physical properties of the enzyme and its subunits. *J. Biol. Chem.* **1970**, *245*, 4335−4341.

(91) Castellino, J.; Barker, R. Examination of the dissociation of multichain proteins in guanidine hydrochloride by membrane osmometry. *Biochemistry* **1968**, *7*, 2207−2217.

(92) Stellwagen, E.; Schachman, H. K. The dissociation and reconstitution of aldolase. *Biochemistry* **1962**, *1*, 1056−1069.

(93) Hass, L. F. Aldolase dissociation into subunits by reaction with succinic anhydride. *Biochemistry* **1964**, *3*, 535−541.

(94) Glikina, M. V.; Finogenov, P. A. Investigation of muscular aldolase in various stages of isolation. *Biokhimiya* **1950**, *15*, 457−464.

(95) Kawahara, K. Evaluation of diffusion coefficients of proteins from sedimentation boundary curves. *Biochemistry* **1969**, *8*, 2551−2557.

(96) Taylor, J. F.; Green, A. A.; Cori, G. T. Crystalline aldolase. *J. Biol. Chem.* **1948**, *173*, 591−604.

(97) Christen, P.; Göschke, H.; Leuthardt, F.; Schmid, A. Über die aldolase der kaninchenleber molekulargewicht, dissoziation in untereinheiten. *Helv. Chim. Acta* **1965**, *48*, 1050−1056.

(98) Tanford, C.; Lovrien, R. Dissociation of catalase into subunits. *J. Am. Chem. Soc.* **1962**, *84*, 1892−1896.

(99) Samejima, T.; Yang, J. T. Reconstitution of acid-denatured catalase. *J. Biol. Chem.* **1963**, *238*, 3256−3261.

(100) Sumner, J.; Gralén, N. The molecular weight of crystalline catalase. *Science* **1938**, *87*, 284.

(101) Samejima, T. Splitting of the catalase molecule by alkali treatment. *J. Biochem.* **1959**, *46*, 155−159.

(102) Sund, H.; Weber, K. Studies on the lactose-splitting enzyme. XIII. Quantity and configuration of beta-galactosidase from *E. coli*. *Biochem. Z.* **1963**, *363*, 24−34.

(103) Tellam, R.; Winzor, D. J. Self-association of alpha-chymotrypsin at low ionic strength in the vicinity of its pH optimum. *Biochem. J.* **1977**, *161*, 687−694.

(104) McEvily, A. J.; Harrison, J. H. Subunit equilibria of porcine heart citrate synthase. Effects of enzyme concentration, pH, and substrates. *J. Biol. Chem*. **1986**, *261*, 2593−2598.

(105) Avaeva, S.; Grigorjeva, D.; Mitkevich, V.; Sklyankina, V.; Varfolomeyev, S. Interaction of *Escherichia coli* inorganic pyrophosphatase active sites. *FEBS Lett.* **1999**, *464*, 169−173.

(106) Malinowski, D. P.; Fridovich, I. Subunit association and side-chain reactivities of bovine erythrocyte superoxide dismutase in denaturing solvents. *Biochemistry* **1979**, *18*, 5055−5060.

(107) Lebherz, H. G. Stability of quaternary structure of mammalian and avian fructose diphosphate aldolases. *Biochemistry* **1972**, *11*, 2243−2250.

(108) Senear, D. F.; Teller, D. C. Thermodynamics of concanavalin a dimer−tetramer self-association: Sedimentation equilibrium studies. *Biochemistry* **1981**, *20*, 3076−3083.

(109) Garen, A.; Levinthal, C. A fine-structure genetic and chemical study of the enzyme alkaline phosphatase of *E. coli*. I. Purification and characterization of alkaline phosphatase. *Biochim. Biophys. Acta* **1960**, *38*, 470−483.

(110) Borchert, T. V.; Abagyan, R.; Jaenicke, R.; Wierenga, R. K. Design, creation, and characterization of a stable, monomeric triosephosphate isomerase. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1515−1518.

(111) Chiancone, E.; Gilbert, L. M.; Gilbert, G. A.; Kellett, G. L. Dissociation of hemoglobin into subunits. II. Human oxyhemoglobin: Gel filtration studies. *J. Biol. Chem*. **1968**, *243*, 1212−1219.

(112) Lukin, J. A.; Kontaxis, G.; Simplaceanu, V.; Yuan, Y.; Bax, A.; Ho, C. Quaternary structure of hemoglobin in solution. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 517−520.

(113) Silva, M. M.; Rogers, P. H.; Arnone, A. A third quaternary structure of human hemoglobin A at 1.7-Å resolution. *J. Biol. Chem*. **1992**, *267*, 17248−17256.

(114) Rutenber, E.; Katzin, B. J.; Montfort, W.; Villafranca, J. E.; Ernst, S.; Collins, E. J.; Mlsna, D.; Ready, M. P.; Robertus, J. D. Crystallographic refinement of ricin to 2.5 Å. *Proteins: Struct., Funct., Genet.* **1991**, *10*, 240−250.

(115) Bushnell, G. W.; Louie, G. V.; Brayer, G. D. High-resolution structure of horse heart cytochrome C. *J. Mol. Biol.* **1990**, *214*, 585−595.

(116) Sweeney, E. C.; Tonevitsky, A. G.; Temiakov, D. E.; Agapov, I. I.; Saward, S.; Palmer, R. A. Preliminary crystallographic characterization of ricin agglutinin. *Proteins* **1997**, *28*, 586−589.

CT600062Y

J. Chem. Theory Comput. **2006,** *2,* 1429−1434

JCTC Journal of Chemical Theory and Computation

Automatic Control of Solvent Density in Grand Canonical Ensemble Monte Carlo Simulations

Joshua A. Speidel,[†] Jason R. Banfelder,[†,‡] and Mihaly Mezei*[,§]

Department of Physiology and Biophysics and HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York 10021, and Department of Molecular Physiology and Biophysics, Box 1218, Mount Sinai School of Medicine, NYU, One Gustave L. Levy Place, New York, New York 10029

Received February 1, 2006

Abstract: We present automated methods for determining the value of Adams' *B* parameter corresponding to a target solvent density in grand canonical ensemble Monte Carlo simulations. The method found to work best employs a proportional-integral control equation commonly used in industrial process control applications. We show here that simulations employing this method rapidly converge to the desired target density. We further show that this method is robust over a wide range of system sizes. This advance reduces the overall CPU time and user effort in determining the equilibrium excess chemical potential in these systems.

Introduction

Simulations in the grand canonical ensemble have two unique and attractive features: they can be used to identify the chemical potential without additional costly free-energy simulations, and the combination of insertions and deletions results in large molecular displacements during the simulation that are not possible in any of the closed ensembles at condensed phase densities. These large displacements are particularly useful for solvating isolated pockets, as exist in most proteins. This simulation method can also be used to solvate lipid membranes, allowing penetration of water molecules deep into the interior of the lipid far more rapidly than with molecular dynamics.[1]

Systems solvated this way can serve as an initial configuration of a molecular dynamics run.[2] Alternatively, the simulation can be extended, and solvation sites can be deduced from it using, e.g., the generic site approach.[3] It has also been demonstrated that potential of mean force calculations benefit from the GCE framework: the changes in the distance between bulky solutes that these simulations require can be enabled through the removal or insertion of intervening solvents, rather than waiting for them to diffuse out/into the region between the solutes.[3]

A limitation of this simulation method is that insertions and deletions are currently feasible only for small, neutral molecules such as water. This is because the probability of accepting a random insertion of a bulky or charged molecule in a solvated system is very low. The insertion of water molecules has only become practical with the introduction of cavity biased sampling. A second limitation of this method is that when simulating the solvation of a system, the chemical potential yielding the target density is initially unknown. Therefore, prior to simulating at a target density, a tuning phase is necessary to identify the chemical potential parameter that yields the target density. In practice, this required several runs with the chemical potential parameter adjusted each time based on the results of previous run(s). Such manual interventions not only consume human time but also in general lengthen the overall time of the simulation. This article presents three procedures to perform this tuning without user intervention and compares their performances for systems of varying sizes and compositions.

---

* Corresponding author phone: (212)241-2186; fax: (212)860-3369; e-mail: Mihaly.Mezei@mssm.edu.

[†] Department of Physiology and Biophysics, Weill Cornell Medical College.

[‡] HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College.

[§] Department of Molecular Physiology and Biophysics, Mount Sinai School of Medicine, NYU.

10.1021/ct0600363 CCC: $33.50

Background

Monte Carlo simulations in the Grand Canonical (T,V,μ) Ensemble (GCE) are conveniently performed with the introduction of the parameter B which is related to the excess chemical potential μ' as

$$\mu' = kTB - kT\ln\langle N\rangle \tag{1}$$

where $\langle N\rangle$ is the average number of particles,[4] k is the Boltzmann constant, and T is the absolute temperature. Note that B depends not only on the excess chemical potential but also on the system size and composition as well. In addition to the conventional translations and rotations, simulations in the GCE require periodic insertions and deletions of molecules. The Cavity Biased variant (CB/GCE)[5,6] attempts insertion of a molecule only if a cavity of an appropriate radius is found and accepts the insertion with probability

$$P_i = \min\{1, P_N^{\mathrm{cav}} \exp[B + (E(r_{N+1}) - E(r_N))/kT]/(N+1)\} \tag{2}$$

where $E(r_N)$ is the potential energy of the system of N particles at configuration r_N, and P_N^{cav} is the probability of finding a cavity of a specific size. To maintain microscopic reversibility, the probability of a deletion of a particle is given by

$$P_d = \min\{1, N\exp[-B + (E(r_N) - E(r_{N-1}))/kT]/P_{N-1}^{\mathrm{cav}}\} \tag{3}$$

The robustness of the CB/GCE technique was demonstrated by its robustness in modeling solvent molecules in crystal hydrates and protein active sites.[3,5,7,8] In these simulations, the value of B directly affects the probability of a successful insertion or deletion. An automated method to identify B must correctly change B when the calculated density differs from the target density. At a given temperature, there is a natural fluctuation of the density that will occur for any given chemical potential. Therefore, we are primarily concerned with identifying the correct B parameter such that the mean density equals the target density over an equilibrated portion of the simulation.

Methods

Iterative Tuning Based on Fluctuations. Fluctuation (F) in the number of molecules is related to the mean number of molecules and the B parameter through eq 4:[9]

$$\Delta\langle N\rangle/\Delta B \approx F_{N,B} = \langle N^2\rangle - \langle N\rangle^2 \tag{4}$$

B is tuned in an iterative process that assumes that the relation of eq 4 is constant to a reasonable extent over a finite range of $\langle N\rangle$'s. Initially, $\Delta\langle N\rangle/\Delta B$ is either calculated or estimated experimentally from the isothermal compressibility of the pure liquid at the target density[4] ($F_{N,B}^{\mathrm{initial}}$). Each iteration, i, first simulates the CB/GCE system for X^{e}_{MC} steps to allow equilibration with the newly chosen B value, followed by X^{s}_{MC} steps to gather statistics for $\langle N_i\rangle$. The sensitivity coefficient $S_{N,B}^{i}$ is calculated as a linear combination of the pure liquid value and the value calculated from $\langle N^2\rangle - \langle N\rangle^2$ at the end of iteration i, using the whole run:

$$S_{N,B}^{i} = (F_{N,B}^{\mathrm{initial}} + i{*}F_{N,B}^{i})/(i+1) \tag{5}$$

The change in B is determined by eq 6:

$$\Delta B = (N_{\mathrm{target}} - \langle N_i\rangle)/(S_{N,B}^{i}) \tag{6}$$

The maximum $|\Delta B|$ value is limited by default to 1.0 to further dampen oscillations in B. The current implementation gathers statistics for the fluctuation cumulatively over the whole simulation. The use of the fluctuation in $\langle N\rangle$ calculated separately in each iteration is precluded by the slow convergence of fluctuations, i.e., the simulation would not converge in X^{s}_{MC}, steps unless it was very long.

Iterative Tuning with Empirical Estimates of the Sensitivity Coefficient. In this method the sensitivity coefficient is a scaling factor that incorporates the effect of a change in B on the mean number of particles (eq 7).

$$S_{N.B}^{i} = \sum_{j=0}^{i}(\langle N_j\rangle - \langle N_{j-1}\rangle)/(B_j - B_{j-1}) \tag{7}$$

The change in B is then determined according to eq 8:

$$\Delta B = (N_{\mathrm{target}} - \langle N_i\rangle)/S_{N,B}^{i} \tag{8}$$

As with the fluctuation method, the $|\Delta B|$ is limited to the default value of 1.0. Also, there is a filter whereby if in the previous iteration the change in $\langle N\rangle$ is of opposite sign of the change in B, in the next iteration B is unchanged. Such an occurrence clearly indicates inadequate equilibration and/or statistics because by definition as B increases, so should $\langle N\rangle$. This filter implicitly increases X^{e}_{MC} in these instances allowing the system more time to adjust to the new value of B. This has the benefit of simplifying the choice of X^{e}_{MC} and X^{s}_{MC} for iteration lengths that are providing adequate statistics for $S_{N,B}^{\mathrm{i}}$ but allow for frequent enough changes to reach convergence as fast as possible.

Tuning Using Process Control Principles. The canonical proportional-integral-derivative control equation (PID) commonly used in engineering control applications is eq 9[10]

$$\mathrm{MV}(t) = K_{\mathrm{C}}\epsilon(t) + \frac{K_{\mathrm{C}}}{\tau_{\mathrm{I}}}\int_0^t \epsilon(t)\mathrm{d}t + K_{\mathrm{c}}\tau_{\mathrm{D}}\frac{\mathrm{d}\epsilon}{\mathrm{d}t} + c_{\mathrm{s}} \tag{9}$$

where MV is the manipulated variable, K_{C} is the proportional gain, τ_{I} is the integral time, τ_{D} is the derivative time, c_{s} is the controller bias, and ϵ is the current deviation of the process variable from its target. It has been shown that the dynamic stability of a controlled system can be sensitive to the selection of the derivative time. Therefore, the derivative term is frequently omitted unless empirically shown to be necessary.[11] When applying this equation to our simulation systems, the value we seek to control is the density of bulk water. In eq 9, ϵ is the deviation away from the desired bulk density. To achieve the target density, the manipulated variable in our system is B, whose effect is described in the Introduction.

We further implement the differential form of the control equation:

$$\frac{dB}{dt} = K_C\frac{d\epsilon}{dt} + \frac{K_C}{\tau_I}\epsilon \tag{10}$$

This transformation solves the so-called ‘integral windup’ problem, whereby an historic period of large deviation dominates the integral term.[11] It has the added benefit of eliminating the controller bias as another tuning parameter.

The process control equation has an implicit time dependence that is not present in Monte Carlo simulations. To emphasize this, we rewrite the control equation once more in finite difference form in terms of the simulation step number, i:

$$B_{i+1} = B_i + K_c(\epsilon_i - \epsilon_{i-1}) + \frac{K_C}{\tau_I}\epsilon_i \tag{11}$$

Unlike the first two methods, this technique changes the B parameter at every insertion/deletion attempt.

Calculations

All three methods were implemented into the Monte Carlo program MMC.[12]

Grand Canonical Ensemble Simulation Systems. For each of the three tuning methods, four simulations of differing size and components were monitored for their ability to identify the value of Adams’ B parameter that yields the target density.

All our systems targeted equilibrium with bulk water at 310 K, 0.997 g/mL. The first three simulations were small, medium, and large pure TIP3P water simulations. With the cubic edge length of the boxes equal to 16.48, 34.99, and 75.26 Å, respectively, the simulations targeted 142, 1420, and 14 200 waters, respectively. To mimic systems with a solute, one water was treated as the solute, and the density was monitored in the region outside a cubic cell centered at the center-of-mass of this water. The position of this water molecule was kept fixed throughout the simulation. The edge of the cube, enclosing the waters perturbed by the solute, was set to 6 Å.

The fourth simulation included a trypsin protein with a bound benzamidine.[13] The protein simulation volume had an edge length of 75.26 Å. The center of mass of the protein was placed at the center of the simulation unit cell. The density was monitored in the region outside a rectangular volume centered at the origin with minimum x, y, z and maximum x, y, z coordinates of −26, −24, −25, 24, 26, and 24, respectively. At the target density, approximately 10 124 water molecules fill this monitored volume.

All simulations used scaled force biased sampling[14] where the level of biasing was scaled down in the vicinity of the solute. The water potential used was TIP3P,[15] and the solute−solvent interactions were described by the CHARMM force field.[16] The solvent−solvent potential was cut off at 12 Å. Fifty water molecules were randomly placed in the simulation cell at the start of the simulation. The solute was kept fixed in all simulations. At each MC step a randomly selected solvent was translated with a maximum step size of 0.275 Å and rotated with a maximum angle of 60°. Initially the B parameter was set to 1.0.

The small water simulations were run for 5M steps. The medium and large water simulations were run for 25M and 50M steps, respectively. The large protein simulation was run for 50M steps.

Sensitivity Coefficient Based Tuning Simulations. As the system size increases, generally more time is needed to average the properties of the system and to reach equilibrium after a change in the control parameter. For these simulations, the small water simulation used 10K steps of equilibration followed by 10K steps to gather statistics. The medium water simulation (1420 water molecule target) used 50K steps of equilibration followed by 50K steps to gather statistics. The large water and protein simulations used 100K steps for equilibrium and 100K steps to gather statistics. The density and B parameter were reported at 10 000 equally spaced steps in each simulation, i.e., every 200 steps for the small water simulation.

Fluctuation Based Tuning Simulations. These simulations are the same as those of the sensitivity coefficient based simulations, except that the initial configuration has a number of waters approximately equal to the target number of waters. This was done because the accumulated fluctuation generated by starting with 50 water molecules was thought to prevent the B parameter from correctly responding to changes in $\langle N \rangle$. In an attempt to minimize that influence, we therefore performed an initial short CB/GCE simulation with B set to 10.0 until the desired number of waters was reached. The system that resulted from that simulation was used as the starting structure of the fluctuation based tuning simulations.

Proportional Integral Control Based Tuning Simulations. Simulations were run for the same lengths as the fluctuation and sensitivity coefficient based simulations. The B parameter in these simulations is changed at each step of the simulation. The two controller tuning parameters (K_C and τ_I) were determined to be −112 and 13,000 respectively, by following the open loop protocol of Ziegler and Nichols for the large protein−ligand system.[17] This method of tuning the control parameters is based on the response of the system to a step change in the manipulated variable (B, in our case). However, instead of considering the steady-state result (i.e., the equilibrated value of the density) in response to the step change, it considers the delay of the response of the controlled variable (the so-called ‘dead-time’ in the vernacular of control engineering) and the initial rate of the response. This method has the advantage that it can be applied even to nonself-regulating systems (i.e., those systems that have an unbound response to the step change) because it only depends on the dynamic character of the *initial* response to the perturbation.

Once the dead-time and initial response rate for a process in known, the Ziegler-Nichols method provides rules for determining the PID controller constants that appear in eq 9. Evaluation of the dead-time and initial response rate is somewhat subjective as it involves an assessment of how long the system took to ‘appreciably respond’ after the perturbation was applied. However, as long as the dead-time/initial response rate pairs are self-consistent, the Ziegler-Nichols tuning method will yield acceptable (and usually indistinguishable) empirical performance.

Table 1. Testing Convergence

	A. convergence to target density[a]			B. convergence in *B*[b]		
	fluctuation	sensitivity coefficient	PIC	fluctuation	sensitivity coefficient	PIC
small water	+	+	+	/	×	−4.46 ± 0.35
medium water	−	×	+	×	×	−2.42 ± 0.14
large water	−	+	+	×	−0.23 ± 1.07	−0.18 ± 0.08
protein water	−	×	+	×	×	−0.22 ± 0.15

[a] An '×' indicates that the density was not normal for either the last 50 or 25% of the simulation. A '−' indicates that the density was normal but was not converged for the last 50 or 25% of the simulation. A '+' indicates that the density converged to the target density by the criteria described in the text. [b] A '/' indicates that *B* was not normally distributed in one of the block average segments analyzed. An '×' indicates that the simulation did not converge to the target density. A '−' indicates that the simulation was converged to the target density but not converged to a single *B* value. Simulations that converged to the target and a single *B* have the value of *B* ± the standard deviation in the table.

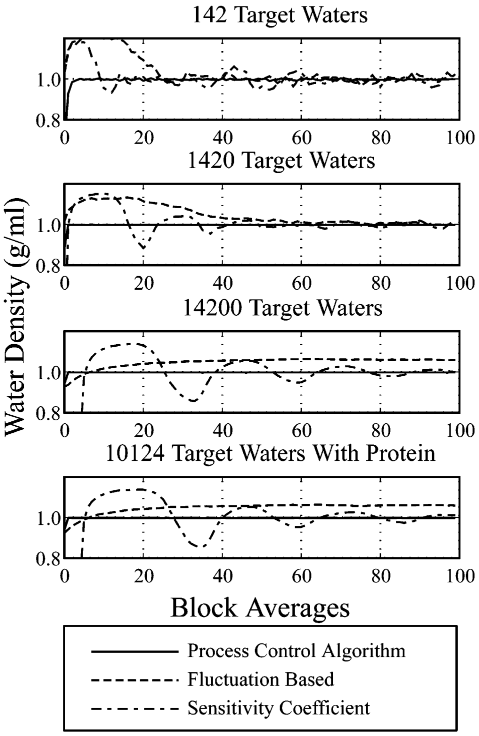


Figure 1. Density block average. The block average of the density is plotted for each simulation. The solid line is the PIC method, the dashed line is the fluctuation based method, and the dash−dot line is the sensitivity coefficient method. Each simulation is divided into 100 equally sized blocks. The block averages for the small, medium, and large water simulation were 50 000 steps, 250 000, and 500 000 steps per block average. The large protein and water simulation used 500 000 steps per block average.

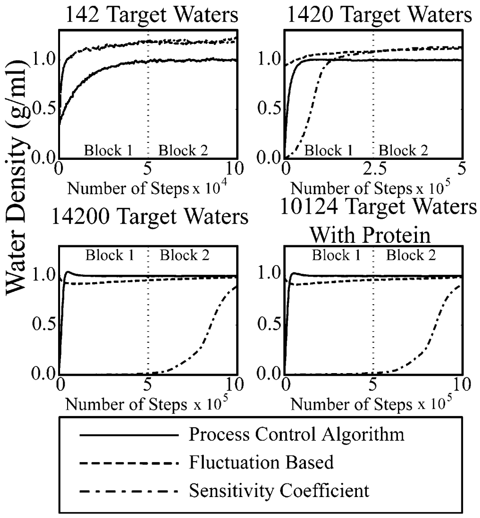


Figure 2. Early system density. The instantaneous density during the first 2% of each simulation is shown. The labels 'Block 1' and 'Block 2' indicate the portion of the plot that would be averaged to produce the block average densities used in Figure 1.

In engineering control scenarios, proportional-integral controller (PIC) performance is typically robust over a wide range of operating conditions.[11] By analogy, the parameters determined here ($K_c = -112$ and $\tau_I = 13\,000$) are expected to be effective in a variety of simulation systems without modification. Although the range of systems for which a given set of parameters is effective is difficult to predict, we note that they perform quite well on all systems tested here (which vary over 2 orders of magnitude in size). Additionally, these parameters produce the targeted density within 1M steps when applied in other simulation systems (1SDO, 80 Å by 80 Å by 80 Å system;[18] 1TY6, 100 Å by 80 Å by 70 Å[19]) with equal effect. In systems that are many orders of magnitude larger (or otherwise different in their dynamic character) a retuning of the parameters may give enhanced performance. However, systems of such size are outside of the purview of most GCE simulation methodologies at the present time.

Grand Canonical Ensemble Simulations at Fixed *B*. The last configuration of the PIC protein-water tuning simulation was used as the starting point of these simulations. The density initially was 0.997 g/mL in the region monitored in the previous simulations. All parameters detailed for the protein-water simulation above were used, with two exceptions, the tuning keyword was off, and the *B* parameter was fixed throughout the simulations. Simulations were allowed to evolve for 20M steps, and the density was reported every 2000 steps.

Results and Discussion

Achieving the Target Density. The effectiveness of the three tuning methods in approaching and maintaining the targeted density of water is shown in Figure 1.

The densities of the steps corresponding to the first two block averages are shown in Figure 2. The process control

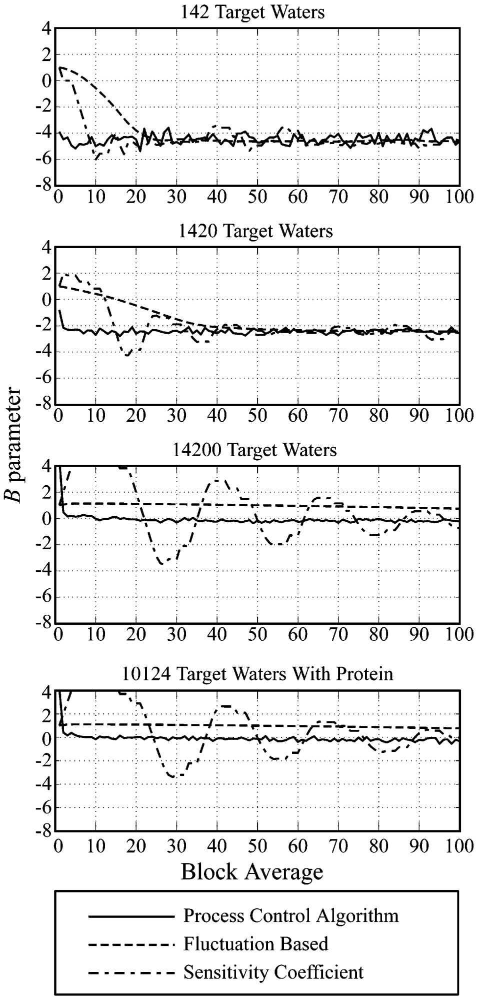


Figure 3. *B* parameter block average. The block average of the *B* parameter is plotted for each simulation. The solid line is the PIC method, the fluctuation based method is indicated by the dashed line, and the dash−dot line is the sensitivity coefficient method.

algorithm approaches the target density within these two blocks and maintains it throughout the simulation. In the small water simulation, the sensitivity coefficient algorithm and the fluctuation based algorithm have the exact same densities for the first 20K steps of the simulation due to the 10K of equilibration and 10K of statistical gathering time that they require. After that, they independently modify the *B* parameter and their densities diverge. Also, in the medium water, large water, and large water−protein simulations, the fluctuation based algorithm starts with a density near the target as noted in the computational methods section.

Generally, the standard deviation of the density in the PIC simulations decreases with increased targeted *N*. This is due to the discrete method of adding and subtracting water molecules. For example, in a simulation targeting 142 water molecules, a single water molecule addition increases the density by 0.007 g/mL, while for simulations with target *N* equal to 1420, a single water molecule addition increases the density by 0.002 g/mL.

Testing for Convergence. In equilibrated systems, the distribution of the density is expected to be close to normal. First, we apply the Lilliefors test for normality of the density distribution in the last 50 and 25% of the simulations. If the distribution is normal, we then apply the Student's *t*-test at 95% confidence with a mean of 0.997 and unknown variance. A simulation that passes all of these tests is considered converged to the target density. All methods converge to the target density in the small water simulation. Table 1A indicates that only the PIC method converges to the target density in all simulation systems.

Inspection of Figure 1 shows that the sensitivity coefficient method appears to be converging to the correct density in many of the simulations but has not achieved convergence by our criteria in the medium water and protein/water simulations. Figure 1 also shows that the fluctuation method uses the most steps to approach the target density in the small and medium simulations. In the large simulation with and without protein, the fluctuation method does not even reach the target density. This is possibly due to an overestimation of the fluctuation. As the target system gets larger, changes in $\langle N \rangle$ produce an increasing error of this estimate.

In all four simulation conditions, the PIC algorithm identifies and maintains the target density the soonest. This may be due to the method with which the PIC method adjusts the *B* parameter at each step with no inherent lag time for averaging or equilibration. The PIC algorithm in all four simulations reliably identifies the target density and maintains that state. Figure 2 shows the evolution of *B* for all simulations.

Identifying B. The ultimate goal of the protocol is to identify the correct *B* parameter. The convergence of the simulation to the target density is a necessary, but not sufficient, condition for correct determination of this parameter. We consider *B* converged when the simulation has passed two criteria for block averages 51−75 and 76−100. The first criteria is a Lilliefors test for normality in these sections individually. While normality is not an absolute requirement for convergence in *B*, this test determines whether the paired Student's *t*-test between the two segments of the simulations will be reliable. A 5% significance level is used in the Student's *t*-test. Cases where the target density is not converged are not considered. Table 1B shows that only the PIC method produces a converged *B* value for all simulation systems.

In all cases, the PIC simulations converged to the target density within 500K steps, while the *B* parameter required more steps to reach equilibrium. In all PIC simulations, the *B* parameter appears steady by the midpoint of the simulations. The target *N* of the simulation has some effect on this convergence, with larger target numbers requiring more steps to converge to a value in *B*. Further tests for convergence used the mean value of *B* from the block averages of the last 50% of the PIC simulation for the protein-water simulation system where the mean density was $0.9970 \pm 8.6 \times 10^{-5}$.

Testing the Identified B. To determine if we had identified the correct value of *B* in the large protein simulation, and that the fluctuations of *B* around its mean did not introduce a bias, we ran a series of simulations each with constant *B*. The starting system from these simulations used the final system from the PIC simulations and varied

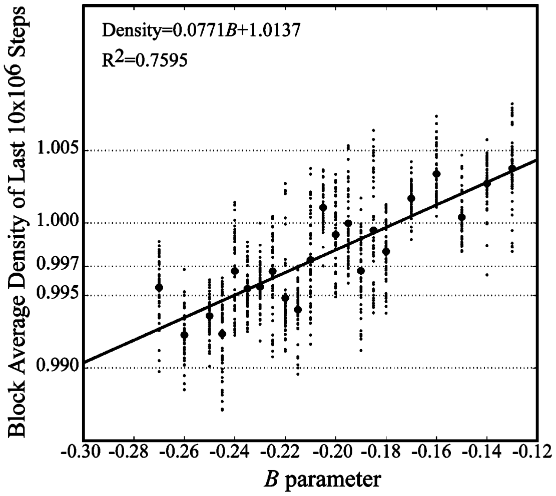


Figure 4. Fixed *B* parameter simulation results. The small black points show the block averages plotted for the last 50% (1×10^7 steps) of the simulations. Each block is the average of 200 000 steps. The larger black dots mark the mean of those block averages. The line shows the linear regression through those means.

B around the mean of the last 50% of the simulation. Each simulation ran for 20M steps, and data analysis was performed on 100 equally sized block averages. In Figure 4, these block averages are plotted against *B* for the last 10M steps (50 block averages). The larger black dot marks the mean of those block averages. The linear regression through those mean values yields a line producing an idealized value of *B* for this system. The value of *B* from the PIC simulations deviated from this ideal by 0.0046 and was well within the standard deviation of *B* (±0.1470) for the last 50% of the PIC simulation.

The R^2 correlation of the line though the means of the fixed *B* simulations indicates that there is a fair amount of uncertainty in the density at a constant *B*. This effect can be minimized through longer simulation times, increasing the statistical sampling of the ensemble of densities and *B* parameters.

Conclusions

We show that it is possible to tune the *B* parameter of the CB/GCE simulations efficiently and without user intervention to a good approximation of the actual value. Of the three methods tested, the PIC method performed best by far, and it is the recommended method. Also, the parameters of the PIC method established in this work are robust enough to be applicable to many diverse system sizes.

Tuning the B parameter without user intervention streamlines the simulation process and allows completion of a project in significantly shorter time. The time saving comes both in terms of CPU time used and elapsed time because tuning can be completed in fewer simulation steps than with manual tuning, and there is no need to interrupt the runs to manually adjust the *B* parameter.

Acknowledgment. This work was supported in part by NIH Grant P01 GM66531.

References

(1) Mezei, M. Efficient Monte Carlo sampling for long molecular chains using local moves, tested on a solvated lipid bilayer. *J. Chem. Phys.* **2003**, *118* (8), 3874−3879.

(2) Ebersole, B. J.; Visiers, I.; Weinstein, H.; Sealfon, S. C. Molecular basis of partial agonism: orientation of indoleamine ligands in the binding pocket of the human serotonin 5-HT2A receptor determines relative efficacy. *Mol. Pharmacol.* **2003**, *63* (1), 36−43.

(3) Marrone, T. J.; Resat, H.; Hodge, C. N.; Chang, C. H.; McCammon, J. A. Solvation studies of DMP323 and A76928 bound to HIV protease: analysis of water sites using grand canonical Monte Carlo simulations. *Protein Sci.* **1998**, *7* (3), 573−9.

(4) Adams, D. J. Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid. *Mol. Phys.* **1975**, *29*, 307−311.

(5) Resat, H.; Mezei, M. Grand canonical ensemble Monte Carlo simulation of the dCpG/proflavine crystal hydrate. *Biophys. J.* **1996**, *71* (3), 1179−90.

(6) Mezei, M., Grand-canonical ensemble Monte Carlo study of dense liquid Lennard-Jones, soft spheres and water. *Mol. Phys.* **1987**, *61* (3), 565−582.

(7) Resat, H.; Mezei, M. Grand canonical Monte Carlo simulation of water positions in crystal hydrates. *J. Am. Chem. Soc.* **1994**, *116*, 7451−7452.

(8) Woo, H. J.; Dinner, A. R.; Roux, B. Grand canonical Monte Carlo simulations of water in protein environments. *J. Chem. Phys.* **2004**, *121* (13), 6392−400.

(9) McQuarrie, D. *Statistical Thermodynamics*; University Science Books: Mill Valley, CA 94941, 1973.

(10) Stephanopolous, G. *Chemical Process Control, an introduction to Theory and practice*; Prentice Hall: Englewood Cliff, NJ, 1984.

(11) St. Clair, D. *Controller Tuning and Control Loop Performance*, 2nd ed.; Straight Line Control: 1993.

(12) MMC is available at the URL http://inka.mssm.edu/∼mezei/mmc (last accessed: 05/16/2006).

(13) Chamorro, J. A.; Cuesta-Seijo, J. A.; Garca-Granda, S. Pancreatic bovine trypsin native and inhibited with benzamidine from synchotron data. To be published **2006**; PDB ID: 1S0R.

(14) Mezei, M. Distance-scaled force biased Monte Carlo simulation for solutions containing a strongly interacting solute. *Mol. Simul.* **1991**, *5*, 405−408.

(15) Jorgensen, W.; Chandrashekar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.

(16) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, W.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4* (2), 187−217.

(17) Zeigler, J. G.; Nichols, N. B. Optimum settings for automatic controllers. *Trans. ASME* **1942**, 759.

(18) Niv, M. Y., personal communication.

(19) Filizola, M., personal communication.

CT0600363

J. Chem. Theory Comput. **2006,** *2,* 1435−1443

JCTC Journal of Chemical Theory and Computation

Calculation of the Relative Binding Affinity of Enzyme Inhibitors Using the Generalized Linear Response Method

Xin Chen[†,‡] and Alexander Tropsha*[,†]

Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, and Computer Assisted Drug Discovery, Research and Early Development, Johnson & Johnson Pharmaceutical Research and Development, L.L.C., 920 Route 202, Raritan, New Jersey 08869

Received February 20, 2006

Abstract: The generalized linear response (GLR) method initially developed for *hydration* free energy calculations has been adapted for *binding* free energy calculations. The calculations employ the concept of thermodynamic cycle. To obtain the value of the relative binding free energy between two ligands, we run molecular dynamics simulations at only four "midpoint" states along the thermodynamic pathways connecting the two ligands in the unbound and bound states, respectively. This approach significantly simplifies and accelerates the calculations as compared to the traditional free energy simulations where significantly more intermediate states are usually sampled. We show that each of these "midpoint" states can be approximately defined by a modified force field function in which both the van der Waals and electrostatic interactions between the variant part of a ligand and its environment, either binding site or aqueous solution, are scaled by a factor of 0.5. We tested this new approach to relative binding affinity calculations on the HIV-1 protease complex with its inhibitor JG365 as a starting point for the following two structural transformations: (a) the critical chiral center on the central residue was changed from (*S*) to (*R*) configuration, and (b) the C terminal valine residue was deleted. In both cases, the GLR method afforded calculated values that were in good agreement with the experimental data.

Introduction

The success of structure-based drug design approach relies on our ability to predict, rapidly and accurately, the binding modes and affinities of ligands that interact with their macromolecular receptors.[1] In general, current computational approaches to modeling ligand−receptor interaction can be divided into two categories: qualitative (empirically based) and quantitative (first-principle based) methods.[2] Qualitative modeling methods typically rely on empirical measures of geometric and/or chemical complementarities between ligands and receptors.[3] These methods have not only the advantages of speed and simplicity but also the limited ability to differentiate between structurally similar ligands despite some recent efforts to improve their accuracy by taking into account the conformational flexibility[4−6] and the solvation effect.[7,8] As a result, they may be only capable of identifying weak active hits from database screening.

Quantitative modeling methods are usually based on molecular simulations, such as free energy perturbation (FEP) and thermodynamic integration (TI) that use molecular dynamics or Monte Carlo simulations as the sampling technique.[9−12] These methods are theoretically sound and in principle able to afford accurate calculation of relative

---

* Corresponding author fax: (919)966-0204; e-mail: alex_tropsha@unc.edu.

[†] University of North Carolina at Chapel Hill.

[‡] Johnson & Johnson Pharmaceutical Research and Development.

10.1021/ct600071z CCC: $33.50

binding affinity of structurally similar ligands. Typically, they can provide calculated values within 1.5 kcal/mol range of the experimental data.[12] Adequate sampling is a key problem for these methods due to the extremely high computational cost usually required to sample all microscopic states accessible to the system to obtain a convergent calculated value. Another problem is the “force field problem”, which implies an insufficiently accurate description of van der Waals and electrostatic interactions in ligand−receptor systems, although certain progress has been made to circumvent these problems as well.[13,14]

Several years ago, Aqvist et al.[15] and Jorgensen et al.[16] proposed a new approach to accelerate free energy calculations based on the linear response approximation. This approach avoids the exhaustive sampling at multiple intermediate states along the transformation path from the initial to final state as in traditional free energy simulations and therefore provides a computationally more efficient route for free energy calculations. According to Aqvist et al., the binding free energy $\Delta G_{\mathrm{binding}}$ can be calculated from the change of electrostatic and van der Waals interaction energies between a ligand and its environment as it is transferred from aqueous solution to binding site, as follows:[15]

$$\Delta G_{\mathrm{binding}} = 1/2 \cdot \Delta\langle V_{\mathrm{elec}}\rangle + \alpha \cdot \Delta\langle V_{\mathrm{vdw}}\rangle \qquad (1)$$

Here, $\langle V_{\mathrm{elec}}\rangle$ and $\langle V_{\mathrm{vdw}}\rangle$ are the ensemble averages of the electrostatic and van der Waals interaction energies, respectively, between the ligand and its environment, and Δ refers to the difference between the ensemble averages in the aqueous solution and in the binding site. The 1/2 coefficient before the electrostatic contribution term can be derived from the standard theory behind the linear response approximation, whereas the α coefficient before the van der Waals contribution term is determined by empirical calibration against the experimental data. Equation 1 was first applied to calculate the binding affinities of five structurally different inhibitors of endothiapepsin, and it was found that the α value of 0.162 gave the best agreement with the experimental data with a mean unsigned error of ca. 0.3 kcal/mol.[15] Subsequently, Aqvist’s group has successfully applied this approach to several other ligand−receptor systems.[17−22]

Carlson and Jorgensen[16] proposed a slightly different version as an extended linear response equation for the hydration free energy calculations. They introduced an additional cavity contribution term and included adjustable coefficients before all terms. Thus, the hydration free energy ΔG_{hyd} is calculated as[16]

$$\Delta G_{\mathrm{hyd}} = \beta \cdot \langle V_{\mathrm{elec}}\rangle + \alpha \cdot \langle V_{\mathrm{vdw}}\rangle + \gamma \cdot (\mathrm{cavity\_term}) \qquad (2)$$

Here, $\langle V_{\mathrm{elec}}\rangle$ and $\langle V_{\mathrm{vdw}}\rangle$ are the ensemble averages of the electrostatic and van der Waals interaction energies between a solute and aqueous solvents. The coefficients, α, β, and γ, are adjustable parameters, which are obtained by fitting to the experimental hydration free energies. The cavity contribution term could represent one of three different molecular properties: molecular surface area, solvent-accessible surface area, or molecular volume. Equation 2 was first applied to 14 diverse organic compounds. It was found that when the cavity term represented the solvent accessible surface area, the best agreement with the experimental data was obtained with RMS deviations of ca. 0.8 kcal/mol for the all-atom model and 0.9 kcal/mol for the united-atom model. Jorgensen et al. then expanded the test data set to 35 organic compounds and obtained an RMS deviation of ca. 1.0 kcal/mol from the experimental data.[23] They have also successfully applied their approach to several ligand−receptor binding systems.[24−27]

The published work on extending the standard linear response approximation to free energy calculations has demonstrated the efficiency of this approach. However, the empirical parameters, such as α, β, and γ used in eqs 1 and 2, were derived from the analysis of specific training sets. Consequently, as with any empirically derived parameters, there is always a question concerning their generality and applicability to other systems. We proposed a more generalized linear response (GLR) method[28] for free energy calculations, which does not include empirically derived parameters such as α, β, and γ used in eqs 1 and 2 and, consequently, has an inherently greater potential to be applied to diverse molecular systems. Previously, we have successfully applied this method to calculate hydration free energies for a set of organic compounds.[28] In this report, we extend the GLR method to include relative binding free energy calculations, using HIV-1 protease complexes with several of its inhibitors as the test system. We show that the calculated relative binding free energies of the inhibitor JG-365 vs two of its analogues agree closely with the experimental binding free energies yet computed in a much more efficient way than using traditional free energy simulation approaches.

Methodology

Linear Response Approximation (LRA). Linear response approximation for treating electrostatic interaction has been considered in many cases including the Marcus theory of electron-transfer reactions.[29] According to LRA, a system may respond linearly to the change of its electrostatic field, yielding a free energy curve with parabolic form as illustrated in Figure 1.[15,30] The reaction coordinate X describes the transformation of the whole system between the initial state A and the final state B, which are defined by their potential functions V_{A} and V_{B}, respectively. It is assumed that the free energy functions of the system at states A and B can be approximately described by two parabolic curves, $\Delta g_{\mathrm{A}}(x)$ and $\Delta g_{\mathrm{B}}(x)$, respectively, and that these two parabolic curves have the same force constant, i.e., the whole system responds linearly to the change of electrostatic field with the same dielectric constant along the path. From these postulates, it can be easily inferred that the change of the free energy between states A and B, ΔG_{AB}, can be expressed as[15,30]

$$\Delta G_{\mathrm{AB}} = 1/2 \cdot (\langle V_{\mathrm{B}} - V_{\mathrm{A}}\rangle_{\mathrm{A}} - \langle V_{\mathrm{A}} - V_{\mathrm{B}}\rangle_{\mathrm{B}}) \qquad (3a)$$

and[31,32]

$$\Delta G_{\mathrm{AB}} = \langle V_{\mathrm{B}} - V_{\mathrm{A}}\rangle_{\mathrm{mid}} \qquad (3b)$$

Here, $\langle \cdot\cdot\cdot \rangle_{\mathrm{A}}$ and $\langle \cdot\cdot\cdot \rangle_{\mathrm{B}}$ denote the ensemble averages at states A and B, respectively, while $<...>_{\mathrm{mid}}$ stands for the ensemble average at the “midpoint” state between states A

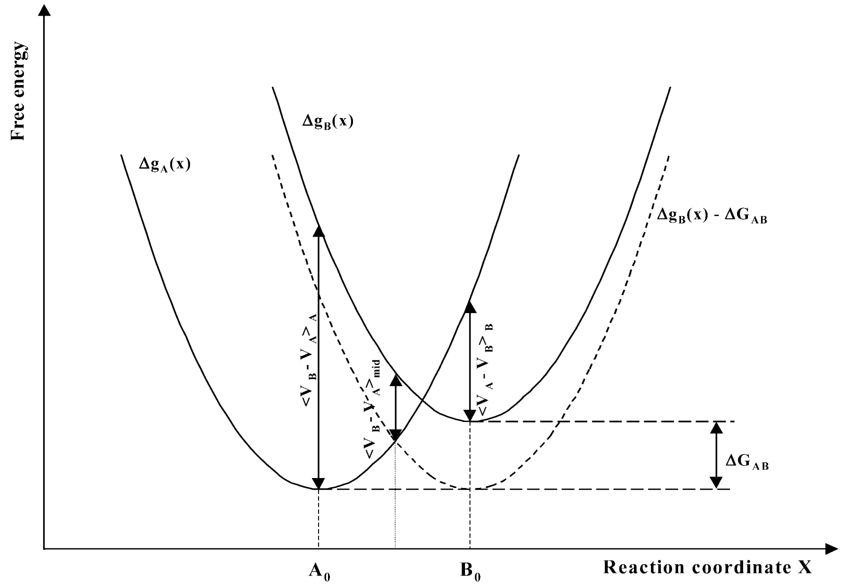


Figure 1. Illustration of the free energy curves for a system that obeys the linear response approximation (LRA). The vertical dotted line indicates the "midpoint" state between the equilibrium coordinates A_0 and B_0 of the initial and final states. The dashed curve $\Delta g_B(x)$-ΔG_{AB} is a copy of the $\Delta g_B(x)$ curve obtained by shifting it downward by ΔG_{AB}. Since $\Delta g_A(x)$ and $\Delta g_B(x)$-ΔG_{AB} have the equal curvature, it is easy to infer that $\langle V_B - V_A \rangle_{mid}$ is equal to ΔG_{AB}. See text for the additional discussion.

and B. The geometric descriptions of all the terms in eqs 3a and 3b are given in Figure 1.

LRA and Free Energy Simulations. The free energy calculations as applied to studying intermolecular interaction are often associated with the physicochemical processes characterized by the transfer of a compound between two phases, e.g., from vacuum to aqueous solution (hydration process) or from aqueous solution to binding site (binding process). These processes can be theoretically perceived and described as the disappearance of a compound from one phase followed by its appearance in the other phase. In a physical sense, it means that both van der Waals and electrostatic interactions between a compound and its environment are eliminated in the initial phase and then re-established in the final phase. Although LRA has been successfully applied to calculate the free energy change resulting from the change of electrostatic interaction, to the best of our knowledge prior to our previous publication[28] there were no reports on the direct (i.e., without introducing any additional empirically derived constants) application of LRA to describe the contribution of van der Waals interaction to the free energy change.

From the mathematical point of view, both the van der Waals and electrostatic terms of a molecular force field have a singularity point at the interaction distance of zero, which may lead to the failure of LRA as illustrated by the following example. Assume that the state A in Figure 1 represents a system consisting of a solute atom surrounded by water molecules, and the state B represents the system after the solute atom has been removed from the aqueous solution. After both van der Waals and electrostatic interactions between the solute atom and the surrounding water molecules disappear, the water molecules have a chance to occupy the former position of the solute atom. If that happens, the value of V_A and therefore $V_A - V_B$ in eq 3a will approach infinity and obviously break the LRA assumption (cf. Figure 1). In fact, this problem has been encountered in traditional free energy simulations, where it is often referred to as the "singularity problem". However, if one only allows the electrostatic interaction between the solute atom and the water solvents disappear, the presence of the remaining van der Waals interaction will prevent the water molecules from approaching the solute atom too closely (or overlapping with it) and thereby avoid the "singularity problem". This is exactly the case in the electron transfer processes and explains why LRA has been successfully applied to these processes.

Generalized Linear Response (GLR) Method. In the previous publication,[28] we suggested that the linear response approximation could also be applied to the change of van der Waals interaction in the course of free energy simulations provided that one could find a proper way to circumvent the "singularity problem". In the case of hydration of small molecules, we have demonstrated that the hydration free energy ΔG can be expressed as[28]

$$\Delta G = \langle V \rangle_{mid} + \Delta G^c \tag{4}$$

Here, V is the interaction energy between the hydrated solute molecule and water molecules; $\langle \cdot\ \cdot\ \cdot \rangle_{mid}$ stands for the ensemble average at the "midpoint" state of the hydration process, which can be approximately defined by a modified molecular force field in which the interaction between the hydrated molecule and the surrounding water molecules are decreased by a half from the original one. The ΔG^c term is similar to a cavity contribution term, representing the free energy required for introducing n independent point particles into the aqueous solution, where n is the number of atoms in the hydrated molecule. Based on the scaled particle theory,[33] we have shown[28] that this term can be approximated as $1.49k_BnT$, if the SPC model[34] is used to describe water molecules (k_B is the Boltzmann constant; T is the system

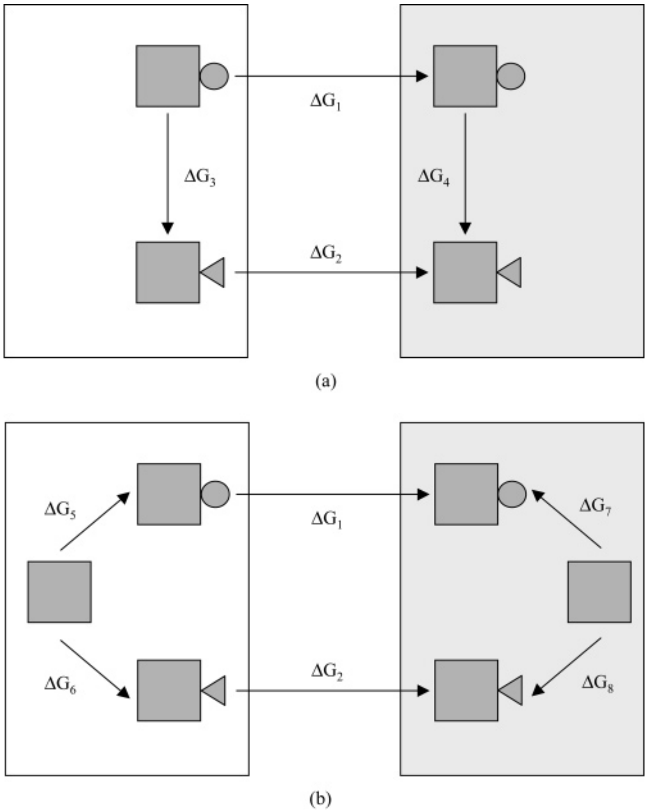


Figure 2. (a) Thermodynamics cycle used in the traditional free energy simulation methods. (b) New thermodynamics cycle used in the GLR method for relative binding free calculations. The large empty box represents the aqueous solution, while the large gray box represents the binding site

temperature). Equation 4 was tested on a data set of 14 chemically diverse organic compounds, and the results were in a good agreement with those obtained from the traditional free energy simulations with an average deviation of only 0.4−0.5 kcal/mol.[28]

GLR and Relative Binding Free Energy Calculations. In drug design, we are often concerned with estimating the relative binding free energy between two different ligands rather than with their absolute binding constants. Free energy simulations in the context of thermodynamic cycle, as shown in Figure 2(a), are frequently used to compute relative binding free energies of two ligands. According to this cycle, the difference between two binding free energies, $\Delta\Delta G$, is defined as $\Delta G_2 - \Delta G_1$ and can be expressed as $\Delta G_4 - \Delta G_3$ (cf. Figure 2(a)) as well. ΔG_3 and ΔG_4 are the free energy changes associated with the transformation of one ligand into another, in solution and in binding site, respectively; these can be relatively easily calculated by free energy simulations if the structural difference between these two ligands is relatively small. To extend eq 4 to relative binding free energy calculations, we introduce two intermediate states into the simple thermodynamics cycle, in which the pseudoligands are composed of the common structural part of the two original ligands. This modified thermodynamic cycle is shown in Figure 2(b). ΔG_5, ΔG_6, ΔG_7, and ΔG_8 represent the free energies required for changing the pseudoligands into the original ligands, respectively. Then, we assume that eq 4 can be used for calculating each of these four values,

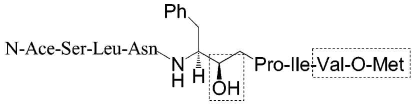


Figure 3. Structure of HIV-PR inhibitor, JG365(S). The dashed boxes indicate the part of the structure subjected to transformation in the course of simulations.

and therefore the relative binding free energy can be expressed as

$$\begin{aligned}\Delta\Delta G &= \Delta G_5 - \Delta G_6 - \Delta G_7 + \Delta G_8\\ &= (\langle V_5\rangle_{\mathrm{mid}} + \Delta G_5{}^{\mathrm{c}}) - (\langle V_6\rangle_{\mathrm{mid}} + \Delta G_6{}^{\mathrm{c}}) - \\ &\qquad (\langle V_7\rangle_{\mathrm{mid}} + \Delta G_7{}^{\mathrm{c}}) + (\langle V_8\rangle_{\mathrm{mid}} + \Delta G_8{}^{\mathrm{c}}) \end{aligned} \quad (5)$$

Here, V_5, V_6, V_7, and V_8 are the interaction energies between the variant parts of the ligands and their environments, either the aqueous solution or the binding site. $\langle\cdot\cdot\cdot\rangle_{\mathrm{mid}}$ stands for the ensemble average at the "midpoint" state of each of the four transformation paths. This state can be approximately defined by a modified molecular force field in which the interactions between the variant part of a ligand and its environment are decreased by a half from the original ones. $\Delta G_5{}^{\mathrm{c}}$, $\Delta G_6{}^{\mathrm{c}}$, $\Delta G_7{}^{\mathrm{c}}$, and $\Delta G_8{}^{\mathrm{c}}$ are the cavity contribution terms for each transformation step, and we speculate that their net contribution is close to zero due to the following considerations. First, the water density in the binding site should be similar to that in the bulk water, if not identical. Since water density is the major variable to determine the free energy required to introduce a point particle into water,[28,33] this will lead to similar free energy contributions from the cavity terms in the aqueous solution and in the binding site, and therefore they may cancel each other. Second, $\Delta G_7{}^{\mathrm{c}}$ and $\Delta G_8{}^{\mathrm{c}}$ very probably occur in the same local environment inside the binding pocket. If they are not structurally highly dissimilar to each other, these two terms may also cancel each other. Consequently, eq 5 can be simplified as

$$\Delta\Delta G = \langle V_5\rangle_{\mathrm{mid}} - \langle V_6\rangle_{\mathrm{mid}} - \langle V_7\rangle_{\mathrm{mid}} + \langle V_8\rangle_{\mathrm{mid}} \quad (6)$$

This is the master equation that was used in this work for the relative binding free energy calculations.

HIV-1 Protease and Its Inhibitors as a Test System. HIV-1 protease (HIV-PR) is one of the key enzymes that activate the human immunodeficiency virus (HIV). Many efforts went into the design of specific inhibitors based on the crystallographic structure of this enzyme. One of the well-known specific inhibitors of HIV-PR, JG365, is a non-hydrolyzable heptapeptide analogue of the protease substrate. It incorporates a modified phenylanaline residue named Phs, in which a backbone carbonyl group is substituted with $-\mathrm{CH(OH)CH_2}-$ as illustrated in Figure 3. Tropsha and Hermans[35] and Ferguson et al.[36] used molecular simulations independently to calculate the relative binding affinity of (*S*) vs (*R*) isomers of JG365. Reddy et al.[37] used a thermodynamic cycle approach to calculate the effect of deleting the C terminal valine residue of JG365 on the binding affinity of the modified inhibitor. Our GLR method was tested on these two model systems.

In the first case, the configuration of the Phs residue of JG365 is changed from (*S*) to (*R*). Thus, the small circle in

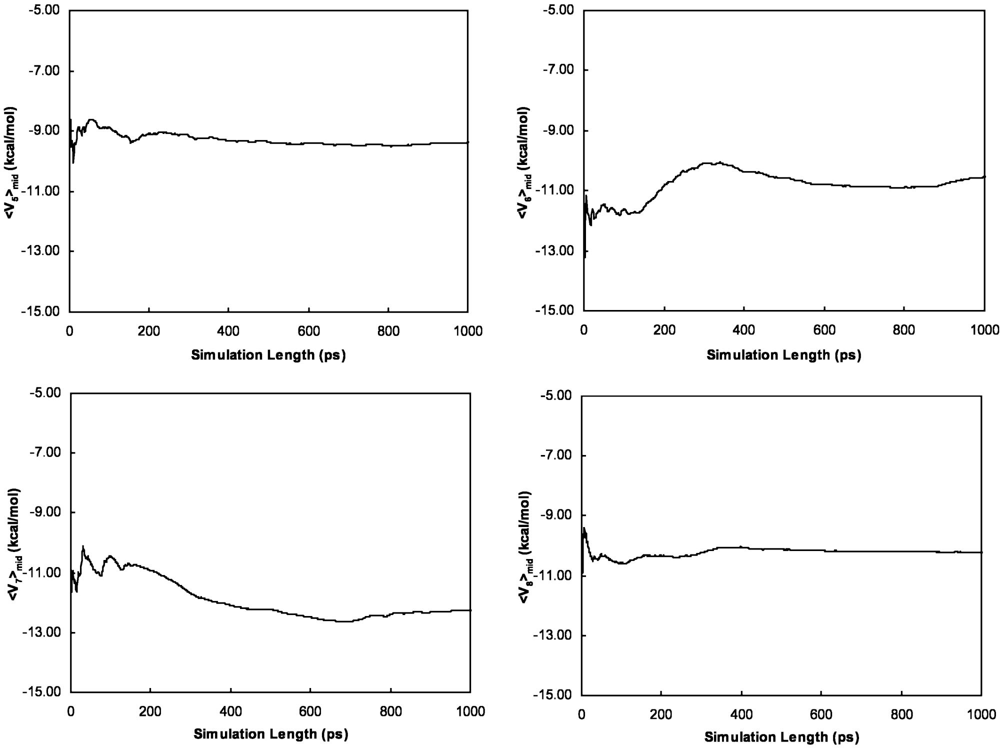


Figure 4. Values of the $\langle V_5\rangle_{\mathrm{mid}}$, $\langle V_6\rangle_{\mathrm{mid}}$, $\langle V_7\rangle_{\mathrm{mid}}$, and $\langle V_8\rangle_{\mathrm{mid}}$ terms in eq 6 as the function of simulation time, for calculating the relative binding free energy between JG365(S) and JG365(R).

Figure 2(b) represents the −CH(OH)− group in the (*S*) configuration, while the small triangle in Figure 2(b) represents the same group in the (*R*) configuration. In the second case, the valine residue of JG365 is removed so the terminal methoxy group is directly linked to the preceding isoleucine residue at the end of transformation. In this case, the small circle in Figure 2(b) represents the valine residue and the terminal methoxy group, while the small triangle in Figure 2(b) represents the terminal methoxy group only. For each case, four independent molecular dynamics simulations were run at the "midpoint" states of each transformation path, in which the interactions between the variant parts of the ligands and their environments, either binding site or aqueous solution, were decreased by one-half. The ensemble average values of the original interaction energies between the variant parts of the ligands and their environments were recorded, and the final $\Delta\Delta G$ value was calculated according to eq 6.

Computational Details

The X-ray structure of HIV-PR in complex with JG365(S) (PDB code 7HVP[38]) was used as the reference structure, from which the structures of HIV-PR and JG365 analogues were modeled. The charge distributions for JG365(S), JG365(R), and their hexapeptide analogues were the same as used in the earlier work.[35] Similar to the procedure reported earlier,[35] JG365 and its analogues were considered in the protonated states, both in the aqueous solution and in the binding site of HIV-PR. For the HIV-PR/ligand complexes, a sufficiently large box of water was defined to guarantee that there was at least 10 Å from each side of the box to the nearest protein atom. Thus, the dimensions of the water box were 77 × 56 × 61 Å, and it contained over 7000 water molecules. The same criterion was used to determine the size of the water box for the ligands alone in aqueous solution. The box dimensions were 44 × 27 × 29 Å for JG365(S) and JG365(R) and 42 × 27 × 29 Å for their hexapeptide analogues.

The molecular dynamics program package Sigma[39] was used for the all simulations. Sigma uses the same description of the nonbonded interactions as GROMOS molecular force field[40] and the bonded interaction parameters that were developed independently.[41] The cutoff radius for the nonbonded interactions was set to 10 Å. The Shake algorithm was used to maintain the bond lengths close to the standard values.[42] The SPC model was used to describe water molecules,[34] and periodic boundary conditions were used. All the simulations were done at 298 K.

For JG365(S), JG365(R), and their hexapeptide analogues in the HIV-PR binding sites, 50 steps of energy minimization followed by 20 ps molecular dynamics simulations were used to equilibrate the whole systems. Then, atoms with an unconstraint motion were limited to those within a 12 Å sphere around the C_α atoms of the Phs residue, its two adjacent residues on the ligand and the critical residues Asp25 and Asp125 on the protein. One nanosecond molecular dynamics simulation was run for calculating each of the four ensemble average terms in eq 6, to guarantee the convergence of each calculated value. For JG365(S), JG365(R), and their hexapeptide analogues in the aqueous solution, the same protocol of simulation was followed, except that all the water molecules were allowed to move.

Results

In the case of the transformation from JG365(S) to JG365(R), the values of the four ensemble average terms in eq 6 as a function of simulation time are shown in Figure 4. With 1

Table 1. Calculated and Experimental Relative Binding Free Energies (kcal/mol)

ligand structures	ΔΔ*G* (calc)	ΔΔ*G* (expt)
JG365(S) vs JG365(R)	3.1	2.6[a]
JG365(S) vs its hexapeptide analog	4.0	
JG365(R) vs its hexapeptide analog	2.0	
mixture of JG365(S) and JG365(R) vs their hexapeptide analogues	3.0[b]	3.8[c]

[a] Reference 43. [b] Average calculated value, assuming the equal-molar mixture of (*S*) and (*R*) configurations of JG365 and their correspondent hexapeptide analogues. [c] Reference 37. Experimental value for the equilibrium mixture of (*S*) and (*R*) configurations of JG365 and their hexapeptide analogues, whose ratio was undetermined.

ns molecular dynamics simulation, the ensemble average values are −9.4 kcal/mol for $\langle V_5\rangle_{\mathrm{mid}}$, −10.5 kcal/mol for $\langle V_6\rangle_{\mathrm{mid}}$, −12.2 kcal/mol for $\langle V_7\rangle_{\mathrm{mid}}$, and −10.2 kcal/mol for $\langle V_8\rangle_{\mathrm{mid}}$, respectively. Thus, according to eq 6, the value of ΔΔ*G* is 3.1 kcal/mol (cf. Table 1), meaning that the (*S*) configuration of JG365 binds to HIV-PR stronger than the (*R*) configuration by about 3.1 kcal/mol. This result is in a reasonable agreement with the experimental data of the difference of the binding energies of these two isomers of JG365, ∼2.6 kcal/mol.[43]

Since the available experimental data for the relative biding affinity between JG365 and its hexapeptide analogue were obtained from the equilibrium mixture of both (*S*) and (*R*) configurations, we did two independent series of simulations to determine the relative binding free energies for both (*S*) and (*R*) configurations. In the case of JG365(S) and its hexapeptide analogue, the values of the four ensemble average terms in eq 6 are plotted as a function of simulation time in Figure 5. Based on 1 ns molecular dynamics simulation, the ensemble average values are −6.0 kcal/mol for $\langle V_5\rangle_{\mathrm{mid}}$, −4.9 kcal/mol for $\langle V_6\rangle_{\mathrm{mid}}$, −10.4 kcal/mol for $\langle V_7\rangle_{\mathrm{mid}}$, and −5.3 kcal/mol for $\langle V_8\rangle_{\mathrm{mid}}$. According to eq 6, we obtain the ΔΔ*G* value of 4.0 kcal/mol for the (*S*) configuration (cf. Table 1). In the case of JG365(R) and its hexapeptide analogue, the values of the four ensemble average terms in eq 6 are shown as a function of the simulation time in Figure 6. Based on 1 ns molecular dynamics simulation, the ensemble average values are −6.8 kcal/mol for $\langle V_5\rangle_{\mathrm{mid}}$, −4.8 kcal/mol for $\langle V_6\rangle_{\mathrm{mid}}$, −9.4 kcal/mol for $\langle V_7\rangle_{\mathrm{mid}}$, and −5.4 kcal/mol for $\langle V_8\rangle_{\mathrm{mid}}$. According to eq 6, we obtain a value of ΔΔ*G* as 2.0 kcal/mol for the (*R*) configuration (cf. Table 1). Due to the fact that the ratio of (*S*) and (*R*) configurations in the mixture has not been experimentally determined, we arbitrarily assume a 50:50 ratio of these two configurations. It leads to an average ΔΔ*G* value of ∼3.0 kcal/mol, indicating that the binding free energy of the racemic mixture of JG365 isomers to HIV-PR is more favorable than that of their hexapeptide analogues by about 3.0 kcal/mol. This result is still in a reasonable agreement with the experimental result of 3.8 kcal/mol.[37] Actually, some evidences indicated that the (*S*) configuration might be the dominant component in the mixture,[37] so if we assume a higher proportion of the (*S*) configuration, we will get a calculated value even closer to the experimental data.

1 ns molecular dynamics simulation was conducted for each of the ensemble average terms in eq 6 to monitor the convergence behavior. We can see from Figures 4−6 that in most cases the ensemble average value is practically converged after 400−600 ps. This gives us the confidence

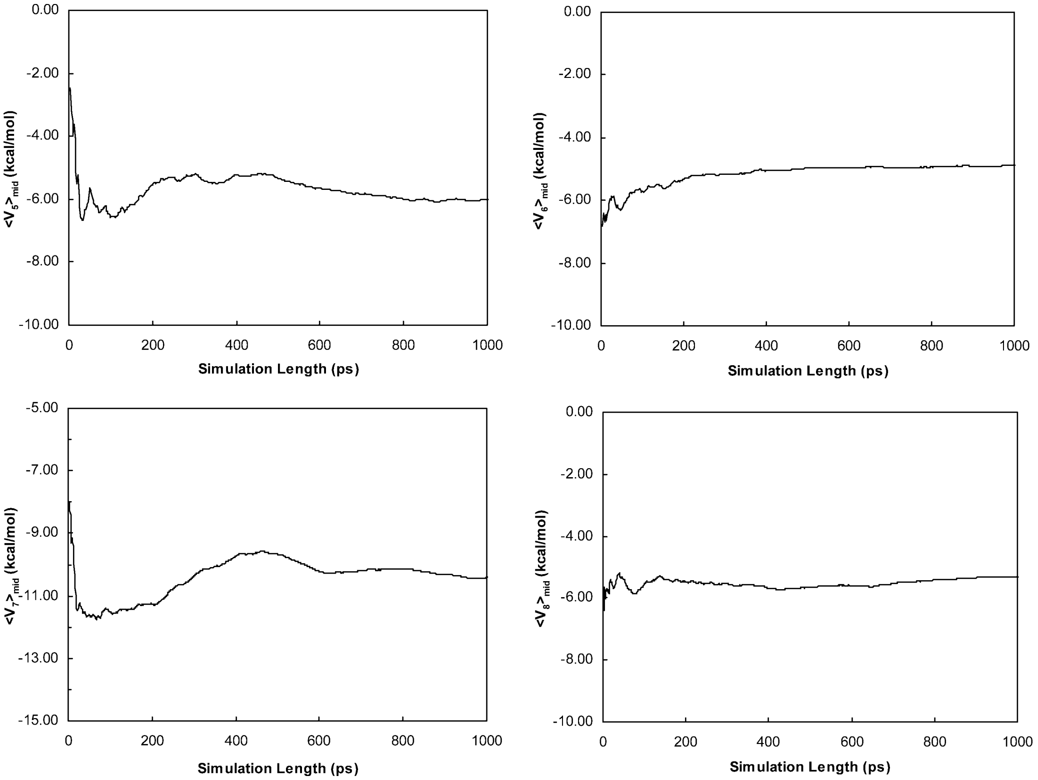


Figure 5. Values of the $\langle V_5\rangle_{\mathrm{mid}}$, $\langle V_6\rangle_{\mathrm{mid}}$, $\langle V_7\rangle_{\mathrm{mid}}$, and $\langle V_8\rangle_{\mathrm{mid}}$ terms in eq 6 as the function of simulation time, for calculating the relative binding free energy between JG365(S) and its hexapeptide analogue.

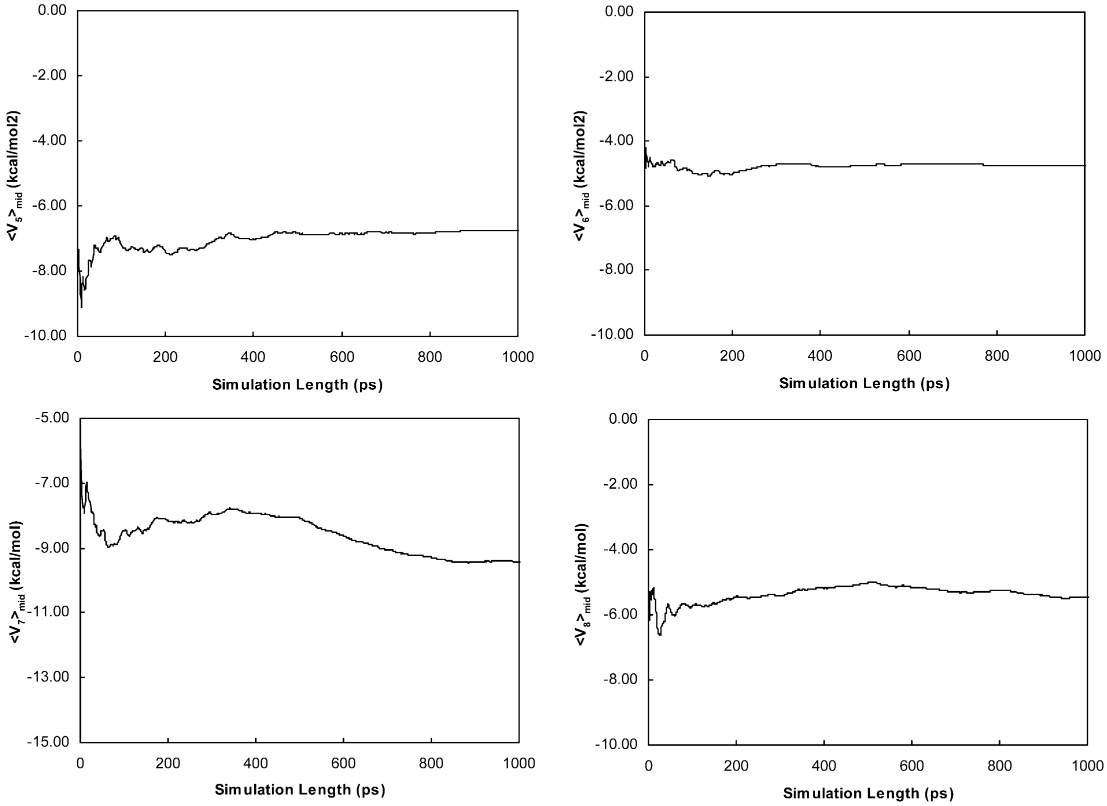


Figure 6. Values of the $\langle V_5\rangle_{mid}$, $\langle V_6\rangle_{mid}$, $\langle V_7\rangle_{mid}$, and $\langle V_8\rangle_{mid}$ terms in eq 6 as the function of simulation time, for calculating the relative binding free energy between JG365(R) and its hexapeptide analogue.

in the reliability of these calculated values. This is an important observation because the convergence of calculated values is not always guaranteed in the traditional free energy simulations where more than several nanoseconds simulation time is usually required.[44,45]

Discussion

The GLR method is a novel and simple approach for relative binding free energy calculations. It can be regarded as a method somewhere between the traditional qualitative and quantitative modeling methods. It differs from the qualitative methods in that it utilizes molecular dynamics or Monte Carlo simulation technique to sample the complex system, thereby directly taking into account the effects such as flexibility and entropy. On the other hand, it is also different from the quantitative methods, mainly the traditional free energy simulation methods, in that it only requires sampling at a particular “midpoint” state, instead of multiple (usually, dozens) intermediate states along the transformation pathway. Consequently, the GLR method is computationally much more efficient as compared with the traditional free energy simulation methods. Compared with other methods in the same class,[19,46] our GLR method does not require the calibration parameters that are usually used to fit the known experimental data and therefore may be more general.

As reported here, we have tested the GLR method on the HIV-1 protease complex with its inhibitor JG365. Two independent series of calculations were done: one for the relative binding free energy between the (*S*) and (*R*) configurations of JG365 and the other for the relative binding free energy between JG365 and its hexapeptide analogues. These two cases provide us with two representative examples that may well reflect the application scope of the current free energy simulation techniques. In the first case, the structural change is very small, only involving the conversion of a single configuration center that is critical for the binding of HIV-PR inhibitors. In the second one, the structural change is much larger: the entire valine residue disappears during the simulation. Such a structural change probably represents the upper limit of the current capability of free energy simulations. In both cases, our GLR method affords the results in good agreement with the experimental data, with the deviation values of 0.5 and 0.8 kcal/mol, respectively. This accuracy is also comparable with those in the studies conducted by Aqvist et al.[17−22] and Jorgensen et al.;[24−27] however, it should be emphasized that unlike our approach, their studies used several specially derived empirical parameters to fit the experimental data for each compound series.

At the present level, the GLR method requires about several hundred picoseconds of molecular dynamics simulation to guarantee a convergent calculated value, and therefore its application is still limited to the lead optimization rather than lead identification stage. More efficient sampling techniques may further increase the computational efficiency of this method. However, the GLR method does provide a novel and efficient alternative to the traditional free energy simulation techniques for calculating relative binding free energies. We suggest that it can be effectively used in structure-based lead optimization projects.

References

(1) Blaney, J. M.; Dixon, J. S. A Good Ligand Is Hard to Find: Automated Docking Methods. *Perspect. Drug Discovery Des.* **1993**, *1*, 301−319.

(2) Kollman, P. A. Theory of Macromolecule-ligand Interactions. *Curr. Opin. Struct. Biol.* **1994**, *4*, 240−245.

(3) Ajay; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand−receptor Complexes. *J. Med. Chem.* **1995**, *38*, 4953−4967.

(4) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(5) Taylor, J. S.; Burnett, R. M. DARWIN: a Program for Docking Flexible Molecules. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 173−191.

(6) Diller, D. J.; Merz, K. M., Jr. High Throughput Docking for Library Design and Library Prioritization. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 113−124.

(7) Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of Solvation in Ligand Binding Free Energy Calculations using the Generalized-born Model. *J. Am. Chem. Soc.* **1999**, *121*, 8033−8043.

(8) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient Electrostatic Solvation Model for Protein-fragment Docking. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 256−268.

(9) Karplus, M.; Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nature* **1990**, *347*, 631−639.

(10) Van Gunsteren, W. F.; Berendsen, H. J. C. Molecular Dynamics Computer Simulation. Method, Application and Perspectives in Chemistry. *Angew. Chem.* **1990**, *102*, 1020−1055.

(11) Jorgensen, W. L. Free Energy Calculations: a Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184−189.

(12) Kollman, P. A. Free Energy Calculations: Application to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(13) York, D. M.; Darden, T. A.; Pedersen, L. G. The Effect of Long-Range Electrostatic Interactions in Simulations of Macromolecular Crystals: a Comparison of Ewald and Truncated List Methods. *J. Phys. Chem.* **1993**, *99*, 8345−8348.

(14) Caldwell, J. W.; Kollman, P. A. Cation-pi Interactions - Nonadditive Effects are Critical in Their Accurate Representation. *J. Am. Chem. Soc.* **1995**, *117*, 4177−4178.

(15) Aqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method for Predicting Binding Affinity in Computer-aided Drug Design. *Protein Eng.* **1994**, *7*, 385−391.

(16) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667-.

(17) Hansson, T.; Aqvist, J. Estimation of Binding Free Energies for HIV Protease Inhibitors by Molecular Dynamics Simulations. *Protein Eng.* **1995**, *8*, 1137−1144.

(18) Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Aqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 Protease Inhibitors Derived from Mannitol: Synthesis, Inhibitory Potencies, and Computational Predictions of Binding Affinities. *J. Med. Chem.* **1997**, *40*, 885−97.

(19) Hansson, T.; Marelius, J.; Aqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.

(20) Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Aqvist, J.Computation of Affinity and Selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline Inhibitors to Dihydrofolate Reductases. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 119−131.

(21) Ljungberg, K. B.; Marelius, J.; Musil, D.; Svensson, P.; Norden, B.; Aqvist, J. Computational Modelling of Inhibitor Binding to Human Thrombin. *Eur. J. Pharm. Sci.* **2001**, *12*, 441−446.

(22) Brandsdal, B. O.; Aqvist, J.; Smalas, A. O. Computational Analysis of Binding of P1 Variants to Trypsin. *Protein Sci.* **2001**, *10*, 1584−1595.

(23) McDonald, H.; Carlson, A.; Jorgensen, W. L. Free Energies of Solvation in Chloroform and Water from a Linear Response Approach. *J. Phys. Org. Chem.* **1997**, *10*, 563−576.

(24) Jones-Hertzog, D. K.; Jorgensen, W. L.Binding Affinities for Sulfonamide Inhibitors with Human Thrombin using Monte Carlo Simulations with a Linear Response Method. *J. Med. Chem.* **1997**, *40*, 1539−1549.

(25) Smith, R. H., Jr.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Smith, M. B. K. Prediction of Binding Affinities for TIBO Inhibitors of HIV-1 Reverse Transcriptase using Monte Carlo Simulations in a Linear Response Method. *J. Med. Chem.* **1998**, *41*, 5272−5286.

(26) Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of the Binding Affinities of FKBP12 Inhibitors using a Linear Response Method. *Bioorg. Med. Chem.* **1999**, *7*, 851−860.

(27) Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. New Linear Interaction Method for Binding Affinity Calculations using a Continuum Solvent Model. *J. Phys. Chem. B* **2001**, *105*, 10388−10397.

(28) Chen, X.; Tropsha, A. A Generalized Linear Response Method: Application to the Hydration Free Energy Calculations. *J. Comput. Chem.* **1999**, *20*, 749−759.

(29) Marcus, R. A. Chemical and Electrochemical Electron-transfer Theory. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155−196.

(30) Lee, F. S.; Chu, Z.-T.; Bolger, M. B.; Warshel, A. Calculations of Antibody−antigen Interactions: Microscopic and Semi-microscopic Evaluation of the Free Energies of Binding of Phosphorylcholine Analogs to McPC603. *Protein Eng.* **1992**, *5*, 215−228.

(31) King, G.; Barford, R. A. Calculation of Electrostatic Free Energy Differences with a Time-saving Approximate Method. *J. Phys. Chem.* **1993**, *97*, 8798−8802.

(32) Dejaegere, A.; Karplus, M. Analysis of Coupling Schemes in Free energy Simulations: a Unified Description of Nonbonded Contribution to Solvation Free Energies. *J. Phys. Chem.* **1996**, *100*, 11148−11164.

(33) Pierotti, R. A. A Scaled Particle Theory of Aqueous and Nonaqueous Solutions. *Chem. Rev.* **1976**, *76*, 717−726.

(34) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces*; Reidel: Dordrecht, Holland, 1981.

(35) Tropsha, A.; Hermans, J. Application of Free Energy Simulations to the Binding of a Transition-state-analogue Inhibitor to HIV Protease. *Protein Eng.* **1992**, *5*, 29−33.

(36) Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. Determination of the Relative Binding Free Energies of Peptide Inhibitors to the HIV-1 Protease. *J. Med. Chem.* **1991**, *34*, 2654−2659.

(37) Reddy, M. R.; Viswanadhan, V. N.; Weinstein, J. N. Relative Differences in the Binding Free Energies of Human Immunodeficiency Virus 1 Protease Inhibitors: a Thermodynamic Cycle-perturbation Approach. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10287−10291.

(38) Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. Structure at 2.5-.ANG. Resolution of Chemically Synthesized Human Immunodeficiency Virus Type 1 Protease Complexed with a Hydroxyethylene-based Inhibitor. *Biochemistry* **1991**, *30*, 1600−1609.

(39) Hermans, J. *Sigma Manual*; University of North Carolina at Chapel Hill, 1994.

(40) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. A Consistent Empirical Potential for Water-protein Interactions. *Biopolymers* **1984**, *23*, 1513−1518.

(41) Ferro, D. R.; McQueen, J. E.; McCown, J. T.; Hermans, J. Energy Minimizations of Rubredoxin. *J. Mol. Biol.* **1980**, *136*, 1−18.

(42) Van Gunsteren, W. F.; Karplus, M. A Method for Constrained Energy Minimization of Macromolecules. *J. Comput. Chem.* **1980**, *1*, 266−274.

(43) Rich, D. H.; Sun, C. Q.; Vara Prasad, J. V. N.; Pathiasseril, A.; Toth, M. V.; Marshall, G. R.; Clare, M.; Mueller, R. A.; Houseman, K. Effect of Hydroxyl Group Configuration in Hydroxyethylamine Dipeptide Isosteres on HIV Protease Inhibition - Evidence for Multiple Binding Modes. *J. Med. Chem.* **1991**, *34*, 1222−1225.

(44) Pearlman, D. A. A Comparison of Alternative Approaches to Free Energy Calculations. *J. Phy. Chem.* **1994**, *98*, 1487−1493.

(45) Reddy, M. R.; Appelt, K. HIV-1 Protease: Structure-based Drug Design Using the Free Energy Perturbation Approach. *Free Energy Calculations Rational Drug Des.* **2001**, 317−334.

(46) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. Examining Methods for Calculations of Binding Free Energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA Calculations of Ligands Binding to an HIV Protease. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 393−407.

CT600071Z

1444 *J. Chem. Theory Comput.* **2006,** *2,* 1444−1452

JCTC Journal of Chemical Theory and Computation

Influence of Stacking on the Hydrogen Bond Donating Potential of Nucleic Bases

K. Vanommeslaeghe,*[†,‡] P. Mignon,[†,§] S. Loverix,[†,||] D. Tourwé,[‡] and P. Geerlings[†]

General Chemistry Group, Vrije Universiteit Brussel, Pleinlaan 2,
B-1050 Brussel, Belgium, Organic Chemistry Group, Vrije Universiteit Brussel,
Pleinlaan 2, B-1050 Brussel, Belgium, Centre for Surface Chemistry and Catalysis,
Katholieke Universiteit Leuven, Kasteelpark Arenberg 23, B-3001 Heverlee, Belgium,
and Department of Molecular and Cellular Interactions, Vlaams Interuniversitair
Instituut voor Biotechnologie, Pleinlaan 2, B-1050 Brussel, Belgium

Received April 25, 2006

Abstract: Hydrogen bonding is the dominant interaction in the pairing of nucleic bases and largely determines the stability of the double-helical structure of DNA. In a previous study, we used the molecular electrostatic potential (MEP) near a hydrogen-bond (HB) *acceptor* to demonstrate that the *intra*strand $\pi-\pi$ stacking interaction influences the *inter*strand HB *accepting* capacity of DNA/RNA bases. In the present work, we first examined at the MP2/6-31G(d) level whether the MEP near a HB *donating* site of an aromatic or nucleic base can be used as a computationally inexpensive measure for its HB *donating* potential, quantified as the interaction energy with an HB acceptor probe, and whether this also holds in the presence of a stacking partner. A good correlation was found for substituted anilines in a vacuum, and this seemed to hold for cytosine, stacked with substituted benzenes. However, when stacked pairs of nucleic bases were studied, no correlation between the MEP and the HB strength was found. This turned out to be caused by the direct interaction of the HB donor's stacking partner with the probe molecule as well as its influence on the MEP. After this perturbation was eliminated, a significant correlation was found. The influence of stacking on the HB donating potential was shown to be dominated by the stacking geometry and not by the nature of the stacking partner. The present findings suggest that the $\pi-\pi$ interaction on itself does not have an overall strengthening on H bonding in DNA.

Introduction

Hydrogen bonds (HBs) are highly important in biological systems. In particular, hydrogen bonding is the dominant interaction in the pairing of nucleic bases and largely determines the stability of the double-helical structure of DNA.[1] Likewise, $\pi-\pi$ stacking interactions play an important role in the structure,[2] catalysis,[3−5] and inhibition[6−8] of proteins, and the DNA double helix exhibits prominent stacking interactions between the nucleic bases.[1] Thus, it can be said that, apart from the covalent bonds in the backbone, the double-helical structure of DNA is essentially held together by polar *inter*strand HBs[9] and apolar (London) *intra*strand $\pi-\pi$ interactions.[1,10−12] These different interactions are thought to be cooperative; however, to the best of our knowledge, no theoretical study has been able to confirm this. In the study of stair motifs at protein−DNA interfaces, cooperativity was investigated for H-bonded and stacked trimers of nucleic acid bases and a charged amino acid.[13] The three-body term contribution was found to vary from −0.4 to +7.4 kcal/mol, where significantly negative values

---

* Corresponding author fax: +32-2-26293317; e-mail: Kenno.Vanommeslaeghe@vub.ac.be.

[†] General Chemistry Group, Vrije Universiteit Brussel.

[‡] Organic Chemistry Group, Vrije Universiteit Brussel.

[§] Katholieke Universiteit Leuven.

[||] Vlaams Interuniversitair Instituut voor Biotechnologie.

10.1021/ct600150n CCC: $33.50

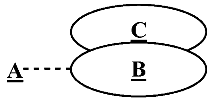


Figure 1. Naming conventions. **A**: ammonia. **B**: hydrogen-bonded base. **C**: stacking partner.

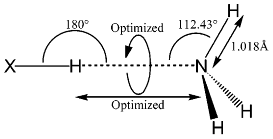


Figure 2. Partial optimization of an ammonia molecule as an HB acceptor.

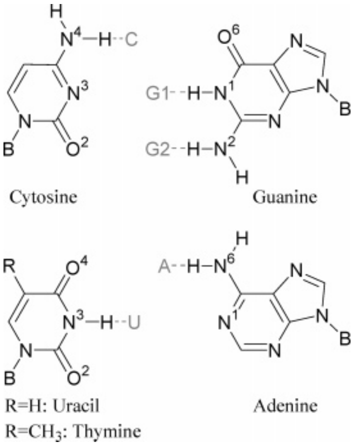


Figure 3. Four nucleic bases and their five biologically relevant HB donating sites, henceforward dubbed C, U, A, G1, and G2 (colored gray). B: sugar−phosphate backbone. R: variable group. DNA contains thymine (R = CH_3), while RNA contains uracil (R = H). Throughout this study, uracil was used for the sake of computational simplicity.

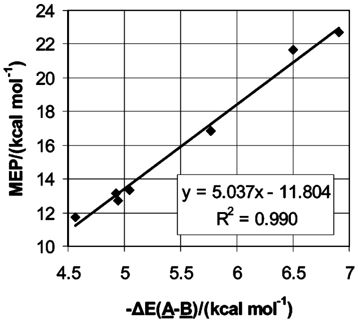


Figure 4. Correlation between the MEP and the strength of the HB between ammonia and a series of substituted anilines.

would have corresponded to cooperative binding. Also, electrostatics-based studies by Sivanesan et al. on hydrated nucleic bases show that stacked base pairs hydrate better than H-bonded ones.[14,15] They observed the most negative values of the molecular electrostatic potential (MEP) for the stacked pairs. Likewise, by calculating the MEP, we have demonstrated in a previous study that these *intra*strand stacking interactions influence the *inter*strand HB *accepting* capacity of DNA/RNA bases.[16]

As this influence is thought to be mainly the result of charge transfer, the impact of stacking on HB *donating* capacity may differ significantly. Also, the MEP has been known for a long time[17] as a reliable descriptor of the hydrogen-bond *accepting* strength: the deeper the electrostatic potential, the stronger the electrostatic interaction with water molecules and with hydrogen-bond donors in general.[18−21] However, we presently wish to use the MEP as a descriptor of the HB *donating* strength. An indication that this approach might be rewarding is the work by Politzer et al, showing successful correlations in the case of relatively simple systems.[22−24] Nevertheless, we cannot be fully confident that it will be accurate for the more complicated stacked systems presently studied. Consequently, our first goal is to find out whether the MEP near a hydrogen-bond donating site of a nucleic base in the presence of a stacking partner can be used as a computationally inexpensive way to estimate its HB donating potential, quantified as the binding affinity for a probe HB acceptor. Then, the influence of stacking on this HB donating potential will be investigated. To meet these goals, we first considered substituted anilines in order to test whether the correlation between the MEP and the HB strength is linear over a relevant range. Then,

Table 1. Strength of the HB between Ammonia and a Series of Substituted Anilines, $-\Delta E(\underline{A}-\underline{B})$, and the MEP at 4 a_0 from the HB Donating Hydrogen Atom[a]

	$-\Delta E(\underline{A}-\underline{B})$	MEP
p-aminoaniline	4.57	11.71
aniline	5.05	13.31
p-hydroxyaniline	4.93	13.16
m-methylaniline	4.95	12.73
m-chloroaniline	5.76	16.87
m-nitroaniline	6.50	21.63
p-nitroaniline	6.90	22.68

[a] All energies are in kcal mol^{-1}.

Table 2. Strength of the HB between Ammonia and Cytosine, Stacked with Substituted Benzenes, and the MEP at 4 a_0 from the HB Donating Hydrogen Atom[a]

	$-\Delta E(\underline{A}-\underline{BC})$	MEP
aniline	6.36	5.68
toluene	6.51	6.91
benzene	6.53	7.38
phenol	6.60	7.96
benzaldehyde	6.65	9.26
fluorobenzene	6.72	8.98
nitrobenzene	6.88	13.30

[a] All energies are in kcal mol^{-1}.

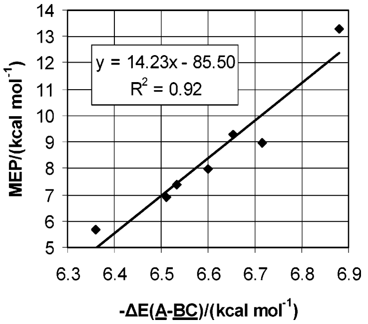


Figure 5. Correlation between the MEP and the strength of the HB between ammonia and cytosine stacked with a series of substituted benzenes.

Table 3. Strength of the HB between Ammonia and the Five Biologically Relevant HB Donating Sites on Four Nucleic Bases and the MEP at 4 a_0 from the HB Donating Hydrogen Atom[a]

$\underline{B}$	$-\Delta E^{AB}(\underline{A}-\underline{B})$	MEP	μ/D	AVG(ΔHB)	σ(ΔHB)	AVG(ΔHB/HB)
C	7.33	6.57	7.41	−0.52	0.52	−6.83%
U	8.66	7.22	5.03	−0.65	0.19	−7.86%
A	6.82	8.93	2.53	−0.69	0.22	−8.70%
G1	9.11	19.55	7.34	−0.71	0.35	−8.19%
G2	9.66	27.94	7.34	−0.71	0.35	−8.19%

[a] Also included are dipole moments (μ) of the isolated bases, as well as averages (AVG) and standard deviations (σ) of the influence of stacking on the HB donating potential; see the section "Corrected Results for Stacked Nucleic Bases" and Figure 12. All energies are in kcal mol^{-1}.

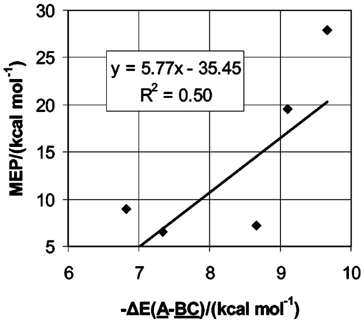


Figure 6. Poor correlation between the MEP and the strength of the HB between ammonia and the five biologically relevant HB donating sites on four nucleic bases.

this approach is extended to systems comprising cytosine, stacked with substituted benzenes. Finally, stacked pairs of nucleic bases are studied.

Methodology

Because stacking interactions are mainly determined by dispersion forces, they can only be properly described at a level of theory that includes electron correlation.[25−27] As density functional theory methods do not correctly reproduce the dispersion component of stacking interactions, at least second-order Møller−Plesset (MP2) theory must be used.[25,28] Consequently, all calculations were carried out at the MP2/6-31G(d) level of theory, using the Gaussian 03 software.[29] Ammonia was used as a HB acceptor and placed adjacent to the HB donating sites under investigation. After optimization of the hydrogen-bond geometry, the interaction energy was calculated as a measure for the HB strength, rigorously applying the counterpoise method to correct the basis set superposition error.[30] Wherever a two-body subset of the complete three-body system is considered (e.g., for the calculation of the $\Delta E(\underline{A}-\underline{C})$ term, see below), the basis set of the whole system was used, as recommended in the literature.[5,31,32] (The alternative approach, using the two-body basis set for the two-body interactions, was also attempted. The results deviated only 0.06 kcal mol^{-1} on average from those presently presented, with a maximum deviation of 0.15 kcal mol^{-1}. Nevertheless, the correlations with the MEP were slightly worse.)

Additionally, the MEP of the complex at a distance of 4 a_0 from the hydrogen atom of the same HB donating sites was determined, as earlier studies demonstrate that this descriptor correlates well with the acidity.[6,33,34] (Although it is likely that the Hartree−Fock level would be sufficiently accurate to calculate the MEP, MP2 results were used for the sake of consistency.)

To test whether the correlation between the MEP and the HB strength is linear over a relevant range, a series of substituted anilines was constructed and optimized. Then, the MEP was calculated at a distance of 4 a_0 from one of the hydrogen atoms attached to the aniline nitrogen along the N−H bond axis. The same hydrogen atom was subsequently used as a HB donor for a newly added NH_3 molecule. The resulting complex was fully optimized, whereupon the interaction energy between ammonia and the substituted aniline was calculated. As this yielded an excellent correlation (see "Results and Discussion"), a similar procedure was applied to cytosine, stacked with a series of substituted benzenes. Cytosine-substituted benzene complexes were already constructed and optimized in a previous study.[16] On each of these complexes, the MEP at a distance of 4 a_0 from the hydrogen atom of the biochemically relevant HB donating site of cytosine, was calculated. In parallel, an ammonia molecule was added to all of these HB donating sites (Figure 1), using the geometry from the fully minimized complex with *p*-nitroaniline. Then, the hydrogen-bond length and torsion angle were optimized (see Figure 2). Finally, the interaction energy between the ammonia molecule and the pair of stacked bases was calculated.

Using the resulting linear fit as a calibration curve should make it possible to infer the HB donating potential from the MEP, which is desirable because adding and optimizing an ammonia molecule and calculating the interaction energy becomes computationally expensive for larger systems. However, although a fair correlation was obtained, we cannot yet be sure that the HB donating potential can always be inferred from the MEP. Therefore, both descriptors were calculated for the biologically relevant HB donating sites of stacked pairs of nucleic bases (Figure 3). All possible combinations of bases were studied this way, each time considering both an optimized and an experimentally observed stacking orientation. The optimized geometries were taken from a study by Šponer et al., wherein pairs of separately optimized nucleic bases were positioned at a fixed distance, after which the twist angle and the parallel displacement between the rigid bases were optimized.[35] The experimental orientations were based on X-ray structures from the Research Collaboratory for Structural Bioinformatics Protein Data Bank,[36] on which the above-mentioned preoptimized nucleic bases were superimposed, as described in an earlier publication.[16] The MEP and the ammonia affinity were calculated for all biologically relevant HB

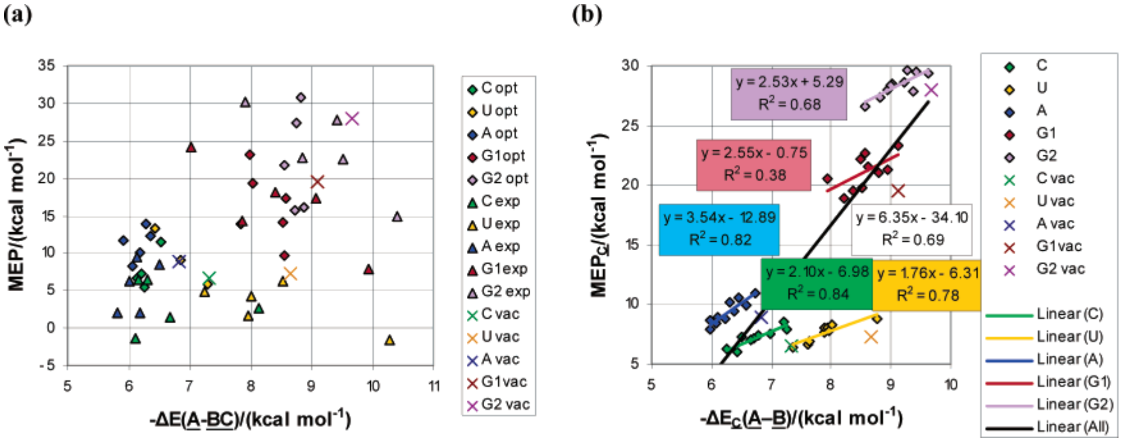


Figure 7. Correlation between the MEP and the interaction energy between ammonia and stacked pairs of nucleic bases before (a) and after (b) correction of the direct influence of base **C** on the MEP and the interaction energy. See the section "Corrected Results for Stacked Nucleic Bases" for an explanation.

donating sites (Figure 3), applying exactly the same procedure as that for cytosine stacked with substituted benzenes.

Throughout this article, the ammonia molecule will be named **A**, the base from which it receives a hydrogen bond **B**, and its stacking partner **C**, as schematized in Figure 1.

Results and Discussion

Substituted Anilines. The preliminary study of substituted anilines, aimed at establishing the range in HB strengths and MEPs and the linear correlation between both, yielded encouraging results. Indeed, Figure 4 shows an excellent correlation ($R^2 = 0.99$) for this series. Furthermore, the fact that a 12 kcal mol^{-1} range in MEP is translated into a 2.5 kcal mol^{-1} range in binding affinity (Table 1) indicates that possible inaccuracies in the MEP (see below) are not likely to have a significant influence on the predicted HB strengths.

Cytosine Stacked with Substituted Benzenes. As mentioned in the Introduction, a calibration curve was established for cytosine by calculating ammonia affinities and MEPs for this base, stacked with substituted benzenes with a wide range of electron-donating properties (Table 2 and Figure 5).[16] Nevertheless, the range in HB strengths is a factor 5 smaller than that for the substituted anilines. Although this results in a lower correlation coefficient ($R^2 = 0.92$), it should still be useful for extrapolation. Remarkably, the ratio between the MEP and the binding affinity is about a factor of 3 higher than that for the substituted anilines (Figure 5).

Isolated Nucleic Bases. Prior to starting the calculations on the stacked pairs of nucleic bases, HB strengths and MEPs were calculated for isolated bases, as a reference for calculating the influence of stacking on the HB donating potential. Table 3 shows that the exocyclic amine of adenine is the weakest HB donor and the exocyclic amine of guanine the strongest. As the bases differ significantly in chemical nature, there is a poor correlation with the MEP (Figure 6).

Stacked Nucleic Bases. As described under the Methodology section, ammonia affinities and MEPs were calculated for the 48 biologically relevant HB donating sites in 20 geometrically different stacked pairs of bases. The results are summarized in Figure 7a. Surprisingly, correlation between the MEP and the HB strength is completely absent for the whole data set ($R^2 = 0.18$), as well as for each nucleic base on its own. Also remarkable is the presence of two negative MEPs in the data set, which can be explained by observing that a lone electron pair of base **C** points roughly in the direction of the point at which the MEP is calculated. Further inspection of the NH_3 binding geometries reveals a few cases where the ammonia molecule *donates* a HB *to base* **C**. Although these undesired HBs feature such a strongly distorted geometry that they may more accurately be called "HB-like polar interactions", we deem them capable of ruining the correlation on their own right, given the narrow range of interaction energies observed for each base **B**.

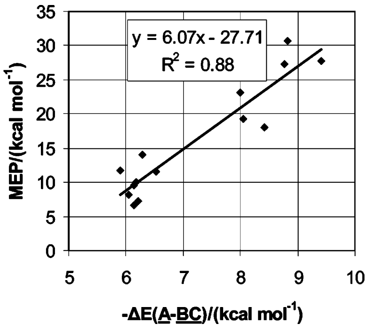


Figure 8. Correlation between the MEP and ΔE(**A**−**BC**) for **A**−**C** distances larger than 5 Å.

To test whether the bad correlation is indeed caused by direct interference of the base **C** with both the ammonia affinities and the MEP, correlation was sought in two subsets of the original data set, where only the points were retained in which the distance between base **C** and ammonia is larger than 4 and 5 Å, respectively. Although the former yields a definite improvement, the correlation is still rather poor ($R^2 = 0.57$). Indeed, to get a reasonable correlation, the cutoff distance must be set as high as 5 Å ($R^2 = 0.88$; Figure 8). The ratio is comparable to the value for substituted anilines but not to the value for cytosine stacked with substituted benzenes.

Corrected Results for Stacked Nucleic Bases. Although Figure 8 exhibits a reasonable correlation, it contains only 14 of our 48 original data points. Moreover, only three of the experimental—and thus biologically relevant—geometries

Table 4. Corrected Interaction Energies, ΔHBs, and MEPs for the Optimized (a) and Experimental (b) Geometries[a]

Part a

B	**C**	−Δ*E*(**A**−**BC**)	−Δ*E*(**A**−**C**)	−ΔE_{C}(**A**−**B**)	ΔHB	ΔHB/HB	MEP(**BC**)	MEP(**C**)	MEP$_{\mathrm{C}}$(**B**)	ΔMEP$_{\mathrm{C}}$(**B**)
C	C	6.52	0.09	6.43	−0.90	−12.34%	11.50	5.45	6.05	−0.52
C	U	6.25	−0.54	6.79	−0.54	−7.41%	5.49	−1.94	7.43	0.87
C	A	6.14	−0.56	6.70	−0.63	−8.57%	6.56	−0.58	7.14	0.58
C	G	6.22	−0.44	6.66	−0.67	−9.16%	7.20	0.21	6.99	0.42
U	C	6.44	−0.92	7.37	−1.29	−14.92%	13.29	6.86	6.43	−0.79
U	U	7.83	0.22	7.61	−1.04	−12.06%	13.92	7.23	6.69	−0.53
U	A	7.31	−0.58	7.89	−0.77	−8.92%	5.86	−1.83	7.69	0.48
U	G	6.86	−0.79	7.65	−1.01	−11.68%	9.06	2.13	6.93	−0.29
A	C	6.29	0.31	5.98	−0.85	−12.39%	13.96	6.04	7.93	−1.01
A	U	6.36	0.12	6.23	−0.59	−8.69%	12.33	3.52	8.80	−0.13
A	A′	6.18	0.07	6.11	−0.71	−10.47%	9.99	1.13	8.87	−0.07
A′	A	6.05	−0.33	6.39	−0.44	−6.38%	8.21	−1.24	9.45	0.51
A	G	5.90	−0.17	6.07	−0.75	−10.98%	11.75	3.37	8.39	−0.55
G1	C	8.52	−0.26	8.79	−0.32	−3.55%	14.00	−7.06	21.07	1.52
G1	U	8.59	0.06	8.53	−0.58	−6.36%	17.27	−2.47	19.74	0.19
G1	A	8.05	−0.33	8.37	−0.73	−8.06%	19.24	−0.30	19.54	−0.01
G1	G′	7.99	−0.23	8.22	−0.88	−9.71%	23.05	4.11	18.95	−0.60
G1′	G	8.55	−0.39	8.94	−0.17	−1.82%	9.67	−11.66	21.33	1.79
G2	C	8.87	−0.55	9.42	−0.25	−2.54%	16.08	−13.47	29.54	1.60
G2	U	8.56	−0.47	9.03	−0.64	−6.60%	21.69	−6.77	28.47	0.52
G2	A	8.76	−0.18	8.94	−0.72	−7.48%	27.29	−0.63	27.92	−0.03
G2	G′	8.83	0.00	8.83	−0.84	−8.66%	30.71	3.35	27.36	−0.58
G2′	G	8.73	−0.55	9.28	−0.39	−3.99%	15.76	−13.82	29.58	1.63

Part b

B	**C**	−Δ*E*(**A**−**BC**)	−Δ*E*(**A**−**C**)	−ΔE_{C}(**A**−**B**)	ΔHB	ΔHB/HB	MEP(**BC**)	MEP(**C**)	MEP$_{\mathrm{C}}$(**B**)	ΔMEP$_{\mathrm{C}}$(**B**)
C	C′	6.15	−0.10	6.25	−1.08	−14.70%	6.45	0.25	6.20	−0.36
C′	C	8.13	0.88	7.25	−0.08	−1.11%	2.66	−5.26	7.92	1.36
C	U	6.67	−0.30	6.97	−0.36	−4.87%	1.45	−6.09	7.54	0.98
C	A	6.31	−0.21	6.52	−0.81	−11.08%	6.38	−0.84	7.22	0.65
C	G	6.12	−1.09	7.21	−0.12	−1.63%	−1.41	−9.96	8.55	1.99
U	C	10.28	1.50	8.77	0.11	1.31%	−1.49	−10.22	8.73	1.51
U	U′	8.02	0.14	7.88	−0.78	−8.97%	4.21	−3.83	8.04	0.83
U′	U	8.53	0.55	7.98	−0.68	−7.87%	6.31	−1.41	7.72	0.50
U	A	7.24	−0.69	7.94	−0.72	−8.35%	4.91	−3.16	8.07	0.85
U	G	7.97	−0.05	8.03	−0.63	−7.29%	1.59	−6.68	8.28	1.06
A	C	6.49	0.20	6.30	−0.53	−7.72%	8.48	−1.71	10.19	1.26
A	U	6.01	−0.45	6.46	−0.37	−5.38%	6.17	−4.36	10.53	1.59
A	A′	6.14	0.17	5.97	−0.85	−12.47%	9.48	0.84	8.65	−0.29
A′	A	6.19	−0.39	6.58	−0.25	−3.62%	1.98	−7.98	9.96	1.03
A	G	5.81	−0.92	6.73	−0.09	−1.35%	2.01	−8.97	10.99	2.05
G1	C	9.93	0.81	9.12	0.01	0.10%	7.87	−15.38	23.25	3.70
G1	U	9.08	0.58	8.50	−0.61	−6.66%	17.28	−4.90	22.18	2.64
G1	A	7.85	−0.77	8.63	−0.48	−5.31%	14.37	−7.15	21.52	1.98
G1	G′	8.41	−0.15	8.56	−0.55	−6.05%	18.03	−4.63	22.66	3.12
G1′	G	7.02	−0.92	7.94	−1.17	−12.86%	24.11	3.63	20.48	0.93
G2	C	10.41	0.78	9.63	−0.03	−0.35%	14.83	−14.53	29.36	1.42
G2	U	9.52	0.29	9.23	−0.44	−4.53%	22.61	−5.96	28.57	0.62
G2	A	8.86	−0.53	9.38	−0.28	−2.91%	22.81	−5.03	27.84	−0.10
G2	G′	9.41	0.40	9.01	−0.66	−6.80%	27.79	−0.57	28.35	0.41
G2′	G	7.91	−0.66	8.57	−1.09	−11.30%	30.20	3.58	26.62	−1.32

[a] If the nature of the two bases is the same, a choice has to be made as to which base is **B** and which is **C**. In these cases, both possibilities were considered, where the base **B** that has the longest distance between its HB donating site(s) and base **C** is notated as X and the other, "most hindered" base is notated as X′. Note that, for the optimized structures of C−C and U−U, this is irrelevant because their geometries are perfectly symmetrical. All energies are in kcal mol^{-1}.

satisfied the criterion for this data set. Clearly, if any relevant insights into the influence of stacking on the HB donating potential are to be obtained, a method should be devised to eliminate the *direct* influence of the stacking partner, so that the influence *via the stacking interaction* becomes apparent. The interaction energy between **A** and **B** *in the presence of*

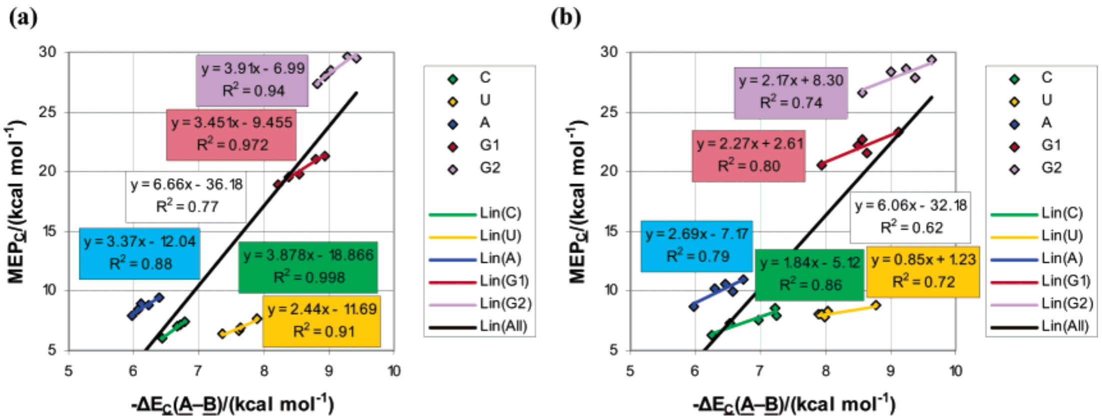


Figure 9. Correlation between the corrected MEP and the final corrected interaction energy for the optimized (a) and the experimental (b) structures.

$\underline{\mathbf{C}}$, $\Delta E_{\underline{\mathbf{C}}}(\underline{\mathbf{A}}-\underline{\mathbf{B}})$, was calculated as $\Delta E(\underline{\mathbf{A}}-\underline{\mathbf{BC}}) - \Delta E(\underline{\mathbf{A}}-\underline{\mathbf{C}})$. Similarly, the corrected MEP was calculated by subtracting the MEP of base $\underline{\mathbf{C}}$ from the MEP of the $\underline{\mathbf{BC}}$ system. The results of the final corrected calculations of the ammonia affinity and the MEP are summarized in Table 4 and Figure 7b. In contrast with Figure 7a, reasonable correlations are obtained for all of the biologically relevant HB donating sites except the endocyclic N^1-H of guanine (G1), with ratios between 1.8 and 3.5. Also, there is a low but significant correlation across the whole data set, with a ratio between the MEP and the HB strength that is on the same order of magnitude as the ratio for the substituted anilines.

Remarkably, separating the optimized and the experimental data significantly improves correlation in nearly all cases, as can be seen by comparing Figures 9a and b to Figure 7b. This improvement is especially pronounced for site 1 of guanine (G1); here, the extremely bad correlation in Figure 7b results from the fact that the optimized and the experimental geometries are clearly located on two distinct lines, each with a rather good correlation (Figure 9a and b). It is also apparent that, for each base, the range in HB strengths is significantly smaller for the optimized structures than for the experimental data. Nevertheless, the correlations are much better for the optimized structures. This can be attributed to the fact that the experimental geometries are imposed by the overall structure of the RNA/DNA double helix. From the point of view of an isolated duo of stacked bases in a vacuum, these experimental geometries are rather arbitrary. In particular, the tilt angle between the two bases cannot be justified without considering the double-helical structure.[16] This can be regarded as a perturbation of the optimized geometry, which results in a perturbation of the correlations. When comparing parts a and b of Figure 9, it is tempting to speculate that a tilt angle differing slightly from zero has a smaller effect on the MEP than on the “internal electronical structure” of base $\underline{\mathbf{B}}$ and thus on the HB strength. In any case, one should be careful when trying to correlate the HB strength with the MEP of nonoptimized structures.

As the data sets for the different bases $\underline{\mathbf{B}}$ occupy discrete regions in Figure 7b, we tried to correlate the influence of stacking on the HB strength ΔHB, defined as $\Delta E^{AB}(\underline{\mathbf{A}}-\underline{\mathbf{B}}) - \Delta E_{\underline{\mathbf{C}}}(\underline{\mathbf{A}}-\underline{\mathbf{B}})$, with the influence of stacking on the MEP: $\Delta\text{MEP} = \text{MEP}_{\underline{\mathbf{C}}}(\underline{\mathbf{B}}) - \text{MEP}(\underline{\mathbf{B}})$. However, as can be seen in Figure 10, there is no clear advantage to this approach. Nevertheless, ΔHB is an interesting quantity in its own right, as it is the most direct measure for the influence of stacking on the HB donating potential. First, we tried to correlate ΔHB with the nature of the stacking partner $\underline{\mathbf{C}}$ (Figure 11). Overall, the range of ΔHBs, quantified as its standard deviation [σ(ΔHB) in Table 3], is largest when base $\underline{\mathbf{C}}$ is cytosine and smallest in the case of uracil. This range might be related to the diversity of relative orientations, on one hand, and on the dipole moment of the stacking partner $\underline{\mathbf{C}}$, on the other hand. The former effect is difficult to quantify; for the latter effect, Figure 12 suggests a weak correlation, although this result is inconclusive. Likewise, it is difficult to draw general conclusions about the effect of the nature of stacking partner $\underline{\mathbf{C}}$ on the HB strength itself. Although it is clear that stacking generally weakens the hydrogen-bond donating capacity—on average, by 0.59 kcal mol^{-1} [see AVG(ΔHB) in Table 3]—the influence of the nature of molecule $\underline{\mathbf{C}}$ seems to be subordinate to the influence of its orientation, to the extent that it cannot be inferred from our data which of the nucleic bases $\underline{\mathbf{C}}$ has the strongest intrinsic electron-donating effect on its stacking partner $\underline{\mathbf{B}}$.

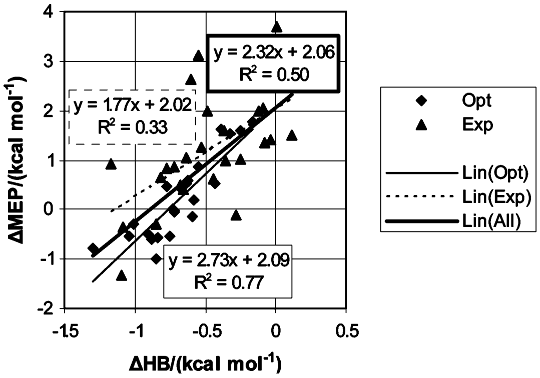


Figure 10. Correlation between ΔMEP and ΔHB, for the optimized, the experimental, and all geometries of the stacked pairs of nucleic bases.

In our previous work,[16,37] a comparable approach yielded evidence that the HB accepting capacity increases upon stacking. In the present study, we observe a decrease of the HB donating capacity. Both observations are consistent with the notion that the stacking partner “pushes electron density into the ring”, as proposed in our previous work.[16] As a result,

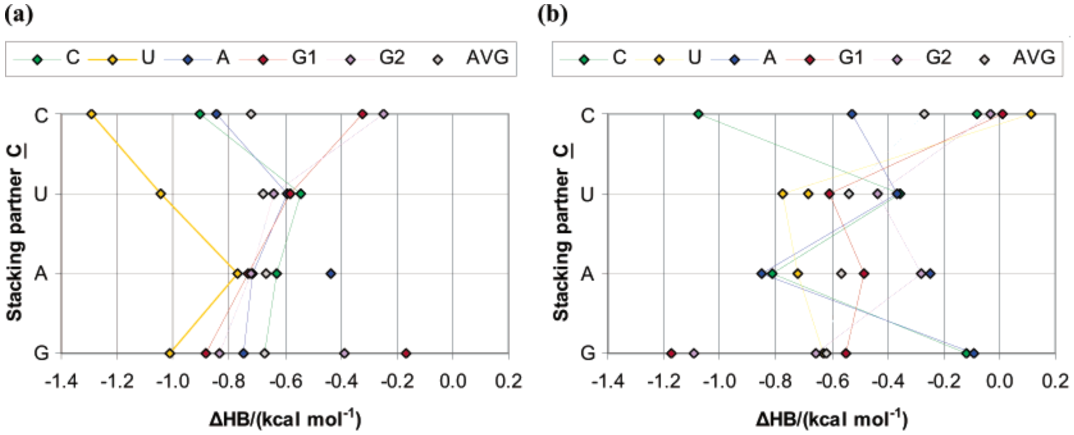


Figure 11. Influence of stacking on the HB donating potential for the optimized (a) and the experimental (b) structures. The difference in HB strength with respect to the unstacked base **B** is plotted in the horizontal axis, so that negative values represent a decrease in hydrogen-bond strength upon stacking. The color coding represents the nature of the base **B** as defined in Figure 1, while the nature of the base **C** is plotted on the vertical axis.

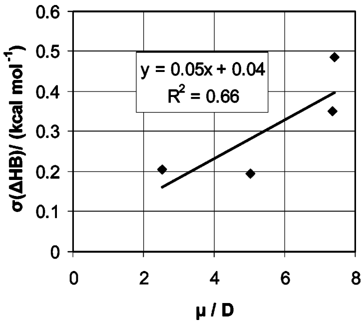


Figure 12. Correlation between the standard deviation of the influence of stacking on the HB strength and the dipole moment μ.

the influence of the stacking interaction on the HB donating ability counteracts its effect on the HB accepting capacity. To test the relative strength of these opposing influences, NH_3 was used as a HB donor for both of the HB accepting sites of cytosine and for the single HB accepting site of adenine. To determine the influence of stacking on the resulting HB strengths, these calculations were repeated with adenine and guanine as stacking partners for cytosine and with uracil as the stacking partner for adenine, using the optimized stacking geometries. The results for this set are summarized in Table 5 and show only small increases in HB strength (0.21 kcal mol^{-1} at most).

However, one could argue that this result is biased because ammonia is a better HB acceptor than a HB donor, as the absolute HB strengths are about half as high when ammonia is used as a HB donor instead of an acceptor. If, for the sake of argument, we suppose that the influence of stacking on the HB strength is proportional to the HB strength itself (thus ignoring the cases where the influence is opposite in sign), we should consider percentages rather than absolute differences in HB strength. The ΔHB/HB values in Tables 4 and 5 and the averages [AVG(ΔHB/HB)] in Table 3 indicate that, even under this bold supposition, the average negative influence of stacking on the HB donating capacity of nucleic bases is higher than its positive influence on their HB accepting capacity. Although more data are needed to obtain a definitive conclusion in this respect, the present findings suggest that the electron-transfer accompanying $\pi-\pi$ stacking on itself does not strengthen interstrand hydrogen bonding in DNA.

Corrected Results for Cytosine Stacked with Substituted Benzenes. The same corrections as discussed above were applied on cytosine, stacked with substituted benzenes (Table 6). Here, a slight worsening of the correlation is observed ($R^2 = 0.88$; see Figure 13), which can be explained by considering that the range of HB strengths decreased from 0.52 to 0.16 kcal mol^{-1}. More importantly, the (previously anomalous) ratio is now in line with the results for the substituted anilines and the stacked nucleic bases. Thus, it can be concluded that, in the uncorrected calculations, we were mainly seeing the *direct effects* of the stacking partner on the MEP and the HB strength, while the *effects via the stacking interaction* are much more subtle.

Conclusions

In line with previous results,[24,33,34] it was shown that the MEP is an excellent measure for the HB donating potential of substituted anilines. Likewise, a good correlation between this MEP and the HB donating potential was found for a fixed nucleic base (cytosine) stacked with substituted benzenes in similar orientations. However, the observed differences were found to be mostly the result of the direct influence of the stacking partner **C** on the HB acceptor's binding site, while the electronic influence of **C** on the hydrogen-bonded base **B** was shown to be weaker by factors on the order of 3 (HB strength) and 15 (MEP). Moreover, the correlation completely breaks down when substantially differing stacking partners such as different nucleic bases are used, or if geometric variations of the stacking partner are allowed.

Correlation was partially restored by eliminating the direct influence of **C** out of the MEP as well as the binding energy of the HB accepting probe. Yet, for the pairs of stacked nucleic bases that were studied, no definite conclusions could be drawn regarding the influence of the nature of base **C** on the hydrogen-bond donating capacity of a parallel stacked base **B**. Nevertheless, the present results suggest that this

Table 5. Interaction Energies for Stacked Pairs of Nucleic Bases and Isolated Nucleic Bases *Receiving* a HB from NH_3

$\underline{\mathbf{B}}$	$\underline{\mathbf{C}}$	$-\Delta E(\underline{\mathbf{A}}-\underline{\mathbf{BC}})$	$-\Delta E(\underline{\mathbf{A}}-\underline{\mathbf{C}})$	$-\Delta E_{\underline{\mathbf{C}}}(\underline{\mathbf{A}}-\underline{\mathbf{B}})$	$-\Delta E^{AB}(\underline{\mathbf{A}}-\underline{\mathbf{B}})$	ΔHB	ΔHB/HB
C2	A	4.08	0.85	3.23	3.11	0.12	3.88%
C2	G	4.30	0.99	3.31	3.11	0.21	6.61%
C3	A	4.40	−0.43	4.83	4.78	0.05	1.06%
C3	G	4.33	−0.51	4.84	4.78	0.06	1.25%
A	U	3.27	−0.19	3.46	3.50	−0.04	−1.08%

Table 6. Corrected Interaction Energies and MEPs for Cytosine, Stacked with Substituted Benzenes[a]

	$-\Delta E(\underline{\mathbf{A}}-\underline{\mathbf{BC}})$	$-\Delta E(\underline{\mathbf{A}}-\underline{\mathbf{C}})$	$-\Delta E_{\underline{\mathbf{C}}}(\underline{\mathbf{A}}-\underline{\mathbf{B}})$	MEP($\underline{\mathbf{BC}}$)	MEP($\underline{\mathbf{C}}$)	$\mathrm{MEP}_{\underline{\mathbf{C}}}(\underline{\mathbf{B}})$
aniline	6.36	−0.22	6.58	5.68	−1.34	7.02
toluene	6.51	−0.05	6.56	6.91	0.02	6.89
benzene	6.53	−0.08	6.61	7.38	0.30	7.08
phenol	6.60	0.01	6.59	7.96	1.03	6.92
benzaldehyde	6.65	0.06	6.59	9.26	2.26	7.00
fluorobenzene	6.72	0.06	6.65	8.98	1.93	7.05
nitrobenzene	6.88	0.15	6.73	13.30	5.85	7.45

[a] All energies are in kcal mol^{-1}.

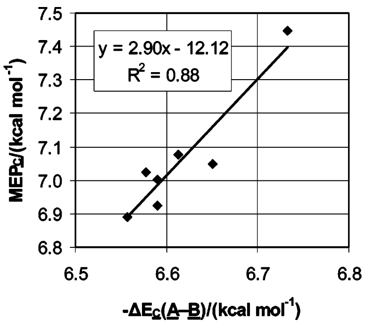


Figure 13. Correlation between the corrected MEP and the final corrected interaction energy between ammonia and cytosine stacked with a series of substituted benzenes.

influence varies more with the stacking geometry if the stacking partner $\underline{\mathbf{C}}$ has a larger dipole moment. Also, the set of optimized geometries behaved qualitatively differently from the set of experimental geometries. Stacking was shown to weaken the HB *donating* capacity in 46 out of 48 cases, with a mean weakening of 0.59 kcal mol^{-1}. Conversely, the previously observed strengthening effect on the HB *accepting* capacity[16,37] was found to be 0.21 kcal mol^{-1} at most, albeit on a very limited data set and using a poor HB donor probe. Although a more thorough study is required to obtain a definitive conclusion in this respect, the present findings suggest that the $\pi-\pi$ interaction on itself has no net potentiating effect on the interstrand hydrogen bonds in DNA. It should also be noted that both the weakening effect on the donating capacity and the strengthening effect on the accepting capacity are consistent with the notion that the stacking partner “pushes electron density into the ring”, as proposed in our previous work.[16]

On a practical note, the present study demonstrates that optimized or at least similar stacking geometries are required in order to have any meaningful correlation between the MEP and the HB donating potential. Even under these premises, final correlations are clearly not good enough for predicting hydrogen-bond strengths based on the MEP. Nevertheless, for a series of optimized geometries with the same HB donor $\underline{\mathbf{B}}$ and similar stacking partners $\underline{\mathbf{C}}$, a ranking of the (corrected) MEPs should be useful as a rough measure for the ranking of the HB donating potentials, where a difference of two to four energy units in the MEP very approximately corresponds to a difference of one energy unit in the HB donating potential.

Acknowledgment. P.G. thanks the “F.W.O.-Vlaanderen” and the “Vrije Universiteit Brussel” for continuous support of his research group.

References

(1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: Berlin, Germany, 1984.

(2) McGaughey, G. B.; Gagné, M.; Rappé, A. K. *J. Biol. Chem.* **1998**, *273*, 15458−15463.

(3) Versées, W.; Loverix, S.; Vandemeulebroucke, A.; Geerlings, P.; Steyaert, J. *J. Mol. Biol.* **2004**, *338*, 1−6.

(4) Steyaert, J. *Eur. J. Biochem.* **1997**, *247*, 1−11.

(5) Raines, R. T. *Chem. Rev.* **1998**, *98*, 1045−1065.

(6) Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. *Nature* **1999**, *401*, 188−193.

(7) Wang, D.-F.; Wiest, O.; Helquist, P.; Lan-Hargest, H.-Y.; Wiech, N. L. *J. Med. Chem.* **2004**, *47*, 3409−3417.

(8) Nielsen, T. K.; Hildmann, C.; Dickmanns, A.; Schwienhorst, A.; Ficner, R. *J. Mol. Biol.* **2005**, *354*, 107−120.

(9) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem.*−*Eur. J.* **1999**, *5*, 3581−3594.

(10) Suzuki, M.; Amano, N.; Kakinuma, J.; Tateno, M. *J. Mol. Biol.* **1997**, *274*, 421−435.

(11) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911−940.

(12) Bommarito, S.; Peyret, N.; SantaLucia, J., Jr. *Nucleic Acids Res.* **2000**, *28*, 1929−1934.

(13) Wintjens, R.; Biot, C.; Rooman, M.; Lievin, J. *J. Phys. Chem. A* **2003**, *107*, 6249−6258.

(14) Sivanesan, D.; Babu, K.; Gadre, S. R.; Subramanian, V.; Ramasami, T. *J. Phys. Chem. A* **2000**, *104*, 10887−10894.

(15) Sivanesan, D.; Sumathi, I.; Welsh, W. J. *Chem. Phys. Lett.* **2003**, *367*, 351−360.

(16) Mignon, P.; Loverix, S.; Steyaert, J.; Geerlings, P. *Nucleic Acids Res.* **2005**, *33*, 1779−1789.

(17) Kollman, P.; McKelvey, J.; Johansson, A.; Rothenberg, S. *J. Am. Chem. Soc.* **1975**, *97*, 955−965.

(18) Baeten, A.; De Proft, F.; Geerlings, P. *Chem. Phys. Lett.* **1995**, *235*, 17−21.

(19) Baeten, A.; De Proft, F.; Geerlings, P. *Int. J. Quantum Chem.* **1996**, *60*, 931−939.

(20) Mishra, P. C.; Kumar, A. In *Molecular Electrostatic Potentials: Concepts and Applications*; Murray, J. S., Sen, K., Eds.; Elsevier: Amsterdam, The Netherlands, 1996; Chapter 6, pp 257−296.

(21) Kushwaha, P. S.; Mishra, P. C. *Int. J. Quantum Chem.* **2000**, *76*, 700−713.

(22) Murray, J. S.; Politzer, P. *J. Org. Chem.* **1991**, *56*, 6715−6717.

(23) Murray, J. S.; Politzer, P. *J. Chem. Res. Synop.* **1992**, 110−111.

(24) Hagelin, H.; Murray, J. S.; Brinck, T.; Berthelot, M.; Politzer, P. *Can. J. Chem.* **1995**, *73*, 483−488.

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

(26) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887−10893.

(27) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200−10207.

(28) Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624−1626.

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

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

(31) Van Duijneveldt, F. B.; Van Duijneveldt-Van De Rijdt, J. G. C. M.; Van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873−1885.

(32) Turi, L.; Dannenberg, J. J. *J. Phys. Chem.* **1993**, *97*, 2488−2490.

(33) De Proft, F.; Amira, S.; Choho, K.; Geerlings, P. *J. Phys. Chem.* **1994**, *98*, 5227−5233.

(34) Olasz, A.; Mignon, P.; De Proft, F.; Veszprémi, T.; Geerlings, P. *Chem. Phys. Lett.* **2005**, *407*, 504−509.

(35) Šponer, J.; Leszczyński, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590−5596.

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

(37) Guo, D. W.; Sijbesma, R. P.; Zuilhof, H. *Org. Lett.* **2004**, *6*, 3667−3670.

CT600150N

J. Chem. Theory Comput. **2006,** *2,* 1453−1463 **1453**

JCTC Journal of Chemical Theory and Computation

Molecular Dynamics Simulations of the Orientation of Ni(II)•Gly-Gly-His and Ni(II)•Arg-Gly-His Metallopeptide−DNA Association

Ya-Yin Fang,[§] Kenny B. Lipkowitz,[‡] and Eric C. Long*[,†]

Department of Biochemistry and Molecular Biology and Department of Chemistry,
Howard University, 520 W Street, NW, Washington, D.C. 20059, and
Department of Chemistry and Chemical Biology, Purdue School of Science,
Indiana University−Purdue University Indianapolis (IUPUI),
402 North Blackford Street, Indianapolis, Indiana 46202-3274

Received February 21, 2006

Abstract: Ni(II)•Xaa-Gly-His metallopeptides (where Xaa is any α-amino acid) bind selectively to the minor groove of A/T-rich DNA regions as a function of their amino acid composition and chirality. Molecular dynamics simulations were performed to clarify the most likely binding orientations of Ni(II)•Gly-Gly-His and Ni(II)•L-Arg-Gly-His upon association with the B-form oligonucleotide d(CGCGAATTCGCG)$_2$. Upon examination of four possible docking orientations (**I**−**IV**), these studies indicated that both metallopeptides favor association with DNA via **I**, involving insertion of the edge of the metallopeptide containing the amino-terminal N−H and the imidazole pyrrole N−H group of His into the minor groove. These metallopeptide moieties play important roles in this DNA recognition mode by functioning as H-bond donors to minor groove acceptors such as the N3 of adenine or the O2 of thymine located on the floor of the minor groove. The positively charged side chain of L-Arg was found to enhance DNA recognition relative to that exhibited by Ni(II)•Gly-Gly-His through an increased electrostatic interaction, its favorable stereochemistry, and by providing a third point of contact with the minor groove floor. The simulation of orientation **I** was found to reproduce the experimentally supported DNA−metallopeptide orientation, revealing factors that are important for the further development of DNA-binding ligands.

Introduction

Sequence-selective DNA minor groove binding ligands are of ongoing interest due to their ability to act as nucleic acid conformational probes, footprint reagents, and antitumor or antiparasitic drugs.[1−4] In this general area, metallopeptides such as M(II)•Gly-Gly-His (where M = Cu or Ni) have attracted attention as DNA-interactive agents because they display the same chemical functional groups employed by proteins or peptide-derived natural products for DNA and RNA recognition.[5,6] In addition, the metal center of a metallopeptide produces a well-defined complex shape while also giving the peptide redox reactivity, enabling metallopeptides to cleave nucleic acids oxidatively in a fashion reminiscent of the bleomycin group[7] of clinically employed antitumor agents. Indeed, recent results indicate that these systems are very efficient single-[8] and double-strand[9] DNA cleavage agents (via C4′−H abstraction) and, earlier, that Cu(II)•Gly-Gly-His displays activity against Ehrlich ascites tumor cells.[10] Metallopeptides derived from the basic Gly-Gly-His peptide thus provide models for the study of fundamental DNA-small molecule interactions.[5,6,11]

---

* Corresponding author phone: (317)274-6888; fax: (317)274-4701; e-mail: long@chem.iupui.edu.

[†] IUPUI.

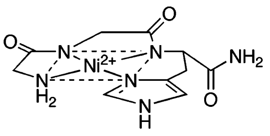
[‡] Department of Chemistry, Howard University.

[§] Department of Biochemistry and Molecular Biology, Howard University.

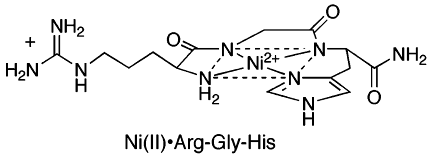
10.1021/ct600074r CCC: $33.50

To date, studies of M(II)•Gly-Gly-His-derived metallopeptides have focused on their ability to recognize and cleave DNA.[5,6] Through a variety of experimental techniques it has been determined that these systems orient stereospecifically[12,13] in the DNA minor groove and that Ni(II)•Gly-Gly-His displays a slight preference for A/T-rich DNA regions;[14] however, DNA cleavage becomes more efficient and focused to A/T-rich regions when Arg and Lys, containing positively charged side chains, are included in the tripeptide ligand.[14] DNA cleavage recognition is also altered by the stereochemistry of the amino acids present in the tripeptide ligand. For example, recent experimental studies have shown that Ni(II)•L-Arg-Gly-His has an increased and altered DNA site-selectivity relative to Ni(II)•D-Arg-Gly-His.[15]

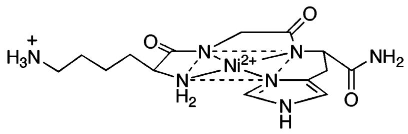
Unfortunately, information concerning the exact orientation of DNA−metallopeptide association, and factors leading to their efficient DNA cleavage, is still insufficient due to a lack of high-resolution structures and the dynamic nature of their DNA interaction. While experimental evidence available at the present time, including results from an NMR investigation,[13] point to a likely DNA binding orientation in which the His imidazole and terminal amine "edge" of each square-planar structure is inserted into the minor groove, alternative DNA association modes that are not detected by NMR cannot be definitively ruled out. Given the tetradentate fashion in which the central metal ion is bound through the amino terminal amine, two intervening, deprotonated peptide nitrogens, and the His imidazole,[16] each side of the square planar complex formed exposes several combinations of functional groups that could potentially mediate an alternative interaction with DNA. It is thus difficult to identify unambiguously which of four possible edges of the equatorial plane of a Ni(II)•Gly-Gly-His-derived complex inserts into the minor groove and which atoms interact with the floor of the minor groove. Further, without knowledge of a well-defined DNA binding orientation, it is difficult to assess the origin of the site-selectivities observed and how DNA cleavage activity is enhanced in some cases.



Ni(II)•Gly-Gly-His



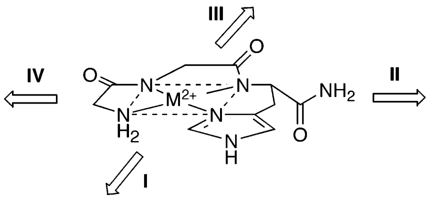
Ni(II)•Arg-Gly-His



Ni(II)•Lys-Gly-His

To address these questions, four potential minor groove association modes of two metallopeptides were investigated and compared for the first time through molecular dynamics simulations using Ni(II)•Gly-Gly-His, the least substituted metallopeptide, and Ni(II)•L-Arg-Gly-His, which displays enhanced A/T site-selectivity. To probe and compare their most likely DNA binding orientations, these metallopeptides were docked without constraints into the minor groove of DNA in four possible orientations as defined in Scheme 1. These two particular metallopeptides were chosen for this examination because they have been studied experimentally;[5,6] therefore, information exists to guide and validate the interpretation of the simulations presented herein.

Scheme 1. Binding Orientations Employed upon Metallopeptide-Minor Groove Docking



As will be presented, our simulations indicate that the most likely Ni(II)•Xaa-Gly-His metallopeptide-minor groove binding orientation is **I**, in which the N−H group of the His imidazole plays an important role in DNA recognition by functioning as an H-bond donor to the floor of the minor groove. In this orientation, the amino-terminal N−H protons assist in the recognition process in a fashion dependent upon the stereochemistry and identity of the amino acid present in this terminal residue. When present, a positively charged Arg residue assists in maintaining the favored binding orientation and enhances DNA association relative to Gly through an increased electrostatic interaction and by providing a third transient point of contact with the DNA. In the favored orientation (**I**) the simulations also support the DNA−metallopeptide orientation suggested through previous NMR studies.[13] In addition, new insights into the origins of the observed C4′−H abstraction chemistry, leading to direct DNA strand scission, were revealed. Overall, these studies have disclosed factors that are involved in metallopeptide−DNA association and impact the development of DNA binding ligands, in general. Perhaps as important, these studies provide an example to indicate that the judicious use of MD simulations can lead to meaningful insights pertaining to drug−DNA orientation phenomena in the absence of abundant experimental data.

Methods

General Considerations. Molecular modeling and molecular dynamics simulations of Ni(II)•Gly-Gly-His and Ni(II)•L-Arg-Gly-His bound to the minor groove of d(CGCGAATTCGCG)$_2$ were carried out using the protocols employed by Wellenzohn and co-workers.[17−19] Calculations were performed on an SGI Octane using MacroModel 7.4,[20] SPARTAN 5.11,[21] and AMBER 7.[22] Default settings for these programs were used unless specified otherwise. Force-field parameters for the tripeptide ligand bound to Ni^{2+} were developed as reported elsewhere;[13] these force-field parameters were derived from a crystallized Gly-Gly-His metallopeptide by optimizing missing parameters until the RMS deviation between the calculated and observed structure was

Table 1. Box Dimensions, Number of Waters, and Γ Values of Each Simulation

system	binding orientation	box dimension (Å)			number of waters	Γ value
		X	Y	Z		
DNA-Ni(II)•Gly-Gly-His	**I**	47.67	69.33	60.81	4810	200.41
	II	49.75	69.70	59.63	4951	206.29
	III	49.75	68.06	64.48	5232	218.00
	IV	51.73	67.01	62.33	5303	220.95
DNA-Ni(II)•L-Arg-Gly-His	**I**	57.63	69.84	50.12	4871	202.95
	II	54.78	67.01	65.03	5822	242.58
	III	51.04	68.06	63.40	5783	240.96
	IV	56.74	67.26	61.99	5847	243.62

minimized. As reported previously, the RMS deviation between the crystal structure and the AMBER minimized structure for those charges and geometry parameters is 0.257 Å, and the average RMS deviation between the X-ray structure and the computed structure along the testing simulation for those charges and geometry parameters is 0.283 Å. While there are no guarantees that the existing parameter set is optimal, as noted previously, the molecular dynamics simulations of the DNA−metallopeptide complex with the AMBER force field and the parameter set for Ni(II)•Gly-Gly-His not only generated adequate structural information without major structural deformations but also yielded reliable relative binding energies which corresponded to the site selectivities exhibited by the different metallopeptides and DNA.[13]

Starting Structures. Taking advantage of the noted similarities that occur between netropsin and Ni(II)•Xaa-Gly-His metallopeptides with regard to their A/T-rich site-selectivities and other features,[13] the crystal structure of netropsin-bound d(CGCGAATTCGCG)$_2$ (PDB reference code 1D86)[23] was used to create a minor groove binding site by replacing netropsin with Ni(II)•Gly-Gly-His or Ni(II)•L-Arg-Gly-His. The Ni(II)•Gly-Gly-His metallopeptide starting structure was derived from a crystal structure,[24] while that of Ni(II)•L-Arg-Gly-His was generated based on the same crystal structure of Ni(II)•Gly-Gly-His substituted appropriately in the first peptide position followed by molecular mechanics minimization; this substitution does not lead to changes in the rigid Ni(II)•Gly-Gly-His-like “core” of Ni(II)•Arg-Gly-His. The assembly of each DNA−metallopeptide complex was accomplished by docking a metallopeptide manually at the dyad axis of the oligonucleotide substrate in one of the four orientations defined in Scheme 1. Bump-checking was turned on to ensure that no overlapping atoms resulted during the docking process. Subsequently, each assembled metallopeptide−DNA complex was exported to AMBER. Given that each strand of DNA has 11 PO_4^- and Ni(II)•L-Arg-Gly-His possesses one positive charge, to achieve electroneutrality, 22 and 21 Na^+ counterions were added separately to the DNA-Ni(II)•Gly-Gly-His and DNA-Ni(II)•L-Arg-Gly-His complex systems, respectively, through use of the add-Ions command in the Xleap facility of AMBER. Solvation with TIP3P water boxes (resulting in a 10 Å solvent shell in all directions) completed each system. Table 1 lists the dimensions, number of waters, and corresponding relative Γ values (water/nucleotide) for each system studied.

Molecular Dynamics Simulations. The docked metallopeptide−DNA complexes described earlier, with associated Na^+ ions and water bath, were energy-minimized through 500 steps of a conjugate gradient, with 100 kcal $(\mathrm{mol}\bullet\text{Å})^{-1}$ restraints on the DNA and counterion positions. During the following three, 500-step minimizations, the restraints were relaxed stepwise by 25 kcal $(\mathrm{mol}\bullet\text{Å})^{-1}$ per step. Thus, the fifth, 500-step minimization was performed without restraints. Each optimized structure was heated from 100 to 300 K over a time period of 125 ps (with a temperature coupling of 0.2 ps), while positional restraints of 100 kcal$(\mathrm{mol}\bullet\text{Å})^{-1}$ were used for the DNA and the counterions. A constant volume was maintained during this process. Subsequently, three sequential 25 ps MD steps at 300 K were carried out accompanied by the gradual loosening of restraints at a rate of 25 kcal$(\mathrm{mol}\bullet\text{Å})^{-1}$ per step. The resulting restraint-free systems were allowed to equilibrate for an additional 15 ps; the temperature was allowed to fluctuate around 300 K with a temperature coupling time of 0.2 ps, and the pressure was allowed to fluctuate around 1 bar with a pressure coupling of 0.2 ps. Finally, production runs in excess of 1400 ps were carried out following this protocol. For each simulation, 5000 structures were saved to disk for postprocessing by uniformly sampling the trajectory during the production run. All analyses were carried out using the CARNAL and ANAL programs in AMBER.

MM Energies. Gas-phase molecular mechanical (MM) energies were averaged over all snapshots acquired between 500 and 1500 ps and included contributions from electrostatic and van der Waals interactions for the complex (DNA + metallopeptide), receptor (DNA), and ligand (metallopeptide). All MM calculations (AMBER) were performed with a nonbonded cutoff of 99 Å and a dielectric constant of 1 in the absence of any solvent or counterions. The receptor and ligand geometries were taken from those of the complex, and thus there is no internal energy contribution (i.e., from bonds, angles, and dihedrals) to the net MM average.

Potential Sources of Error. Potential errors in our computational methods that could influence the results of this study are worthy of note. The first possible error is the potential function itself. No force field is perfect, and they all have inherent errors. In this study, however, we have carried out simulations on exactly the same guest molecule (metallopeptide) binding to the same host (DNA) in the same binding site but with different docking orientations. Hence if the force field underestimates, for example, hydrogen bonding or overestimates electrostatic effects, it does so for all orientations, thus leading to a cancellation of errors and providing meaningful relative energies. The next possible error to consider is that the simulation times, albeit quite lengthy in terms of CPU time (multiple weeks per trajectory), correspond to exceedingly short times in reality. In this study we probed energies, structures, and dynamical features on the nanosecond time scale, and that might be a source of error. However, we note that in structures we deem as being stable docking orientations the guest remains bound to the host for the entire trajectory, but, for unstable orientations, the guest dissociates from the DNA and migrates away from its docking site or it reorients itself from that unstable orientation to the stable orientation. Hence, we are observing important structural transitions within the simulation time period even though this time period is brief. Finally, we point out that by selecting another water potential or a different force field or a different set of simulation conditions, it is possible that we could derive different results from those presented here. However, previous simulation results using the same methods were found to be fully consistent with NOE data showing which protons of the metallopeptide are adjacent to a given set of protons on the DNA, and the results are consistent with other experimental results, the most notable of which is the agreement with the DNA degradation results described in an earlier paper.[15] Hence we summarize by noting that while possible errors could influence the results, those errors are marginal in scope and the computational protocols we have implemented have provided results that are consistent with experimental facts and allow us to extract meaningful information that is not amenable to experimentation.

Result and Discussion

Initially, the root-mean-square deviations (RMSDs) observed for each simulation were examined to determine if each system had attained equilibrium. The RMSDs of each DNA-Ni(II)•Gly-Gly-His and DNA-Ni(II)•L-Arg-Gly-His complex system, with respect to their starting structures, attained equilibrium after 250 ps, and their energies were found to be stable during the course of the remainder of each simulation. Hence, a time of 500 ps after thermal warm-up was selected as a starting point for data collection and for the analysis of each simulation.

Simulations of DNA + Ni(II)•Gly-Gly-His. Small RMSD values of each metallopeptide−DNA complex indicate stable structures. Figure 1 shows that the stabilities of the DNA-Ni(II)•Gly-Gly-His complexes initiated from orientations **I**, **II**, and **IV** were comparable during the simulations. In particular, with **I** and **IV**, the RMSDs of the DNA−metallopeptide complexes changed within the same range (2.0−4.0 Å). Worthy of note, the RMSDs observed for the complex initiated from **I** changed in a somewhat narrower range (2.5−3.5 Å), suggesting that the structure of this particular complex had less fluctuation during the course of the simulation.

Visual analysis of each trajectory obtained from these four simulations supports the general observations described above. When the simulation was initiated from orientation **I**, Ni(II)•Gly-Gly-His was found to maintain the key characteristics of its starting structure. More specifically, (1) Ni(II)•Gly-Gly-His remained located at the dyad of the A/T-rich minor groove region; (2) the Ni(II) ion remained in close proximity to the C4′−H's of the nearest thymine residues on both antiparallel strands of the oligonucleotide; and (3) the equatorial plane of the metallopeptide remained parallel to the walls of the minor groove. However, while positioned as such, the metallopeptide−DNA complex initiated in **I** displayed dynamic motions relative to the DNA. Of note, Ni(II)•Gly-Gly-His was observed to rotate clockwise and counterclockwise around the C_n axis of the central Ni(II) ion of the complex. This clockwise and counterclockwise rotation produced changes in the identity and number of functional groups employed in the metallopeptide-minor groove interaction; that is, the starting structure alternated from an initial structure (**I**) in which the amino-terminal N−H and the pyrrole N−H of the His imidazole were involved simultaneously in minor groove binding [which takes place as single H-bonding events formed by the amino-terminal N−H (39% of the simulation time) as well as simultaneous multiple H-bonding events (simultaneous double and simultaneous triple H-bonds, 11% of the simulation time)] to structures involving either the amino-terminal N−H alone (21% of time) or the carboxy-terminal amide alone (22% of time). These two latter structures are similar to the starting structures assembled for orientations **IV** and **II**, respectively. Thus, Ni(II)•Gly-Gly-His remains in the center of the A/T-rich region and rocks back and forth between orientations that resemble the starting orientations defined previously as **II** and **IV**.

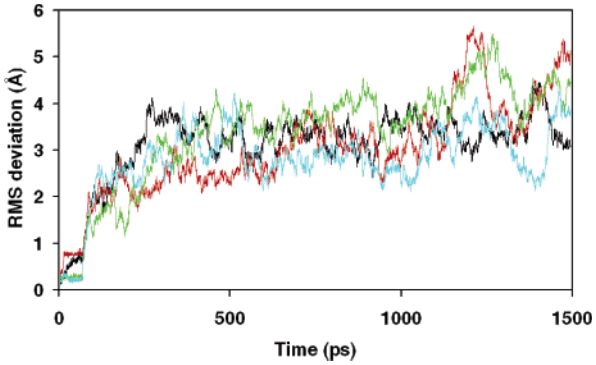


Figure 1. RMS deviations with respect to the starting structure in the simulation of Ni(II)•Gly-Gly-His bound to $d(CGCGAATTCGCG)_2$ in four orientations: **I**, black; **II**, red; **III**, green; and **IV**, blue.

Similar to the above, when the simulation of Ni(II)•Gly-Gly-His was initiated from **II**, the metallopeptide immediately turned from its starting position in the minor groove to an orientation that was analogous to **I**. Upon attainment of equilibrium after this reorientation, Ni(II)•Gly-Gly-His behaved dynamically as described above for the simulation initiated from **I**. Similarly, the simulation that was initiated from **IV** rotated toward orientation **I** and, upon equilibration, remained in this orientation and again behaved dynamically as described for orientation **I**. Contrary to these findings is the case for **III** where the ligand moved out of the minor groove, suggesting either a very weak attraction or a repulsion between host and guest in that orientation. In

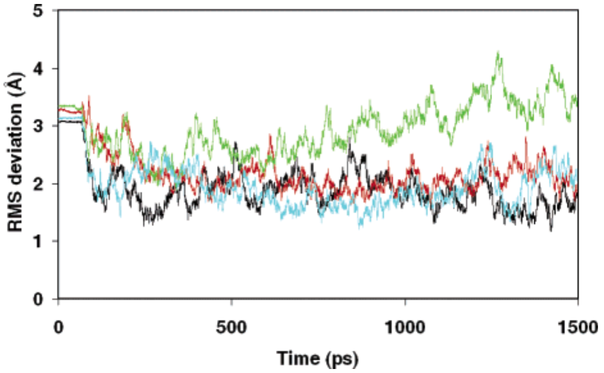


Figure 2. RMS deviations of Ni(II)•Gly-Gly-His bound to d(CGCGAATTCGCG)$_2$ in four directions (**I**, black; **II**, red; **III**, green; and **IV**, blue) with respect to the average structure generated in the simulation initiated from orientation **I**.

Table 2. Relative Binding Energies of DNA + Ni(II)•Gly-Gly-His[a]

orientation	$\Delta E_{\text{complex}}$[b]	ΔE_{DNA}[b]	ΔE_{ligand}[b]	$\Delta\Delta E$[c]
I	0	0	0	0
II	9	6.7	−3.9	6.2
III	88.5	−9.5	7.9	90.1
IV	−18.4	−26.3	7.7	0.2

[a] Energies are reported in kilocalories per mole. [b] $\Delta E_{\text{complex}}$, ΔE_{DNA}, and ΔE_{ligand} denote the molecular mechanical (MM) energies for the complex, DNA, and ligand, respectively, relative to orientation **I**. [c] Relative binding energies, calculated by $\Delta\Delta E = \Delta E_{\text{complex}} - \Delta E_{\text{DNA}} - \Delta E_{\text{ligand}}$.

support of these observations, Figure 2 illustrates the RMS deviations found for each of the above four DNA-Ni(II)•Gly-Gly-His simulations with respect to the average structure generated in the simulation resulting from **I**. This analysis shows that the simulations initiated from **I**, **II**, and **IV** all correspond quite closely to this particular final average structure, while aberrant behavior is noted from orientation **III**.

The above observations indicate that the most favored orientation of minor groove-bound Ni(II)•Gly-Gly-His initiates from an orientation as described for **I**. In this orientation, Ni(II)•Gly-Gly-His forms H-bonding interactions with the minor groove floor via the amino-terminal N−H protons and the pyrrole N−H of the His imidazole. However, with these initiating intermolecular contacts alone (in contrast to Ni(II)•L-Arg-Gly-His, as described later), the interaction between the DNA and the metallopeptide is not strong enough to hold this relatively small ligand in the minor groove firmly. As a consequence, the metallopeptide rotates as described. In addition, the relatively weak interactions that occur between Ni(II)•Gly-Gly-His and the minor groove also allow the metallopeptide to turn and take on this favored binding orientation even when it is docked initially via the adjacent edges of the equatorial plane (i.e., via **II** and **IV**).

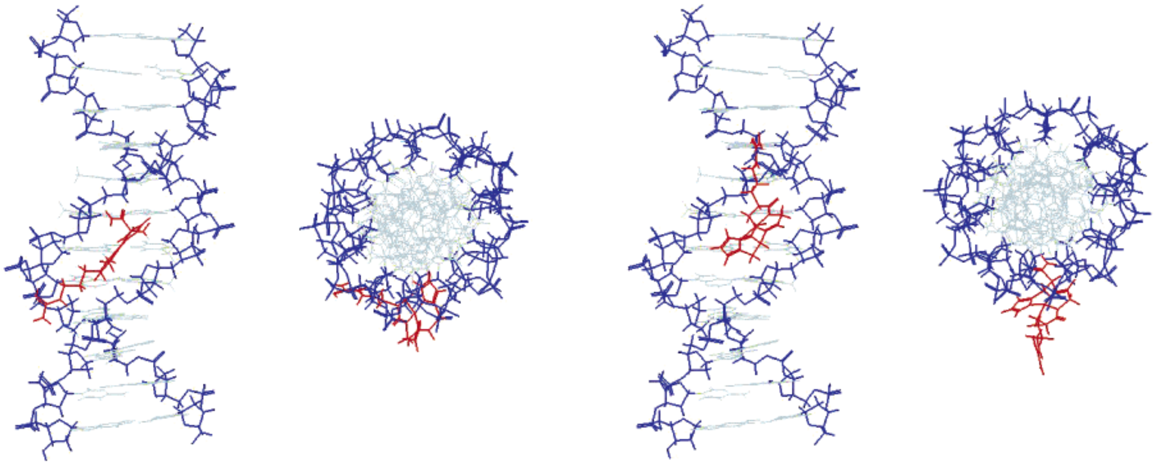
The relative binding energies from each of the four simulations of Ni(II)•Gly-Gly-His (Table 2) are in good agreement with the qualitative observations made above. The simulation initiated from **I** has the lowest relative binding energy, and the simulation initiated from **III** has the highest relative binding energy; the relative binding energies of **IV** and **II** are close to the values found for **I**, with **IV** being slightly lower than **II**. In stark contrast, the energies determined as a consequence of binding orientation **III** were found to be much higher than those resulting from the other orientations. Indeed, over the lifetime of the trajectory measured, orientation **III** leads to a destabilizing, repulsive interaction.

Simulations of DNA-Ni(II)•L-Arg-Gly-His. Illustrated in Figure 3 are the four starting structures in orientations **I**−**IV** of DNA + Ni(II)•L-Arg-Gly-His as viewed from the side and above (orthogonal and parallel to the helix axis). In contrast to the starting structures assembled for Ni(II)•Gly-Gly-His, it is clear that the Arg side chain of Ni(II)•L-Arg-Gly-His introduces an extension that could either complement or complicate minor groove binding by the main equatorial plane of the metallopeptide; however, this appendage remains quite flexible and, as observed and noted below, only forms transient contacts with the DNA that are secondary to the main metallopeptide−DNA recognition elements: the His imidazole N−H and the terminal peptide amine N−H. The representations in Figure 3 also emphasize the similarities and differences between these different binding orientations.

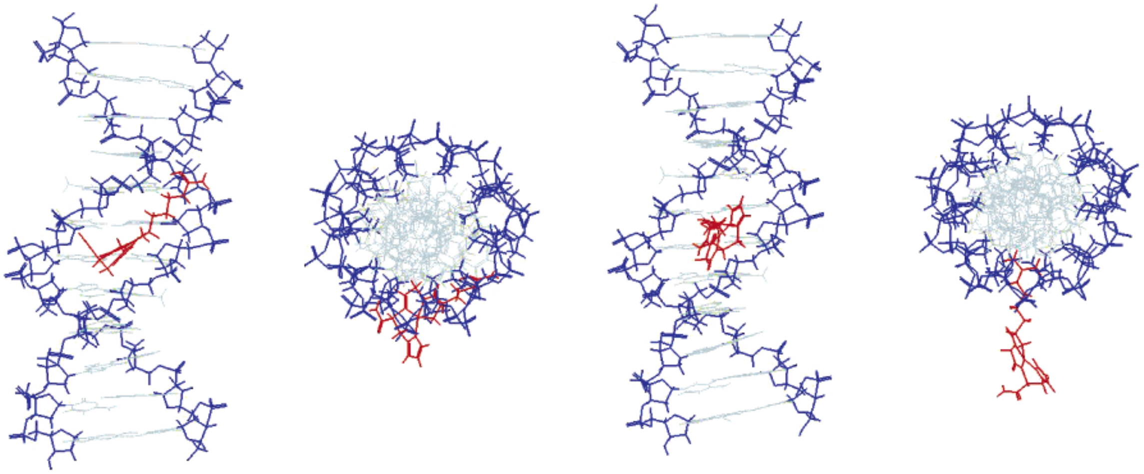
As in the analysis of Ni(II)•Gly-Gly-His, Figure 4 combines plots of the RMS deviations observed for simulations of DNA-Ni(II)•L-Arg-Gly-His oriented as starting structures in orientations **I**−**IV**. These results indicate that the DNA-Ni(II)•L-Arg-Gly-His system oriented initially as **I** exhibits the lowest RMSD, suggesting that this particular DNA−metallopeptide complex orientation is most stable throughout the simulation. In contrast, the complex oriented initially as **II** has the highest RMS value with a maximum deviation at ∼1000 ps indicating a relatively large structural change at this point in the simulation; the RMSD resulting from the simulations of metallopeptide binding initiated from **III** and **IV** were found to be intermediate between those observed for **I** and **II**.

Visual analyses of the trajectories from the four simulations of Ni(II)•L-Arg-Gly-His support the general observations described above and reveal dynamic details of their DNA-Ni(II)•L-Arg-Gly-His interactions. In the simulation initiated from **I**, Ni(II)•L-Arg-Gly-His was found to be stable in its starting orientation. The metallopeptide remained bound in the minor groove, and its motion paralleled the motions of the DNA backbone. At the same time, (1) the width of the minor groove narrowed at the dyad of the A/T-rich region and widened at the junction between the A/T and G/C base pairs, and (2) the positively charged Arg side chain was localized within the minor groove and interacted transiently with H-bond acceptors located on the floor of the groove. These observations are reminiscent of the results of similar studies performed with Hoechst 33258 and netropsin.[18]

In the simulation initiated from **II**, the coordination plane alone was inserted into the minor groove with the positively charged Arg side chain projecting away from the helix axis. Without a firm anchoring interaction with the minor groove, the relative long, positively charged side chain of Arg remained in motion, flexed at the surface of the DNA, and, at times, bent toward the minor groove. These translational movements account for the high RMSD values overall, and the bending motion contributed to the large RMSD observed



Orientation **I** Orientation **II**



Orientation **III** Orientation **IV**

Figure 3. Starting structures of Ni(II)•L-Arg-Gly-His (red) bound to the minor groove of d(CGCGAATTCGCG)$_2$ in orientations **I**−**IV** as viewed orthogonal and parallel to the DNA axis.

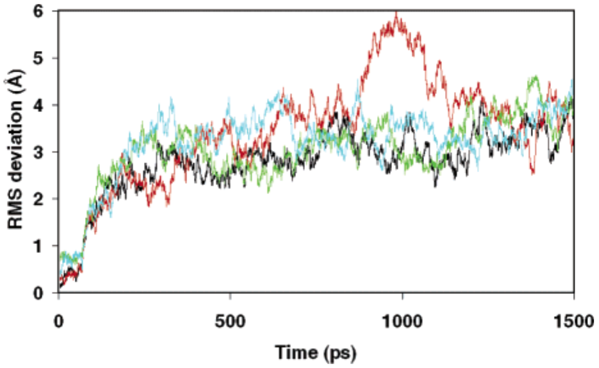


Figure 4. RMS deviations with respect to the starting structure in the simulations of Ni(II)•L-Arg-Gly-His bound to d(CGCGAATTCGCG)$_2$ in four starting orientations: **I**, black; **II**, red; **III**, green; and **IV**, blue.

at ~1000 ps. In comparison to this is the simulation resulting from **IV** in which the positively charged side chain of Ni(II)•L-Arg-Gly-His was inserted into the minor groove, while the coordination plane was projecting outwardly from the minor groove. The results of this simulation were similar to those observed for **II**: The equatorial plane of the metallopeptide flexed at the end of the Arg side chain 'tether' and also bent, at times, toward the minor groove. However, because of the heavier mass associated with the metal complex core compared to the side chain, the movements of the projecting equatorial plane were slower and smaller and consequently did not result in a high RMSD as was found for **II**.

Finally, the simulation initiated from **III** revealed that this metallopeptide did not change its orientation within the simulation time with both the equatorial plane and the Arg side chain remaining in the minor groove; however, unlike the stable orientation observed as a result of orientation **I**, the metallopeptide bound initially as **III** interacted only at the surface of the minor groove and exhibited a strong tendency to move out of the groove. Also, in this simulation the DNA backbone was not affected as it was with **I** indicating a much lower extent of intermolecular interaction. The binding behavior of Ni(II)•L-Arg-Gly-His in orientation **III** is accounted for by the presence of the positively charged Arg side chain that helped to maintain the metallopeptide in proximity to the minor groove, a finding that is in contrast to the effect of orientation **III** found for Ni(II)•Gly-Gly-His, a charge-neutral system.

In a parallel analysis aimed at investigating and comparing

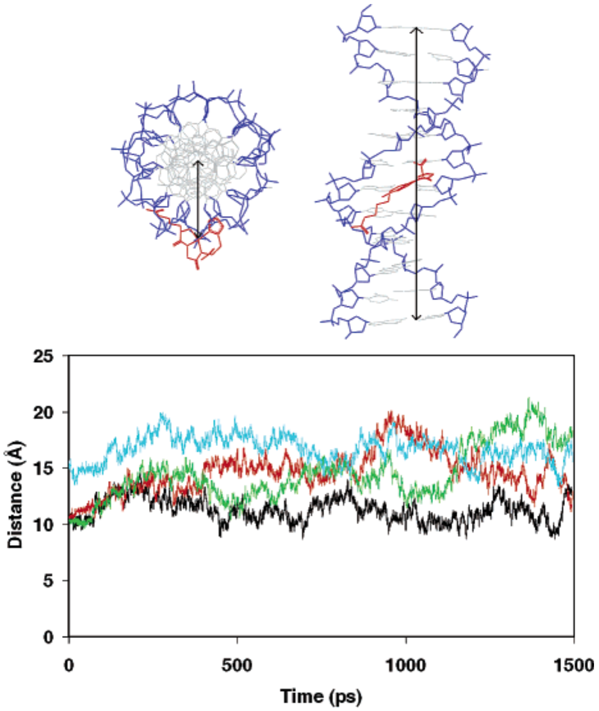


Figure 5. Definition of the distance between the Ni^{2+} of Ni(II)•L-Arg-Gly-His and the central axis of d(CGCGAATTCGCG)$_2$ (upper panel) and changes in the distance between Ni^{2+} and the central axis from the simulation of Ni(II)•L-Arg-Gly-His bound to d(CGCGAATTCGCG)$_2$ in four orientations (lower panel): **I**, black; **II**, red; **III**, green; and **IV**, blue.

the depths of metallopeptide-minor groove binding, the changes in the distance between the central Ni^{2+} atom of Ni(II)•L-Arg-Gly-His and the helix axis of DNA (as defined in Figure 5) were determined from each simulation. A plot of this distance versus time (Figure 5) shows that Ni(II)•L-Arg-Gly-His is inserted deepest into the minor groove via orientation **I** compared to all other orientations. In conjunction with this measurement, an analysis of the distances between the DNA A6 C2 proton and the His imidazole ring C4 proton, the NOE partners reported earlier,[13] indicates that only in orientation **I** is this distance (∼3−5 Å) compatible with the observed NOE; in orientations **II**, **III**, and **IV**, these distances are ∼8, 10, and 15 Å, respectively. Further, an examination of the distances between the Ni^{2+} center of the metallopeptide and proximal C4′−H protons, likely targeted during DNA cleavage, indicates, again, that in orientation **I** these distances are most compatible with observed C4′−H abstraction chemistry; the averaged distances measured for orientations **I**, **II**, **III**, and **IV** were 3.6, 4.2, 4.4, and 7.6 Å, respectively.

Taken together, these observations suggest that the favored binding orientation of DNA-Ni(II)•L-Arg-Gly-His, like Ni(II)•Gly-Gly-His, is **I**, an orientation in which the metallopeptide equatorial plane is associated with the minor groove via the amino-terminal N−H, the pyrrole N−H of the His imidazole, and the side chain of Arg. Oriented as such, and with an occasional third point of contact provided by the Arg side chain, the interaction between the metallopeptide and the DNA is stronger than with Ni(II)•Gly-Gly-His, and this orientation not only holds the metallopeptide deeply in the minor groove but also narrows the width of the DNA minor groove.

Table 3. Relative Binding Energies of DNA + Ni(II)•L-Arg-Gly-His[a]

orientation	$\Delta E_{complex}$[b]	ΔE_{DNA}[b]	ΔE_{ligand}[b]	$\Delta\Delta E$[c]
I	0	0	0	0
II	62.6	−11.7	20.5	53.8
III	133.4	−15.3	19.1	129.6
IV	53.2	−53.1	5.4	100.9

[a] Energies are reported in kilocalories per mole. [b] $\Delta E_{complex}$, ΔE_{DNA}, and ΔE_{ligand} denote the molecular mechanical (MM) energies for the complex, DNA, and ligand, respectively, relative to orientation **I**. [c] Relative binding energies, calculated by $\Delta\Delta E = \Delta E_{complex} - \Delta E_{DNA} - \Delta E_{ligand}$.

The outcome described above is consonant with the relative binding energies of Ni(II)•L-Arg-Gly-His shown in Table 3 where the simulation resulting from **I** is most stable. The binding energy of **III** is much higher than all others, indicating that the intermolecular interaction resulting from this initial orientation, in which H-bond *acceptors* are presented by the metallopeptide to the floor of the minor groove, a completely opposite scenario in comparison to **I**, is weaker. As shown in Table 3, the relative binding energies resulting from orientations **II** and **IV** show that they are significantly less stable than that obtained for **I**.

Further Analysis and Comparison of the Simulations of DNA + Ni(II)•Gly-Gly-His and DNA + Ni(II)•L-Arg-Gly-His Initiated from Orientation I. Further analyses of the simulations of DNA-Ni(II)•Gly-Gly-His and DNA-Ni(II)•L-Arg-Gly-His initiated from **I** were carried out in light of our initial findings. Illustrated in Figure 6 are the RMS deviations of the DNA-Ni(II)•Gly-Gly-His complex, the DNA alone, and the metallopeptide alone relative to the starting structure of Ni(II)•Gly-Gly-His bound to the DNA minor groove in orientation **I**. These results indicate that (1) the RMSDs of the DNA-Ni(II)•Gly-Gly-His complex and of the DNA alone have identical patterns of change, suggesting that Ni(II)•Gly-Gly-His did not move substantially along the minor groove and that the major structural changes attributable to the RMS deviations of the metallopeptide + DNA complex system derive from the flexibility of the DNA backbone and (2) the RMSDs of the DNA + metallopeptide

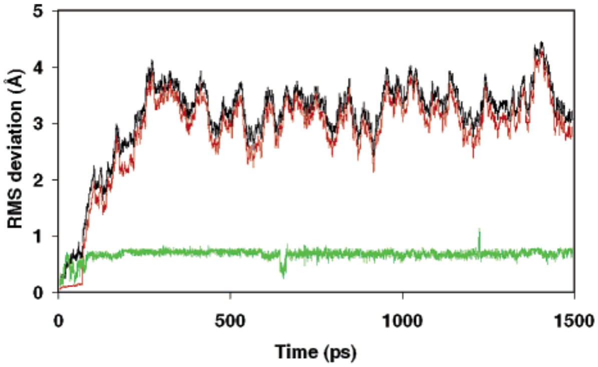


Figure 6. RMS deviations of DNA + Ni(II)•Gly-Gly-His (black), DNA alone (red), and metallopeptide alone (green) with respect to the starting structure in orientation **I** in the simulation of Ni(II)•Gly-Gly-His bound to d(CGCGAATTCGCG)$_2$.

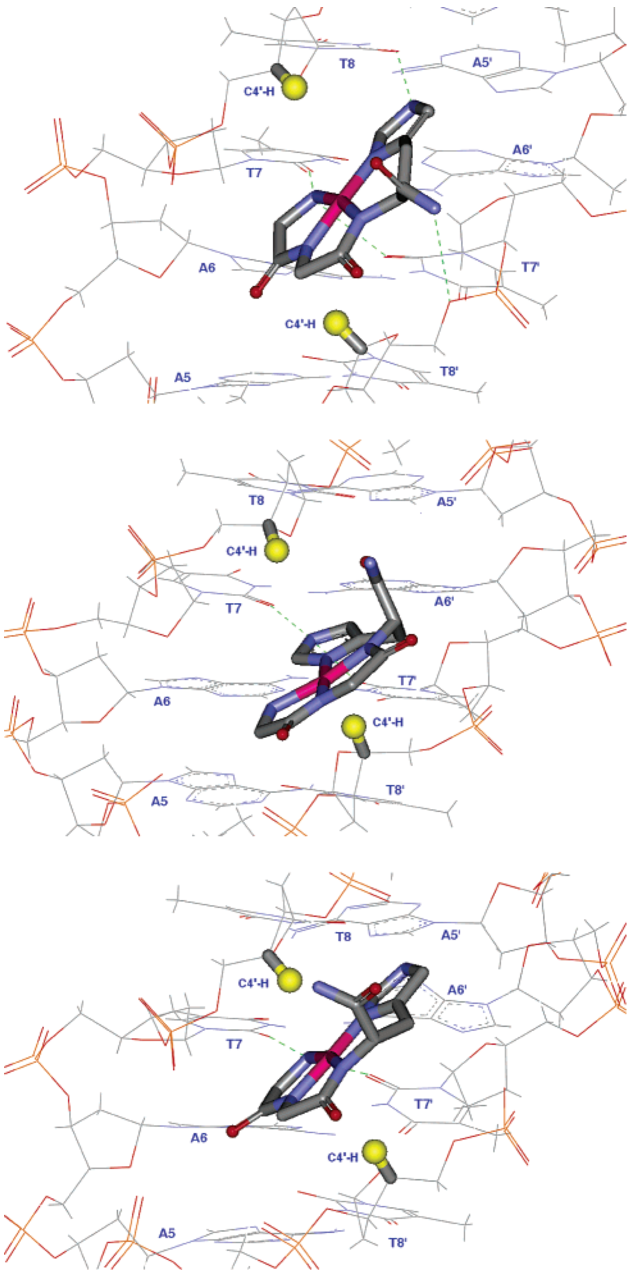


Figure 7. Snapshots of structures illustrating the H-bonds observed upon simulation of Ni(II)•Gly-Gly-His bound to d(CGCGAATTCGCG)$_2$ in orientation **I**. In these snapshots (1) the His imidazole and the amino-terminal amine are involved simultaneously in H-bonding to the O2 of T8 and O2 of T7 and T7′, respectively, while the carboxy-terminal amide N−H is occasionally associated with the O3′ of T7′ (upper panel); (2) the His imidazole pyrrole N−H interacts alone, as described above, with the floor of the minor groove (middle panel); and (3) the amino-terminal N−H interacts alone, as described above, with the floor of the minor groove (lower panel).

complex deviated consistently (~0.2 Å) from that of DNA alone, while the average RMS value of Ni(II)•Gly-Gly-His is small (~0.69 Å), indicating that the motion of the DNA-bound Ni(II)•Gly-Gly-His remained somewhat independent of the DNA groove, a result arising from a weakened intermolecular interaction. This observation is entirely consistent with the random and inefficient DNA cleavage data found in previous experimental studies.[5]

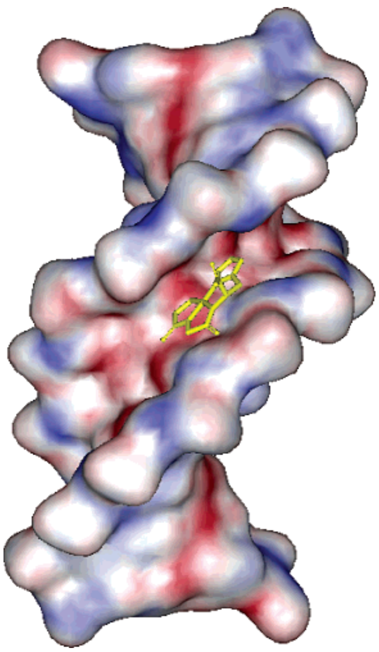


Figure 8. Space filling DNA model of the average structure of Ni(II)•Gly-Gly-His bound to d(CGCGAATTCGCG)$_2$ in orientation **I** emphasizing the binding orientation of the metallopeptide within the DNA minor groove and the resulting groove widths.

In light of the dynamic motions described earlier, an analysis of the intermolecular contacts that form between the DNA and the metallopeptide in orientation **I** indicated that Ni(II)•Gly-Gly-His formed H-bonds to acceptors on the floor of the DNA minor groove via the amino-terminal N−H protons to O2 of T7, O2 of T7′ and O4′ and O5′ of T8 (32% of the simulation time), the N−H of the His imidazole pyrrole to O2 and O2P of T8 (7% of the simulation time), and the carboxyl-terminal amide protons to O2P, O5′ of T8′, and O3′ of T7 or the phosphate moiety of T7 and T8′ (17% of the simulation time). Figure 7 illustrates the above through snapshots of structures that occur in the simulation initiated from **I** showing the H-bonds formed between the DNA minor groove and Ni(II)•Gly-Gly-His. In addition, as shown in Figure 8, we note that the DNA backbone is relatively stable during the course of the simulation without visible bending, but, as a consequence of metallopeptide binding, the minor groove width slightly widens to 11.4 Å in comparison to the DNA alone (10.4 Å).[18]

For comparison to Ni(II)•Gly-Gly-His, we illustrate in Figure 9 the RMSDs of the DNA-Ni(II)•L-Arg-Gly-His complex, the DNA alone, and the metallopeptide alone relative to the starting structure of Ni(II)•L-Arg-Gly-His bound to the DNA minor groove in orientation **I**. The deviations found for this system differ from those of DNA + Ni(II)•Gly-Gly-His discussed earlier. The RMSDs of Ni(II)•L-Arg-Gly-His and the DNA alone coincided precisely during the simulation without the consistent deviation observed in the case of Ni(II)•Gly-Gly-His. This observation suggests that Ni(II)•L-Arg-Gly-His fits snugly into the minor groove with very little independent motion. The metallopep-

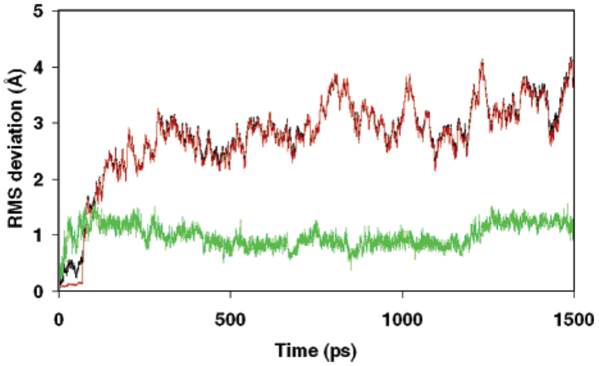


Figure 9. RMS deviations of Ni(II)•L-Arg-Gly-His + DNA (black), DNA alone (red), and Ni(II)•L-Arg-Gly-His alone (green) with respect to the starting structure in orientation **I** in the simulations of Ni(II)•L-Arg-Gly-His bound to d(CGCGAATTCGCG)$_2$.

tide moves and flexes with the motion of the DNA backbone during the course of the simulation and the structural changes of both the metallopeptide and the DNA backbone contributed to the RMS deviations of the whole complex. The DNA-Ni(II)•L-Arg-Gly-His complex has a lower averaged RMSD (∼3.11 Å), and Ni(II)•L-Arg-Gly-His alone has a higher RMSD (∼1.46 Å) compared to the corresponding values for Ni(II)•Gly-Gly-His (∼3.37 and 0.69 Å, respectively), indicating that the L-Arg isomer maintained a better overall complementary fit with the DNA during the course of the simulation, i.e., as the DNA flexed, the Ni(II)•L-Arg-Gly-His metallopeptide was capable of moving with it due to its stronger interactions with the minor groove.

Illustrating the above, Figure 10 shows a space filling DNA model of the average structure resulting from the simulation of Ni(II)•L-Arg-Gly-His bound initially to d(CGCGAATTCGCG)$_2$ in orientation **I**. As shown, Ni(II)•L-Arg-Gly-His fits in the minor groove at the dyad of the A/T-rich region with the equatorial plane of the metallopeptide residing parallel to the walls of the minor groove and with the central Ni^{2+} of the complex equidistant from two proximal cross-strand C4′−H protons, as noted earlier. While the equatorial plane of the complex is docked in the minor groove, the Arg side chain extends along the minor groove and interacts transiently with H-bond acceptors at the junction between the A/T-rich central core of the oligonucleotide and the flanking G/C-rich termini. As also noted earlier, the width of the minor groove at the center of the A/T-rich region is narrowed (from 10.4 to 9.7 Å)[18] upon interaction of the DNA and metallopeptide, while the width of the minor groove at the G/C−A/T junction was widened (from 10.4 to 16.8 Å) due to the transition of the phosphate diesters at that location (A$_6$ and G$_{10}'$) from B$_{\text{II}}$ to B$_{\text{I}}$.[15,19]

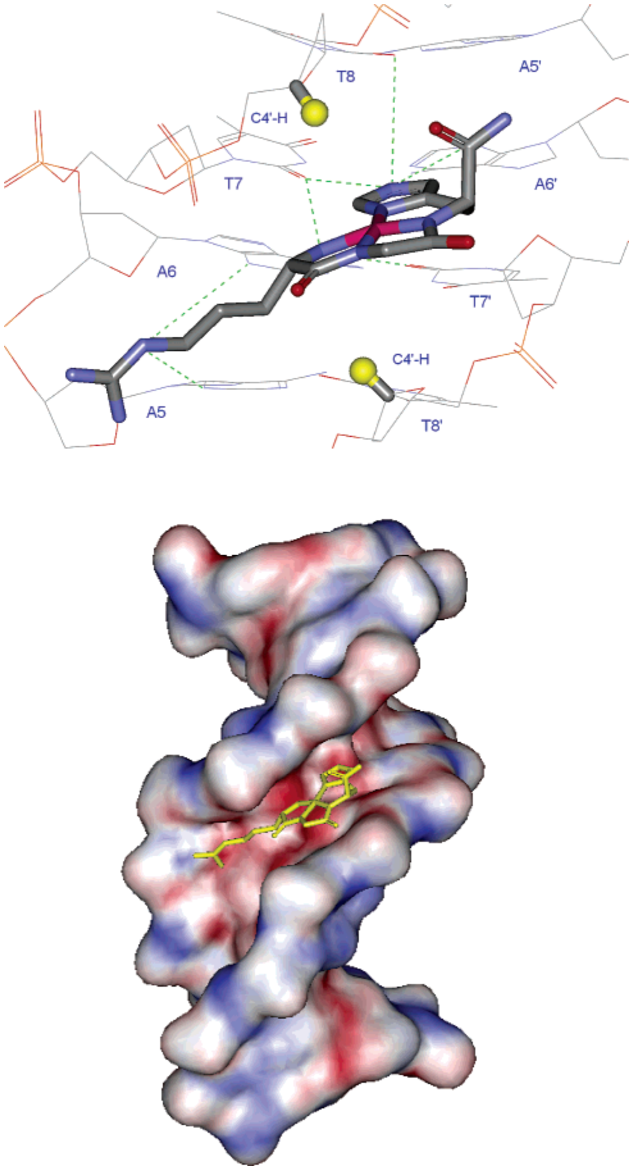


Figure 10. Average structures illustrating the H-bonds observed upon simulation of Ni(II)•L-Arg-Gly-His bound to d(CGCGAATTCGCG)$_2$ via orientation **I** (upper panel) and a space filling DNA model (lower panel) from the same simulation emphasizing the binding orientation of the metallopeptide within the DNA minor groove and the resulting groove width.

Along with the gross structure changes, an analysis of the H-bonds that form between Ni(II)•L-Arg-Gly-His and the A/T region of the minor groove (Figure 10) revealed differences when compared to those observed for Ni(II)•Gly-Gly-His + DNA. The key difference is that the longest-lived H-bonds formed between the DNA and Ni(II)•L-Arg-Gly-His were due to the imidazole pyrrole N−H and the amino-terminal peptide N−H protons to H-bond acceptors on the floor of the minor groove, which include the N3 of adenine and O2 of thymine on both DNA strands (up to 80% of the simulation time). The stability of these interactions is greater than the same, albeit transient, interactions that occur between the minor groove and Ni(II)•Gly-Gly-His. Due to the L-stereochemistry of the Arg side chain, the overall curvature of Ni(II)•L-Arg-Gly-His complemented well the curvature of the minor groove and allowed the amino-terminal N−H protons to explore and form H-bonds to the O2 of two thymine residues in the center of the AATT sequence. Along with the imidazole N−H and amino-terminal N−H interactions, which locked Ni(II)•L-Arg-Gly-His deeply in the AATT core region, the positively charged side chain guanidinium group extended beyond the AATT core region and interacted with the DNA at the junction between the G4•C9′ and A5•T8′ base pairs (via the N3 and phosphate moiety O2P, O3′, and O5′ of A5 and A6). The Arg side chain also

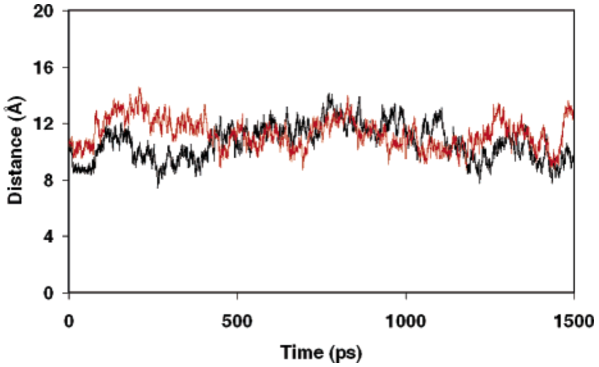


Figure 11. Changes in the distance between metallopeptide-bound Ni^{2+} and the central axis of DNA during the simulation of Ni(II)•Gly-Gly-His (black) and Ni(II)•L-Arg-Gly-His (red) bound to $d(CGCGAATTCGCG)_2$ in orientation **I**.

assisted in maintaining the core equatorial plane of the metallopeptide in orientation **I**. Indeed, the above two factors, stereochemistry and electrostatics, confer a distinct stabilizing advantage to the L-Arg-containing metallopeptide system in comparison to the Gly-only metallopeptide.

With the interactions described above, the distance (noted earlier) between the Ni(II)•L-Arg-Gly-His imidazole C4 proton and the minor groove A6-C2 proton ranges from 3 to 5 Å, consistent with the intermolecular NOE observed to occur between these two protons in the NMR study of DNA-Ni(II)•L-Arg-Gly-His.[13] In addition, the distances from the Ni(II) center to the T7 and T8 C5-methyl groups were 5.5 and 6.7 Å, respectively, again corresponding to the NMR data. Previous NMR studies also indicated a slight perturbation in the G•C base pairs flanking the AATT core of the oligonucleotide consistent with the point of Arg side chain contact observed in the present study.

Even though the binding of Ni(II)•Gly-Gly-His and Ni(II)•L-Arg-Gly-His in orientation **I** to the minor groove of DNA have apparent differences, they both insert into the minor groove to comparable depths (as shown in Figure 11). This insertion leads to a positioning of the central Ni^{2+} ion of each complex approximately equidistant between the T8 and T8′ C4′−H protons located on complementary DNA strands with the square-plane parallel to the walls of the minor groove. These observations account for the fact that despite the instability of the minor groove interaction displayed by Ni(II)•Gly-Gly-His vs Ni(II)•L-Arg-Gly-His, both metallopeptides share a common pathway leading to DNA cleavage via C4′−H abstraction; indeed, the C4′−H is prominently displayed in the minor groove[25] and readily accessible by either metallopeptide relative to other deoxyribose positions.

The above observations are consistent with, and assist in explaining, the ability of both Ni(II)•Gly-Gly-His and Ni(II)•L-Arg-Gly-His to interact with DNA, albeit with differing stabilities and cleavage efficiencies. These differences also rationalize the experimentally determined site-selectivites of these two distinct metallopeptides: Ni(II)•Gly-Gly-His results in relatively nonselective, inefficient DNA strand scission via C4′−H abstraction, while Ni(II)•L-Arg-Gly-His cleaves a subset of available nucleotides within a given A/T-rich site, indicating a more selective and stabilized association with the minor groove.[13]

Conclusions

Implications for Understanding DNA Recognition and Cleavage. Molecular dynamics simulations were carried out to examine all reasonably possible orientations of DNA minor groove binding by Ni(II)•Xaa-Gly-His metallopeptides. This investigation indicated that the most likely minor groove binding orientation for both metallopeptides is orientation **I**, through which both metallopeptides bind to the DNA minor groove deeper and with lower binding energies in comparison to other possibilities. Among the interactions found to occur between the minor groove of DNA and a metallopeptide, the N−H group of the His imidazole plays an important role by functioning as an H-bond donor to the N3 of adenine residues and the O2 of thymine residues located on the floor of the minor groove. The amino-terminal N−H also assists in metallopeptide-minor groove recognition and association in conjunction with the stereochemistry and chemical nature of the amino acid side chain substituted at this position.[14,15] In the presence of a positively charged L-Arg side chain, DNA recognition and affinity is enhanced through an increased electrostatic interaction, favorable stereochemical interactions, and the ability to mediate a third point of contact with the floor of the minor groove; this stabilizing factor is absent in Ni(II)•Gly-Gly-His.

As a consequence of binding orientation (**I**), our analyses indicate that the reactive metal center of each Ni(II)•Xaa-Gly-His metallopeptide is held in close proximity to known target C4′−H protons of the DNA. Abstraction of these protons via 1 or 2 e^- oxidation[8] leads either to a radical or cationic C4′ nucleotide lesion, respectively, and ultimately DNA strand scission. The observations made in the current investigation also correlate well to the efficiencies of DNA strand scission observed experimentally: Ni(II)•Arg-Gly-His is more efficient at C4′−H abstraction than Ni(II)•Gly-Gly-His. Noteworthy also is the fact that Ni(II)•Arg-Gly-His [or Ni(II)•Lys-Gly-His] generates fewer single electron abstraction products relative to Ni(II)•Gly-Gly-His;[8] single electron abstraction products require rapid recombination with O_2, which is likely blocked by the more stable DNA-Ni(II)•Arg-Gly-His interaction found in orientation **I**. Overall, these findings reinforce the importance of complementary H-bonding, multiple points of drug−DNA contact, and stable positioning of a catalytic center relative to a target atom in the design of reactive minor groove binding agents. Low molecular weight examples illustrating these factors are important toward the development of improved DNA targeted agents and our understanding of DNA recognition phenomena.

Acknowledgment. We thank the National Institutes of Health for financial support of this work (GM 62831 to E.C.L.). In addition, one of us (Y.-Y. F.) would like to acknowledge funding from the National Institutes of Health (G12 RR003048 from the RCMI program, Division of

Research Infrastructure, National Center for Research Resources to Howard University).

References

(1) Sigman, D. S.; Mazumder, A.; Perrin, D. M. Chemical nucleases. *Chem. Rev*. **1993**, *93*, 2295−2316.

(2) Pyle, A. M.; Barton, J. K. Probing nucleic acids with transition metal complexes. *Prog. Inorg. Chem*. **1990**, *38*, 413−475.

(3) Long, E. C. The DNA helical biopolymer: A template for the binding, assembly, and reactivity of metal ions and complexes. *J. Inorg. Organomet. Polym*. **1993**, *3*, 3−39.

(4) Neidle, S. DNA minor-groove recognition by small molecules. *Nat. Prod. Rep*. **2001**, *18*, 291−309.

(5) Long, E. C.; Claussen, C. A. DNA and RNA recognition and modification by Gly-Gly-His-derived metallopeptides. In *DNA and RNA Binders: From Small Molecules to Drugs*; Demeunynk, M., Bailly, C., Wilson, W. D., Eds.; Wiley-VCH: New York, 2003; pp 88−125.

(6) Long, E. C. Ni(II)•Xaa-Xaa-His metallopeptide DNA/RNA interactions. *Acc. Chem. Res*. **1999**, *99*, 827−836.

(7) Claussen, C. A.; Long, E. C. Nucleic acid recognition by metal complexes of bleomycin. *Chem. Rev*. **1999**, *99*, 2797−2816.

(8) Liang, Q.; Ananias, D. C.; Long, E. C. Ni(II)•Xaa-Xaa-His induced DNA cleavage: Deoxyribose modification by a common “activated” intermediate derived from $KHSO_5$, MMPP, or H_2O_2. *J. Am. Chem. Soc*. **1998**, *120*, 248−257.

(9) Jin, Y.; Cowan, J. A. DNA cleavage by copper-ATCUN complexes. Factors influencing cleavage mechanism and linearization of dsDNA. *J. Am. Chem. Soc*. **2005**, *127*, 8408−8415.

(10) Kimoto, E.; Tanaka, H.; Gyotoku, J.; Morishige, F.; Pauling, L. Enhancement of antitumor activity of ascorbate against Ehrlich ascites tumor cells by the copper:glycylglycylhistidine complex. *Cancer Res*. **1983**, *43*, 824−828.

(11) Licini, G.; Scrimin, P. Metal-ion binding peptides: From catalysis to protein tagging. *Angew. Chem., Int. Ed*. **2003**, *42*, 4572−4575.

(12) Nagane, R.; Koshigoe, T.; Chikira, M.; Long, E. C. The DNA-bound orientation of Cu(II)•Xaa-Gly-His metallopeptides. *J. Inorg. Biochem*. **2001**, *83*, 17−23.

(13) Fang, Y.-Y.; Ray, B. D.; Claussen, C. A.; Lipkowitz, K. B.; Long, E. C. Ni(II)•Arg-Gly-His-DNA interactions: Investigation into the basis for minor-groove binding and recognition. *J. Am. Chem. Soc*. **2004**, *126*, 5403−5412.

(14) Liang, Q.; Eason, P. D.; Long, E. C. Metallopeptide-DNA interactions: Site-selectivity based on amino acid composition and chirality. *J. Am. Chem. Soc*. **1995**, *117*, 9625−9631.

(15) Fang, Y.-Y.; Claussen, C. A.; Lipkowitz, K. B.; Long, E. C. Diastereoselective DNA cleavage recognition by Ni(II)•Gly-Gly-His derived metallopeptides. *J. Am. Chem. Soc*. **2006**, *128*, 3198−3207.

(16) Harford, K.; Sarkar, B. Amino terminal Cu(II)- and Ni(II)-binding (ATCUN) motif of proteins and peptides: Metal binding, DNA cleavage, and other properties. *Acc. Chem. Res*. **1997**, *30*, 123−130.

(17) Wellenzohn, B.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. Simulation of *Eco*RI dodecamer netropsin complex confirms class I complexation mode. *J. Am. Chem. Soc*. **2000**, *122*, 3927−3931.

(18) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. Significance of ligand tails for interaction with the minor groove of B-DNA. *Biophys. J*. **2001**, *81*, 1588−1599.

(19) Flader, W.; Wellenzohn, B.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. $B_I - B_{II}$ substrate transitions induce changes in the hydration of B-DNA, potentially mediating signal transduction from the minor to major groove. *J. Phys. Chem. B* **2001**, *105*, 10379−10387.

(20) Mohamadi, F.; Richardson, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel − an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem*. **1990**, *11*, 440−467.

(21) *Spartan02*; Wavefunction, Inc.: Irvine, CA, 2002.

(22) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K. and Kollman, P. A. *AMBER 7*; University of California: San Francisco, CA, 2002.

(23) Sriram, M.; van der Marel, G. A.; Roelen, H. L.; van Boom, J. H.; Wang, A. H. J. Structural consequences of a carcinogenic alkylation lesion on DNA: Effect of O6-ethylguanine [e6G] on the molecular structure of d(CGC[e6G]AATTCGCG)-netropsin complex. *Biochemistry* **1992**, *31*, 11823−11834.

(24) Bal, W.; Djuran, M. I.; Margerum, D. W.; Gray, E. T.; Mazid, M. A.; Tom, R. T.; Nieboer, E.; Sadler, P. J. Dioxygen-induced decarboxylation and hydroxylation of [Ni^{II}(glycyl-glycyl-L-histidine)] occurs via Ni^{III}: X-ray crystal structure of [Ni^{II}(glycyl-glycyl-hydroxy-D,L-histamine)]•$3H_2O$. *J. Chem. Soc., Chem. Commun*. **1994**, 1889−1890.

(25) Knapp, W.; Tullius, T. D. Oxidative strand scission of nucleic acids: Routes initiated by hydrogen abstraction from the sugar moiety. *Chem. Rev*. **1998**, *98*, 1089−1107.

CT600074R

JCTC Journal of Chemical Theory and Computation

Effects of Zero-Point and Thermal Vibrational Averaging on Computed NMR Properties of a Model Compound for Purine Nucleosides

Jeffrey N. Woodford

Department of Chemistry and Biochemistry, Eastern Oregon University,
One University Boulevard, La Grande, Oregon 97850-2899

Gerard S. Harbison*

Department of Chemistry, University of Nebraska−*Lincoln,*
Lincoln, Nebraska 68588-0304

Received February 21, 2006

Abstract: A method for the incorporation of thermal averaging into the calculation of nuclear magnetic resonance properties is given. These properties are computed using density functional theory, anharmonic first-order perturbation corrections to a normal-mode analysis, and standard statistical mechanical averaging. The method is applied to the calculation of chemical shieldings and spin−spin coupling constants (*J* couplings) of 1′-imidazolyl-2′-deoxy-β-ribofuranose (IDR), a model compound for purine nucleosides, at the B3LYP/6-311++G(2d,p) level of theory. Thermal averaging causes substantial changes in the values of computed parameters. The calculated harmonic normal modes of IDR are also investigated; we find reasonable agreement with published results from vibrational spectroscopy on DNA fragments. Finally, the calculated magnetic and structural data regarding the reported hydrogen bond between H8 on the imidazole ring and O5′ on the sugar ring are investigated; we find that such data do not strongly support the formation of a hydrogen bond between these two atoms.

Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a primary tool for investigating the structure and dynamics of important biochemical systems.[1] One instance is the study of deoxyribonucleosides with NMR, to obtain data about the conformation and dynamics of the sugar moiety. Many methods have been employed to correlate NMR data with conformations of the deoxyribofuranose ring, including methods based on the Karplus equation,[2,3] the CUPID method,[4,5] and the more recent DORCO method.[6] In all of the above methods, to extract the desired conformational data, it is necessary to have knowledge of the molecule's potential energy surface (PES). With the advent of modern computer equipment and software, it is now common practice to rely on high-quality ab initio or density functional (DFT) calculations to obtain such data. This procedure generally consists of generating a molecular structure of the system of interest, optimizing its structure using a high level of theory and a large basis set, and then adjusting the desired torsion angles to obtain a high-quality PES.

However, the molecular geometry obtained in this way is that of the isolated, vibration-free molecule at 0 K in the gas phase. For typical solid- or liquid-phase samples at room temperature, many vibrational levels are populated, especially low-frequency vibrations, and the measured NMR quantities are actually thermal averages over the zero-point motion and over a Boltzmann distribution of thermally accessible vibrational states. Thus, for a proper comparison between theory and the experiment, a correction accounting for this vibrational averaging must be applied to the theoretical results. This phenomenon is of course intimately related to

---

* Corresponding author phone: (402) 472-9346; fax: (402) 472-9402; e-mail: gerry@setanta.unl.edu.

10.1021/ct6000736 CCC: $33.50

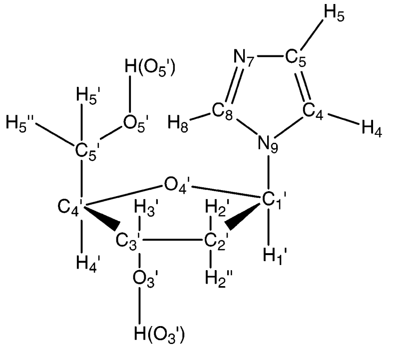


Figure 1. Molecular structure of 1′-imidazolyl-2′-deoxyribofuranose (IDR) and numbering system used.

the well-known primary isotope effect on NMR chemical shifts and *J* couplings, which arises from differences in quantum mechanical vibrational amplitudes and thermal populations resulting from changes in the mass of the nucleus, for which there is vast experimental (for a recent review, see ref 7) and computational[8−11] literature.

It has been known for some time that ensemble averaging over a distribution of microstates should be properly employed in order to achieve suitable agreement with the experiment.[12−14] Most of the previous methods used for averaging have relied upon a combination of molecular dynamics simulations and ab initio or semiempirical calculations. In this work, we present an alternative method for accomplishing the thermal averaging process. In this method, the equilibrium geometry of the molecule of interest, calculated by ab initio means, is displaced along each vibrational coordinate, and the relevant NMR properties are calculated at each displaced geometry. Then, a combination of time-independent perturbation theory and statistical mechanical averaging is applied to the equilibrium and displaced computational results. We present a detailed theoretical treatment of this method as it is applied to the calculation of thermally averaged chemical shieldings and spin−spin coupling constant (SSCC) values, along with the results of this method as they are applied to 1′-imidazolyl-2′-deoxy-β-ribofuranose (IDR; Figure 1), an analogue of purine DNA bases. It has been shown that IDR is the minimum structural unit required to correctly model the furanose moiety of purine nucleosides, because of extensive coupling between the base and the sugar ring.[15] Furthermore, IDR does not have charged atoms, such as those in a phosphate group, to avoid complications arising from the selection and proper placement of the counterion(s).[16]

Many previous authors have conducted theoretical investigations on DNA base model systems. Typically, however, only a small part of a nucleoside is modeled, appropriate for the particular study. For instance, Lluch et al.[17−19] have theoretically examined dimers of the model compound 7-azaindole as a model system for excited-state proton transfer between DNA base pairs. The formamidine−formamide complex has been used to mimic the A−T hydrogen-bonding interaction.[20] Dimethyl phosphate is often used as a model compound to represent the DNA/RNA backbone.[21,22] To model entire nucleosides, either a fragmentation scheme is employed[23] or a relatively low level of theory, either HF or semiempirical methods, must be used.[24] Similarly, many previous workers have employed vibrational averaging techniques on ab initio derived magnetic properties. For instance, Del Bene et al.[25,26] have theoretically examined SSCCs for hydrogen-bonded complexes, with the inclusion of full vibrational averaging; Bühl and Grigoleit[27] have examined vibrational corrections to chemical shifts of transition metal complexes; and Barone et al.[28] used a hybrid ab initio/semiempirical method to incorporate vibrational averaging into two different conformers of β-alanine. These studies, however, have typically involved either a small number of atoms or selectively including only certain vibrations in the averaging process. (A notable exception is the work by Case.[29]) The present work, on the other hand, represents the result of advanced ab initio computations on a model compound representing an entire nucleoside unit, including full vibrational averaging.

Theory

To obtain the thermally averaged chemical shielding $\bar{\sigma}(T)$, we proceed according to the following strategy. First, the molecule of interest is fully optimized with a large basis set, and a harmonic frequency analysis is performed at the equilibrium geometry. Then, the molecule is displaced along each normal mode between $-\alpha$ and $+\alpha$, where α is the characteristic length of the vibration ($\alpha = \sqrt{\hbar/\mu\omega_0}$ with μ being the reduced mass and ω_0 the harmonic frequency of the vibration), and the potential energy is fitted to a fourth-order polynomial.

$$E_j(y) = \beta_j y^2 + \gamma_j y^3 + \delta_j y^4 \tag{1}$$

Here $y = x/\alpha$ and the index j runs from 1 to $M = 3N - 6$, the number of normal modes. The harmonic portion of the vibration is contained in the β_j term, while the anharmonicity is contained in the γ_j and δ_j terms.

Next, for each mode, the chemical shielding for each atom is calculated at each displaced geometry and fitted to a second-order polynomial. For the *A*th atom displaced along the *j*th vibrational mode, the chemical shielding is given by the following expression.

$$\sigma_j^{(A)}(y) = \sigma_0^{(A)} + \sigma_{1j}^{(A)} y + \sigma_{2j}^{(A)} y^2 \tag{2}$$

The quantity $\sigma_0^{(A)}$ represents the electronic contribution to the chemical shielding and is independent of molecular vibration; $\sigma_1^{(A)}$ and $\sigma_2^{(A)}$ represent the linear and quadratic corrections to the chemical shielding, respectively, due to vibration. In principle, additional correction terms (e.g., cubic) may be required in order to correctly model the data; however, in the present study, we found that truncation at the quadratic term yielded excellent results. Of course, in higher-symmetry molecules, only the fully symmetric modes would possess linear coefficients; the present molecule is unfortunately C_1.

Following the determination of these coefficients, first-order perturbation theory is applied to determine the expectation values of E_j, $\sigma_1^{(A)}y$, and $\sigma_2^{(A)}y^2$ in the *n*th vibrational state. When $\hat{H}'_j = \gamma_j\hat{y}^3 + \delta_j\hat{y}^4$ is used as the perturbation

Hamiltonian and the harmonic oscillator wave functions are used as a basis, the following results are obtained.

$$E_{jn} = E_{jn}^{(0)} + E_{jn}^{(1)} = \left(n + \frac{1}{2}\right)\hbar\omega_{0j} + \left(n^2 + n + \frac{1}{2}\right)\delta_j \quad (3)$$

$$\langle\sigma_{1j}^{(A)}\rangle_n = \frac{\langle\psi_n^{(0)} + \psi_n^{(1)}|\sigma_{1j}^{(A)}\hat{y}|\psi_n^{(0)} + \psi_n^{(1)}\rangle}{\langle\psi_n^{(0)} + \psi_n^{(1)}|\psi_n^{(0)} + \psi_n^{(1)}\rangle} = \frac{\sigma_{1j}^{(A)}\left(\sum_{m\neq n} c_{j,nm}^{(1)}\langle\psi_n^{(0)}|\hat{y}|\psi_m^{(0)}\rangle + \sum_{m\neq n} c_{j,nm}^{(1)}\langle\psi_m^{(0)}|\hat{y}|\psi_n^{(0)}\rangle + \sum_{m\neq n}\sum_{b\neq n} c_{j,nm}^{(1)}c_{j,nb}^{(1)}\langle\psi_m^{(0)}|\hat{y}|\psi_b^{(0)}\rangle\right)}{1 + \sum_{m\neq n}|c_{j,nm}^{(1)}|^2} \quad (4)$$

$$\langle\sigma_{2j}^{(A)}\rangle_n = \frac{\langle\psi_n^{(0)} + \psi_n^{(1)}|\sigma_{2j}^{(A)}\hat{y}^2|\psi_n^{(0)} + \psi_n^{(1)}\rangle}{\langle\psi_n^{(0)} + \psi_n^{(1)}|\psi_n^{(0)} + \psi_n^{(1)}\rangle} = \frac{\sigma_{2j}^{(A)}\left[\left(n + \frac{1}{2}\right) + \sum_{m\neq n} c_{j,nm}^{(1)}\langle\psi_n^{(0)}|\hat{y}^2|\psi_m^{(0)}\rangle + \sum_{m\neq n} c_{j,nm}^{(1)}\langle\psi_m^{(0)}|\hat{y}^2|\psi_n^{(0)}\rangle + \sum_{m\neq n}\sum_{b\neq n} c_{j,nm}^{(1)}c_{j,nb}^{(1)}\langle\psi_m^{(0)}|\hat{y}^2|\psi_b^{(0)}\rangle\right]}{1 + \sum_{m\neq n}|c_{j,nm}^{(1)}|^2} \quad (5)$$

In the above equations, $c_{j,nm}^{(1)}$ is the usual set of expansion coefficients of the first-order wave function in terms of the basis functions for the *j*th mode.

$$c_{j,nm}^{(1)} = \frac{\langle\psi_m^{(0)}|\hat{H}_j'|\psi_n^{(0)}\rangle}{E_{jn}^{(0)} - E_{jm}^{(0)}} = \frac{\gamma_j\langle\psi_m^{(0)}|\hat{y}^3|\psi_n^{(0)}\rangle}{\hbar\omega_{0j}(n-m)} + \frac{\delta_j\langle\psi_m^{(0)}|\hat{y}^4|\psi_n^{(0)}\rangle}{\hbar\omega_{0j}(n-m)} \quad (m\neq n) \quad (6)$$

While the sums in eqs 4 and 5 are infinite, they can be solved in closed form; expressions for these sums are given in the Appendix.

The correction to the chemical shielding due to zero-point vibration (ZPV) is then calculated by summing the contributions of the ground vibrational state for each mode.

$$\bar{\sigma}_{\mathrm{ZPV}}^{(A)} = \sum_{j=1}^{M}(\langle\sigma_{1j}^{(A)}\rangle_0 + \langle\sigma_{2j}^{(A)}\rangle_0) \quad (7)$$

The correction to the chemical shielding due to the thermal population of excited vibrational states is then calculated by taking the statistical mechanical average over the thermally accessible vibrational levels. Assuming no coupling between modes, this correction is given by

$$\bar{\sigma}_{\mathrm{thermal}}^{(A)}(T) = \sum_{j=1}^{M}\frac{1}{q_j}\sum_{n=0}^{\infty}[(\langle\sigma_{1j}^{(A)}\rangle_n - \langle\sigma_{1j}^{(A)}\rangle_0) + (\langle\sigma_{2j}^{(A)}\rangle - \langle\sigma_{2j}^{(A)}\rangle_0)]\exp(-\Delta E_{jn}/kT) \quad (8)$$

where *k* is Boltzmann's constant, *T* is the absolute temperature, ΔE_{jn} is the difference in energy for the *j*th mode from the ground vibrational state to the *n*th state, and q_j is the thermal part of the vibrational partition function for the *j*th mode.

$$q_j = \sum_{n=0}^{\infty}\exp(-\Delta E_{jn}/kT) \quad (9)$$

Finally, the overall thermal average of the chemical shielding for the *A*th atom is taken by summing the electronic contribution, the zero-point vibrational contribution, and the thermal excitation contribution.

$$\bar{\sigma}^{(A)}(T) = \sigma_0^{(A)} + \bar{\sigma}_{\mathrm{ZPV}}^{(A)} + \bar{\sigma}_{\mathrm{thermal}}^{(A)}(T) \quad (10)$$

For purposes of comparison, these values were also calculated using the unperturbed, purely harmonic Hamiltonian. These expressions are

$$\bar{\sigma}_{\mathrm{ZPV,harmonic}}^{(A)} = \frac{1}{2}\sum_{j=1}^{M}\sigma_{2j}^{(A)} \quad (11)$$

$$\bar{\sigma}_{\mathrm{thermal,harmonic}}^{(A)} = \sum_{j=1}^{M}\left[\sigma_{2j}^{(A)}\frac{\exp(-\hbar\omega_{0j}/kT)}{1 - \exp(-\hbar\omega_{0j}/kT)}\right] \quad (12)$$

It is $\bar{\sigma}^{(A)}(T)$ which is then compared to the experiment. To compute a thermally averaged scalar coupling value $\bar{J}^{(a)}(T)$ for the *a*th coupling constant, the same procedure was followed. The corresponding equations are given as eqs 13−15.

$$\bar{J}_{\mathrm{ZPV}}^{(a)} = \sum_{j=1}^{M}(\langle J_{1j}^{(a)}\rangle_0 + \langle J_{2j}^{(a)}\rangle_0] \quad (13)$$

$$\bar{J}_{\mathrm{thermal}}^{(a)}(T) = \sum_{j=1}^{M}\frac{1}{q_j}\sum_{n=0}^{\infty}[(\langle J_{1j}^{(a)}\rangle_n - \langle J_{1j}^{(a)}\rangle_0) + (\langle J_{2j}^{(a)}\rangle - \langle J_{2j}^{(a)}\rangle_0)]\exp(-\Delta E_{jn}/kT) \quad (14)$$

$$\bar{J}^{(a)}(T) = J_0^{(a)} + \bar{J}_{\mathrm{ZPV}}^{(a)} + \bar{J}_{\mathrm{thermal}}^{(a)}(T) \quad (15)$$

Computational Details

All calculations were performed on a 16-node Beowulf cluster running Linux. Geometry optimizations and vibrational frequency calculations were performed with GAMESS.[30] The structure of IDR was optimized using delocalized internal coordinates[31] and DFT with the B3LYP functional[32] and the 6-311++G(2d,p) basis set. Optimization was continued until each member of the Cartesian gradient vector was less than or equal to 1.0×10^{-6} Hartree/bohr. Chemical shift and scalar coupling values were calculated using Gaussian 03[33] also at the B3LYP/6-311++G(2d,p) level of theory and the "tight" criterion for the self-consistent field density.

The chemical shielding values calculated at the equilibrium geometry were used to obtain σ_0 of eq 1. To obtain γ_j, δ_j, $\sigma_1^{(A)}$, $\sigma_2^{(A)}$, $J_1^{(a)}$, and $J_2^{(a)}$, first, four displacements between $-\alpha$ and $+\alpha$ were calculated along each normal coordinate and the energy, chemical shielding, and SSCC values were

Table 1. Selected Calculated Bond Lengths and Valence Angles for IDR[a]

property	value	property	value
C1′−O4′	1.428	C5′−C4′−C3′	115.1
C4′−O4′	1.432	O4′−C1′−N9	109.7
C1′−N9	1.447	C2′−C1′−N9	114.9
C1′−C2′	1.533	O5′−C5′−C4′	109.9
C2′−C3′	1.520	C1′−N9−C4	125.9
C3′−C4′	1.544	τ_0 (C4′−O4′−C1′−C2′)	−15.4
C3′−O3′	1.431	τ_1 (O4′−C1′−C2′−C3′)	29.8
C4′−C5′	1.520	τ_2 (C1′−C2′−C3′−C4′)	−31.9
C5′−O5′	1.427	τ_3 (C2′−C3′−C4′−O4′)	23.7
C1′−C2′−C3′	103.1	τ_4 (C3′−C4′−O1′−C2′)	−5.3
C2′−C3′−C4′	102.8	β	162.5
C3′−C4′−O4′	106.5	γ	51.5
C4′−O4′−C1′	111.1	δ	146.0
O4′−C1′−C2′	105.8	ϵ	−66.1
C2′−C3′−O3′	107.6	χ	243.9
C4′−C3′−O3′	111.8		

[a] Bond lengths in angstroms; valence angles and torsion angles in degrees.

calculated at the displaced coordinates. For the 69 normal modes of IDR, this procedure resulted in 276 separate chemical shielding/SSCC calculations. In addition, for a few modes, more than four displacements were required to obtain reliable fitted values of the energy; these additional displacements were evaluated between -0.5α and $+0.5\alpha$. Then, the calculated energy values for each mode were fitted to a fourth-order polynomial with the linear term set equal to zero, and the calculated chemical shielding and SSCC values for each mode were fitted to quadratic polynomials. In all cases, the fitting procedures led to correlation coefficients (R^2) greater than or equal to 0.99, thereby justifying the expansion of the chemical shielding/SSCC data only through second order in vibrational displacement. Equations 7−10 were then applied to find the thermally averaged chemical shielding values. An analogous procedure was used for the SSCC data.

In eqs 8 and 9, because the perturbed energy difference ΔE_{jn} contains terms proportional to n^2, the infinite sums over n are only conditionally convergent. Thus, the sum does not have a closed-form solution and must be truncated. We found that truncation at $n \sim 50$ yielded values that were sufficiently converged. In all cases, the calculations were repeated for a maximum value of n of 100; we found the error due to truncation was negligible.

All of the post-electronic structure calculations were performed with an in-house program written in C. Data files and the computer program code may be found in the Supporting Information.

Results and Discussion

Figure 1 shows the B3LYP/6-311++G(2d,p)-optimized geometry of IDR, along with the numbering scheme employed in the current study. The numbering scheme was chosen to be consistent with the numbering for purine DNA bases. In general, several conformers of IDR are possible; we have chosen the lowest-energy conformation found in B-DNA, of C2′-endo/anti.[34] Table 1 shows selected structural data obtained for IDR. One can see that DFT predicts reasonable values for the bond lengths, valence angles, and torsion angles presented. The bond lengths tend to be higher than those predicted at the MP2/6-31G(d) level of theory[10] and also higher compared to crystal structure data on B-DNA; however, the discrepancy is not bad considering the strong crystal packing forces present in DNA crystals which are not accounted for in the theoretical calculations. The agreement in valence angles between the DFT, MP2, and crystal structure results is excellent.

The predicted torsion angles are also in line with those found for B-DNA, with the exception of β and ϵ; these values are outside the ranges traditionally found for B-DNA. This is most likely due to the lack of phosphate groups at the O3′ and O5′ positions in our model compound. In particular, the important glycosidic torsion angle, χ, is well-predicted by the present calculations; B-DNA normally exhibits χ values in the range of 252° ± 23°, and the calculated value of 243.9° is well within these error limits. Furthermore, the pseudorotation amplitude and phase angle, obtained using the method of Altona and Sundaralingam,[35] are 32.3° and 170.6°, respectively; these values are reasonable for a C2′-endo (south) structure and within values normally found for B-DNA.

Previous authors have commented on the C8−H8···O5′ “hydrogen bond” conferring stability on the anti conformer as opposed to the syn conformer for rotation about χ.[36−38] In the present work, we find a H8···O5′ distance of 2.685 Å and a C8···O5′ distance of 3.576 Å. In our estimation, these interatomic distances are too large in order for this interaction to be properly considered a “hydrogen bond”. This view is bolstered by the fact that the calculated C8−H8 bond length, 1.078 Å, is virtually identical with the calculated C4−H4 and C5−H5 bond lengths (1.078 and 1.077 Å, respectively), which do not participate in any sort of hydrogen bond. If the H8−O5′ interaction were a true hydrogen bond, one would expect a lengthening of the C8−H8 bond compared to comparable C−H bonds as a result of this interaction. Indeed, it is this structural feature that is a key indicator of a traditional hydrogen bond. One might argue that the C8−H8···O5′ interaction is a recently discovered “blue-shifting hydrogen bond”,[39] in which the C−H bond actually shortens and the C−H stretching frequency, ν(C−H), increases upon complexation. We do not find any shortening of the C8−H8 bond either, and ν(C8−H8) lies directly between the C4−H4 and C5−H5 symmetric and asymmetric stretching frequencies, where an “ordinary” C−H bond stretch would be expected to lie. If this is a “blue-shifting” hydrogen bond, the blue shift in stretching frequency in this case is exceedingly small. Finally, our calculated NMR SSCC data do not indicate the presence of any significant through-bond coupling between H8 and C5′. In our opinion, while this interaction may technically satisfy many criteria of a rigorous hydrogen-bonding analysis based on atoms-in-molecules principles,[40] we believe this interaction should more properly be characterized as a weak dipolar interaction between a slightly electropositive hydrogen atom (H8) and an electronegative oxygen atom (O5′), and the stability of the anti conformer over that of the syn conformer in DNA is most likely due to a number of contributions, including this weak dipolar interaction, steric effects, and numerous other structural modifications associated with binding in DNA. It is worth pointing out that, in the gas phase, it appears that the syn conformer of 2′-deoxyguanosine is energetically more stable,[41] again owing to the weakness of the C8−H8···O5′ interaction.

Table 2 displays the calculated B3LYP/6-311++G(2d,p) harmonic vibrational frequencies for IDR above 800 cm^{-1}. The lack of an imaginary harmonic frequency illustrates that the optimized structure is at a minimum on the potential energy surface. Also presented in Table 2 are the contributions to each mode from the molecule's natural internal coordinates, as determined by the method of Boatz and Gordon[42] and by visual inspection of each mode. Because the molecule lacks symmetry, there are very few modes consisting of only a small number of internal coordinates; modes below 800 cm^{-1} tend to be highly delocalized, with no individually large contribution from even a small number of internal coordinates. The exceptions are ν_1, calculated at 39 cm^{-1}, which is identified as the glycosidic rotation mode (χ), and ν_3, at 96 cm^{-1}, which is identified as the sugar ring puckering mode. Even these two modes, however, are not "pure" χ-rotation or pseudorotation modes; there is substantial mixing in all of the low-energy modes with many low-energy vibrational features, including C1′−N9 torsion and furanose ring torsions. Surprisingly, there are few published theoretical calculations of purine deoxyribonucleosides at the DFT level with large basis sets. Our results are in reasonable agreement with the ones already published,[43,44] considering the different levels of theory and different basis sets employed in those studies.

However, to truly assess the validity of the calculated numbers, especially with regard to the deoxyribose moiety, a comparison to experimental data is required. The proper comparison should be to deoxyribose vibrations of purine bases in DNA, especially because gas-phase vibrational data on individual purine deoxyribonucleosides are unfortunately scarce.[45−47] (However, our results are in substantial agreement with the existing gas-phase vibrational data that appear in the literature.) There are many well-known DNA vibrations arising from the sugar ring; most of the prominent ones appear in the 800−1500 cm^{-1} frequency range, with vibrations in the higher frequencies of this range corresponding to ones that are coupled between the base and sugar rings. A compendium of solution-phase Fourier transform infrared results on DNA fragments has recently been published.[48] The prominent vibrations arising from the deoxyribose ring occur at 1221, 1135, 1116−1119, 1044−1069, 1010−1020, 950−970, 890−899, and 820−842 cm^{-1}. Of these eight vibrational bands, we are able to successfully reproduce the first seven of them, considering that DFT, and ab initio calculations in general, tend to overestimate vibrational frequencies. For the lowest-energy experimental vibrational band, we are unable to find a corresponding one predicted from the calculations; however, because this vibration involves vibration between the oxygen atoms and the phosphate groups, such disagreement is not surprising.

We calculate a deoxyribose vibration at 1229 cm^{-1}, corresponding to the experimental 1221 cm^{-1} vibration. This

Table 2. Calculated Harmonic Vibrational Frequencies above 800 cm^{-1} and Their Descriptions[a]

mode	harmonic frequency (cm^{-1})	description
ν_{21}	807	δ(C1′−C2′−C3′)
ν_{22}	838	H8, H5 out-of-plane bends (symmetric)
ν_{23}	859	H8, H5 out-of-plane bends (asymmetric)
ν_{24}	880	τ(C2′−C3′) + τ(C3′−C4′) (symmetric)
ν_{25}	888	τ(C2′−C3′) + τ(C3′−C4′) (asymmetric)
ν_{26}	914	imidazole ring breathing mode (asymmetric)
ν_{27}	923	τ(C1′−C2′) + τ(C3′−C4′)
ν_{28}	945	furanose ring breathing mode
ν_{29}	993	furanose ring bending mode (asymmetric)
ν_{30}	1005	furanose ring bending mode (symmetric)
ν_{31}	1050	imidazole ring bending mode
ν_{32}	1057	ν(C1′−C2′) + ν(C2′−C3′) (asymmetric)
ν_{33}	1088	ν(C3′−O3′) + ν(C5′−O5′) (asymmetric)
ν_{34}	1108	ν(C3′−O3′) + ν(C5′−O5′) (symmetric)
ν_{35}	1109	ν(C1′−O4′) + ν(C4′−O4′) (symmetric)
ν_{36}	1116	ν(C1′−O4′) + ν(C4′−O4′) (asymmetric)
ν_{37}	1130	ν(C−O) in furanose ring coupled to ν(C−N) in imidazole ring
ν_{38}	1196	δ(H−O3′−C3′) + δ(H−O5′−C5′) (asymmetric)
ν_{39}	1210	δ(H−O3′−C3′) + δ(H−O5′−C5′) (symmetric)
ν_{40}	1229	C2′−H_2 twist + ν(C3′−O3′)
ν_{41}	1247	ν(C−N) in imidazole ring
ν_{42}	1278	ν(C1′−N9) + ν(C8−N9) + ν(C4−N9) (asymmetric)
ν_{43}	1285	ν(C1′−N9) + ν(C8−N9) + ν(C4−N9) (asymmetric)
ν_{44}	1312	coupled furanose + imidazole stretch mode
ν_{45}	1332	H3′ rock
ν_{46}	1337	coupled furanose + imidazole stretch mode
ν_{47}	1354	furanose CH rocking + asymmetric ν(C3′−C4′) + ν(C4′−O4′)
ν_{48}	1360	furanose ring deformation
ν_{49}	1386	ν(C1′−N9)+ imidazole ring breathing
ν_{50}	1390	ν(C2′−C3′)+ ν(C3′−C4′) (asymmetric)
ν_{51}	1424	ν(C2′−C3′)+ ν(C3′−C4′) (symmetric)
ν_{52}	1454	ν(C4′−C5′)
ν_{53}	1470	C2′−H_2 scissor mode
ν_{54}	1489	C2′−H_2 scissor mode
ν_{55}	1505	C5′−H_2 scissor mode
ν_{56}	1515	ν(N7−C8)+ ν(N7−C6) (asymmetric)
ν_{57}	1541	ν(N7−C8)+ ν(C4−C5)
ν_{58}	2979	ν(C5′−H5′)+ ν(C5′−H5″) (symmetric)
ν_{59}	3010	ν(C4′−H4′)
ν_{60}	3034	ν(C5′−H5′)+ ν(C5′−H5″) (asymmetric)
ν_{61}	3043	ν(C3′−H3′)
ν_{62}	3058	ν(C1′−H1′)
ν_{63}	3086	ν(C2′−H2′)+ ν(C2′−H2″) (symmetric)
ν_{64}	3134	ν(C2′−H2′)+ ν(C2′−H2″) (asymmetric)
ν_{65}	3235	ν(C4−H4)+ ν(C5−H5) (asymmetric)
ν_{66}	3250	ν(C8−H8)
ν_{67}	3264	ν(C4−H4)+ ν(C5−H5) (symmetric)
ν_{68}	3819	ν(O3′−H)
ν_{69}	3842	ν(O5′−H)

[a] δ = valence angle, ν = linear displacement (stretch), τ = torsion.

Table 3. Selected Calculated Carbon and Hydrogen Chemical Shielding (σ) Values for IDR at 298 K

atom	$\sigma_0^{(A)}$	$\bar{\sigma}_{\mathrm{ZPV}}^{(A)}$	$\bar{\sigma}_{\mathrm{thermal}}^{(A)}$	$\bar{\sigma}^{(A)}$	$\bar{\sigma}_{\mathrm{ZPV,harmonic}}^{(A)}$	$\bar{\sigma}_{\mathrm{thermal,harmonic}}^{(A)}$	$\bar{\sigma}_{\mathrm{harmonic}}^{(A)}$
C1′	88.8259	−1.9918	−0.1608	86.6733	−1.2695	−0.2489	87.3075
C2′	137.4698	−4.4233	−0.1180	132.9285	−2.4228	−0.5198	134.5272
C3′	101.5792	−4.0187	−0.4640	97.0965	−2.0890	−0.7185	98.7718
C4′	87.7574	−6.6294	−0.4139	80.7140	−2.5422	−1.2407	83.9744
C5′	113.4706	−7.9713	−0.7875	104.7118	−2.8484	−1.9584	108.6638
N9	47.2250	−3.1595	−1.2149	42.8506	−2.8739	−1.8335	42.5176
H1′	25.7170	−0.4174	−0.0290	25.2706	−0.1970	−0.0321	25.4879
H2′	29.2808	−0.5247	−0.0789	28.6772	−0.2696	−0.0696	28.9416
H2″	29.6026	−0.4345	−0.0568	29.1113	−0.2570	−0.0632	29.2823
H3′	27.2119	−0.5948	−0.0496	26.5675	−0.2632	−0.0549	26.8938
H(O3′)	31.5791	−4.1908	−0.1429	27.2454	−0.9588	−0.7009	29.9195
H4′	27.9921	−0.3482	−0.0296	27.6143	−0.2188	−0.0117	27.7616
H5′	27.8037	−0.9491	−0.0605	26.7940	−0.3463	−0.1822	27.2753
H5″	27.9183	−0.8501	−0.0825	26.9857	−0.3470	−0.1606	27.4107
H(O5′)	31.5991	−10.2818	−0.2810	21.0363	−1.5921	−2.8101	27.1969

vibration consists mostly of a C3′−O3′ stretch and a twisting motion of the hydrogen atoms on C2′. We calculate three vibrations clustered at 1108, 1109, and 1116 cm^{-1}, having large calculated IR intensities, corresponding to the two experimental IR bands at 1116−1119 and 1135 cm^{-1}. These bands are experimentally observed to be strong, and they consist mainly of C−C and C−O stretches. We calculate two vibrations at 1057 and 1088 cm^{-1}, corresponding to the two experimental IR bands at 1044−1069 and 1010−1020 cm^{-1}. Again, these vibrations consist of C−O vibrations, both symmetric and asymmetric. We find two vibrations at 1005 and 993 cm^{-1} that consist of symmetric and asymmetric furanose ring-bending modes, which we assign to the observed IR bands between 950 and 970 cm^{-1}. Finally, we calculate several lower-energy vibrations involving the ribose ring, mainly involving stretches and torsions involving the C2′−C3′, C3′−C4′, and C4′−C5′ bonds; their energies are at 945, 923, 888, and 880 cm^{-1}. We assign these modes to the ones observed between 890 and 899 cm^{-1}. It is also noteworthy that we are able to predict some vibrational modes in the purine ring, even though IDR is not an actual purine deoxyribonucleoside. These modes are at 1515 cm^{-1}, involving the C5/C8−N7 symmetric stretch (observed: 1476−1495 cm^{-1}), and at 1386 cm^{-1}, involving an imidazole breathing mode coupled to the C1′−N9 stretch (observed: 1369−1381 cm^{-1}).

In Table 3 are presented the calculated chemical shielding values for carbon, hydrogen, and nitrogen atoms at the equilibrium geometry of IDR and the thermally averaged chemical shielding values obtained from eq 10 using both harmonic and anharmonic contributions to the shielding. Also presented are the chemical shielding values assuming each vibration is purely harmonic, calculated using eqs 11 and 12. Considering first the equilibrium geometry values, we see that the chemical shieldings of C1′ and C4′, the carbons bonded to O4′, are large compared to those of C2′, C3′, and C5′, with C1′ being the most deshielded because it is adjacent to two electronegative atoms, O4′ and N9. The trend for the hydrogen chemical shifts also follows the usual pattern, with the hydroxyl protons being the most shielded and H1′, the proton adjacent to the strongly deshielded C1′, itself being the most deshielded.

Table 4 presents selected calculated one-bond, two-bond, and three-bond J coupling data for the equilibrium geometry of IDR, as well as their thermally averaged values using eq 15 employing first-order perturbation theory and using the harmonic oscillator approximation. We focus mainly on the NMR properties of the furanose ring, so J couplings involving just the imidazole ring are excluded. Among one-bond coupling constants, as expected, atom C1′ exhibits both the strongest homonuclear and heteronuclear coupling constants because C1′ is so deshielded compared to other carbons. Consequently, $^1J_{\mathrm{C2'C3'}}$ is comparatively small, and the heteronuclear couplings $^1J_{\mathrm{C2'H2''}}$ and $^1J_{\mathrm{C3'H3''}}$ are large. The hydrogen ^3J values also show evidence of strong coupling between C1′ and C2′, with $^3J_{\mathrm{H1'H2'}} = 5.7$ Hz and $^3J_{\mathrm{H1'H2''}} = 8.4$ Hz, both values being unusually large compared to the other H−H ^3J values.

As a typical example of the thermal averaging process, the carbon and hydrogen chemical shift data are presented in Table 5 for IDR displaced along ν_{69}, the O5′−H5′ stretch mode. Clearly, the effect of vibrational displacement is greatest on H5′, the atom which is the most substantially displaced, and on C5′, H5′, and H5″, the methylene group adjacent to O5′. Also shown in Table 5 are $\sigma_{1,j=69}^{(A)}$ and $\sigma_{2,j=69}^{(A)}$, the fitting parameters obtained after fitting the chemical shift data to eq 2 for this mode. Shown in Figure 2 is the chemical shift for H5′ as a function of vibrational displacement, along with the quadratic fitting function. The correlation coefficient for this fit is greater than 0.9999. In this particular example, $\sigma_j^{(A)}(y)$ is a decreasing function of α; however, this is not generally the case. An inspection of the data in Table 5 reveals that $\sigma_j^{(A)}(y)$, in general, does not have any preferred dependence on α, other than approximately quadratic.

Once the linear and quadratic fitting coefficients are obtained for each chemical shielding value and each J coupling, eqs 7 and 13 may then be applied to obtain $\bar{\sigma}_{\mathrm{ZPV}}^{(A)}$ and $\bar{J}_{\mathrm{ZPV}}^{(a)}$, the corrections to the chemical shielding and J coupling, respectively, due to ZPV. These ZPV corrections are presented in column three of Tables 3 and 4, respectively. Turning first to $\bar{\sigma}_{\mathrm{ZPV}}^{(A)}$, we see that the effect of ZPV correction is to decrease σ compared to its vibration-free,

Table 4. Selected Calculated *J* Coupling Data for IDR at 298 K[a]

N	bond path	$J_0^{(a)}$	$\bar{J}^{(a)}_{\mathrm{ZPV}}$	$\bar{J}^{(a)}_{\mathrm{thermal}}$	$\bar{J}^{(a)}$	$\bar{J}^{(a)}_{\mathrm{ZPV,harmonic}}$	$\bar{J}^{(a)}_{\mathrm{thermal,harmonic}}$	$\bar{J}^{(a)}_{\mathrm{harmonic}}$
1	$^1J_{C1'C2'}$	35.773	0.650	0.053	36.477	0.284	0.065	36.122
2	$^1J_{C1'H1'}$	150.667	1.966	0.315	152.948	2.474	0.306	153.447
3	$^1J_{C1'N9}$	8.437	0.082	−0.180	8.338	0.002	−0.236	8.202
4	$^1J_{C2'C3'}$	36.470	0.094	0.011	36.575	0.187	−0.085	36.572
5	$^1J_{C2'H2''}$	129.931	2.694	0.094	132.719	2.127	0.172	132.230
6	$^1J_{C2'H2'}$	127.123	1.262	0.147	128.532	1.703	0.155	128.982
7	$^1J_{C3'C4'}$	31.813	1.724	0.085	33.622	0.528	0.265	32.606
8	$^1J_{C3'H3'}$	142.163	1.720	0.218	144.100	2.133	0.428	144.724
9	$^1J_{C4'C5'}$	41.282	−1.038	−0.124	40.120	0.011	−0.581	40.711
10	$^1J_{C4'H4'}$	135.707	−0.602	0.217	135.322	1.474	−0.182	137.000
11	$^1J_{C5'H5'}$	134.794	5.671	0.484	140.949	2.924	1.390	139.109
12	$^1J_{C5'H5''}$	130.096	3.906	0.607	134.609	2.486	0.847	133.429
13	$^1J_{C8N9}$	6.370	−0.115	−0.225	6.029	−0.242	−0.333	5.795
14	$^1J_{C5N9}$	10.610	−0.221	−0.178	10.211	−0.219	−0.286	10.105
15	$^1J_{C8N7}$	−0.890	−0.486	−0.272	−1.648	−0.399	−0.458	−1.747
16	$^1J_{C8H8}$	200.018	2.714	1.670	204.402	2.673	2.964	205.654
17	$^1J_{C5N7}$	−3.195	−0.397	−0.356	−3.947	−0.344	−0.631	−4.170
18	$^1J_{C4C5}$	72.371	−0.134	−0.090	72.147	−0.734	−0.161	71.477
19	$^1J_{C5H5}$	179.193	2.884	1.082	183.159	2.371	1.982	183.545
20	$^1J_{C4H4}$	178.125	1.923	1.261	181.309	1.526	2.198	181.849
21	$^2J_{C1'C3'}$	1.246	0.030	−0.009	1.267	−0.074	−0.023	1.148
22	$^2J_{C1'H2''}$	1.939	−0.295	0.000	1.644	−0.282	0.004	1.661
23	$^2J_{C1'H2'}$	−6.095	−0.365	0.033	−6.427	−0.371	0.015	−6.450
24	$^2J_{C1'C4'}$	1.995	−0.029	−0.074	1.892	−0.018	−0.059	1.918
25	$^2J_{C1'C8}$	1.586	−0.032	−0.053	1.501	−0.043	−0.123	1.420
26	$^2J_{C1'C4}$	4.302	0.069	−0.094	4.278	0.072	−0.045	4.329
27	$^2J_{C2'H1'}$	−0.133	−0.119	−0.027	−0.280	−0.196	0.056	−0.274
28	$^2J_{C2'C4'}$	0.210	0.037	0.030	0.277	−0.019	0.000	0.190
29	$^2J_{C2'N9}$	0.373	−0.052	0.009	0.330	−0.023	−0.010	0.340
30	$^2J_{C2'H3'}$	−0.434	0.182	0.022	−0.230	−0.181	0.082	−0.533
31	$^2J_{H2'H2''}$	−12.941	−0.447	−0.040	−13.428	−0.538	0.033	−13.446
32	$^2J_{C3'H2''}$	−4.811	−0.212	−0.024	−5.047	−0.301	0.019	−5.094
33	$^2J_{C3'H2'}$	−0.581	−0.649	0.029	−1.200	−0.366	0.002	−0.944
34	$^2J_{C3'H(O3')}$	−2.382	−1.665	−0.022	−4.068	−0.516	−0.227	−3.124
35	$^2J_{C3'C5'}$	0.816	0.087	0.037	0.940	−0.020	0.047	0.843
36	$^2J_{C3'H4'}$	−4.443	−0.265	−0.017	−4.725	−0.302	0.007	−4.739
37	$^2J_{C4'H3'}$	5.947	−0.525	−0.005	5.417	−0.208	−0.126	5.613
38	$^2J_{C4'H5'}$	−5.350	0.054	0.015	−5.281	−0.310	0.140	−5.521
39	$^2J_{C4'H5''}$	3.179	0.254	0.085	3.519	−0.166	0.109	3.123
40	$^2J_{C5'H(O5')}$	−2.472	−4.448	−0.081	−7.001	−0.861	−1.114	−4.447
41	$^2J_{H5'H5''}$	−9.481	−1.279	0.052	−10.707	−0.450	−0.280	−10.210
42	$^2J_{N9H8}$	5.337	0.178	0.100	5.616	0.083	0.168	5.589
43	$^2J_{N9H4}$	3.058	0.011	0.105	3.175	0.005	0.193	3.256
44	$^2J_{N9N7}$	−1.001	0.002	0.006	−0.994	0.006	0.016	−0.979
45	$^2J_{N9C5}$	4.626	−0.059	−0.013	4.554	−0.067	−0.020	4.538
46	$^2J_{C5C8}$	−3.413	−0.448	−0.113	−3.974	−0.423	−0.173	−4.008
47	$^2J_{C4C8}$	11.748	0.274	0.025	12.047	0.128	0.021	11.897
48	$^2J_{N7H8}$	8.584	0.352	0.072	9.009	0.186	0.141	8.911
49	$^2J_{N7H5}$	7.321	0.361	0.067	7.748	0.171	0.144	7.635
50	$^2J_{N7C4}$	−1.545	−0.104	−0.007	−1.656	−0.085	−0.010	−1.640
51	$^2J_{C5H4}$	9.088	−0.174	0.025	8.939	−0.151	0.009	8.947
52	$^2J_{C4H5}$	16.862	0.320	−0.050	17.132	0.027	−0.096	16.793
53	$^3J_{C1'H3'}$	6.643	0.512	−0.044	7.110	0.276	−0.016	6.903
54	$^3J_{C1'C5'}$	0.127	0.023	−0.015	0.135	−0.003	0.028	0.152
55	$^3J_{C1'H4'}$	3.494	0.041	0.121	3.656	0.091	−0.011	3.574
56	$^3J_{C1'H8}$	0.530	0.011	−0.029	0.511	0.014	−0.039	0.505
57	$^3J_{C1'H4}$	2.011	0.080	−0.091	2.000	0.064	−0.156	1.919
58	$^3J_{C1'N7}$	−0.149	−0.005	−0.007	−0.161	0.001	−0.014	−0.163
59	$^3J_{C1'C5}$	4.273	0.060	−0.034	4.299	0.025	−0.039	4.259
60	$^3J_{C2'H(O3')}$	8.482	−0.056	0.001	8.426	0.072	−0.097	8.457

Table 4. Continued

N	bond path	$J_0^{(a)}$	$\bar{J}^{(a)}_{ZPV}$	$\bar{J}^{(a)}_{\mathrm{thermal}}$	$\bar{J}^{(a)}$	$\bar{J}^{(a)}_{\mathrm{ZPV,harmonic}}$	$\bar{J}^{(a)}_{\mathrm{thermal,harmonic}}$	$\bar{J}^{(a)}_{\mathrm{harmonic}}$
61	$^3J_{C2'J4'}$	0.999	−0.095	0.046	0.949	0.028	0.001	1.028
62	$^3J_{C2'C8}$	1.096	0.107	0.071	1.274	0.089	0.051	1.236
63	$^3J_{C2'C4}$	0.930	0.011	0.062	1.003	0.028	0.009	0.967
64	$^3J_{H1'H2''}$	5.690	0.328	0.053	6.071	0.274	−0.044	5.920
65	$^3J_{H1'H2'}$	8.356	1.097	−0.041	9.412	0.735	−0.017	9.074
66	$^3J_{C3'H1'}$	0.072	0.051	0.006	0.130	0.049	0.032	0.153
67	$^3J_{C3'H5'}$	7.010	0.367	0.015	7.392	0.290	−0.019	7.281
68	$^3J_{C3'H5''}$	0.874	0.217	0.022	1.113	0.108	0.048	1.031
69	$^3J_{C3'N9}$	0.900	0.007	−0.010	0.897	0.014	−0.005	0.908
70	$^3J_{H3'H(O3')}$	1.368	1.506	0.038	2.913	0.370	0.252	1.989
71	$^3J_{C4'H2''}$	3.766	0.057	−0.049	3.775	0.143	−0.039	3.870
72	$^3J_{C4'H2'}$	0.229	0.125	0.036	0.390	0.071	0.038	0.338
73	$^3J_{C4'H(O3')}$	0.255	0.357	−0.022	0.590	0.033	0.050	0.337
74	$^3J_{C4'H(O5')}$	7.146	−0.536	−0.186	6.424	−0.033	−0.266	6.848
75	$^3J_{C4'N9}$	0.550	0.013	−0.027	0.537	−0.003	−0.001	0.547
76	$^3J_{C5'H3'}$	2.053	0.164	0.007	2.224	0.074	0.018	2.145
77	$^3J_{H4'H5'}$	2.284	−0.125	0.017	2.176	0.037	−0.045	2.276
78	$^3J_{H4'H5''}$	1.878	−0.030	−0.021	1.827	−0.019	−0.004	1.855
79	$^3J_{H5'H(O5')}$	4.211	3.374	0.221	7.806	0.528	0.745	5.484
80	$^3J_{H5''H(O5')}$	−0.557	7.782	0.017	7.242	1.175	2.094	2.712
81	$^3J_{H2''H3'}$	0.307	0.102	0.002	0.412	0.078	0.022	0.408
82	$^3J_{H2'H3'}$	5.388	0.179	0.025	5.593	0.188	−0.051	5.525
83	$^3J_{H3'H4'}$	0.024	0.137	0.026	0.187	0.082	0.035	0.142
	$^{2h}J_{C5'H8}$	0.008	−0.012	−0.002	−0.006	−0.003	−0.004	0.001

[a] All values are in units of Hz.

Table 5. Calculated Fitting Parameters for the Chemical Shielding of Selected Atoms in IDR Displaced along the ν_{69} Vibrational Coordinate

atom	$\sigma^{(A)}$ (−α)	$\sigma^{(A)}$ (−0.5α)	$\sigma^{(A)}$ (+0.5α)	$\sigma^{(A)}$ (+α)	$\sigma_1^{(A)}$	$\sigma_2^{(A)}$
C1′	88.8392	88.8324	88.8196	88.8137	-1.28×10^{-2}	5.71×10^{-4}
C2′	137.5414	137.5044	137.4377	137.4080	-6.67×10^{-2}	4.89×10^{-3}
C3′	101.5085	101.5429	101.6175	101.6579	7.47×10^{-2}	4.00×10^{-3}
C4′	88.0081	87.8842	87.6261	87.4879	-2.60×10^{-1}	-9.46×10^{-3}
C5′	114.5899	114.0504	112.8476	112.1773	−1.21	-8.71×10^{-2}
H1′	25.7179	25.7174	25.7169	25.7171	-4.20×10^{-4}	4.86×10^{-4}
H2′	29.6025	29.6024	29.6031	29.6040	6.62×10^{-3}	2.14×10^{-3}
H2″	29.2763	29.2781	29.2846	29.2896	7.40×10^{-4}	6.57×10^{-4}
H3′	27.2229	27.2170	27.2079	27.2050	-8.98×10^{-3}	2.03×10^{-3}
H(O3′)	31.5940	31.5865	31.5717	31.5643	-1.48×10^{-2}	5.71×10^{-5}
H4′	27.9821	27.9873	27.9964	28.0002	9.06×10^{-3}	-9.43×10^{-4}
H5′	27.9060	27.8582	27.7422	27.6728	-1.16×10^{-1}	-1.43×10^{-2}
H5″	28.0140	27.9689	27.8620	27.7995	-1.07×10^{-1}	-1.16×10^{-2}
H(O5′)	35.4773	33.3906	30.0602	28.7362	−3.36	5.08×10^{-1}

purely electronic value. This trend can be rationalized by considering that a vibrating, moving atom will be less able to effectively screen an applied magnetic field than a stationary atom; hence, we expect $\bar{\sigma}^{(A)}_{\mathrm{ZPV}} < 0$ in all cases. The ZPV effect is comparatively larger on carbon atoms than hydrogen atoms, because of the larger number of vibrational degrees of freedom available to carbon atoms than that to hydrogen atoms.

Upon inspection of the J_0 and $\bar{J}^{(a)}_{\mathrm{ZPV}}$ data in Table 4, we see that $^1J_{\mathrm{CH}}$ and $^1J_{\mathrm{CC}}$ tend to increase (except for $^1J_{\mathrm{C4C5}}$), and $^1J_{\mathrm{CN}}$ tends to decrease, upon ZPV correction. For the two-bond coupling constants, all ^2J values tend to decrease; however, there are more exceptions. In evaluating three-bond coupling, there is no consistent upward or downward trend,

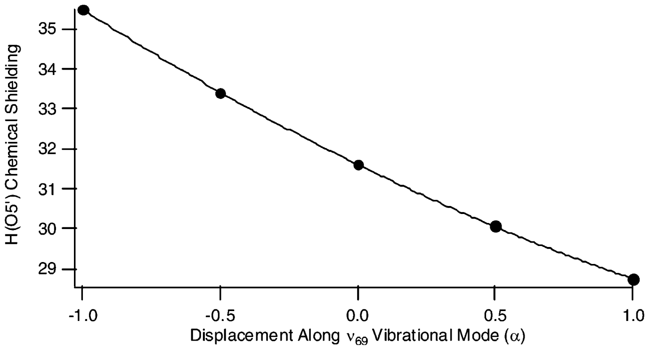


Figure 2. H(O5′) chemical shielding vs displacement along ν_{69}, in units of α. Circles: Calculated points. Line: Fit to eq 2.

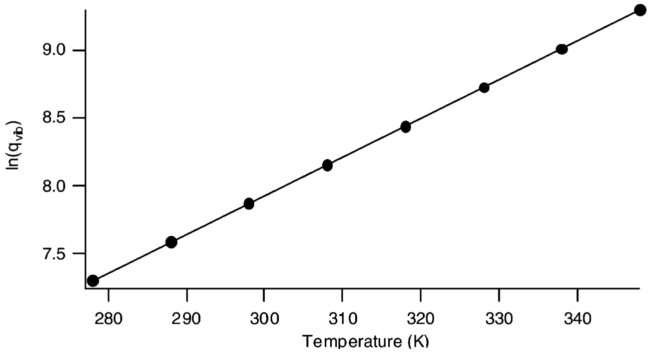


Figure 3. Dependence of the vibrational partition function on temperature for IDR in the range $T = 278-348$ K. Circles: Calculated points. Line: Fit to a straight line.

one way or another. This lack of a trend may simply be due to the fact that three-bond coupling is relatively weak, and any consistent influence on ^3J values may not manifest itself in the data.

The next two columns in Tables 4 and 5 display the thermal excitation contribution and the thermal average of the chemical shielding and J coupling values, respectively. We calculate $q_{\mathrm{vib}} = 2611$ at 298 K; the dependence of the vibrational partition function on temperature is shown in Figure 3. This large value for q_{vib} is the result of several low-lying vibrational states; as expected, states with energies below 2000 cm^{-1} overwhelmingly dominate the vibrational average. As expected, $\ln(q_{\mathrm{vib}})$ depends linearly on the temperature in the region near room temperature; we find a slope of 0.0284 K^{-1}. When thermal averaging is included to obtain average values of σ and J, we see that, as expected, σ decreases even further; all $\bar{\sigma}^{(A)}_{\mathrm{thermal}}$ values are negative. Again, more vibrational activity leads to less efficient screening of the magnetic field. The trend among J couplings is mixed, as before, but in general, larger vibrational effects are observed for couplings across flexible bonds (e.g., $^1J_{\mathrm{C8H8}}$, $^2J_{\mathrm{C5'H(O5')}}$). We also note that the calculated value of $^{2h}J_{\mathrm{C5'H8}}$ is 0.008 Hz, decreasing to -0.006 Hz upon ZPV and thermal excitation correction, which we consider to be, for all intents and purposes, negligible.

Also presented in Tables 4 and 5 are the same thermally averaged chemical shielding/J coupling data for IDR, assuming each vibration is purely harmonic. These data are presented to illustrate the effect of the application of first-order perturbation theory to the vibrational wave functions. For the chemical shielding, we see that the effect is to increase the magnitude of $\bar{\sigma}^{(A)}_{\mathrm{ZPV}}$ but to decrease the magnitude of $\bar{\sigma}^{(A)}_{\mathrm{thermal}}$. This is due to the fact that the perturbed energy levels are increased compared to their harmonic values. As a result, because the vibrational ground states increase in energy, the ZPV correction becomes larger in magnitude, but because the vibrational excited states also increase in energy, fewer excited states are thermally accessible, and the thermal excitation correction becomes smaller in magnitude. Overall, the net effect is to reduce $\bar{\sigma}^{(A)}$ compared to its harmonic value (with only one exception, N9). It is also notable that the range of vibrational corrections is very wide. The correction to the carbon shielding is as small as 2.2 ppm for C1′ and as large as 8.8 ppm for C5′.

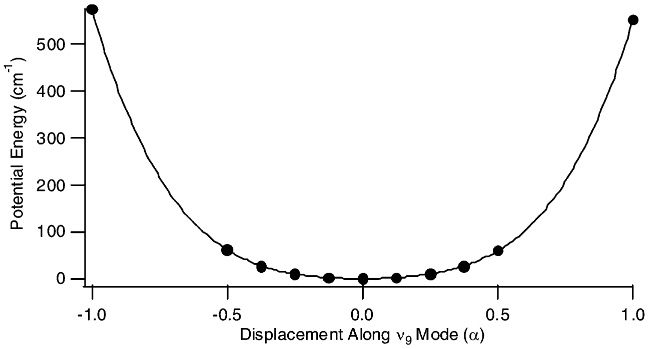


Figure 4. Potential energy curve for displacement along ν_9. Circles: Calculated points. Line: Fit to eq 1 with $\beta_9 = 134.4$ cm^{-1}, $\gamma_9 = -11.7$ cm^{-1}, and $\delta_9 = 429.1$ cm^{-1}.

The biggest single contributor to the chemical shielding vibrational correction for C1′ is the ZPV contribution from ν_{62}, corresponding to the C1′−H1 stretch mode, a comparatively high-energy vibrational mode. Conversely, the biggest contributors to the chemical shielding vibrational correction for C5′ are low-energy vibrational modes, particularly ν_6, which is a C5′−O5′−H(O5′) bending mode. Corrections to the proton shielding are smaller and more uniform, except for the OH protons. The effect of first-order perturbation theory on the chemical shielding is particularly significant for H(O3′) and H(O5′) because of two particular modes, ν_6 and ν_9. These two modes represent the C−O−H bending modes for H(O5′) and H(O3′), respectively. Shown in Figure 4 is the potential energy for displacement along the ν_9 mode. Clearly, the data indicate that this mode is highly anharmonic; a fit to a quadratic function is very poor.

In comparing the calculated thermally averaged J coupling data with the experiment, it is most fruitful to compare the $^3J_{\mathrm{HH}}$ data, because these couplings have been the most thoroughly studied in connection with the furanose ring conformation. It would be impossible in this limited space to compare the present results with the wealth of experimental data that exist for J couplings of DNA fragments. Hence, we limit our comparisons only to free 2′-deoxyadenosine (dA) and 2′-deoxyguanosine (dG) in solution[49] and to a typical set of experimental data obtained from purine nucleosides in a self-complementary double-stranded DNA decamer, d(CCGAATCGG)$_2$.[50] This DNA fragment has a B-DNA conformation in solution and is not known to form exotic structures such as hairpin loops. The terminal G residue was omitted from the comparison owing to the conformational flexibility available for terminal groups. Shown in Table 6 is a comparison of the relevant $^3J_{\mathrm{HH}}$ values. For the H1′−H2′ and H1′−H2″ coupling, we see that there is reasonable agreement with the experiment, both with free dA and dG and with the DNA decamer. However, for the H2′−H3′ and H2″−H3 couplings, agreement with the free dA and dG data is poor. The same poor agreement is seen with the H4′−H5′ and H4′−H5″ three-bond couplings. We believe that these discrepancies arise mainly from the neglect of contributions from other pseudorotational conformers. In solution, free nucleosides are able to adopt many conformations, including C3′-endo and -syn; the measured J couplings are therefore statistical averages over all of the conformations

Table 6. Comparison of Calculated $\bar{J}$ with Experiment at $T = 298$ K

bond path	calculated $\bar{J}$ (Hz)	dA[a] (Hz)	dG[a] (Hz)	dA[b] (Hz)	dA[c] (Hz)	dG[d] (Hz)	dG[e] (Hz)
$^3J_{H1'H2''}$	6.071	7.7	7.3	5.4	6.2	5.0	5.3
$^3J_{H1'H2'}$	9.412	6.2	6.5	9.5	8.5	10.4	9.6
$^3J_{H2''H3'}$	0.412	6.1	6.3	2.7	2.4	1.5	2.0
$^3J_{H2'H3'}$	5.593	3.3	3.6	n/a	5.3	5.6	6.2
$^3J_{H1'H2'} + {}^3J_{H1'H2''}$	15.483	13.9	13.8	14.9	14.7	15.4	14.9
$^3J_{H2'H3'} + {}^3J_{H2''H3'}$	6.005	9.4	9.9	n/a	7.7	7.1	8.2
$^3J_{H3'H4'}$	0.187	3.0	3.2				
$^3J_{H4'H5'}$	2.176	3.3	3.6				
$^3J_{H4'H5''}$	1.827	4.3	4.7				

[a] Free dA/dG in solution, from ref 49. [b] dA(4) in a d(CCGAATTCGG)$_2$ duplex decamer, from ref 50. [c] dA(5) in a d(CCGAATTCGG)$_2$ duplex decamer, from ref 50. [d] dG(3) in a d(CCGAATTCGG)$_2$ duplex decamer, from ref 50. [e] dG(9) in a d(CCGAATTCGG)$_2$ duplex decamer, from ref 50.

Table 7. Comparison of Calculated Vicinal $^3J_{HH}$ Coupling Constants with the Haasnoot Equation

bond path	HCCH dihedral angle (deg)	calculated $\bar{J}$ (Hz)	J[a] (Hz)
$^3J_{H1'H2''}$	32.16	6.071	6.25
$^3J_{H1'H2'}$	154.27	9.412	9.47
$^3J_{H2''H3'}$	86.95	0.412	0.53
$^3J_{H2'H3'}$	−34.53	5.593	6.10
$^3J_{H3'H4'}$	−99.13	0.187	1.20
$^3J_{H4'H5'}$	−64.87	2.176	3.57
$^3J_{H4'H5''}$	52.68	1.827	1.38

[a] From eq 16.

present. In the present study, on the other hand, we have considered only the C2′-endo (south) conformation of the furanose ring. These *J* couplings are obviously strongly dependent on the conformation of the ring; indeed, it is somewhat surprising that the H1′−H2′ and H1′−H2″ *J* coupling agreement is so good. When the calculated data are compared to the experimental DNA *J* coupling data, agreement is much better, although quantitative agreement for $^3J_{H2''H3'}$ is still lacking.

It is also instructive to compare the calculated values with those obtained from empirical, Karplus-like equations in common use. The most relevant of these is the Haasnoot equation for $^3J_{HH}$ between two sp^3 carbon centers,[51] given by eq 11.

$$^3J_{HH} = 13.70\cos^2\phi - 0.73\cos\phi + \sum_i \Delta\chi_i[0.56 - 2.47\cos^2(s_i\phi + 16.9|\Delta\chi_i|)] \quad (16)$$

In this equation, ϕ is the H−C−C−H dihedral angle and $\Delta\chi_i = \Delta\chi_i^{\alpha} - 0.14\sum \Delta\chi_i^{\beta}$ is the difference in group electronegativity between the *i*th substituent and hydrogen (1.3 for O, 0.4 for C, and 0.85 for N), corrected for α and β substituents; the sum over *i* runs over all carbon group substituents, and s_i is the “substituent orientation factor” (cf. Figure 2 of ref 51), either +1 or −1. Shown in Table 7 is a comparison of the relevant $^3J_{HH}$ values. We see that, for these $^3J_{HH}$ values, the Haasnoot equation tends to overestimate the *J* coupling compared to the quantum chemical results. This discrepancy between the calculated DFT values and those predicted by eq 16 may be attributable to several factors: (1) the lack of inclusion of solvent effects in the DFT calculations, (2) the incomplete inclusion of electron correlation, and (3) the inability of the current parametrization of the Haasnoot equation to correctly model the unique torsional behavior in nucleosides.

Conclusion

We have developed a method, based on ab initio calculations and first-order perturbation theory, to include thermal averaging in calculated magnetic properties, such as chemical shielding and *J* couplings. We have computed chemical shielding values and many *J* couplings for 1′-imidazolyl-2′-deoxy-β-ribofuranose, a model compound for purine nucleosides in DNA. For IDR, as a part of our analysis, we have performed a detailed vibrational analysis at the density functional level of theory with a large basis set, an analysis which was lacking previously. We find reasonable agreement between the calculated harmonic vibrational frequencies and measured solution values in DNA. We have shown that the inclusion of thermal averaging can drastically affect the calculated values of the magnetic properties, as compared to the purely electronic values. A comparison of the *J* coupling data to typical results from DNA fragments shows good agreement; a comparison to isolated dA/dG molecules in solution, however, is poor. We believe this poor agreement is due to a lack of consideration of more low-energy conformers (e.g., C3′-endo and -syn) in our model system. A comparison of our results with empirical Karplus-like relationships, such as the Haasnoot equation, illustrates a systematic underestimation of present results with the empirically derived ones. It is not clear whether this trend is due to a failure of the DFT calculations or a failure of the empirical relationship to correctly predict the *J* coupling properties of sugar protons in nucleosides.

As a side benefit of our work, in IDR, we have also analyzed the reported “hydrogen bond” that exists between H8 of the imidazole ring and O5′ on the sugar moiety. We find that the structural and magnetic data lending support to the notion of a hydrogen bond are weak; we believe that this interaction should more properly be characterized as a dipole−dipole interaction.

To improve upon our results, we believe more structural features should be added to the model compound, such as phosphate groups, to more closely model actual B-DNA strands. Also, even though a particular DNA strand may consist predominantly of B-DNA conformers, there may be a small population of A-DNA conformers that also contribute to the measured NMR response; an analysis of the A-DNA conformer, along with suitable statistical averaging, should also improve the agreement between theory and the experiment.

One additional problem arises because a Cartesian normal-mode analysis leads to errors in computing the effects of torsional modes. Treatment of the torsion about bonds as strictly linear Cartesian displacements normal to the bond leads to alteration in bond distances at high displacements, and corrections for such effects would require the introduction of anharmonic mode couplings, which greatly increases the computational time needed. However, in the present instance, the glycosidic and exocyclic torsions both involve heavy atom displacements, which means the vibrational amplitudes are rather small. The most serious errors, therefore, are likely to be on O−H proton chemical shifts, where the torsional modes introduce a second-order increase in the O−H bond length, which likely results in an artifactual reduction in the computed chemical shielding of the proton. However, OH proton chemical shifts are often difficult to measure, and two of the three OH protons are absent in RNA as a result of esterification. This error, therefore, is likely of limited practical impact, but we caution that any comparison of the present result with computed OH chemical shifts should bear this effect in mind.

Acknowledgment. J.N.W. would like to acknowledge the assistance of Professor Stephen Tanner, Department of Mathematics at Eastern Oregon University, and Mr. Xiongjian Wu, Department of Chemistry at University of Nebraska−Lincoln, in the computation of thermal averages. We gratefully acknowledge the financial support of the National Institutes of Health (R01 GM 065252).

Appendix

To evaluate the infinite sums in eqs 4 and 5, one must first compute the integrals $\langle\psi_m^{(0)}|\hat{y}|\psi_n^{(0)}\rangle$, $\langle\psi_m^{(0)}|\hat{y}^2|\psi_n^{(0)}\rangle$, $\langle\psi_m^{(0)}|\hat{y}^3|\psi_n^{(0)}\rangle$, and $\langle\psi_m^{(0)}|\hat{y}^4|\psi_n^{(0)}\rangle$ that appear repeatedly. These integrals may be solved using standard quantum-mechanical operator and integration techniques to yield the following expressions, found in common quantum-mechanical textbooks.

$$\langle\psi_m^{(0)}|\hat{y}|\psi_n^{(0)}\rangle = \sqrt{\frac{n}{2}}\delta_{m,n-1} + \sqrt{\frac{n+1}{2}}\delta_{m,n+1} \qquad (17)$$

Then, one substitutes these integrals into the expressions for

$$\langle\psi_m^{(0)}|\hat{y}^2|\psi_n^{(0)}\rangle = \frac{1}{2}\sqrt{n(n-1)}\delta_{m,n-2} + \left(n+\frac{1}{2}\right)\delta_{m,n} + \frac{1}{2}\sqrt{(n+1)(n+2)}\delta_{m,n+2} \qquad (18)$$

$$\langle\psi_m^{(0)}|\hat{y}^3|\psi_n^{(0)}\rangle = \sqrt{\frac{n(n-1)(n-2)}{8}}\delta_{m,n-3} + \frac{3n}{2}\sqrt{\frac{n}{2}}\delta_{m,n-1} + \frac{3(n+1)}{2}\sqrt{\frac{n+1}{2}}\delta_{m,n+1} + \sqrt{\frac{(n+1)(n+2)(n+3)}{8}}\delta_{m,n+3} \qquad (19)$$

$$\langle\psi_m^{(0)}|\hat{y}^4|\psi_n^{(0)}\rangle = \frac{1}{4}\sqrt{n(n-1)(n-2)(n-3)}\delta_{m,n-4} + \sqrt{n(n-1)}\left(n-\frac{1}{2}\right)\delta_{m,n-2} + \frac{3}{2}\left(n^2+n+\frac{1}{2}\right)\delta_{m,n} + \sqrt{(n+1)(n+2)}\left(n+\frac{3}{2}\right)\delta_{m,n+2} + \frac{1}{4}\sqrt{(n+1)(n+2)(n+3)(n+4)}\delta_{m,n+4} \qquad (20)$$

$c_{j,nm}^{(1)}$, $\langle\sigma_{1j}^{(A)}\rangle_n$, and $\langle\sigma_{2j}^{(A)}\rangle_n$ and evaluates the sums directly. Because of the Kronecker δ's, only a limited number of terms remain in each sum. The result for each sum is a complex polynomial in n. Define $\gamma_{r,j} = \gamma_j/\hbar\omega_{0j}$ and $\delta_{r,j} = \delta_j/\hbar\omega_{0j}$ as the reduced cubic and quartic potential energy fitting coefficients for the jth mode, respectively. The sums become:

$$\sum_{m\neq n}|c_{j,nm}^{(1)}|^2 = \frac{\gamma_{r,j}^{\;2}}{36}\left(n+\frac{1}{2}\right)(82n^2+82n+87) + \frac{\delta_{r,j}^{\;2}}{128}(65n^4 + 130n^3 + 487n^2 + 422n + 156) \qquad (21)$$

$$\langle\sigma_{1j}^{(A)}\rangle_n = \frac{\sigma_1^{(A)}\,\gamma_{r,j}}{1+\sum_{m\neq n}(c_{j,nm}^{(1)})^2}\left[\frac{\delta_{r,j}}{16}(27n^4 + 54n^3 + 177n^2 + 150n + 52) - 3\left(n+\frac{1}{2}\right)\right] \qquad (22)$$

$$\langle\sigma_{2j}^{(A)}\rangle_n = \frac{\sigma_2^{(A)}}{1+\sum_{m\neq n}(c_{j,nm}^{(1)})^2}\times \left[\left(n+\frac{1}{2}\right) - 3\delta_{r,j}\left(n^2+n+\frac{1}{2}\right) + \frac{\gamma_{r,j}^{\;2}}{144}(202n^4 + 404n^3 + 1478n^2 + 1276n + 393) + \frac{\delta_{r,j}^{\;2}}{256}(146n^5 + 365n^4 + 2704n^3 + 3691n^2 + 3438n + 1116)\right] \qquad (23)$$

Supporting Information Available: Equilibrium coordinates for IDR; isotropic chemical shifts as a function of displacement along each mode; isotropic J couplings as a function of displacement along each mode; equilibrium structure shielding and the zero-point and thermal corrections to the shielding, for each atom; equilibrium structure J couplings and the zero-point and thermal corrections to the couplings, for each atom; equilibrium values of the chemical shift and J coupling for each atom or pair of atoms; and the C program used to calculate the perturbation corrections to the shielding and J couplings. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Cavanaugh, J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: San Diego, CA, 1996. (b) Clore, G. M.; Gronenbom, A. M. *NMR of Proteins*; CRC Press: Boca Raton, FL, 1993. (c) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986. (d) Evans, J. N. S. *Biomolecular NMR Spectroscopy*; Oxford University Press: New York, 1995.

(2) Karplus, M. J. *J. Chem. Phys.* **1959**, *30*, 11−15.

(3) Karplus, M. J. *J. Am. Chem. Soc.* **1963**, *85*, 2870−2871.

(4) Džakula, Ž.; Westler, W. M.; Edison, A. S.; Markley, J. L. *J. Am. Chem. Soc.* **1992**, *114*, 6195−6199.

(5) Džakula, Ž.; DeRider, M. L.; Markley, J. L. *J. Am. Chem. Soc.* **1996**, *118*, 12796−12803.

(6) Wu, A.; Cremer, D. *J. Phys. Chem. A* **2003**, *107*, 1797−1810.

(7) Dziembowska, T.; Hansen, P. E.; Rozwadowski, Z. *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, *45*, 1−29.

(8) Vidossich, P.; Piana, S.; Miani, A.; Carloni, P. *J. Am. Chem. Soc.* **2006**, *128*, 7215−7221.

(9) Limbach, H.-H.; Denisov, G. S.; Golubev, N. S. In *Isotope Effects in Chemistry and Biology*; Kohn, A., Ed.; CRC Press: Boca Raton, FL, 2006; pp 193−230.

(10) Servis, K. L.; Domenick, R. L. *J. Am. Chem. Soc.* **1986**, *108*, 2211−2214.

(11) Freedberg, D. I.; Kopelevich, M.; Anet, F. A. L. *J. Phys. Org. Chem.* **2001**, *14*, 625−635.

(12) Burgi, R.; Pitera, J.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *19*, 305−320 and references therein.

(13) Case, D. A.; Scheurer, C.; Brüschweiler, R. A. *J. Am. Chem. Soc.* **2000**, *122*, 10390−10397.

(14) Markwick, P. R. L.; Sprangers, R.; Sattler, M. *J. Am. Chem. Soc.* **2003**, *125*, 644−645.

(15) Foloppe, N.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **1998**, *102*, 6669−6678.

(16) Leulliot, N.; Ghomi, M.; Scalmani, G.; Berthier, G. *J. Phys. Chem. A* **1999**, *103*, 8716−8724.

(17) Moreno, M.; Douhal, A.; Lluch, J. M.; Castano, O.; Frutos, L. M. *J. Phys. Chem. A* **2001**, *105*, 3887−3893.

(18) Casadesus, R.; Moreno, M.; Lluch, J. M. *Chem. Phys.* **2003**, *290*, 319−336.

(19) Gelabert, R.; Moreno, M.; Lluch, J. M. *J. Phys. Chem. A* **2006**, *110*, 1145−1151.

(20) Bertran, J.; Oliva, A.; Rodriguez-Santiago, L.; Sodupe, M. *J. Am. Chem. Soc.* **1998**, *120*, 8159−8167.

(21) Kuo, I. F.; Tobias, D. J. *J. Phys. Chem. B* **2001**, *105*, 5827−5832.

(22) Schwegler, E.; Galli, G.; Gygi, F. *Chem. Phys. Lett.* **2001**, *342*, 434−440.

(23) Chen, X. H.; Zhang, J. Z. H. *J. Chem. Phys.* **2004**, *120*, 11386−11391.

(24) Barone, G.; Saturnino, C.; DeMartino, G.; Duca, D.; La-Manna, G. *THEOCHEM* **2001**, *572*, 113−119.

(25) Del Bene, J. E.; Jordan, M. J. T.; Perera, A.; Bartlett, R. J. *J. Phys. Chem. A* **2001**, *105*, 8399−8402.

(26) Del Bene, J. E.; Jordan, M. J. T. *J. Phys. Chem. A* **2002**, *106*, 5385−5392.

(27) Grigoleit, S.; Bühl, M. *Chem.*—*Eur. J.* **2004**, *10*, 5541−5552.

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

(29) Case, D. A. *J. Biomol. NMR* **1999**, *15*, 95−102.

(30) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363; version 20 JUN 2002 (R2).

(31) Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192−212.

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

(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.04; Gaussian, Inc.: Pittsburgh, PA, 2003.

(34) Foloppe, N.; Nilsson, L.; MacKerell, A. D., Jr. *Biopolymers* **2002**, *61*, 61−76.

(35) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205−8212.

(36) Rubin, J.; Brennan, T.; Sundaralingam, M. *Biochemistry* **1972**, *11* (1), 3112−3128.

(37) Wahyl, M. C.; Sundaralingam, M. *Trends Biochem. Sci.* **1997**, *22*, 97−102.

(38) Hocquet, A.; Ghomi, M. *Phys. Chem. Chem. Phys.* **2000**, *2*, 5351−5353.

(39) Hobza, P.; Havlas, Z. *Chem. Rev.* **2000**, *100*, 4253−4264.

(40) Hocquet, A. *Phys. Chem. Chem. Phys.* **2001**, *3*, 3192−3199.

(41) Nir, E.; Hünig, I.; Kleinermanns, K.; de Vries, M. *ChemPhysChem* **2004**, *5*, 131−137.

(42) Boatz, J. A.; Gordon, M. S. *J. Phys. Chem.* **1989**, *93*, 1819−1826.

(43) Pelmenschikov, A.; Hovorun, D. M.; Shishkin, O. V.; Leszczynski, J. *J. Chem. Phys.* **2000**, *113*, 5986−5990.

(44) Leulliot, N.; Ghomi, M.; Jobic, H.; Bouloussa, O.; Baumruk, V.; Coulombeau, C. *J. Phys. Chem. B* **1999**, *103*, 10934−10944.

(45) Nir, E.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. *J. Am. Chem. Soc.* **2000**, *122*, 8091−8092.

(46) Nir, E.; Plutzer, C.; Kleinermanns, K.; de Vries, M. *Eur. Phys. J. D* **2002**, *20*, 317−329.

(47) Nir, E.; Hunig, I.; Kleinermanns, K.; de Vries, M. S. *ChemPhysChem* **2004**, *5*, 131−137.

(48) Banyay, M.; Sarkar, M.; Gräslund, A. *Biophys. Chem.* **2003**, *104*, 477−488.

(49) Plavec, J.; Tong, W.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **1993**, *115* (5), 9734−9746.

(50) Rinkel, L. J.; van der Marel, G. A.; van Boom, J. H.; Altona, C. *Eur. J. Biochem.* **1987**, *163*, 175−286.

(51) Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783−2792.

CT6000736